การทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

น.ส.ปานทิพย์ พู่พุฒ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมซอฟต์แวร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Predicting Code Review Results in Open Source Projects on GitHub

Miss Panthip Pooput

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Software Engineering
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2018

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ
โดย	น.ส.ปานทิพย์ พู่พุฒ
สาขาวิชา	วิศวกรรมซอฟต์แวร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.พรศิริ หมื่นไชยศรี

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.ธาราทิพย์ สุวรรณศาสตร์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.พรศิริ หมื่นไชยศรี)

..... กรรมการ
(รองศาสตราจารย์ ดร.วิวัฒน์ วัฒนาวุฒิ)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.อาทิตย์ ทองทักษ์)

..... กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.ภัทรชัย ลลิตโรจน์วงศ์)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ปานทิพย์ พู่พุฒ : การทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ. (Predicting Code Review Results in Open Source Projects on GitHub) อ.ที่ปรึกษาหลัก : รศ. ดร.พรศิริ หมื่นไชยศรี

กระบวนการรีวิวโค้ดบนกิตฮับ เป็นกระบวนการที่สำคัญกระบวนการหนึ่งในการทำงานบนกิตฮับ โดยนักพัฒนาซอฟต์แวร์จะทำการร้องขอการนำเข้าซอร์สโค้ดหลังจากทำการแก้ไขหรือเปลี่ยนแปลงซอร์สโค้ดเสร็จเรียบร้อยแล้ว ผู้ดูแลระบบจะเป็นผู้รีวิวซอร์สโค้ด โดยพิจารณาถึงคุณภาพ และรายละเอียดทั้งหมดของรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ๆ จากการวิเคราะห์ผลการรีวิว พบว่ามีรายการร้องขอการนำเข้าซอร์สโค้ดจำนวนมาก ที่ไม่ผ่านการรีวิว เนื่องด้วยปัจจัยหลายประการ เช่น ความซับซ้อนของซอร์สโค้ด คุณภาพของซอร์สโค้ด รวมไปถึงจำนวนการเปลี่ยนแปลงไฟล์ เป็นต้น กระบวนการแก้ไขรายการร้องขอการนำเข้าซอร์สโค้ดที่ไม่ผ่านการรีวิวต้องใช้ความพยายาม และเวลาอย่างมากในการแก้ไข ซึ่งอาจส่งผลกระทบกับค่าใช้จ่ายของโครงการ หรือกำหนดการของโครงการได้

งานวิทยานิพนธ์นี้จึงออกแบบ แบบจำลองการทำนายผลการรีวิวโค้ดบนกิตฮับ โดยการวิเคราะห์ข้อมูลบนกิตฮับ กำหนดตัวแปรที่คาดว่าจะมีผลกระทบกับการรีวิวโค้ด รวมถึงการหาความสัมพันธ์ที่เกิดขึ้นบ่อยจากกฎความสัมพันธ์ จากนั้นทำการสร้างแบบจำลองการทำนายการรีวิวโค้ดบนกิตฮับด้วยวิธีการทางสถิติโลจิสติก และหลักการเรียนรู้ด้วยเครื่อง การวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น

ผลการทดลองแสดงสิบรายการร้องขอการนำเข้าซอร์สโค้ดที่พบบ่อยที่สุดในชุดข้อมูล และแบบจำลองการทำนายผลการรีวิวโค้ด โดยแบบจำลองที่ใช้ในการทำนายผลการรีวิวโค้ดที่ใช้การวิเคราะห์โลจิสติกในการวิเคราะห์ข้อมูลมีเปอร์เซ็นต์ความแม่นยำ 89.2307% และแบบจำลองที่ใช้การวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น มีเปอร์เซ็นต์ความแม่นยำ 90.7692%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สาขาวิชา วิศวกรรมซอฟต์แวร์ ลายมือชื่อนิสิต .....

ปีการศึกษา 2561 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

5971016021 : MAJOR SOFTWARE ENGINEERING

KEYWORD: Code Review, Pull Request, GitHub, Open Source, Association Rules, Multi Layer Perceptron, Neural Network, Logistic Regression, Data Mining, Machine Learning, Statistical, Factor

Panthip Pooput : Predicting Code Review Results in Open Source Projects on GitHub. Advisor: Assoc. Prof. Pornsiri Muenchaisri

Code review on GitHub is an important method for code contribution that will be submitted when the developers would like to merge their code changed from their local machine to the main repository. The reviewer will check the source code quality and all pull request details. The pull request status may be rejected due to several factors, such as code complexity, code quality, the number of changed files, etc. Fixing the rejected pull requests will take some extra effort and time which may affect the project cost and timeline.

This thesis design the predicting code review results model by analyzing the data on GitHub, assign the related factors and also discover relationships among impact factors by using association rules in Data mining. The predicting code review results on GitHub is created by the logistic regression analysis in Statistical analysis and the multilayer perceptron in Machine learning.

The results show ten associate rules of the source code file format with the related variables. The accuracy of predicting code review results model that used the logistic regression method is 89.2307%. For the multilayer perceptron in Machine learning is 90.7692%.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Field of Study:	Software Engineering	Student's Signature .....
Academic Year:	2018	Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดีจากคำแนะนำของรองศาสตราจารย์ ดร. พรศิริ หมื่นไชยศรี อาจารย์ที่ปรึกษาที่เคารพรัก ผู้ให้ความเมตตา และสละเวลาอันมีค่าเพื่อให้คำแนะนำ รวมถึงชี้แนวทางต่าง ๆ อันเป็นประโยชน์อย่างมากในการทำวิจัย ผู้วิจัยจึงกราบขอบพระคุณอย่างสูงมา ณ ที่นี้

กราบขอบพระคุณ รองศาสตราจารย์ ดร. ธาราทิพย์ สุวรรณศาสตร์ ประธานกรรมการสอบ รองศาสตราจารย์ ดร. วิวัฒน์ วัฒนาวุฒิ ผู้ช่วยศาสตราจารย์ ดร. อาทิตย์ ทองทักษ์ และผู้ช่วยศาสตราจารย์ ดร. ภัทรชัย ลลิตโรจน์วงศ์ คณะกรรมการสอบวิทยานิพนธ์ ที่ได้สละเวลา ให้ความกรุณาในการให้คำแนะนำ รวมถึงชี้นำแนวทางเกี่ยวกับการทำวิจัย เพื่อเติมเต็มความสมบูรณ์ของวิทยานิพนธ์ฉบับนี้

กราบขอบพระคุณคณาจารย์ ภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย ผู้ถ่ายทอดความรู้เกี่ยวกับวิศวกรรมซอฟต์แวร์ รวมถึงคำแนะนำต่าง ๆ มากมาย ประโยชน์ที่ได้จากการเรียนภาควิชานี้นั้น จะถูกนำไปใช้ประโยชน์สูงสุดให้แก่องค์กร และวงการวิศวกรรมซอฟต์แวร์ต่อไป

ขอบคุณแรงขับเคลื่อนสำคัญ ครอบครัว ผู้ให้กำลังใจ และอยู่เบื้องหลังความสำเร็จทั้งปวง คุณค่าจากสิ่งที่ได้รับในช่วงเวลานี้ ไม่สามารถประเมินค่าได้ ขอบคุณคุณอิศเรศผู้ให้กำลังใจ คอยอยู่เคียงข้าง และให้คำแนะนำในการแก้ไขปัญหาต่าง ๆ มาตลอดการทำวิจัย และสุดท้ายขอบคุณเพื่อน ๆ พี่ ๆ น้อง ๆ ทุกคนที่ให้การสนับสนุนทั้งด้านการเรียน และคอยเติมแรงใจในการทำวิจัย จนงานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี

ปานทิพย์ พู่พุฒ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ค

บทคัดย่อภาษาอังกฤษ ..... ง

กิตติกรรมประกาศ ..... จ

สารบัญ ..... ฉ

สารบัญภาพ 10

สารบัญตาราง 13

บทที่ 1 บทนำ 16

1.1 ที่มาและความสำคัญของปัญหา 16

1.2 วัตถุประสงค์ 17

1.3 ขอบเขตการดำเนินงาน 17

1.4 ขั้นตอนการดำเนินงาน 18

1.5 ประโยชน์ที่คาดว่าจะได้รับ 18

1.6 บทความวิจัยที่ได้รับการตีพิมพ์ 19

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 20

2.1 ทฤษฎีที่เกี่ยวข้อง 20

2.1.1 โอเพนซอร์ส 20

2.1.2 การพัฒนาซอฟต์แวร์แบบการร้องขอการนำเข้าซอร์สโค้ด 21

2.1.3 การรีวิวโค้ดบนกิตฮับ 24

2.1.4 กระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่ 25

2.1.5 กระบวนการหาความสัมพันธ์ 28

2.1.6 การใช้หลักการทางสถิติในการวิเคราะห์ข้อมูล 30

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.1.7 การวิเคราะห์ความถดถอยโลจิสติก 33
2.1.8 การวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น 36
2.1.9 เครื่องมือราดอน 38
2.1.10 เครื่องมือซอร์สมิเตอร์ 40
2.1.11 เครื่องมือสำเร็จรูปทางสถิติเอสพีเอสเอส 41
2.1.12 เครื่องมือเวก้า 44
2.2 งานวิจัยที่เกี่ยวข้อง 46
2.2.1 งานวิจัย “What Factor Influence The Reviewer Assignment To Pull Requests?” 46
2.2.2 งานวิจัย “Acceptance Factors of Pull Requests in Open-Source Projects” 47
2.2.3 งานวิจัย “Rejection Factors of Pull Requests Field by Core Team Developers in Software Projects with High Acceptance Rates” 48
บทที่ 3 แบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 50
3.1 ภาพรวมวิธีการออกแบบการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 50
3.2 วิธีการสร้างแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 51
3.2.1 กำหนดและวิเคราะห์โครงการโอเพนซอร์สที่นำมาวิเคราะห์ในงานวิจัย 51
3.2.2 กำหนดตัวแปรที่คาดว่าจะมีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 63
3.2.3 รวบรวมข้อมูลตัวแปรที่มีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 70
3.2.4 วิเคราะห์ความสัมพันธ์ระหว่างตัวแปรด้วยกฎความสัมพันธ์ 77
3.2.5 สร้างแบบจำลองการทำนายด้วยการวิเคราะห์การถดถอยโลจิสติก 84
3.2.6 สร้างแบบจำลองการทำนายด้วยการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น 93
3.3 การประเมินแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 98
3.3.1 วิธีการประเมินแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 99

3.3.2 การวิเคราะห์ผลลัพธ์การประเมินผลแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 100
3.4 การประยุกต์ใช้แบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 105
บทที่ 4 การออกแบบและการพัฒนาเครื่องมือ 108
4.1 ข้อกำหนดความต้องการ 108
4.1.1 ข้อกำหนดความต้องการที่เป็นหน้าที่หลัก 108
4.1.2 ข้อกำหนดความต้องการที่ไม่เป็นหน้าที่หลัก 109
4.2 แบบจำลองการวิเคราะห์ 110
4.2.1 แบบจำลองเชิงหน้าที่ 110
4.2.2 แบบจำลองเชิงโครงสร้าง 116
4.2.3 แบบจำลองเชิงพฤติกรรม 116
4.3 สถาปัตยกรรมระบบ 117
4.3.1 ภาพรวมเชิงสถาปัตยกรรม 117
4.3.2 แบบจำลองเชิงสถาปัตยกรรม 118
4.4 การออกแบบโครงสร้างการเชื่อมต่อ 118
4.4.1 แผนภาพวินโดว์ 118
4.4.2 ส่วนต่อประสานผู้ใช้งาน 119
4.5 การทดสอบเครื่องมือการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 119
4.5.1 สภาพแวดล้อมที่ใช้ในการทดสอบ 120
4.5.2 การทดสอบเครื่องมือ 120
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 126
5.1 สรุปผลการวิจัย 126
5.2 ข้อจำกัดของงานวิจัย 127
5.3 แนวทางการพัฒนางานวิจัยต่อ 127

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บรรณานุกรม.....129
ภาคผนวก.....132
ภาคผนวก ก.....133
ภาคผนวก ข.....142
ภาคผนวก ค.....146
ภาคผนวก ค.....170
ภาคผนวก ง.....175
ภาคผนวก จ.....197
ประวัติผู้เขียน.....200

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญภาพ

หน้า

รูปที่ 2.] กราฟแสดงอัตรายการเติบโตของการใช้งานการร้องขอการนำเข้าซอร์สโค้ดบนกิตฮับ 17[21

รูปที่ 2.2 ขั้นตอนการทำงานของการร้องขอการนำเข้าบนกิตฮับ [10] 22

รูปที่ 2.3 สถานะของการร้องขอการนำเข้าซอร์สโค้ดบนกิตฮับ]9[..... 24

รูปที่ 2.4 ขั้นตอนการรีวิวโค้ดบนกิตฮับ ผลการรีวิวคือผ่านการรีวิว [1] 25

รูปที่ 2.5 ขั้นตอนการรีวิวโค้ดบนกิตฮับ ผลการรีวิวคือไม่ผ่านการรีวิว [1] 25

รูปที่ 2.6 ขั้นตอนการประยุกต์ใช้กระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่ [11] 28

รูปที่ 2.7 การหากฎความสัมพันธ์ [12] 29

รูปที่ 2.8 แสดงการหาค่าสนับสนุนที่กำหนดและค่าความเชื่อมั่น [12] 29

รูปที่ 2.9 วิธีการคำนวณการวัดแนวโน้มเข้าสู่ศูนย์กลาง 32

รูปที่ 2.10 กราฟและข้อกำหนดการแจกแจกแบบปกติ เบ้ซ้าย และเบ้ขวา 32

รูปที่ 2.11 วิธีการคำนวณส่วนเบี่ยงเบนมาตรฐานและความแปรปรวน 33

รูปที่ 2.กราฟของโมเดลการวิเคราะห์การถดถอยโลจิสติก 12 [17] 34

รูปที่ 2.13 เส้น Linear Probability Model เทียบกับ Logistic Regression Model [17] 34

รูปที่ 2.14 สมการทางคณิตศาสตร์ของค่า Odds 35

รูปที่ 2.15 สมการทางคณิตศาสตร์ที่ใช้ในการทำนายตัวแปรตาม 35

รูปที่ 2.16 หลักการโครงข่ายประสาทเทียม [18] 37

รูปที่ 2.17 ภาพรวมของสถาปัตยกรรมของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ 37

รูปที่ 2.18 ส่วนต่อประสานเครื่องมือเอสพีเอสเอส 41

รูปที่ 2.19 แถบ Variable View ในโปรแกรมเอสพีเอสเอส 42

รูปที่ 2.20 แถบ Data View ในโปรแกรมเอสพีเอสเอส 42

รูปที่ 2.เมนูการวิเคราะห์ข้อมูลสถิติพื้นฐานในโปรแกรมเอส 21พีเอสเอส 43

รูปที่ 2.22 เมนูการวิเคราะห์ข้อมูลสถิติขั้นสูงในโปรแกรมเอสพีเอสเอส 43

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รูปที่ 2. ส่วนหน้าจอ 23Explorer ของเครื่องมือเวก้า 44
รูปที่ 2.24 ตัวอย่างผลลัพธ์ทีแสดงในโปรแกรมเวก้า 46
รูปที่ 3.ภาพรวมวิธีการสร้างแบบจำลองการทำนายผลการรีวิวโ 1ค้ดในโครงการโอเพนซอร์สบนกิตฮับ 50
รูปที่ 3. กระบวนการร้องขอการนำเข้าซอร์สโค้ดของโครงการ 2Ansible 56
รูปที่ 3.3 จำนวนโครงการย่อยและผู้ดูแลระบบของโครงการ Ansible 59
รูปที่ 3.4 ตัวอย่างแผ่นป้ายในเมนู Pull Requests ของโครงการ Anisible บนกิตฮับ 60
รูปที่ 3.5 ตัวอย่างหน้าจอของโปรแกรม Postman ในการดึงข้อมูลรายการ Pull requests 63
รูปที่ 3.6 วิธีการหาตัวแปรจากงานวิจัยที่เกี่ยวข้องกับการรีวิวโค้ด 63
รูปที่ 3.ส่วนต่อประสานของเครื่องมือเวก้าที่ใช้ 7 76
รูปที่ 3.8 เมนู Associate ในโปรแกรมเวก้า 77
รูปที่ 3.9 รายละเอียดการวิเคราะห์ (ข้อมูลส่วนแรก) 79
รูปที่ 3.10 รายละเอียดตัววัดในการวิเคราห์ (ข้อมูลส่วนที่สอง) 79
รูปที่ 3.11 กฎความสัมพันธ์ที่ดีที่สุดสิบกฎ (ข้อมูลชุดที่สาม) 79
รูปที่ 3.ขั้นตอนการวิเคราะห์การถดถอยโลจิสติก 12 85
รูปที่ 3.13 การตั้งค่าตัวแปร หรือเตรียมข้อมูลก่อนการวิเคราะห์ 85
รูปที่ 3.14 การตั้งค่าการวิเคราะห์การถดถอยโลจิสติก 86
รูปที่ 3.สมการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 15 93
รูปที่ 3. เมนู 16Classify เพื่อใช้ในการวิเคราะห์ ในเครื่องมือเวก้า 94
รูปที่ 3.17 การแบ่งช่วงของข้อมูลสำหรับการเตรียมข้อมูลก่อนการวิเคราะห์ 95
รูปที่ 3.18 หน้าจอการตั้งค่าการวิเคราะห์ 95
รูปที่ 3.19 รายละเอียดการวิเคราะห์ (ข้อมูลส่วนแรก) 96
รูปที่ 3.20 หน้าจอผลลัพธ์การวิเคราะห์ข้อมูล (ข้อมูลส่วนที่สอง) 96
รูปที่ 3.21 ขั้นตอนการประเมินผลแบบจำลองการทำนาย 99

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รูปที่ 3.22 ภาพรวมการประยุกต์ใช้แบบจำลองการทำนายก่อนการร้องขอการนำเข้าซอร์สโค้ด 106
รูปที่ 3.23 ภาพรวมการประยุกต์ใช้แบบจำลองการทำนายหลังการร้องขอการนำเข้าซอร์สโค้ด..... 107
รูปที่ 4.แผนภาพยูสเคสเครื่องมือการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 1 ... 110
รูปที่ 4.2 แผนภาพกิจกรรมของการทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 115
รูปที่ 4.3 แผนภาพคลาสของการทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 116
รูปที่ 4.4 แผนภาพลำดับแบบจำลองการทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ 116
รูปที่ 4.5 แผนภาพดีพลอย 118
รูปที่ 4.6 แผนภาพวินโดว์..... 118

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

หน้า

ตารางที่ 2.] คำสั่งการประมวลผลและการแปลความหมายผลลัพธ์ของเครื่องมือราดอน 119[..... 38
ตารางที่ 2.2 การประมวลผล (แถบข้อมูล) ในโปรแกรมเวก้า [22] 45
ตารางที่ 2.3 ลักษณะพิเศษของโครงการที่นำมาพิจารณา]5] 48
ตารางที่ 3.1 รายละเอียดโครงการโอเพนซอร์สสิบอันดับที่มีการรีวิวสูงสุดในกิตฮับ 52
ตารางที่ 3.รายละเอียดโครงการโอเพนซอร์สสิบอันดับที่มีการรีวิวสูงสุดในกิตฮับ 2 53
ตารางที่ 3.3 ภาพรวมวิธีการวิเคราะห์โครงการโอเพนซอร์ส 54
ตารางที่ 3.4 รายการเอพีไอที่ใช้ในการวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านกิตฮับ 58
ตารางที่ 3. รายละเอียดของแผ่นป้ายของโครงการ 5Ansible 60
ตารางที่ 3.6 ตัวแปรที่ได้จากการวิเคราะห์รายละเอียดโครงการ 64
ตารางที่ 3.7 คำสั่งที่ใช้ในการดึงข้อมูลผ่านโปรแกรมราดอน 72
ตารางที่ 3.8 บรรทัดคำสั่งที่ใช้ในการดึงข้อมูลผ่านโปรแกรมซอร์สมิเตอร์ 74
ตารางที่ 3.9 ตัวอย่างการแบ่งช่วงข้อมูล 78
ตารางที่ 3.10 กฎที่ดีที่สุดที่ได้จากการวิเคราะห์ข้อมูล 80
ตารางที่ 3.11 ไอเทมเซตที่เกิดขึ้นบ่อย 82
ตารางที่ 3.12 รายละเอียดการตั้งค่าการวิเคราะห์การถดถอยโลจิสติก 86
ตารางที่ 3.13 ผลลัพธ์สรุปการเตรียมกรณีตัวอย่าง)Case Processing Summary) 88
ตารางที่ 3.14 ผลลัพธ์การกำหนดค่าตัวแปรตาม 88
ตารางที่ 3.15 ผลลัพธ์การวิเคราะห์ Omnibus Tests of Model Coefficients 88
ตารางที่ 3.ผลลัพธ์การสรุปแบบจำลอง 16 89
ตารางที่ 3.17 ตารางการจำแนกข้อมูลแสดงความน่าเชื่อถือข้อมูล 89
ตารางที่ 3.18 ตัวแปรและค่าสัมประสิทธิของตัวแปรที่อยู่ในสมการ 90
ตารางที่ 3.19 เมตริกความสับสนของการวิเคราะห์เพอร์เซปตรอนหลายชั้น 98

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.20 รายละเอียดผลลัพธ์การคัดเลือกข้อมูลการประเมินผล 101

ตารางที่ 3.การพิจารณาผลลัพธ์การวิเคราะห์ด้วยโครงการโอเพนซอร์ส ด้วยวิธีการวิเคราะห์การ 21 ถดถอยโลจิสติก 102

ตารางที่ 3.22 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิวโค้ด ด้วยวิธีการวิเคราะห์การถด ถดถอยโลจิสติก 102

ตารางที่ 3.23 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยโครงการโอเพนซอร์ส ด้วยวิธีการวิเคราะห์ โครงข่ายประสาทเทียมเพอร์เซปตรอน 103

ตารางที่ 3.24 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิวโค้ด ด้วยวิธีการวิเคราะห์โครงข่าย ประสาทเทียมเพอร์เซปตรอน 104

ตารางที่ 3.เมตริกความสับสนของกา 25รประเมินผลแบบจำลองการทำนายการวิเคราะห์การ ถดถอยโลจิสติก 104

ตารางที่ 3.เมตริกความสับสนของการประเมินผลแบบจำลองการทำนายการวิเคราะ 26ห์โครงข่าย ประสาทเทียมเพอร์เซปตรอน 105

ตารางที่ 4.1 ข้อกำหนดความต้องการที่เป็นหน้าที่หลัก 108

ตารางที่ 4.2 ข้อกำหนดความต้องการที่ไม่เป็นหน้าที่หลัก 109

ตารางที่ 4.คำอธิบายยูสเคสนำเข้าข้อมูลซอร์สโค้ด 3 111

ตารางที่ 4.คำอธิบายยูสเคสระบุเลขรายการร้องขอการน 4ำเข้าซอร์สโค้ด 112

ตารางที่ 4.คำอธิบายระบุวิธีการทำนายผลการรีวิวโค้ด 5 112

ตารางที่ 4.คำอธิบายยูสเคสบันทึกผลลัพธ์การทำนายผลการรีวิวโค้ด 6 114

ตารางที่ 4.รายละเอียดสภาพแวดล้อมที่ใช้ในการพัฒนาเครื่องมือ 7 117

ตารางที่ 4.ภาพรวมการทดสอบเ 8ครื่องมือการทำนายผลการรีวิวโค้ด 120

ตารางที่ 4.1 รายละเอียดการทดสอบที่ 9 121

ตารางที่ 4.10 รายละเอียดการทดสอบที่ 2 122

ตารางที่ 4. รายละเอียดการทดสอบที่ 113 122

ตารางที่ 4. รายละเอียดการทดสอบที่ 124 123

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4. รายละเอียดการทดสอบที่ 135..... 124

ตารางที่ 4. รายละเอียดการทดสอบที่ 146..... 125

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ปัจจุบันการรีวิวโค้ดบนกิตฮับ (Code Review on GitHub) [1] มีการใช้งานอย่างแพร่หลายในอุตสาหกรรมซอฟต์แวร์ เนื่องจากโครงการโอเพนซอร์ส (Open Source Project) [2] เป็นซอฟต์แวร์ที่เปิดโอกาสให้บุคคลทั่วไปมีสิทธิในการเข้าถึงซอร์สโค้ดได้ อนุญาตให้ทำการแก้ไข เปลี่ยนแปลง และพัฒนาซอฟต์แวร์นั้น ๆ ร่วมกัน จึงต้องมีศูนย์กลางในการเก็บรวบรวมข้อมูล ไม่ว่าจะเป็นซอร์สโค้ดหรือเอกสารต่าง ๆ ของโครงการ และศูนย์กลางนั้นจะต้องมีฟังก์ชันการสนับสนุนการรีวิวโค้ดที่มีคุณภาพและได้รับการยอมรับ ซึ่งกิตฮับ (GitHub) เป็นระบบควบคุมเวอร์ชันแบบกระจายศูนย์ (Distributed Version Control System) ที่ได้รับความนิยมและยอมรับจากนักพัฒนาซอฟต์แวร์ อีกทั้งยังมีเมนูที่สนับสนุนการรีวิวโค้ด ที่ช่วยอำนวยความสะดวกในการรีวิวโค้ดได้อย่างรวดเร็ว และสามารถรีวิวโค้ดได้ทันทีที่มีนักพัฒนาซอฟต์แวร์ในทีม หรือผู้ร่วมพัฒนาซอฟต์แวร์นำซอร์สโค้ดเข้าระบบ ด้วยเหตุนี้จึงทำให้การรีวิวโค้ดบนกิตฮับได้รับการนิยมและมีการใช้งานอย่างแพร่หลาย

จากข้อมูลผลการรีวิวโค้ดของโครงการโอเพนซอร์สบนกิตฮับ พบว่ามีซอร์สโค้ดจำนวนมากที่ไม่ผ่านการรีวิวโค้ดซึ่งจะส่งผลกระทบกับนักพัฒนาซอฟต์แวร์ที่ต้องทำการแก้ไขซอร์สโค้ด รวมถึงต้องอัพโหลดซอร์สโค้ดนั้นในการรีวิวอีกครั้ง อีกทั้งยังส่งผลกระทบกับระยะเวลาในการพัฒนาซอฟต์แวร์ ทำให้เกิดความล่าช้าในการพัฒนา โดยหากพิจารณาถึงซอร์สโค้ดที่ไม่ผ่านการรีวิวโค้ดแล้วนั้น พบว่ามีปัจจัยต่าง ๆ มากมายที่ส่งผลกระทบกับผลของการรีวิวโค้ด ไม่ว่าจะเป็น คุณภาพของซอร์สโค้ด ขนาดการเปลี่ยนแปลงซอร์สโค้ด ฟังก์ชันการพัฒนาซอฟต์แวร์ และปัจจัยอื่น ๆ อีกมากมาย

งานวิจัยในปัจจุบัน ได้มีการนำเสนอการหาปัจจัยที่มีอิทธิพลกับผู้ที่ได้รับมอบหมายให้มีหน้าที่ในการรีวิวซอร์สโค้ดของรายการร้องขอการนำเข้าซอร์สโค้ด (Pull Requests) ซึ่งเป็นหนึ่งในกระบวนการบนกิตฮับ [3] และนำเสนอการหาปัจจัยที่ทำให้รายการร้องขอการนำเข้าซอร์สโค้ดได้รับการยอมรับในโครงการโอเพนซอร์ส [4] และปัจจัยที่ส่งผลให้รายการร้องขอการนำเข้าซอร์สโค้ดถูกปฏิเสธในโครงการซอฟต์แวร์ที่มีอัตรายการยอมรับการนำเข้าสูง [5] ถึงแม้ว่าปัจจัยที่ได้จากงานวิจัยในปัจจุบันจะมีประโยชน์อย่างมากและสามารถนำเป็นแนวทางในการพัฒนาซอฟต์แวร์เพื่อให้ผ่านการรีวิวโค้ดบนกิตฮับได้ หากแต่การหาปัจจัยเหล่านี้ไม่มีการทำนายผลการรีวิวจากซอร์สโค้ด เป็นเพียงแต่การหาปัจจัยเพื่อใช้ในการพิจารณา ว่าซอร์สโค้ดที่นำเข้าดังกล่าวมีคุณสมบัติของปัจจัยเหล่านี้หรือไม่

ดังนั้นวิทยานิพนธ์นี้จึงนำเสนอการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ โดยการพิจารณาตัวแปรที่มีผลกับการรีวิวโค้ด หาความสัมพันธ์ของตัวแปรนั้น ๆ รวมถึงรูปแบบของความสัมพันธ์ที่เกิดขึ้นบ่อยด้วยหลักการเหมืองข้อมูลโดยกฎของความสัมพันธ์ (Association Rule)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากนั้นนำตัวแปรที่ได้มาคำนวณเพื่อใช้ในการทำนาย ผ่านหลักการทางสถิติด้วยวิธีการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) และหลักการเรียนรู้ของเครื่อง (Machine Learning) โดยโครงข่ายประสาเทียมแบบเพอร์เซปตรอนหลายชั้น (Multi-layer Perceptron Neural Network) ซึ่งตัวแปรที่มีผลกับการรีวิวโค้ด สามารถนำไปประยุกต์ใช้เพื่อเป็นข้อแนะนำในการเขียนโปรแกรม (Coding Guideline) หรือข้อควรระวังในการเขียนโค้ด อีกทั้งยังสามารถนำเอาสมการการทำนายด้วยวิธีดังกล่าวไปทำนายผลการรีวิวโค้ดของซอร์สโค้ดใด ๆ เพื่อใช้ในการประเมินผลการรีวิวโค้ดล่วงหน้าได้ กล่าวคือก่อนที่นักพัฒนาซอฟต์แวร์จะนำซอร์สโค้ดเข้าสู่ระบบโดยการร้องขอการนำเข้าซอร์สโค้ดนั้น นักพัฒนาซอฟต์แวร์สามารถนำซอร์สโค้ดที่จะนำเข้าระบบมาตรวจสอบกับเครื่องมือที่พัฒนาขึ้นมาจากสมการทำนายผลการรีวิว มาช่วยในการวิเคราะห์รวมถึงนำผลการทำนายมาใช้ในการแก้ไขงานก่อนร้องขอการนำเข้าซอร์สโค้ด อีกทั้งยังเป็นการลดปัญหาการทำงานใหม่ (re-work) ในการร้องขอการนำซอร์สโค้ดเข้าระบบอีกครั้ง หากผลการรีวิวเป็นปฏิเสธ ซึ่งอาจส่งผลกระทบกับแผนการดำเนินงานของโครงการได้

1.2 วัตถุประสงค์

1. เพื่อนำเสนอรูปแบบความสัมพันธ์ที่เกิดขึ้นบ่อยของตัวแปรที่มีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ
2. เพื่อนำเสนอวิธีการสร้างแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ
3. เพื่อสร้างเครื่องมือสนับสนุนวิธีการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

1.3 ขอบเขตการดำเนินงาน

1. การวิเคราะห์รูปแบบความสัมพันธ์ที่เกิดขึ้นบ่อยของข้อมูล ด้วยหลักการเหมืองข้อมูล ใช้กฎของความสัมพันธ์ ด้วยขั้นตอนวิธีเอไพรออไร (Apriori)
2. สร้างแบบจำลองการทำนายจากหลักการทางสถิติ ด้วยการวิเคราะห์การถดถอยโลจิสติกและหลักการเรียนรู้ของเครื่อง ด้วยโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น
3. ผลลัพธ์ของการใช้แบบจำลองการทำนายที่นำเสนอได้ผลลัพธ์ 2 ผลลัพธ์ ได้แก่ ยอมรับการนำเข้าซอร์สโค้ด (Accept) และปฏิเสธการนำเข้าซอร์สโค้ด (Reject)
4. ตัวแปรที่นำมาพิจารณาเป็นตัวแปรที่เกี่ยวข้องกับทางเทคนิคเท่านั้น
5. ตัวแปรที่นำมาพิจารณาจะต้องอยู่ในมาตราส่วนแบบช่วง (Interval) หรือมาตราส่วนแบบอัตราส่วน (Ratio) หากอยู่ในช่วงนามบัญญัติ (Ordinal) จะต้องทำการแปลงค่าเป็นตัวแปรหุ่น (Dummy Variable) ก่อนเท่านั้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

6. ข้อมูลที่ใช้ในการวิเคราะห์ กำหนดข้อมูลของโครงการโอเพนซอร์สที่นำมาพิจารณาเป็นโครงการ Ansible ซึ่งอยู่บนกิตฮับเป็นอย่างน้อย

7. ข้อมูลตัวแปรที่ได้มาจากแหล่งข้อมูล 3 แหล่ง ได้แก่ ข้อมูลบนกิตฮับ ข้อมูลที่ประมวลผลได้ผ่านเครื่องมือราดอน และข้อมูลที่ประมวลผลได้ผ่านเครื่องมือซอร์สมิเตอร์

8. เครื่องมือสนับสนุนงานวิจัยต้องสามารถนำเข้าข้อมูลซอร์สโค้ดได้ ผ่านเมนูการอัพโหลดซอร์สโค้ด สามารถระบุวิธีที่ใช้ในการทำนายผลการรีวิวโค้ดได้ สามารถคำนวณผลการทำนายจากแบบจำลองการทำนายที่ได้ถูกระบุไว้ได้ สามารถแสดงผลการทำนาย ได้แก่ ยอมรับและปฏิเสธการนำเข้าซอร์สโค้ดได้ และสามารถจัดทำรายการเพื่อให้ผู้ใช้งานสามารถดาวน์โหลดผลลัพธ์ของการประมวลผลออกมาได้

9. การประเมินผลแบบจำลองการทำนาย ทำโดยการเปรียบเทียบผลลัพธ์ที่ได้จากแบบจำลองการทำนาย และผลลัพธ์การรีวิวโค้ดจริงบนกิตฮับของระบบการจัดเก็บข้อมูลโอเพนซอร์ส ในโครงการ Ansible และโครงการอื่น ๆ กล่าวคือจะต้องได้รับผลลัพธ์ที่เหมือนกัน จึงจะสรุปได้ว่าแบบจำลองที่ได้พัฒนามีความเที่ยงตรง

1.4 ขั้นตอนการดำเนินงาน

1. กำหนดโครงการโอเพนซอร์สที่นำมาพิจารณา
2. วิเคราะห์โครงการโอเพนซอร์สที่กำหนด
3. กำหนดตัวแปรที่มีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ
4. วิเคราะห์รูปแบบความสัมพันธ์ที่เกิดขึ้นบ่อยของข้อมูลโดยกฎความสัมพันธ์ด้วย ขั้นตอนวิธีเอไพรออไร (Apriori)
5. สร้างแบบจำลองการทำนายด้วยหลักการทางสถิติ โดยการวิเคราะห์ความถดถอยโลจิสติก
6. สร้างแบบจำลองการทำนายด้วยหลักการเรียนรู้ของเครื่อง โดยโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น
7. ประเมินผลการทดลอง
8. ออกแบบและพัฒนาเครื่องมือที่ใช้ในการตรวจสอบผลการรีวิวซอร์สโค้ด
9. ทดสอบและประเมินผลเครื่องมือ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถวิเคราะห์ตัวแปรที่มีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับได้
2. สามารถหารูปแบบความสัมพันธ์ที่เกิดขึ้นบ่อยของชุดข้อมูลได้
3. สามารถทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.6 บทความวิจัยที่ได้รับการตีพิมพ์

งานวิจัยนี้ได้รับการคัดเลือกและตีพิมพ์เป็นบทความวิชาการเรื่อง Finding Impact Factors for Rejection of Pull Requests on GitHub โดยนางสาวปานทิพย์ พู่พุฒ และรองศาสตราจารย์ ดร. พรศิริ หมื่นไชยศรี ในการประชุมวิชาการ “2018 VII International Conference on Network, Communication and Computing (ICNCC 2018)” ระหว่างวันที่ 14-16 ธันวาคม 2561 ณ Horward Civil Service International House เมืองไทเป ประเทศไต้หวัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์ในการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ จึงมีความจำเป็นต้องมีองค์ความรู้เกี่ยวกับโครงการโอเพนซอร์ส และลักษณะการทำงานของโอเพนซอร์ส รวมถึงการพัฒนาซอฟต์แวร์แบบการร้องขอการนำเข้าซอร์สโค้ด (Pull Based Development Model) อีกทั้งยังต้องศึกษาทฤษฎีการรีวิวโค้ดบนกิตฮับ (Code Review on GitHub) และในส่วนของการวิเคราะห์ข้อมูลนั้น มีการประยุกต์ใช้หลักการเหมืองข้อมูล ทั้งกระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่ (Knowledge Discovery in Database : KDD) และกระบวนการหาความสัมพันธ์ (Association Algorithm) ซึ่งใช้ในการจัดการข้อมูล รวมถึงวิเคราะห์ข้อมูลเบื้องต้น ซึ่งนอกจากการจัดการข้อมูลโดยใช้หลักการเหมืองข้อมูลแล้วนั้น งานวิจัยนี้ยังมีการนำหลักการทางด้านสถิติมาใช้ในการวิเคราะห์ข้อมูล (Statistical Analysis) รวมถึงการสร้างแบบจำลองการทำนายโดยการวิเคราะห์ความถดถอยโลจิสติก (Logistic Regression Analysis) และมีการนำหลักการเรียนรู้ด้วยเครื่อง ซึ่งใช้ในการสร้างแบบจำลองการทำนายเช่นกัน โดยทฤษฎีโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น (Multi-layer Perceptron Neural Network) รวมถึงเครื่องมือต่าง ๆ ที่ใช้ในการคำนวณค่าตัวแปร และวิเคราะห์ข้อมูล โดยเครื่องมือที่ใช้ในการคำนวณค่าตัวแปรประกอบด้วย เครื่องมือราดอน (Radon Tool) เครื่องมือซอร์สมิเตอร์ (SourceMetter Tool) และเครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลประกอบด้วย เครื่องมือสำเร็จรูปทางสถิติเอสพีเอสเอส (IBM SPSS Statistics 22 Tool) และเครื่องมือเวก้า (Weka Tool) ซึ่งแต่ละหัวข้อมีรายละเอียดดังนี้

Chulalongkorn University

2.1.1 โอเพนซอร์ส

โครงการโอเพนซอร์ส (Open Source Project) [2] คือซอฟต์แวร์ที่เปิดให้บุคคลทั่วไปมีสิทธิในการเข้าถึงซอร์สโค้ดได้ อนุญาตให้ทำการแก้ไข เปลี่ยนแปลง และพัฒนาซอฟต์แวร์นั้น ๆ ร่วมกัน ซอฟต์แวร์โอเพนซอร์สมีโครงสร้างหลักในการพัฒนาขนาดใหญ่ นักพัฒนาจำนวนมากเข้าใช้งานซอฟต์แวร์โอเพนซอร์สในเวลาพร้อม ๆ กัน ซึ่งนักพัฒนาเหล่านั้นไม่จำเป็นต้องเป็นนักพัฒนาซอฟต์แวร์ที่อยู่ในสถานที่เดียวกัน หรือทำงานในองค์กรเดียวกันก็ได้ เนื่องจากโครงการโอเพนซอร์สรองรับการทำงานของนักพัฒนาซอฟต์แวร์ที่อยู่ต่างสถานที่กัน โดยอาศัยการสื่อสารและติดต่อกันทางอินเตอร์เน็ต หรือการทำงานผ่านระบบควบคุมเวอร์ชันแบบกระจายศูนย์

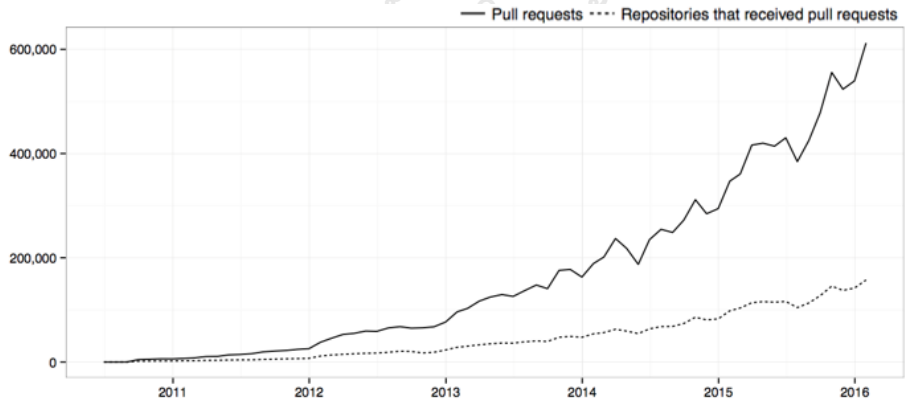
รูปแบบการทำงานของซอฟต์แวร์โอเพนซอร์สเป็นแบบไม่ประสานเวลา (Asynchronous) และกระจาย (Distributed) ซึ่งสนับสนุนการทำงานของนักพัฒนาซอฟต์แวร์ที่อยู่ต่างสถานที่กันได้

อย่างดี อีกทั้งโครงการโอเพนซอร์ส ยังเป็นโครงการที่ไม่มีแผนการพัฒนาที่แน่นอน หากแต่ลักษณะโครงการเป็นการพัฒนาอย่างต่อเนื่อง การมีซอร์สโค้ดหรือฟังก์ชันต่าง ๆ ที่เพิ่มขึ้น เกิดจากการที่นักพัฒนาซอฟต์แวร์ภายใต้โครงการโอเพนซอร์สเหล่านั้น มีการพัฒนาซอฟต์แวร์เพิ่มเข้าไปจากส่วนเดิมที่มีอยู่ โดยในโครงการจะมีทั้งนักพัฒนาหลัก ซึ่งเป็นผู้ทำหน้าที่ในการประกาศเวอร์ชันล่าสุด รวมถึงให้แนวทางในการพัฒนาซอฟต์แวร์นั้น ๆ และมีนักพัฒนาบุคคลอื่น ๆ ที่สามารถเข้าถึงซอร์สโค้ดได้ อีกทั้งยังสามารถนำซอร์สโค้ดเหล่านั้นมาพัฒนาต่อ และเสนอการเปลี่ยนแปลงใด ๆ ไปสู่นักพัฒนาหลักได้ [6]

โครงการโอเพนซอร์ส ไม่ใช่การมอบหมายงานที่ให้นักพัฒนาที่เกี่ยวข้องและสนใจในโครงการโอเพนซอร์สนั้น ๆ หากแต่เป็นการที่นักพัฒนาเป็นผู้เลือกที่จะทำงาน ฟังก์ชัน หรือซอร์สโค้ดไหนด้วยตัวเอง บางโครงการโอเพนซอร์สจะระบุรายงานที่ต้องส่งมอบ (To Do List) ในแต่ละช่วงเวลาเอาไว้ นักพัฒนาซอฟต์แวร์จะเป็นผู้เลือกงานเหล่านั้นมาทำด้วยตัวเอง ซึ่งกระบวนการภายใต้โครงการนั้นเป็นไปอย่างเปิดเผยและเป็นระบบเบ็ดเสร็จ (Self-Organization System) สามารถบริหารจัดการเรื่องต่าง ๆ ได้ภายในโครงการเดียว

2.1.2 การพัฒนาซอฟต์แวร์แบบการร้องขอการนำเข้าซอร์สโค้ด

การพัฒนาซอฟต์แวร์แบบการร้องขอการนำเข้าซอร์สโค้ด (Pull Bases Development Model) เป็นวิธีการพัฒนาซอฟต์แวร์ที่ใช้ในโครงการโอเพนซอร์สรูปแบบหนึ่ง [7] ซึ่งปัจจุบันเป็นที่นิยมอย่างมากในกิตฮับ โดยมีสถิติยืนยันชัดเชนถึงการเติบโตและเป็นที่ยอมรับอย่างรวดเร็วของวิธีการนี้ ดังรูปที่ 2.1



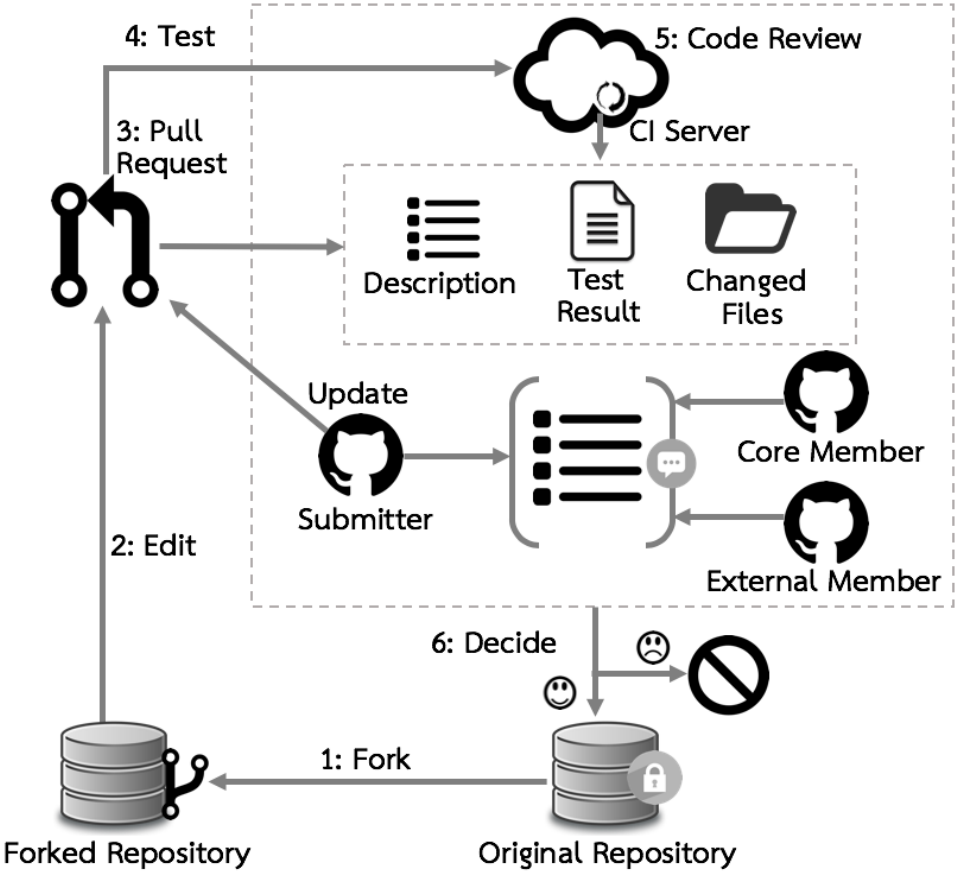
รูปที่ 2.1 กราฟแสดงอัตรายการเติบโตของการใช้งานการร้องขอการนำเข้าซอร์สโค้ดบนกิตฮับ [7]

โดยการพัฒนาซอฟต์แวร์แบบการร้องขอการนำเข้าซอร์สโค้ดเป็นระบบควบคุมเวอร์ชันแบบกระจายศูนย์ [8] และบริการพื้นที่บนอินเตอร์เน็ต โดยครอบคลุมทั้งการแบ่งการควบคุมเวอร์ชันและ

การจัดการซอร์สโค้ด เกินกว่าครึ่งของโครงการโอเพนซอร์สในกิตยอมรับวิธีการแบบการร้องขอการนำเข้าซอร์สโค้ด

รายการร้องขอการนำเข้าซอร์สโค้ดระบุรายละเอียดมากมายเกี่ยวกับรายการร้องขอการนำเข้า ไม่ว่าจะเป็นหมายเลขรายการร้องขอการนำเข้าซอร์สโค้ด ผู้นำเข้า ผู้ดูแลระบบ ข้อมูลต่าง ๆ ของรายการร้องขอการนำเข้านั้น ๆ รวมถึงแผ่นป้ายที่ใช้ในการแสดงข้อมูลของรายการร้องขอการนำเข้าซอร์สโค้ด

รูปแบบการพัฒนาซอฟต์แวร์แบบการร้องขอการนำเข้าซอร์สโค้ด อนุญาตให้ผู้มีส่วนร่วมหรือผู้ที่สนใจ มีสิทธิในการคัดลอกโค้ด (Fork หรือ Clone) ของโครงการโอเพนซอร์สนั้น ๆ และมีสิทธิในการเปลี่ยนแปลง ปรับปรุง และพัฒนา ไม่ว่าจะเป็นการแก้ไขข้อบกพร่องทางซอฟต์แวร์ หรือการเพิ่มฟังก์ชันการทำงานใหม่ ๆ เข้าไป โดยไม่ต้องร้องขอถึงสิทธิในการพัฒนาแต่อย่างใด หากแต่เมื่อโค้ดที่ได้ทำการเปลี่ยนแปลงมีความพร้อมสำหรับการนำกลับเข้าสู่ระบบจัดเก็บ ที่เป็นพื้นที่ในการจัดเก็บซอร์สโค้ดส่วนกลาง และควบคุมเวอร์ชันที่ใช้ร่วมกันทุกคนแล้วนั้น [9] ผู้เขียนโค้ดหรือผู้ทำการแก้ไขโค้ด ๆ จะต้องนำโค้ดที่ผ่านการแก้ไขกลับเข้าสู่ระบบจัดเก็บและควบคุมเวอร์ชันโดยทำการร้องขอการนำเข้า โดยขั้นตอนของรูปแบบการพัฒนาซอฟต์แวร์แบบการร้องขอการนำเข้าซอร์สโค้ดมีดังรูปที่ 2.2



รูปที่ 2.2 ขั้นตอนการทำงานของการร้องขอการนำเข้าบนกิตฮับ [10]

ขั้นตอนที่ 1: Fork คือขั้นตอนการคัดลอกซอร์สโค้ดที่ถูกจัดเก็บอยู่ในพื้นที่การเก็บข้อมูลหลัก (Original Repository) ลงมายังพื้นที่การจัดเก็บส่วนตัว หรือเครื่องคอมพิวเตอร์ที่ใช้ในการพัฒนาซอฟต์แวร์ โดยเราเรียกซอร์สโค้ดที่อยู่ในพื้นที่การเก็บข้อมูลหลักว่าอยู่ในสาขาหลัก (Main Branch) และซอร์สโค้ดที่อยู่ในคอมพิวเตอร์ส่วนตัวของนักพัฒนาซอฟต์แวร์ว่าซอร์สโค้ดที่อยู่ในสาขาท้องถิ่น (Local Branch) [9] หรือสาขาที่คัดลอกซอร์สโค้ดไปยังเครื่องคอมพิวเตอร์ของนักพัฒนาซอฟต์แวร์ (Fork Repository)

ขั้นตอนที่ 2: Edit คือการแก้ไขซอร์สโค้ด โดยการแก้ไขจะเป็นการแก้ไขภายในสาขาของนักพัฒนาซอฟต์แวร์ เรียกอีกชื่อว่าสาขาท้องถิ่น (Local branch) ซึ่งจะไม่กระทบสาขาหลัก นักพัฒนาซอฟต์แวร์สามารถแก้ไขซอร์สโค้ดเพื่อแก้ไขข้อบกพร่องทางซอฟต์แวร์ เพิ่มเติมซอร์สโค้ดเพื่อเพิ่มฟังก์ชันการทำงานให้กับซอฟต์แวร์ และลบซอร์สโค้ดเพื่อลบฟังก์ชันการทำงานหรือแก้ไขข้อบกพร่องที่เกิดจากซอร์สโค้ดได้

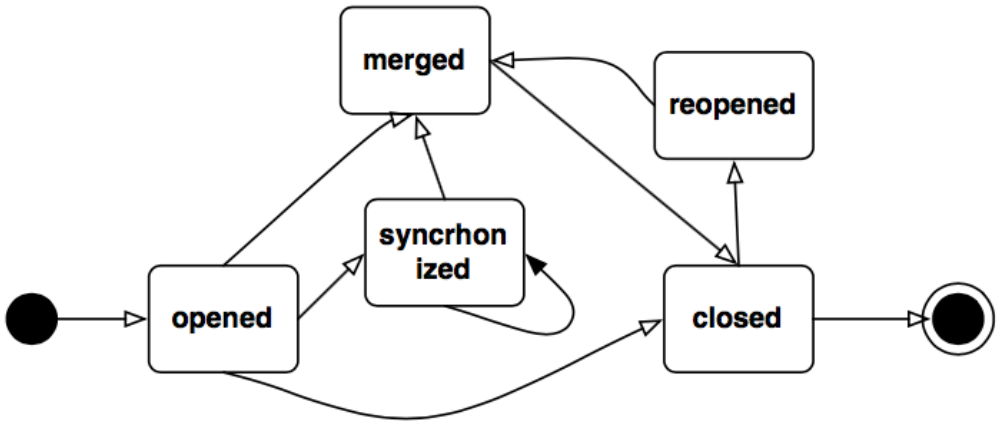
ขั้นตอนที่ 3: Pull Request คือการร้องขอนำซอร์สโค้ดที่ได้แก้ไขกลับเข้าสู่สาขาหลักอีกครั้งเมื่อนักพัฒนาซอฟต์แวร์ทำการแก้ไขซอร์สโค้ดเสร็จเรียบร้อยแล้ว โดยขั้นตอนการร้องขอการนำเข้าทำได้โดย นักพัฒนาซอฟต์แวร์ หรือผู้นำซอร์สโค้ดเข้าระบบ (Submitter) ทำการอัพโหลดไฟล์ซอร์สโค้ดที่แก้ไขแล้ว พร้อมคำอธิบายการแก้ไข เอกสารข้อบังคับต่างๆ รวมถึงผลการทดสอบรายการแก้ไขเข้าสู่ระบบ เพื่ออธิบายถึงวัตถุประสงค์ในการร้องขอนำซอร์สโค้ดเข้าสาขาหลัก ให้ผู้ดูแลระบบหรือนักพัฒนาซอฟต์แวร์คนอื่น ๆ รับรู้

ขั้นตอนที่ 4: Test คือการทดสอบระบบโดยผู้ตรวจสอบซอร์สโค้ดหรือผู้รีวิว เพื่อจะได้มั่นใจว่า การทำการร้องขอดังกล่าวจะไม่ส่งผลกระทบกับระบบ โดยการทดสอบระบบนั้นมีทั้งแบบทำด้วยมือ (Manual Test) และอัตโนมัติ (Automation Test) ขึ้นอยู่กับโครงการโอเพนซอร์สนั้น ๆ จะระบุ

ขั้นตอนที่ 5: Code Review คือการรีวิวโค้ด โดยนักพัฒนาซอฟต์แวร์หรือ ผู้รีวิว ทั้งนักพัฒนาซอฟต์แวร์หลัก (Core Developer) และนักพัฒนาซอฟต์แวร์ภายนอก (External Developer) ทำโดยการอ่านซอร์สโค้ดจากไฟล์ซอร์สโค้ดที่นักพัฒนาได้นำเข้าระบบ รวมถึงตรวจสอบคำอธิบายการแก้ไข รายละเอียด และผลการทดสอบรายการการแก้ไข ซึ่งหากซอร์สโค้ดดังกล่าวไม่เป็นไปตามคาดหวัง หรือไม่ตรงตามที่ระบุไว้ในมาตรฐาน ผู้รีวิวสามารถส่งกลับไปยังนักพัฒนาผู้นำเข้าซอร์สโค้ดอีกครั้งเพื่อทำการแก้ไขตามคำแนะนำ และนำโค้ดกลับเข้ามารีวิวอีกครั้ง

ขั้นตอนที่ 6: Decide คือการตัดสินใจ สามารถทำได้โดยผู้จัดการในกลุ่มนักพัฒนาหลัก โดยมีผลการนำเข้าคือ การอนุมัติให้นำซอร์สโค้ดเข้าระบบ (Approve) และการปฏิเสธการนำเข้าซอร์สโค้ด (Reject) โดยสถานะต่าง ๆ ของการร้องขอการนำเข้าซอร์สโค้ดบนกิตฮับ ดังรูปที่ 2.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 2.3 สถานะของการร้องขอการนำเข้าซอร์สโค้ดบนกิตฮับ [9]

จากรูปที่ 2.3 แผนภาพแสดงสถานะ (State Diagram) พบว่าสถานะของการร้องขอการนำเข้าซอร์สโค้ดบนกิตฮับ สามารถแบ่งได้เป็น 5 สถานะ โดยสถานะเปิดการร้องขอ (Opened) เกิดขึ้นทันที หลังจากนักพัฒนาผู้นำเข้าซอร์สโค้ดเปิดการร้องขอการนำเข้าซอร์สโค้ด โดยหลักจากนั้นการร้องขอจะเปลี่ยนสถานะไปเป็นประสานงาน (Synchronized) เมื่อผู้รีวิวทำการรีวิวซอร์สโค้ด โดยหากซอร์สโค้ดมีความพร้อมในการนำเข้าสาขาหลักแล้วนั้น การร้องขอจะเปลี่ยนสถานะเป็น รวมตัว (Merged) โดยหลังจากรวมตัวสำเร็จแล้วสถานะจะเปลี่ยนเป็น ปิด (Closed) และหากจะนำการร้องขอนั้นกลับเข้ามาแก้ไขข้อมูลบางอย่าง ที่เล็กน้อย (Nitch Data) นักพัฒนาผู้นำเข้าซอร์สโค้ดสามารถร้องขอการนำเข้าได้อีกครั้ง โดยสถานะของการร้องขอจะเปลี่ยนเป็นร้องขอใหม่ (Reopened)

2.1.3 การรีวิวโค้ดบนกิตฮับ

การรีวิวโค้ดบนกิตฮับ (Code Review on GitHub) เป็นหนึ่งในขั้นตอนการร้องขอการนำเข้าซอร์สโค้ด กล่าวคือซอร์สโค้ดที่จะนำเข้าสู่สาขาหลักของโครงการโอเพนซอร์สได้นั้น จะต้องผ่านการรีวิวโค้ดก่อน ซึ่งผลของการรีวิว คือผ่านการรีวิว (Accepted) และไม่ผ่านการรีวิว (Rejected) โดยบุคคลผู้มีส่วนเกี่ยวข้องของการรีวิวโค้ดบนกิตฮับ ซึ่งประกอบด้วยสามกลุ่มคนด้วยกัน ได้แก่ นักพัฒนาซอฟต์แวร์ (ผู้ร้องขอการนำเข้าซอร์สโค้ด) นักพัฒนาซอฟต์แวร์คนอื่น ๆ และผู้รีวิว โดยขั้นตอนการรีวิวโค้ดบนกิตฮับสามารถแบ่งได้เป็น 5 ขั้นตอนหลัก ดังรูป 2.4 และรูป 2.5 ซึ่งมีรายละเอียดดังนี้

ขั้นตอนที่ 1 นักพัฒนาซอฟต์แวร์ทำการร้องขอการนำเข้าซอร์สโค้ด

ขั้นตอนที่ 2 นักพัฒนาซอฟต์แวร์ดำเนินการตามกระบวนการต่าง ๆ ตามที่โครงการโอเพนซอร์สได้กำหนดไว้ เช่น การระบุรายละเอียดของรายการร้องขอการนำเข้าซอร์สโค้ด เป็นต้น

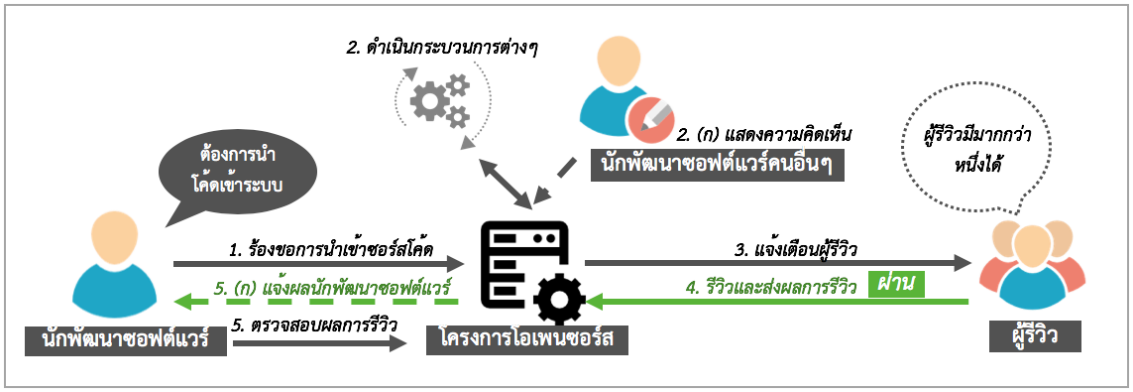
ขั้นตอนที่ 2-ก นักพัฒนาซอฟต์แวร์คนอื่น ๆ ในโครงการ สามารถแสดงความคิดเห็นเกี่ยวกับรายการร้องขอการนำเข้าซอร์สโค้ดได้ โดยกระบวนการนี้ไม่ได้บังคับ หากแต่สามารถแสดงความคิดเห็น หรือข้อแนะนำ เพื่อใช้ในการพัฒนาและแบ่งปันความรู้แก่ผู้ร่วมโครงการได้

ขั้นตอนที่ 3 ระบบจะทำการแจ้งเตือนผู้รีวิว ว่ามีรายการร้องขอการนำเข้าซอร์สโค้ดรายการใหม่เข้ามา โดยผู้รีวิวสามารถมีมากกว่าหนึ่งคนได้

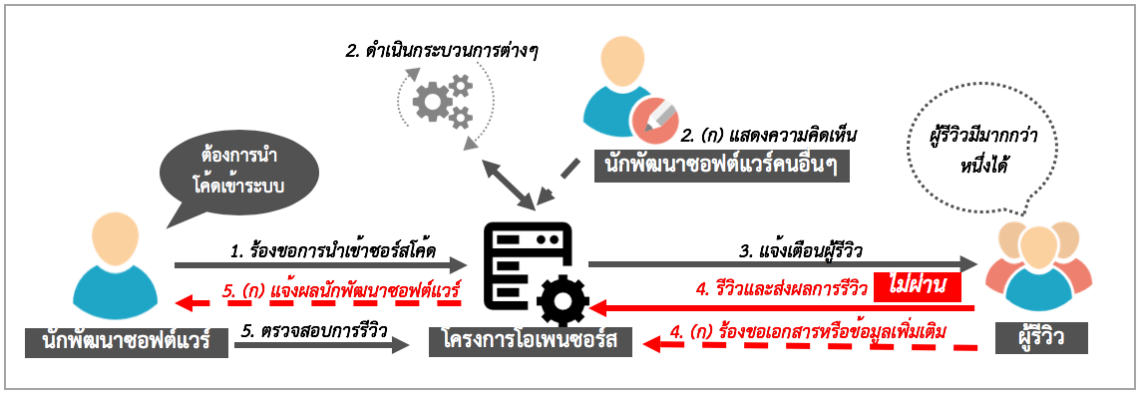
ขั้นตอนที่ 4 ผู้รีวิวทำการรีวิวโค้ด และให้ผลการรีวิว ว่ารายการร้องขอการนำเข้าซอร์สโค้ดนั้นผ่านการรีวิวหรือไม่ผ่านการรีวิว

ขั้นตอนที่ 5-ก ระบบจะทำการแจ้งผลการรีวิวแก่นักพัฒนาซอฟต์แวร์ ซึ่งการแจ้งผลจะเป็นไปตามการตั้งค่าระบบของนักพัฒนาซอฟต์แวร์นั้น ๆ ว่าเปิดให้มีการแจ้งผลการรีวิวหรือไม่

ขั้นตอนที่ 5 นักพัฒนาซอฟต์แวร์ทำการตรวจสอบผลการรีวิว



รูปที่ 2.4 ขั้นตอนการรีวิวโค้ดบนกิตฮับ ผลการรีวิวคือผ่านการรีวิว [1]



รูปที่ 2.5 ขั้นตอนการรีวิวโค้ดบนกิตฮับ ผลการรีวิวคือไม่ผ่านการรีวิว [1]

2.1.4 กระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่

กระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่ (Knowledge Discovery in Database : KDD) มีวัตถุประสงค์ในการจัดการข้อมูลขนาดใหญ่ เนื่องด้วยตัวแปรที่ได้มาจากกิตฮับ มีปริมาณมาก

หรือเป็นข้อมูลขนาดใหญ่ (Big data) งานวิจัยนี้จึงต้องมีการนำหลักการเหมืองข้อมูลมาใช้ ซึ่งหลักการเหมืองข้อมูลเป็นกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ซึ่งการประยุกต์ใช้เหมืองข้อมูลมีประโยชน์หลากหลาย ไม่ว่าจะเป็นการประยุกต์ใช้กับกระบวนการคาดเดา (Prediction Method) หรือกระบวนการอธิบาย (Description Methods) ซึ่งหลักการเหมืองข้อมูล หรือกระบวนการที่นำมาประยุกต์ใช้ในงานวิจัยนี้คือ การประยุกต์ใช้กระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่ (Knowledge Discovery in Database : KDD) [11]

กระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่ เป็นกระบวนการค้นหาความรู้ที่เป็นประโยชน์ในกลุ่มข้อมูล หรือการค้นหารูปแบบ (Pattern) จากปริมาณข้อมูลจำนวนมาก ซึ่งเป็นเทคนิคการจัดการข้อมูลขนาดใหญ่ที่มีการใช้อย่างแพร่หลาย ซึ่งกระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่ มีรายละเอียดและขั้นตอนดังนี้

ขั้นตอนที่ 1 ทำความเข้าใจโดเมนการประยุกต์ (Application Domain) เป็นการเตรียมความพร้อมความเข้าใจ และเข้าใจถึงโดเมนปัญหาเพื่อที่จะได้แก้ไขปัญหาให้ถูกจุด หรือสามารถวิเคราะห์ข้อมูลได้ถูกต้องตามข้อกำหนดความต้องการ รวมถึงการเตรียมความพร้อมของสภาพแวดล้อมที่จะใช้ในการดำเนินกระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่ เพื่อเตรียมความพร้อมในการวิเคราะห์ข้อมูล อีกทั้งยังมีการตั้งเป้าหมาย การตั้งเกณฑ์วัดความสำเร็จ และการวางแผนในการทำเหมืองข้อมูลอีกด้วย

ขั้นตอนที่ 2 รวบรวมข้อมูลและการคัดเลือกข้อมูล (Data Selection) ที่จะทำการวิเคราะห์จากแหล่งข้อมูลต่าง ๆ กำหนดคุณสมบัติของข้อมูลที่ได้รวบรวมมา รวมถึงทำการตรวจสอบข้อมูลขั้นต้น โดยการเลือกข้อมูลจะเลือกเพียงข้อมูลที่เกี่ยวข้องและสนใจเท่านั้น

ขั้นตอนที่ 3 การประมวลผลข้อมูลก่อน (Data Preprocessing) เป็นขั้นตอนสำคัญก่อนการนำข้อมูลไปใช้ในการวิเคราะห์ เนื่องด้วยมีโอกาสสูงที่ข้อมูลที่ใช้งานจริงจะมีความผิดปกติ ไม่ว่าจะเป็นข้อมูลไม่สมบูรณ์ (Incomplete) หรือข้อมูลที่มีข้อมูลรบกวน (Noise) ปริมาณมาก อีกทั้งข้อมูลบางประเภทยังต้องการการวิเคราะห์ก่อนนำไปประมวลผล ซึ่งในขั้นตอนการเตรียมข้อมูลนี้ คือ การทำความสะอาดข้อมูล (Data Cleaning) ซึ่ง ทำโดยการตัดทิ้งรายการซ้ำซ้อน (Duplicated records) ซึ่งเป็นการตรวจหารายการที่ซ้ำซ้อนกันและทำการให้เหลือรายการเพียงรายการเดียวเท่านั้น และการจัดการกับข้อมูลรบกวน (Noise) ซึ่งทำได้โดยการหาข้อมูลที่มีความผิดปกติ (Outlier) ของแต่ละตัวแปรที่ใช้ในการวิเคราะห์ข้อมูล จากนั้นทำการลบค่าความผิดปกตินั้นออกจากการวิเคราะห์ข้อมูล หรือใช้วิธีการจัดการกับข้อมูลรบกวนด้วยวิธีอื่น ๆ ไม่ว่าจะเป็นวิธีบินนิง (Binning Methods) การจัดกลุ่ม (Clustering) วิธีความถดถอย (Regression) เป็นต้น

การจัดการกับข้อมูลด้วยวิธีการบินนิง เป็นการจัดการข้อมูลที่ได้รับความนิยมในปัจจุบัน โดยวิธีบินนิ่งเป็นการเรียงลำดับข้อมูล และทำการแบ่งส่วน (Partition) ข้อมูลออกเป็นส่วน ๆ โดยการทำบินนิงมีหลายวิธี ไม่ว่าจะการแบ่งช่วงเป็นช่วงที่เท่ากัน โดยผู้วิเคราะห์จะเป็นผู้กำหนดจำนวนช่วงที่ต้องการแบ่งส่วน โดยใช้สูตรค่าสูงสุด (Max) ลบด้วยค่าต่ำสุด (Min) และหารด้วยจำนวนช่วงที่ต้องการแบ่งหรือการแบ่งเป็นความลึกที่เท่ากันจะใช้ความถี่ (frequency) แบ่งข้อมูลออกเป็นช่วงโดยแต่ละช่วงจะมีจำนวนตัวอย่างที่เท่ากันหรือใกล้เคียงกัน

ขั้นตอนที่ 4 การแปลงข้อมูล (Data Transformation) คือการแปลงค่าข้อมูลให้อยู่ในช่วงใหม่ เพื่อให้ได้ข้อมูลที่สามารถนำไปวิเคราะห์ได้แม่นยำขึ้น โดยการแปลงข้อมูลที่พบบ่อยในการทำเหมืองข้อมูล ได้แก่ นอร์มอลไลเซชัน (Normalization) หรือการแปลงประเภทข้อมูล (Data Type Conversion) ซึ่งเป็นการแปลงข้อมูลจากประเภทมาตราส่วนแบบนามบัญญัติ (Nominal Data) หรือเรียงลำดับ (Ordinal Data) ให้เป็นมาตราส่วนแบบช่วง (Interval Data) หรือมาตราส่วนแบบอัตราส่วน (Ratio Data) เนื่องด้วยการวิเคราะห์ด้วยหลักการเหมืองข้อมูลบางวิธี ไม่รองรับข้อมูลที่อยู่ในรูปแบบมาตราส่วนแบบนามบัญญัติ หรือมาตราส่วนแบบเรียงลำดับ

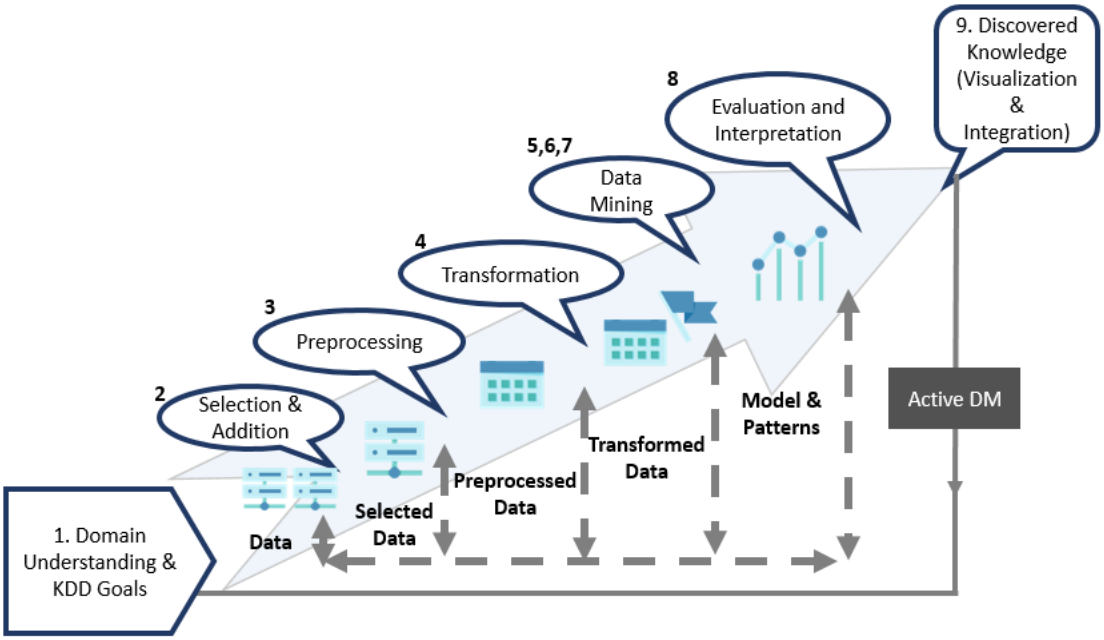
ขั้นตอนที่ 5 เลือกประเภทเหมืองข้อมูลที่เหมาะสม เนื่องด้วยการทำเหมืองข้อมูลมีหลายประเภทด้วยกัน ซึ่งแต่ละประเภทจะมีเป้าหมายและวัตถุประสงค์ในการวิเคราะห์ข้อมูลที่แตกต่างกันไป ขั้นตอนนี้เป็นขั้นตอนที่ใช้ในการเลือกประเภทเหมืองข้อมูลที่เหมาะสมจากเป้าหมายที่ได้ตั้งไว้ ซึ่งการทำเหมืองข้อมูลแบ่งเป็นหลายประเภทด้วยกัน ไม่ว่าจะเป็นประเภทการจำแนกประเภทของข้อมูล (Classification) การทำนายข้อมูล (Prediction) การจัดกลุ่มของข้อมูล (Clustering) และการหาความสัมพันธ์ของข้อมูล (Association)

ขั้นตอนที่ 6 เลือกขั้นตอนวิธีการทำเหมืองข้อมูล (Data Mining Algorithm) คัดเลือกขั้นตอนวิธีการทำเหมืองข้อมูล ให้สอดคล้องกับเป้าหมายการทำและประเภทของเหมืองข้อมูลที่เลือกไว้ตามความเหมาะสม โดยในแต่ละประเภทของเหมืองข้อมูลนั้น จะประกอบไปด้วยขั้นตอนวิธีการต่าง ๆ กันออกไป ตามแต่วัตถุประสงค์และข้อดีข้อเสียของแต่ละขั้นตอนวิธี

ขั้นตอนที่ 7 วิเคราะห์ข้อมูล หรือนำขั้นตอนวิธีการทำเหมืองข้อมูลไปใช้ โดยทำการวิเคราะห์ข้อมูลจากขั้นตอนวิธีการทำเหมืองข้อมูลที่ได้คัดเลือก ด้วยจำนวนการวิเคราะห์อาจจะมากหรือน้อยแล้วแต่เป้าหมายในผลลัพธ์ที่ได้ตั้งไว้ของแต่ละการิเคราะห์ ซึ่งการวิเคราะห์ข้อมูลอาจทำเป็นจำนวนหลายครั้งจนกว่าจะได้รับผลลัพธ์ออกมา รวมถึงการปรับรูปแบบการวิเคราะห์และการตั้งค่าต่างตัวแปรการวิเคราะห์ต่าง ๆ เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด

ขั้นตอนที่ 8 ประเมินผลการวิเคราะห์ที่ได้ โดยการดูว่าตรงกับเป้าหมาย และเกณฑ์การวัดที่ได้ตั้งไว้ก่อนการทำเหมืองข้อมูลหรือไม่ หากไม่ตรงตามเป้าหมายอาจทำการวิเคราะห์ใหม่อีกครั้ง

ขั้นตอนที่ 9 นำหลักการความรู้หรือผลการวิเคราะห์ที่ได้ไปใช้ โดยรายละเอียดขั้นตอนการประยุกต์ใช้กระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่ แสดงดังรูปที่ 2.6



รูปที่ 2.6 ขั้นตอนการประยุกต์ใช้กระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่ [11]

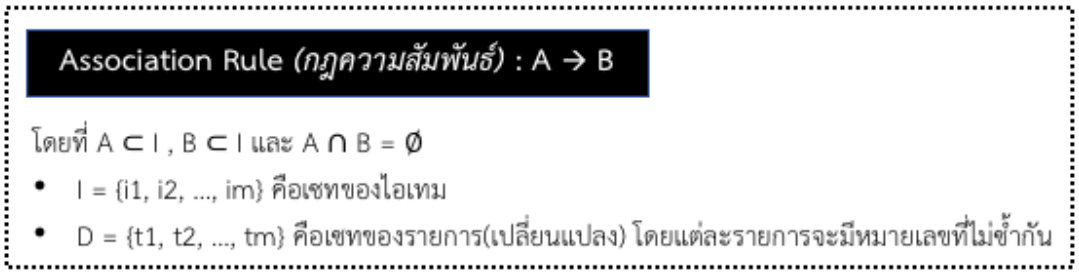
2.1.5 กระบวนการหาความสัมพันธ์

กระบวนการหาความสัมพันธ์ (Association Algorithm) [12] เป็นอัลกอริทึมการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ เพื่อทำไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ ตัวอย่างเช่น การทำเหมืองข้อมูลเพื่อหาความสัมพันธ์ของข้อมูล มักถูกเอามาใช้ในเรื่องของการค้าปลีกการวิเคราะห์ตะกร้าตลาด (Market Basket Analysis) เพื่อศึกษาพฤติกรรมการซื้อสินค้าของลูกค้า เป็นต้น ซึ่งรายละเอียดของกระบวนการหาความสัมพันธ์ สามารถทำได้โดยการหากฎความสัมพันธ์ (Association Rules) เนื่องด้วยตัวแปรที่ได้มาจากกิตฮับ มีปริมาณมากหรือเป็นข้อมูลขนาดใหญ่ (Big data) งานวิจัยนี้จึงต้องมีการนำหลักการเหมืองข้อมูลมาใช้ ซึ่งหลักการเหมืองข้อมูลคือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยประโยชน์ของการทำเหมืองข้อมูลมีหลากหลาย ไม่ว่าจะเป็นกระบวนการคาดเดา (Prediction Method) หรือกระบวนการอธิบาย (Description Methods) [13] ซึ่งกระบวนการหรืออัลกอริทึมที่นำมาใช้ในงานวิจัยนี้คือ กระบวนการหาความสัมพันธ์ (Association Algorithm) เป็นอัลกอริทึมการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่เพื่อทำไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่างๆ ตัวอย่างเช่น การทำเหมืองข้อมูลเพื่อหาความสัมพันธ์ของข้อมูล มักถูก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

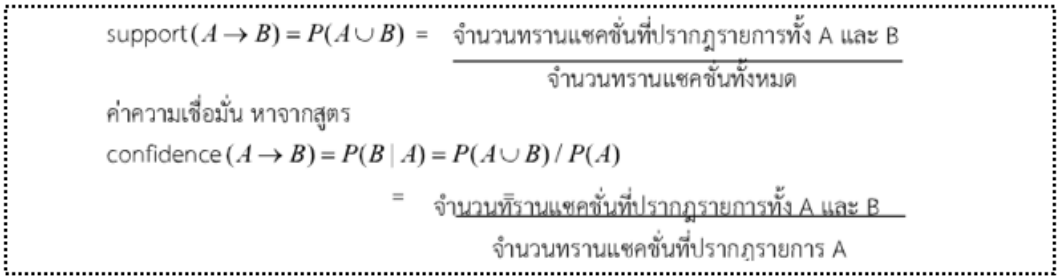
นำมาใช้ในเรื่องของการค้าปลีกการวิเคราะห์ตะกร้าตลาด (Market Basket Analysis) เพื่อศึกษาพฤติกรรมการซื้อสินค้าของลูกค้า เป็นต้น

กฎของความสัมพันธ์ เป็นกระบวนการหนึ่งในการทำเหมืองข้อมูล โดยจะใช้กฎของความสัมพันธ์ในการหาความสัมพันธ์ของข้อมูลสองชุดหรือมากกว่าสองชุดขึ้นไปภายในกลุ่มข้อมูลที่มีขนาดใหญ่ โดยมีรายละเอียดดังรูปที่ 2.7



รูปที่ 2.7 การหากฎความสัมพันธ์ [12]

กฎความสัมพันธ์ ที่มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนที่กำหนด และค่าความเชื่อมั่นมีมากกว่าหรือเท่ากับค่าความเชื่อมั่นที่กำหนด (Minimum Confidence Threshold) โดยค่าสนับสนุนและค่าความเชื่อมั่น สามารถหาได้จากสูตร ดังรูปที่ 2.8



รูปที่ 2.8 แสดงการหาค่าสนับสนุนที่กำหนดและค่าความเชื่อมั่น [12]

ขั้นตอนการหากฎความสัมพันธ์ในการหากฎความสัมพันธ์มีอยู่หลากหลายวิธี หากแต่วิธีที่เป็นที่นิยมในการหาและจะนำมาใช้ในงานวิจัยนี้คือ เอไพรออไรซึ่งเป็นอัลกอริทึมพื้นฐานที่นิยมนำมาใช้ในการหาไอเทมที่เกิดขึ้นบ่อย (Frequent Itemset) โดยประกอบด้วย 2 ขั้นตอนใหญ่ๆ มีรายละเอียดดังนี้

ขั้นตอนที่1 การหาไอเทมที่เกิดขึ้นบ่อย เป็นการหารูปแบบของข้อมูลที่เกิดขึ้นร่วมกันบ่อย ๆ ในฐานข้อมูล หรือมากกว่าค่าความเชื่อมั่นที่กำหนด โดยขั้นตอนนี้แบ่งเป็น 2 ขั้นตอนย่อย ดังนี้

1. การสร้างรูปแบบของไอเทมเซท (Itemset) จะใช้รูปแบบที่มีค่ามากกว่าค่าความเชื่อมั่นที่กำหนด มาทำการสร้างรูปแบบของไอเทมเซท ที่มีขนาดยาวมากขึ้นทีละหนึ่งขั้นไปเรื่อย ๆ
2. การนับค่าสนับสนุน โดยที่ค่าสนับสนุนคือจำนวนเปอร์เซ็นต์ที่พบไอเทมเซท ในฐานข้อมูล

ขั้นตอนที่ 2 การสร้างกฎความสัมพันธ์ ถ้าเอแล้วบี (A --> B) โดยพิจารณาจากเซทของไอเทมและเซทของรายการเปลี่ยนแปลง

2.1.6 การใช้หลักการทางสถิติในการวิเคราะห์ข้อมูล

สถิติ (Statistic) [14] หมายถึง กระบวนการในการเก็บรวบรวม วิเคราะห์ การแปรและการสรุปข้อมูล หรือตัวเลขที่ใช้บรรยายเหตุการณ์หรือข้อเท็จจริง (Facts) ของเรื่องต่าง ๆ ในสิ่งที่ต้องการศึกษา

ตัวแปร (Variable) หมายถึง คุณลักษณะหรือคุณสมบัติใด ๆ ที่มีร่วมกันในทุก ๆ หน่วยของประชากรที่มุ่งศึกษา หรือคุณลักษณะที่สามารถเปลี่ยนแปลงได้หรือแปรค่าได้ตามคุณสมบัติของตัวแปรนั้น ๆ ตัวแปรมีความสำคัญกับงานวิจัยเป็นอย่างมาก เนื่องจากเป็นตัวเชื่อมโยงกับแนวคิดและทฤษฎี อีกทั้งยังใช้เป็นข้อมูลเพื่อใช้ในการพิสูจน์ค่า ประเภทของตัวแปร มีมากมายหลากหลายประเภทขึ้นอยู่กับเกณฑ์ในการพิจารณา อันได้แก่ พิจารณาตามลักษณะของพฤติกรรมมนุษย์ พิจารณาตามลักษณะของคุณสมบัติที่แปรค่าออกมา แต่ในงานวิจัยนี้จะใช้เกณฑ์ในการพิจารณาตามความสัมพันธ์ของตัวแปร เป็นการแบ่งประเภทของตัวแปร ซึ่งประเภทของตัวแปร จะแบ่งออกเป็น 2 ประเภท ได้แก่ ตัวแปรต้น หรือตัวแปรอิสระ (Independent) เป็นตัวแปรที่เป็นสาเหตุที่ก่อให้เกิดผลหรือก่อให้เกิดการแปรผันของปรากฏการณ์ และตัวแปรตาม (Dependent Variable) เป็นตัวแปรที่เป็นผลมาจากการเปลี่ยนแปลงค่าของตัวแปรต้นหรือตัวแปรอิสระ นอกจากนี้ยังมีตัวแปรอื่น ๆ ที่อาจมีผลกระทบต่อตัวแปรตามได้ เรียกว่า ตัวแปรเกิน หรือตัวแปรแทรกซ้อน (Extraneous Variable) หรือตัวแปรควบคุม (Control Variable)

มาตราส่วนของการวัด เป็นสิ่งที่จะจำแนกลักษณะของข้อมูลที่ใช้ในการวิเคราะห์ว่าข้อมูลที่เก็บรวบรวมมามีระดับการวัดอยู่ในระดับใด ซึ่งจะเป็นตัวสำคัญในการเลือกใช้วิธีการทางสถิติแบบใดในการวิเคราะห์ข้อมูลชุดนั้น ๆ ซึ่งมาตรการวัดแบ่งออกเป็น 5 ประเภทดังนี้ [15]

1. มาตราส่วนแบบนามบัญญัติ (Nominal Data)

มาตราส่วนแบบนามบัญญัติ เป็นมาตราส่วนการวัดแบบหยาบ ซึ่งเป็นข้อมูลที่ใช้ในการกำหนดชื่อคุณลักษณะ พฤติกรรมที่ได้จากการวัดหรือการเก็บรวบรวมข้อมูล โดยเป็นระบบข้อมูลจากการสังเกตุประกอบเป็นหลาย ๆ ประเภทที่แตกต่างกัน แต่จะไม่สัมพันธ์กับขนาดของข้อมูลตัวเลขหรือตัวอักษรหมายถึงไม่มีความหมายในเชิงปริมาณที่จะนำไปดำเนินการทางคณิตศาสตร์ได้ ตัวอย่างมาตราส่วนแบบนามบัญญัติ เช่น เพศหญิงและเพศชาย ศาสนา ฯลฯ

2. มาตราส่วนแบบเรียงลำดับ (Ordinal Data)

มาตราส่วนแบบเรียงลำดับ เป็นมาตราส่วนแบบนามบัญญัติที่เพิ่มคุณสมบัติในการเรียงลำดับเป็นข้อมูลที่สามารถนำมาเรียงลำดับจากมากไปหาน้อย หรือจากน้อยไปหามาก หากแต่ไม่สามารถ

เปรียบเทียบความแตกต่างของปริมาณระหว่างลำดับที่ เนื่องจากไม่ทราบความแตกต่าง และยังไม่สามารถนำไปดำเนินการทางคณิตศาสตร์ได้ และสามารถใช้สัญลักษณ์หรือตัวเลข ที่เรียงลำดับกันแทนลำดับที่ต้องการเรียงลำดับ ตัวอย่างมาตราส่วนแบบเรียงลำดับเช่น ความหนาแน่น ระดับง่ายมาก ง่าย ปานกลาง ซับซ้อน และไม่สามารถเข้าใจได้

3. มาตราส่วนแบบช่วง (Interval Data)

มาตราส่วนแบบช่วง เป็นมาตราส่วนที่ตัวเลขจะสามารถเรียงลำดับได้ สามารถบอกความแตกต่างระหว่างสองกลุ่มได้ แต่ไม่สามารถบอกสัดส่วนได้หรือเปรียบเทียบจำนวนเท่าได้ สามารถนำไปดำเนินการทางคณิตศาสตร์ที่เป็นบวกกับลบได้ หากแต่ยังไม่สามารถคูณและหาร และไม่มีระดับศูนย์แท้ (Absolute Zero) ตัวอย่างมาตราส่วนแบบช่วง ตัวอย่างเช่น อุณหภูมิ หนึ่งช่วงองศาทุก ๆ ช่วงจะเท่ากัน แสดงถึงระดับความร้อนที่เพิ่มหรือลด หากแต่อุณหภูมิศูนย์ (0) ไม่ได้หมายความว่าไม่มีระดับความร้อน

4. มาตราส่วนแบบอัตราส่วน (Ratio Data)

มาตราส่วนแบบอัตราส่วน เป็นมาตราส่วนที่ตัวเลขจะสามารถเรียงลำดับได้ บอกความแตกต่างระหว่างสองกลุ่มได้ และสามารถบอกสัดส่วนได้ สามารถนำไปดำเนินการทางคณิตศาสตร์ได้ทั้งบวก ลบ คูณ และหาร ตัวอย่างมาตราส่วนแบบอัตราส่วน ตัวอย่างเช่น ความยาวของสิ่งของ

5. มาตราส่วนแบบสมบูรณ์ (Absolute Data)

มาตราส่วนแบบสมบูรณ์ จะได้จากการนับจำนวนสมาชิกในกลุ่ม คุณลักษณะจะอยู่ในรูปของจำนวนการเกิดขึ้นของเหตุการณ์ใด ๆ หรือเหตุการณ์ x ของเอนทิตี้ มีการเปลี่ยนแปลงเพียงอย่างเดียวคือการนับจำนวน และการคำนวณทางคณิตศาสตร์มีความหมาย ตัวอย่างมาตราส่วนแบบสมบูรณ์ เช่น จำนวนความล้มเหลวระหว่างการทดสอบแบบรวมส่วน (Integration Testing)

การวิเคราะห์เบื้องต้น เป็นการวิเคราะห์เพื่อใช้บรรยายลักษณะของประชากรหรือกลุ่มตัวอย่างและลักษณะของข้อมูล โดยการวิเคราะห์เบื้องต้นที่งานวิจัยนี้ใช้ คือการวัดแนวโน้มเข้าสู่ศูนย์กลาง (Measure of Central Tendency) การวัดการกระจาย (Measure of Variability) และการวัดความแปรปรวน (Variance) โดยแต่ละวิธีมีรายละเอียดดังนี้

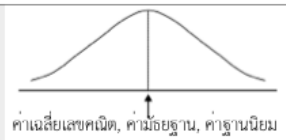
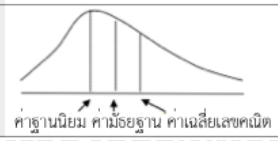
1. การวัดแนวโน้มเข้าสู่ศูนย์กลาง

การวัดแนวโน้มเข้าสู่ศูนย์กลาง เป็นการหาค่ากลางที่เป็นตัวแทนของข้อมูล เพื่อใช้สรุปผล โดยค่ากลางที่งานวิจัยนี้ใช้คือ ค่าเฉลี่ยเลขคณิต (Arithmetic mean หรือ Mean) ค่ามัธยฐาน (Median : Me) และค่าฐานนิยม (Mode : Mo) โดยแต่ละวิธีการวัดแนวโน้มเข้าสู่ศูนย์กลาง มีสูตรในการคำนวณดังรูปที่ 2.9

การหาค่าเฉลี่ยเลขคณิต (Mean)	$\bar{X} = \frac{\sum X}{n}$ หรือ $\mu = \frac{\sum X}{N}$	เมื่อ $\bar{X}$ เป็นค่าเฉลี่ยจากกลุ่มตัวอย่าง μ เป็นค่าเฉลี่ยจากประชากร $\sum X$ เป็นผลรวมของข้อมูล n, N เป็นจำนวนของข้อมูลจากกลุ่มตัวอย่างและประชากร
การหามัธยฐาน (Median)	1. เรียงลำดับข้อมูล 2. หาตัวแหน่งของค่ามัธยฐาน $= \frac{N+1}{2}$	
การหาฐานนิยม (Mode)	พิจารณาข้อมูลที่ซ้ำกันมากที่สุด	

รูปที่ 2.9 วิธีการคำนวณการวัดแนวโน้มเข้าสู่ศูนย์กลาง

หลังจากคำนวณค่าเฉลี่ยเลขคณิต ค่ามัธยฐาน และค่าฐานนิยมแล้ว ให้นำค่าทั้งสามมาพิจารณาเพื่อใช้ในการวิเคราะห์ข้อมูลว่าข้อมูลมีการแจกแจงเป็นโค้งปกติหรือไม่ ซึ่งผลการวิเคราะห์สามารถเป็นไปได้สามกรณี มีรายละเอียด ดังรูปที่ 2.10

การแจกแจงปกติ (Normal Distribution)		• การแจกแจงปกติ • ค่าเฉลี่ยเลขคณิต ค่ามัธยฐาน และค่าฐานนิยมมีค่าเท่ากัน
การแจกแจงเบ้ซ้าย (Negatively Skewness)		• การแจกแจงเบ้ซ้าย • ค่าเฉลี่ยเลขคณิต < ค่ามัธยฐาน < ค่าฐานนิยม
การแจกแจงเบ้ขวา (Positively Skewness)		• การแจกแจงเบ้ขวา • ค่าฐานนิยม > ค่ามัธยฐาน > ค่าเฉลี่ยเลขคณิต

รูปที่ 2.10 กราฟและข้อกำหนดการแจกแจกแบบปกติ เบ้ซ้าย และเบ้ขวา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2. การวัดการกระจาย

การวัดการกระจายเป็นวิธีการหาค่าที่ใช้แสดงความแตกต่างของข้อมูลแต่ละตัวในแต่ละชุด ที่ทำให้การเปรียบเทียบผลค่าวัดแนวโน้มเข้าสู่ส่วนกลางของข้อมูลมีค่าชัดเจนมากขึ้น โดยเฉพาะในกรณีที่ค่าวัดแนวโน้มเข้าสู่ส่วนกลางมีค่าเท่ากัน ต้องนำค่าจากการวัดการกระจายมาเปรียบเทียบเพื่อให้พิจารณาความแตกต่างในระดับมากหรือน้อย โดยงานวิจัยนี้จะนำวิธีการหาส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation : S.D.) ซึ่งเป็นค่าของการวัดการกระจายที่หาได้จากรากที่สองของคะแนนค่าที่เบี่ยงเบนไปจากค่าเฉลี่ยของข้อมูลยกกำลังสอง และความแปรปรวน (Variance) ซึ่งเป็นค่าเฉลี่ยของผลรวมของข้อมูลทุกตัวที่เบี่ยงเบนไปจากค่าเฉลี่ยของข้อมูลยกกำลังสอง หรือเป็นค่าส่วนเบี่ยงเบนมาตรฐานยกกำลังสองมาพิจารณา โดยแต่ละขั้นตอนมีวิธีการคำนวณ ดังรูปที่ 2.11

ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation : SD)	$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$ หรือ $\sigma = \sqrt{\frac{N\sum X^2 - (\sum X)^2}{N^2}}$ $S.D. = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$ หรือ $S.D. = \sqrt{\frac{n\sum X^2 - (\sum X)^2}{n(n-1)}}$ เมื่อ S.D. เป็นส่วนเบี่ยงเบนมาตรฐานของข้อมูลจากกลุ่มตัวอย่าง σ เป็นส่วนเบี่ยงเบนมาตรฐานของข้อมูลจากประชากร $\bar{X}$ เป็นค่าเฉลี่ยเลขคณิตของข้อมูลจากกลุ่มตัวอย่าง μ เป็นค่าเฉลี่ยเลขคณิตของข้อมูลจากประชากร X เป็นข้อมูลแต่ละตัว n เป็นจำนวนข้อมูลของกลุ่มตัวอย่าง N เป็นจำนวนข้อมูลของประชากร
ความแปรปรวน (Variance)	σ^2 แทนความแปรปรวนของข้อมูล

รูปที่ 2.11 วิธีการคำนวณส่วนเบี่ยงเบนมาตรฐานและความแปรปรวน

2.1.7 การวิเคราะห์ความถดถอยโลจิสติก

การวิเคราะห์ความถดถอยโลจิสติก (Logistic Regression Analysis) [17] เป็นการวิเคราะห์วิธีการหนึ่งในการวิเคราะห์ความถดถอย (Regression) ในกรณีที่ตัวแปรตาม เป็นข้อมูลที่ไม่ได้เป็นแบบต่อเนื่อง เป็นแบบหมวดหมู่ (Category) หรืออยู่ในรูปแบบมาตราส่วนแบบนามบัญญัติ ซึ่งการวิเคราะห์ความถดถอยโลจิสติก นำมาใช้ในการทำนายว่าจะเกิดเหตุการณ์หนึ่งขึ้นหรือไม่ หรือมีโอกาสการเกิดมากน้อยเพียงใด โดยมีการรวบรวมตัวแปร หรือกำหนดค่าตัวแปรต้น หรือตัวแปรควบคุมที่จะส่งผลกับตัวแปรตาม โดยเป็นการใช้ตัวแปรต้นในการทำนายโอกาส (Probability) ที่จะเกิดตัวแปรตาม

การวิเคราะห์ความถดถอยโลจิสติก สามารถแบ่งได้เป็น 2 ประเภท คือ การวิเคราะห์การถดถอยโลจิสติกทวิ (Binary Logistic Regression Analysis) และการวิเคราะห์การถดถอยโลจิสติกพหุกลุ่ม (Multinomial Logistic Regression Analysis) โดยการวิเคราะห์สองประเภทนี้แตกต่างกันที่ตัวแปรตาม ซึ่งการวิเคราะห์การถดถอยโลจิสติกทวิ สามารถแบ่งได้เป็น 2 กลุ่มย่อย (Dichotomous Variable) ซึ่งมีค่าเป็น 0 กับ 1 ส่วนการวิเคราะห์โลจิสติกแบบพหุกลุ่มจะมีตัวแปรตามมากกว่า 2 กลุ่ม (Polychromous Variable)

ก่อนการวิเคราะห์การถดถอยโลจิสติกนั้นจะต้องมีการตรวจสอบข้อตกลงการวิเคราะห์การถดถอยโลจิสติกเบื้องต้นก่อน โดยการวิเคราะห์การถดถอยโลจิสติกนั้น มีข้อตกลงเบื้องต้นดังนี้

1. ข้อตกลงเกี่ยวกับตัวแปรที่ใช้ในการวิเคราะห์การถดถอยโลจิสติก

ตัวแปรอิสระ ตัวแปรต้น หรือตัวแปรทำนายต้องเป็นระดับข้อมูลอยู่ในมาตราส่วนแบบช่วงเป็นอย่างต่ำ กรณีที่เป็นข้อมูลเชิงกลุ่มให้แปลงเป็นตัวแปรหุ่น (Dichotomous Variable) โดยกำหนดจากจำนวนกลุ่มของตัวแปร หากเป็นการวิเคราะห์โลจิสติกแบบทวิ ให้กำหนดเป็น 2 ค่าคือ 1 และ 0 หากเป็นการวิเคราะห์โลจิสติกแบบพหุกลุ่ม ให้กำหนดตามจำนวนกลุ่มของตัวแปร

2. ความสัมพันธ์ระหว่างตัวแปร

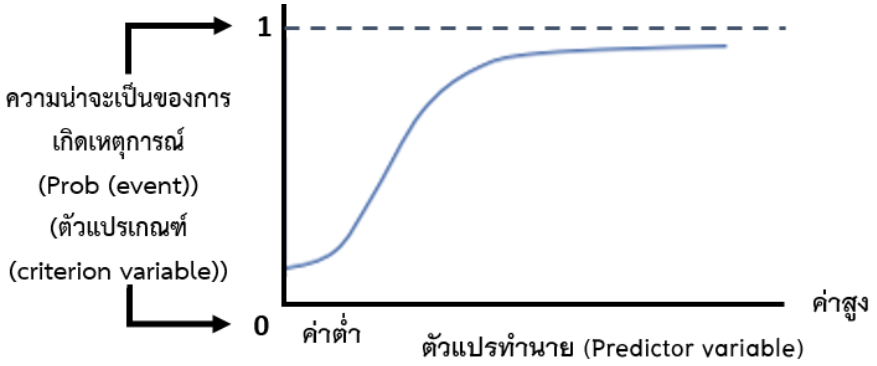
ในการวิเคราะห์การถดถอยโลจิสติกนั้น ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันหรือไม่เกิดปัญหา multicollinearity

จุฬาลงกรณ์มหาวิทยาลัย Chulalongkorn University

3. จำนวนตัวอย่าง

ขนาดตัวอย่างต้องเท่ากับ n มากกว่าหรือเท่ากับ 30x โดยที่ x คือจำนวนตัวแปรทำนาย

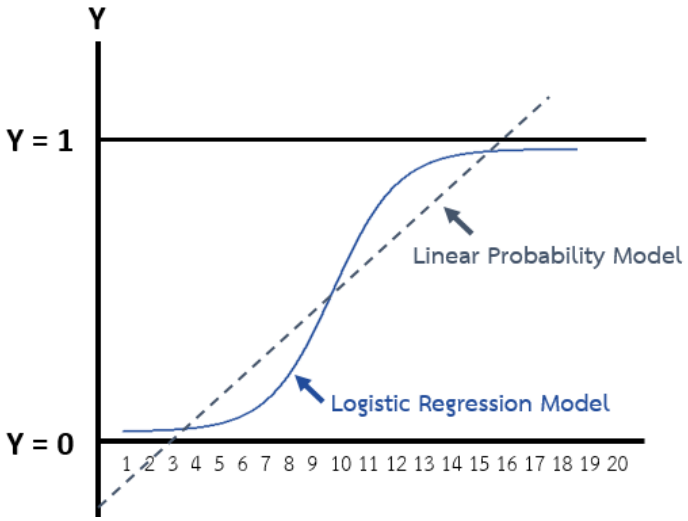
หลังจากตรวจสอบข้อตกลงการวิเคราะห์การถดถอยโลจิสติกเบื้องต้นแล้ว ขั้นตอนต่อไปคือการคำนวณฟังก์ชันการถดถอยโลจิสติกทวิ โดยความสัมพันธ์ของตัวแปรการวิเคราะห์โลจิสติกจะแสดงได้เป็นกราฟรูปตัวเอส (S) ดังรูป 2.12



รูปที่ 2.12 กราฟของโมเดลการวิเคราะห์การถดถอยโลจิสติก [17]

การทำนายการเกิดตัวแปรตาม (ตัวแปร Y) สามารถทำนายได้จากฟังก์ชันโลจิสติก P(Y) โดย P(Y) คือความน่าจะเป็นของการเกิดเหตุการณ์ Y ค่า e คือค่าฟังก์ชันเอ็กซ์โพเนนเชียล (Exponential Function) ซึ่งมีค่าเท่ากับ 2.71828 และ f(x) คือฟังก์ชันของตัวแปรทำนาย

Odd หรือ Odd Ratio คือ อัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์ที่สนใจ (ค่า Y =1) กับโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ (ค่า Y = 0) เป็นการปรับให้ความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามให้อยู่ในรูปเชิงเส้น แสดงเป็นเส้น Linear Probability Model ดังรูป 2.13



รูปที่ 2.13 เส้น Linear Probability Model เทียบกับ Logistic Regression Model [17]

โดยค่า Odds นั้นจะแสดงถึงโอกาสการเกิดเหตุการณ์ที่สนใจ ว่าเป็นกี่เท่าของโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ สามารถนำมาเขียนเป็นสมการทางคณิตศาสตร์ได้ แสดงดังรูป 2.14

$$\text{Odds} = \frac{P_y \text{ (โอกาสเกิดเหตุการณ์ที่สนใจ)}}{Q_y \text{ (โอกาสไม่เกิดเหตุการณ์ที่สนใจ)}}$$

รูปที่ 2.14 สมการทางคณิตศาสตร์ของค่า Odds

สำหรับการทำนายตัวแปรตามหรือ ค่า Y ที่เป็น P(Y) ในการวิเคราะห์การถดถอยโลจิสติกนั้น จะใช้สมการ ดังรูปที่ 2.15 ในการทำนาย

$$P_y = \frac{e^{b_0 + b_1x_1 + \ldots + b_px_p}}{1 + e^{b_0 + b_1x_1 + \ldots + b_px_p}}$$

รูปที่ 2.15 สมการทางคณิตศาสตร์ที่ใช้ในการทำนายตัวแปรตาม

นอกจากสมการทางคณิตศาสตร์ที่ใช้ในการทำนายตัวแปรตามแล้ว การวิเคราะห์การถดถอยโลจิสติก ยังสามารถทดสอบนัยสำคัญของสัมประสิทธิ์การถดถอยโลจิสติกของตัวแปรทำนายแต่ละตัว โดยสามารถทดสอบด้วยสถิติทดสอบ 2 ตัวด้วยกัน ได้แก่ การทดสอบความเหมาะสมของค่าสัมประสิทธิ์ของโมเดล และสถิติทดสอบของวอลด์ (Wald Test) อีกทั้งยังมีสถิติทดสอบระดับความสัมพันธ์ที่ใช้ในการวิเคราะห์การถดถอยโลจิสติก ซึ่งแบ่งเป็น 2 ประเภท ได้แก่ สถิติทดสอบ Cox & Snell R Square และสถิติทดสอบ Nagelkerke R Square และวิธีการเลือกตัวแปรทำนายเข้าวิเคราะห์ มีวิธีเลือก 3 วิธี ได้แก่ วิธีที่ 1 enter method วิธีที่ 2 forward method และวิธีที่ 3 backward method

ขั้นตอนการวิเคราะห์การถดถอยโลจิสติก ประกอบด้วย 4 ขั้นตอนด้วยกัน มีรายละเอียดดังต่อไปนี้

ขั้นตอนที่ 1 เลือกตัวแปรต้นและตัวแปรตามที่ต้องการทำการวิเคราะห์ โดยกำหนดตัวแปรตามเป็นตัวแปรที่สนใจที่จะทำนายค่า และตัวแปรต้น คือตัวเป็นที่เป็นต้นเหตุ หรือตัวแปรที่ทำให้ตัวแปรตามมีค่าเปลี่ยนแปลงไป

ขั้นตอนที่ 2 ตรวจสอบค่าความผิดปกติของตัวแปรทุกตัว และตรวจสอบข้อตกลงเบื้องต้นก่อนการวิเคราะห์การถดถอยโลจิสติก

ขั้นตอนที่ 3 สร้างสมการตามสมการคณิตศาสตร์การทำนายการวิเคราะห์ถดถอยโลจิสติก ดังรูป 2.15 โดย P(Y) ของฟังก์ชันการทำนาย ค่า e คือฟังก์ชันเอ็กซ์โพเนนเชียล และ $b_0+b_1x_1 + \ldots + b_px_p$ คือสมการโลจิสติก จากนั้นตรวจสอบความถูกต้องเหมาะสมของสมการด้วย การทดสอบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

นัยสำคัญของสัมประสิทธิ์ และสถิติทดสอบระดับความสัมพันธ์วิธีใดวิธีหนึ่ง หรืออาจตรวจสอบทุกวิธีก็สามารถทำได้

ขั้นตอนที่ 4 วิเคราะห์ข้อมูล เพื่อสร้างสมการถดถอยโลจิสติก โดยนำค่าที่ได้จากการวิเคราะห์ต่างๆ หรือค่าสัมประสิทธิ์ที่ได้ มาแทนที่สมการถดถอยโลจิสติกและใช้สมการนั้นในการทำนายค่าใหม่ ๆ หรือกรณีใหม่ ๆ ที่เข้ามา

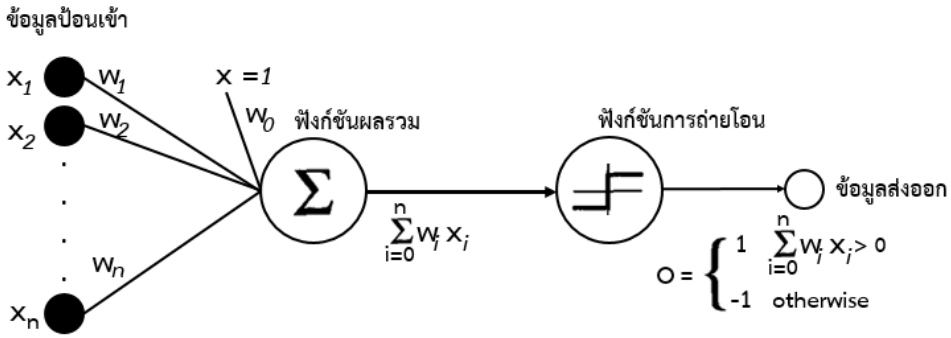
โดยหลังจากที่ค่าการวิเคราะห์ออกมาแล้ว นำค่าผลลัพธ์ที่ได้มาวิเคราะห์ และสรุปเป็นข้อสรุปที่ใช้ในการทำนาย โดยหากค่า P มีค่ามากกว่า 0.5 จะให้ค่า Y เป็น 1 หรือสรุปได้ว่าเกิดเหตุการณ์ที่น่าสนใจ และหาก P มีค่าน้อยกว่าหรือเท่ากับ 0.5 จะให้ Y เป็น 0 หรือไม่เกิดเหตุการณ์ที่สนใจ โดยนอกจาก 0.5 แล้วยังสามารถกำหนดเป็นค่าต่างๆได้ตามความเหมาะสม ตัวอย่างค่าที่อาจใช้ในการกำหนด เช่น 0.7 หรือ 0.4 เป็นต้น

2.1.8 การวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น

การวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น (Multi layer perceptron) ประกอบด้วยโครงข่ายประสาทเทียม (Neural Network) [18] สร้างขึ้นมาจากการจำลองลักษณะของการประมวลผลของสมองมนุษย์ โดยมีส่วนประกอบของการประมวลผลที่เรียกว่า นิวรอน (Neuron) โดยนิวรอนเพียงหนึ่งตัวไม่สามารถแก้ไขปัญหาที่ซับซ้อนได้ อีกทั้งยังไม่สามารถคำนวณได้โดยมือ นิวรอนสามารถมีข้อมูลป้อนเข้าได้หลายค่า หากแต่มีข้อมูลส่งออกได้เพียงค่าเดียว และข้อมูลส่งออก จะเชื่อมโยงไปยังข้อมูลป้อนเข้านิวรอนอื่น ๆ ภายในโครงข่าย

เพอร์เซปตรอน (Perceptron) เป็นรูปแบบในการแสดงความสัมพันธ์ระหว่างโหนดสองชนิด คือโหนดนำเข้า (Input Node) และโหนดส่งออก (Output Node) ซึ่งใช้เป็นตัวแทนแสดงรูปแบบของผลลัพธ์ (Model Output) โดยแต่ละการเชื่อมโยงระหว่างโหนดนำเข้าและโหนดส่งออก จะมีค่าน้ำหนัก (Weight) ที่เป็นค่าแสดงความแข็งแรงในการเชื่อมโยงกัน ระหว่างข้อมูลนำเข้าและข้อมูลส่งออก

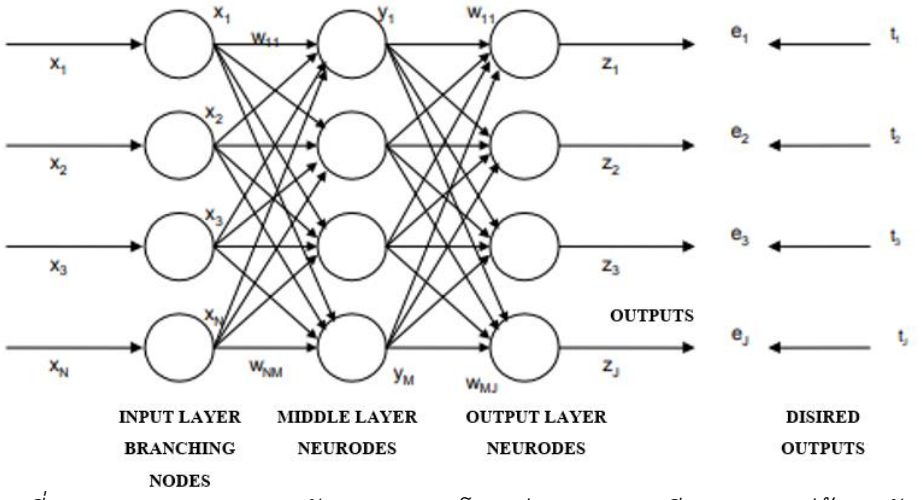
โครงสร้างของโครงข่ายประสาทเทียม ประกอบด้วยชั้นข้อมูลป้อนเข้า (Input Layer) ชั้นข้อมูลส่งออก (Output Layer) และชั้นซ่อน (Hidden Layer) ซึ่งอยู่ระหว่างชั้นข้อมูลป้อนเข้า และชั้นข้อมูลป้อนออก โดยหลักการโครงข่ายประสาทเทียม สามารถอธิบายได้เป็นภาพโดยง่าย รายละเอียดดังรูป 2.16



รูปที่ 2.16 หลักการโครงข่ายประสาทเทียม [18]

ประเภทของโครงข่ายประสาทเทียมสามารถแบ่งได้เป็น 2 ประเภท คือ โครงข่ายประสาทเทียมเพอร์เซปตรอนชั้นเดียว (Single-layer Perceptron Neural Network) และโครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น (Multi-layer Perceptron Neural Network) โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น เป็นรูปแบบของโครงข่ายประสาทเทียมที่สามารถใช้สำหรับงานที่มีความซับซ้อนได้ และมีกระบวนการเรียนรู้แบบมีผู้สอน (Supervised Learning) อีกทั้งยังมีการใช้ขั้นตอนการส่งค่าแบบแพร่ย้อนกลับ (Back Propagation) เพื่อใช้ในการเรียนรู้ โดยประกอบเป็น 2 ส่วนย่อย คือการส่งผ่านไปข้างหน้า (Forward Pass) และการส่งผ่านย้อนกลับ (Backward Pass) โดยสัญญาณที่เกิดขึ้นในโครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น มี 2 ประเภทด้วยกัน ได้แก่ สัญญาณฟังก์ชัน (Function Signal) และสัญญาณข้อผิดพลาด (Error Signal) โดยสัญญาณฟังก์ชันเป็นสัญญาณเข้าที่มาจากโหนดในชั้นก่อนหน้า และจะส่งผ่านไปข้างหน้าจากโหนดหนึ่งไปสู่อีกโหนด ส่วนสัญญาณข้อผิดพลาดเป็นสัญญาณย้อนกลับที่เกิดขึ้นที่โหนดในขั้นข้อมูลออกของโครงข่ายประสาทเทียม และถูกส่งย้อนกลับจากชั้นหนึ่งไปสู่อีกชั้น โดยภาพรวมของสถาปัตยกรรมของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ความหมายและการแปลความค่าต่างๆ สามารถอธิบายได้ดังรูป 2.17

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



รูปที่ 2.17 ภาพรวมของสถาปัตยกรรมของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ

2.1.9 เครื่องมือราดอน

เครื่องมือราดอน (Radon Tool) [19] เป็นเครื่องมือที่ถูกพัฒนาโดยภาษาไพธอน ใช้ในการคำนวณตัววัดที่หลากหลายจากซอร์สโค้ด โดยตัววัดที่ราดอนสามารถคำนวณได้นั้น ได้แก่ (McCabe's Complexity) ตัววัดดิบ (Raw Metrics) ซึ่งประกอบไปด้วย (Source Line of Code : SLOC) จำนวนบรรทัดคอมเม้น (Comment Lines) จำนวนบรรทัดว่างเปล่า หรือจำนวนบรรทัดที่ไม่มีข้อมูล (Blank Lines) (Halstead Metrics) และ (Maintainability Index) โดยเครื่องมือราดอน สามารถประมวลผลได้ใน ไพธอนเวอร์ชัน 2.7 ถึงไพธอนเวอร์ชัน 3.6 เท่านั้น

การติดตั้งเครื่องมือราดอน สามารถติดตั้งได้ผ่านคำสั่ง pip ซึ่งเป็นคำสั่งในคลังโปรแกรมหรือไลบรารีเสริมของไพธอน หรือที่เรียกว่า PyPI (Python Package Index) ที่สร้างขึ้นมาให้สามารถติดตั้งโปรแกรมได้ โดยการใช้งานเครื่องมือราดอนนั้น สามารถใช้ได้ทั้งในคำสั่ง command line และการเขียนโปรแกรม (programmatically) ผ่านเอพีไอ (Application Programming Interface: API)

การใช้เครื่องมือราดอนผ่านคำสั่ง command line นั้น ประกอบไปด้วย 4 คำสั่งด้วยกัน คือ คำสั่ง cc ใช้ในการคำนวณ (Cyclomatic Complexity) คำสั่ง raw ใช้ในการคำนวณตัววัดดิบ คำสั่ง mi ใช้ในการคำนวณค่า Maintainability Index และคำสั่ง hal ใช้ในการคำนวณ Halstead Complexity Metrics ซึ่งผลลัพธ์และการแปลความหมายของการใช้คำสั่ง command line ในการคำนวณค่าตัววัดผ่านเครื่องมือราดอนนั้น แสดงดังตารางที่ 2.1

ตารางที่ 2.1 คำสั่งการประมวลผลและการแปลความหมายผลลัพธ์ของเครื่องมือราดอน [19]

คำสั่ง	วัตถุประสงค์	การแปลความหมาย
cc	ใช้ในการคำนวณตัววัด Cyclomatic Complexity	ผลลัพธ์ที่ได้ออกมาจะเป็นการคำนวณค่าความซับซ้อนในฟังก์ชัน (Function) เมธอด (Method) คลาส (Class) และมีการวัดความซับซ้อนของซอร์สโค้ดโดยรวมจากกลุ่มของความซับซ้อน ประกอบด้วย 6 กลุ่ม ซึ่งแต่ละกลุ่มมีความหมายดังนี้ 1. คะแนน 1-5 กลุ่ม A มีค่าต่ำ เป็นบล็อกแบบง่าย (Simple Block) 2. คะแนน 6-10 กลุ่ม B มีค่าต่ำ มีโครงสร้างการจัดการที่ดี (Well Structured) และเป็นบล็อกที่มีความสเถียร (Stable Block)

ตารางที่ 2.1 คำสั่งการประมวลผลและการแปลความหมายผลลัพธ์ของเครื่องมือราดอน [19] (ต่อ)

คำสั่ง	**วัตถุประสงค์**	**การแปลความหมาย**
cc (ต่อ)	ใช้ในการคำนวณตัววัด Cyclomatic Complexity	3. คะแนน 11-20 กลุ่ม C มีค่าปานกลาง ค่อนข้างเป็นบล็อกที่มีความซับซ้อน (Slightly Complex Block) 4. คะแนน 21-30 กลุ่ม D มีค่ามากกว่าปานกลาง มีความซับซ้อนมากกว่า (More Complex Block) 5. คะแนน 31-40 กลุ่ม E มีค่าสูง บล็อกมีความซับซ้อน (Complex Block) ควรระวัง 6. คะแนน 41 ขึ้นไป มีค่าค่อนข้างสูง เกิด error prone เป็นบล็อกที่ไม่มีความเสถียร (Unstable Block)
raw	ใช้ในการคำนวณตัววัดดิบ	ผลลัพธ์ที่ได้ออกมาเป็นค่าจำนวนจริง ซึ่งค่าที่แสดงนั้นแบ่งเป็น 6 ค่าด้วยกัน ได้แก่ 1. LOC (Lines of Code) จำนวนบรรทัดของโค้ด 2. LLOC (Logical Lines of Code) จำนวนบรรทัดของโค้ดที่มีตรรกะ 3. SLOC (Source Lines of Code) จำนวนบรรทัดของซอร์สโค้ด (ไม่จำเป็นต้องเท่ากับ LLOC) 4. Comments จำนวนบรรทัดของคอมเม้นที่เป็นภาษาไพธอน (Python Comment Lines) 5. Multi จำนวนบรรทัดที่มีความหลากหลายในตัวอักษร (Multi-line Strings) 6. Blank จำนวนบรรทัดที่ว่างเปล่า หรือไม่มีข้อมูล
mi	ใช้ในการคำนวณ Maintainability Index	ผลลัพธ์ที่ได้ออกมาเป็นทั้งค่าจำนวนจริงและกลุ่มอันดับ ซึ่งมีความหมายดังนี้ 1. คะแนน MI 100-200 กลุ่ม A สูง 2. คะแนน 19-10 กลุ่ม B ปานกลาง 3. คะแนน 9-0 กลุ่ม C ต่ำมาก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 2.1 คำสั่งการประมวลผลและการแปลความหมายผลลัพธ์ของเครื่องมือราดอน [19] (ต่อ)

คำสั่ง	วัตถุประสงค์	การแปลความหมาย
hal	ใช้ในการคำนวณตัววัด halstead	ผลลัพธ์ที่ได้ออกมาเป็นค่าจำนวนจริง มีทั้งหมด 12 ค่า ได้แก่ เอช1 (h1), เอช2 (h2), เอ็น1 (N1), เอ็น2 (N2), คำศัพท์ (Vocabulary), ปริมาณ (Volume), ความยาก (Difficulty), ความยาว (Length),ความพยายาม (Effort), ความยาวที่คำนวณโดยโปรแกรม (Calculated Program Length), เวลาที่ใช้ในการโปรแกรม (Time required to Program), จำนวนข้อบกพร่องที่ตรวจสอบเจอ (Number of delivered Bugs)

2.1.10 เครื่องมือซอร์สมิเตอร์

เครื่องมือซอร์สมิเตอร์ (SourceMeter Tool) [20] เป็นเครื่องมือที่สร้างขึ้นมาเพื่อวิเคราะห์ซอร์สโค้ดอย่างแม่นยำ สำหรับซอร์สโค้ดภาษาซี ซีพลัสพลัส จาวา ไพธอน โดยเครื่องมือซอร์สมิเตอร์ช่วยในการหาจุดอ่อนของระบบภายใต้การพัฒนา โดยเป็นการวัดค่ามาจากซอร์สโค้ดเท่านั้น การใช้ผลลัพธ์จากการวิเคราะห์ไปปรับปรุงคุณภาพของซอร์สโค้ด ซึ่งสามารถพัฒนาได้ทั้งระยะสั้น และระยะยาว

เครื่องมือซอร์สมิเตอร์สำหรับไพธอน เป็นการวิเคราะห์ซอร์สโค้ดในระบบที่ซับซ้อนของไพธอน ซึ่งสามารถใช้ได้ในไพธอน เวอร์ชัน 2.7.8 หรือเวอร์ชันก่อนหน้านั้น โดยสำหรับตัววัดซอร์สโค้ดในเครื่องมือซอร์สมิเตอร์นั้น ใช้ในการดูคุณภาพของซอร์สโค้ด โดยการคำนวณส่วนโปรแกรม (Component) ไฟล์ (Source Files) แพคเกจ (Package) ชนิดของคลาส (Class Type) เมธอด (Method) และฟังก์ชัน (Function) โดยการคำนวณแบ่งออกเป็น 6 ประเภทด้วยกัน ได้แก่

ประเภทที่ 1 Cohesion Metrics วัดความเกี่ยวข้องกันระหว่างซอร์สโค้ดส่วนย่อยในระบบ

ประเภทที่ 2 Complexity Metrics วัดความซับซ้อนของซอร์สโค้ดส่วนย่อย

ประเภทที่ 3 Coupling Metrics วัดจำนวนที่มีความเกี่ยวเนื่องกันภายในระหว่างซอร์สโค้ดส่วนย่อย

ประเภทที่ 4 Documentation Metrics วัดจำนวนของการแสดงความคิดเห็นและเอกสารเกี่ยวกับซอร์สโค้ด

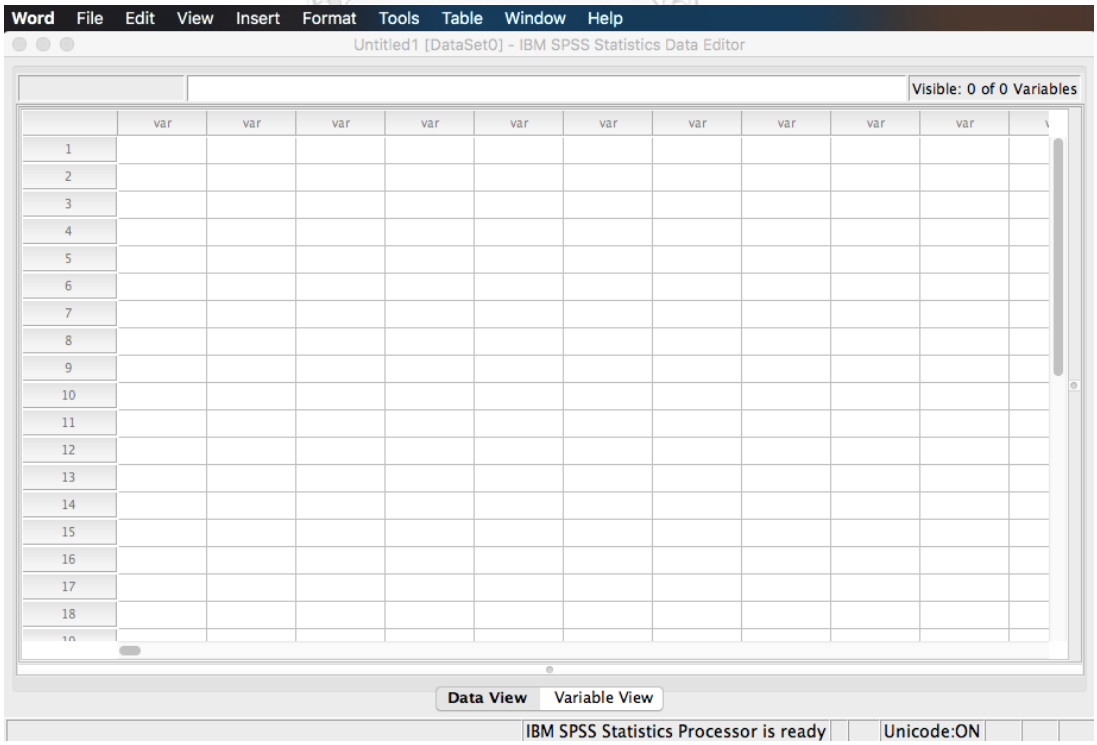
ประเภทที่ 5 ตัววัดการรับทอด (Inheritance Metric) วัดความแตกต่างระหว่างลำดับชั้นการรับทอดของระบบ

ประเภทที่ 6 ตัววัดขนาด (Size Metric) วัดคุณสมบัติพื้นฐานของระบบการวิเคราะห์ในแง่ความแตกต่าง ตัวอย่างเช่น จำนวนบรรทัดของโค้ด จำนวนคลาส จำนวนเมธอด เป็นต้น

โดยนอกจากการคำนวณทั้ง 6 ประเภทแล้วนั้น เครื่องมือซอร์สมิเตอร์ยังสามารถใช้ในการคำนวณการคัดลอกของโค้ด (Code Cloning) ซึ่งเป็นการคัดเลือกของซอร์สโค้ดจากที่หนึ่งไปยังอีกที่ ซึ่งเป็นสาเหตุของร่องรอยที่ไม่ดี (Code Smell)

2.1.11 เครื่องมือสำเร็จรูปทางสถิติเอสพีเอสเอส

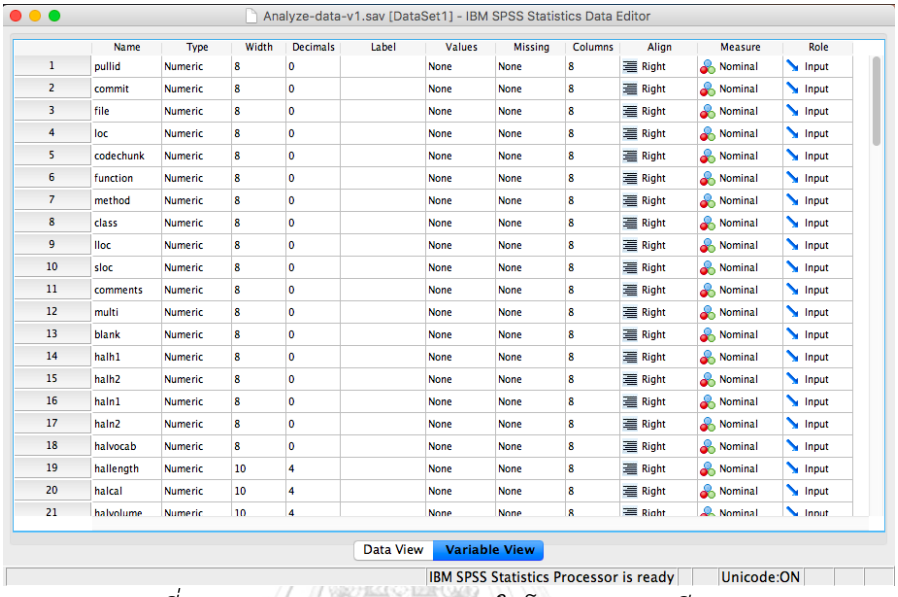
เครื่องมือเอสพีเอสเอส (IBM SPSS Statistics 22 Tool) [21] เป็นเครื่องมือที่ใช้ในการวิเคราะห์ หรือใช้ในการคำนวณหาค่าสถิติของข้อมูลได้หลากหลายรูปแบบ และเป็นที่นิยมมากในการใช้ทำวิจัย เนื่องจากเป็นเครื่องมือที่ใช้ง่ายและสามารวิเคราะห์ข้อมูลปริมาณมากได้ อีกทั้งยังสามารถแสดงผลข้อมูลในรูปแบบต่างๆ ได้ ไม่ว่าจะเป็นกราฟแสดงข้อมูล หรือแผนภาพแสดงข้อมูล ซึ่ง SPSS ย่อมาจาก Statistics Package for The Social Science สามารถใช้ในการวิเคราะห์หลักการสถิติพื้นฐาน ไม่ว่าจะเป็นสถิติเชิงบรรยาย การวิเคราะห์การถดถอยเชิงเส้น การแสดงกราฟ รวมถึงแบบรายงาน และสามารถใช้เอสพีเอสเอสในการวิเคราะห์สถิติอย่างสูงได้เช่นกัน เช่น การใช้เอสพีเอสเอสในการพยากรณ์ (Forecasting) การวิเคราะห์โดยต้นไม้ตัดสินใจ (Decision Trees) การวิเคราะห์องค์ประกอบ (Factor Analysis) โดยส่วนต่อประสานของเครื่องมือเอสพีเอสเอสนั้น แสดงดังรูป 2.18



รูปที่ 2.18 ส่วนต่อประสานเครื่องมือเอสพีเอสเอส

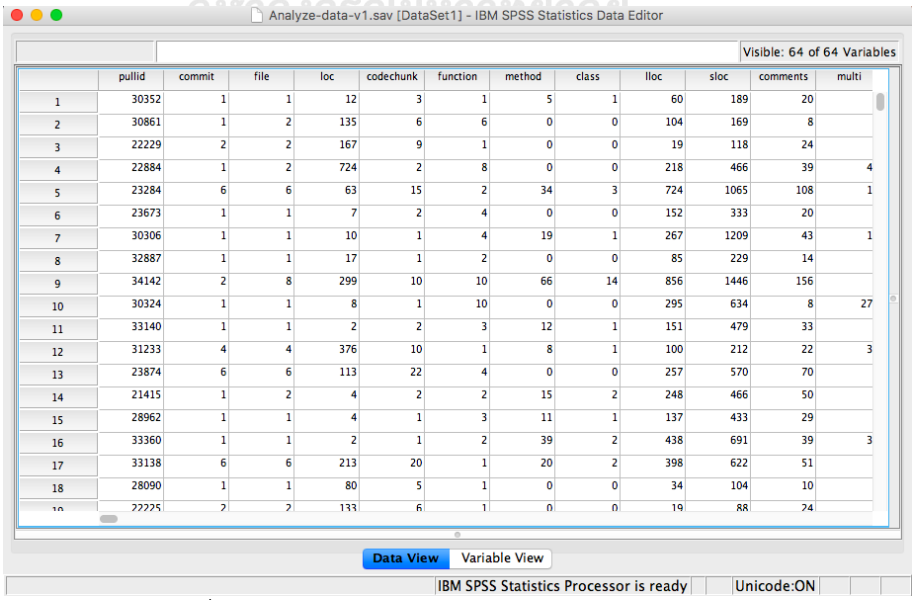
การวิเคราะห์ข้อมูลด้วยเครื่องมือเอสพีเอสเอส สามารถแบ่งได้เป็น 3 ขั้นตอน ด้วยกัน ได้แก่

ขั้นตอนที่ 1 การนำเข้าข้อมูล การนำเข้าข้อมูลในเอสพีเอสเอสนั้น เริ่มจากการระบุรายละเอียดของตัวแปรในข้อมูลก่อน โดยการกรอกข้อมูลไปที่ Variable View หรือหน้าต่างที่ใช้ในการกำหนดชื่อ ลักษณะของตัวแปรแต่ละตัว ดังรูปที่ 2.19



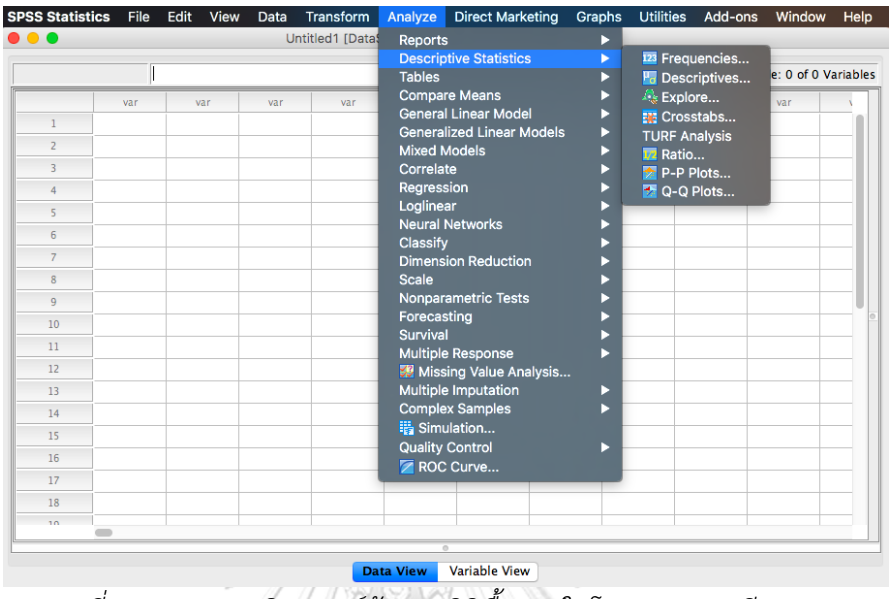
รูปที่ 2.19 แถบ Variable View ในโปรแกรมเอสพีเอสเอส

หลังจากการระบุรายละเอียดของตัวแปรครบถ้วนแล้ว ขั้นตอนต่อมาเป็นการใส่ข้อมูลที่ได้รวบรวมมาลงไปในโปรแกรม โดยระบุลงไปใน Data View หรือหน้าต่างที่จะทำการใส่ข้อมูลเข้าไป ดังรูปที่ 2.20



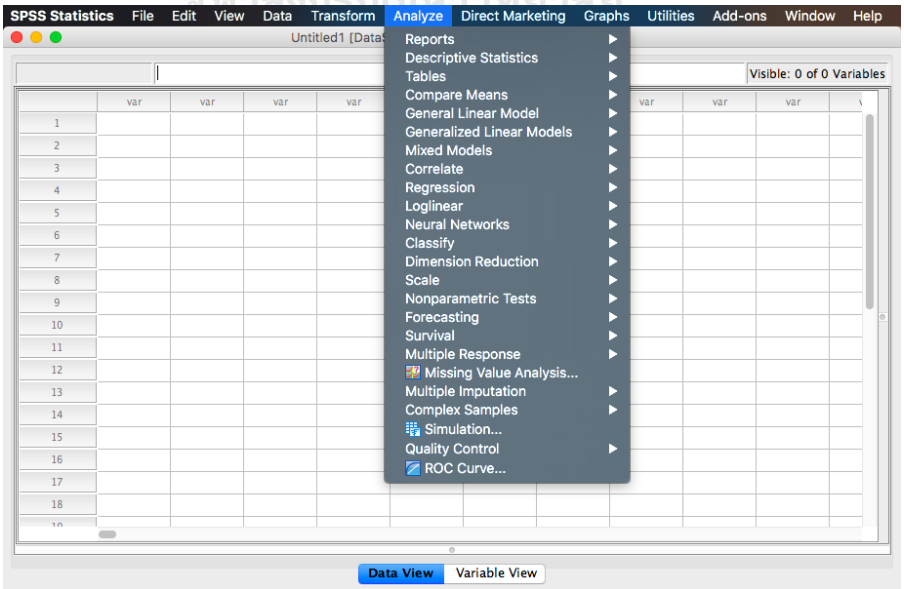
รูปที่ 2.20 แถบ Data View ในโปรแกรมเอสพีเอสเอส

ขั้นตอนที่ 2 การวิเคราะห์ข้อมูลสถิติพื้นฐาน สถิติพื้นฐานในเครื่องมือเอสพีอสเอส สามารถวิเคราะห์ได้โดยการเลือกเมนู Analyze --> Descriptive Statistics ดังรูปที่ 2.21



รูปที่ 2.21 เมนูการวิเคราะห์ข้อมูลสถิติพื้นฐานในโปรแกรมเอสพีเอสเอส

ขั้นตอนที่ 3 การวิเคราะห์ข้อมูลสถิติขั้นสูง นอกจากการวิเคราะห์ข้อมูลสถิติพื้นฐานแล้วนั้น โปรแกรมเอสพีเอสเอส ยังสามารถวิเคราะห์สถิติขั้นสูงได้อีกด้วย ไม่ว่าจะเป็นสถิติที่ใช้ในการทำนาย หรือสถิติที่ใช้ในการหาความสัมพันธ์ของตัวแปรต่าง ๆ สามารถวิเคราะห์ได้โดยเลือกเมนู Analyze จากนั้นเลือกหัวข้อการวิเคราะห์ที่ต้องการ ดังรูปที่ 2.22

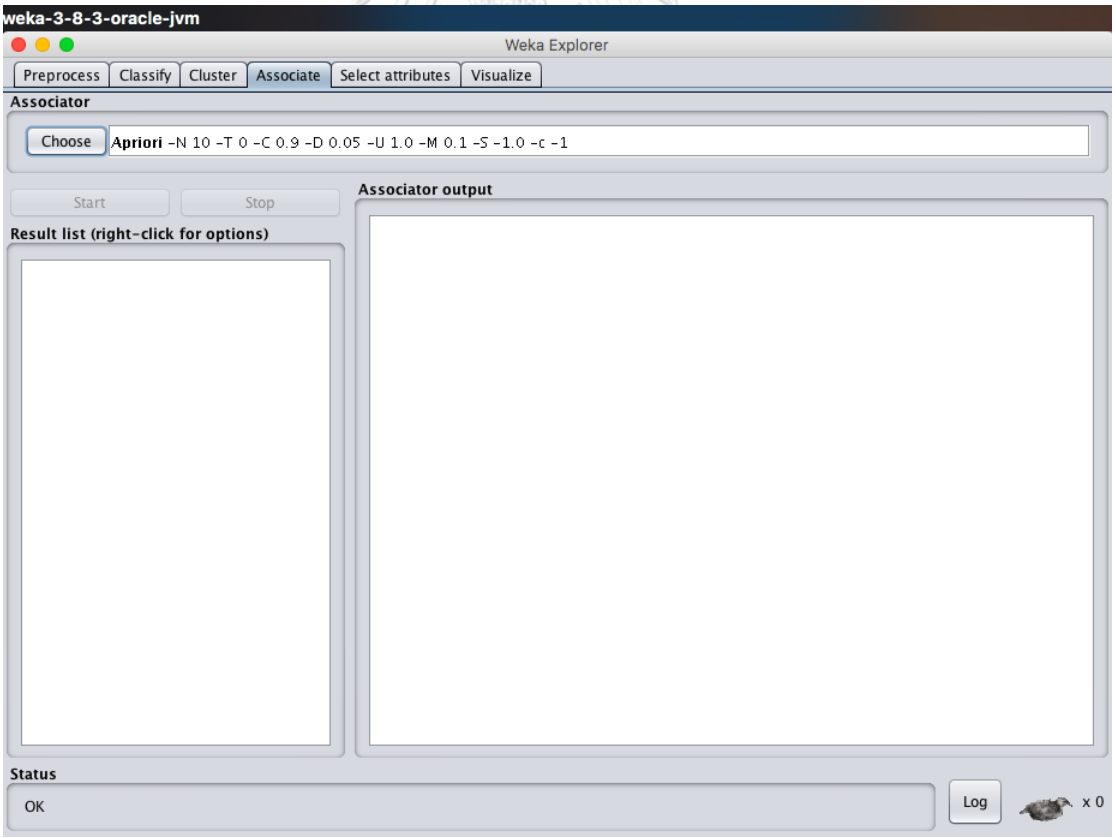


รูปที่ 2.22 เมนูการวิเคราะห์ข้อมูลสถิติขั้นสูงในโปรแกรมเอสพีเอสเอส

2.1.12 เครื่องมือเวก้า

เวก้า (Weka) [22] ย่อมาจาก Waikato Environment for Knowledge Analysis เริ่มพัฒนามาตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นเครื่องมือที่ได้รวบรวมขั้นตอนวิธีของการเรียนรู้ด้วยเครื่องสำหรับการทำเหมืองข้อมูล ซึ่งประกอบไปด้วย การเตรียมข้อมูล การจำแนกประเภท (Classification) การวิเคราะห์ความถดถอย (Regression) การคัสเตอร์ (Clustering) การหากฎความสัมพันธ์ (Association Rules Mining) และการสร้างมโนภาพ (Visualization) ซึ่งเวก้าเป็นโอเพนซอร์สที่อยู่ภายใต้การดูแลขององค์กร GNU General Public License อีกทั้งยังสามารถประมวลผลได้ทั้งในข้อมูลที่มีขนาดใหญ่ และมีความสามารถในการประมวลผลการเรียนเชิงลึก (Deep Learning) อีกด้วย

เวก้าให้การสร้างขั้นตอนวิธีการเรียนรู้ที่ง่ายต่อการใช้กับข้อมูล รวมถึงมีการแปลงข้อมูลที่สามารถใช้กับข้อมูลได้อย่างหลากหลาย การใช้งานในส่วนที่ง่ายที่สุด คือมีส่วนที่สามารถใช้งานส่วนต่อประสานกับผู้ใช้ (Graphic User Interface : GUI) ได้ นั่นคือการเข้าผ่านเมนู Explorer ซึ่งเป็นส่วนที่ให้ผู้ใช้เข้าถึงหน้าจอได้ มีปุ่มและมีเมนูให้สามารถเลือกได้ ซึ่งสะดวกกับการใช้งานในการวิเคราะห์ของผู้ใช้อย่างมาก แสดงดังรูป 2.23



รูปที่ 2.23 ส่วนหน้าจอ Explorer ของเครื่องมือเวก้า

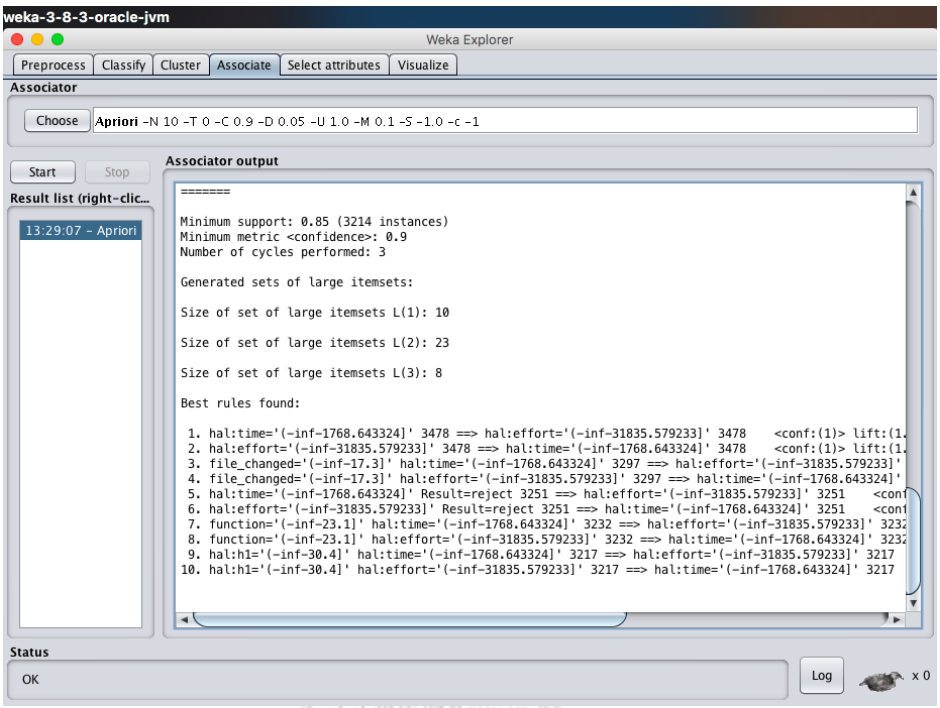
หลังจากที่ได้เข้าเมนู Explorer มาแล้ว แถบด้านบนจะมีทั้งหมด 6 แถบด้วยกัน โดยแต่ละแถบเป็นการประมวลผลในหัวข้อที่แตกต่างกันออกไป แสดงดังตารางที่ 2.2

ตารางที่ 2.2 การประมวลผล (แถบข้อมูล) ในโปรแกรมเวก้า [22]

แถบข้อมูล (ฟังก์ชัน)	**การประมวลผล**
การเตรียมการ (Preprocess)	เป็นฟังก์ชันที่ใช้ในการเตรียมข้อมูล
จำแนกประเภท (Classify)	เป็นฟังก์ชันที่เกี่ยวกับการทำเหมืองข้อมูลด้วยเทคนิคการจำแนกข้อมูล (Classification) ที่สามารถแบ่งแยกข้อมูลออกตามลักษณะประจำตามที่เป้าหมายกำหนด โดยมีตัวแบบจำลอง ดังนี้ เบย์ (Baye) ฟังก์ชัน (Functions) เลซี่ (Lazy) ตัวแบบที่เก็บตัวอย่างการตัดสินใจเมื่อตัวอย่างใหญ่ถูกนำเข้าเท่านั้น เมทา (Meta) การทำตัวแบบให้ดีขึ้นโดยการเรียนข้อมูล ไมล์ (Mise) วิธีการสร้างตัวแบบวิธีอื่น ต้นไม้ (Trees) และการสร้างตัวแบบโดยใช้ต้นไม้
การจัดกลุ่ม (Cluster)	เป็นฟังก์ชันที่เกี่ยวกับการทำเหมืองข้อมูลด้วยเทคนิคการจัดกลุ่ม (Clustering) โดยจะจัดกลุ่มข้อมูลที่มีลักษณะเหมือนกันเข้าไว้ด้วยกัน
ความสัมพันธ์ (Associate)	เป็นฟังก์ชันที่เกี่ยวกับการทำเหมืองข้อมูลด้วยเทคนิคการวิเคราะห์ความสัมพันธ์ เป็นการหาความสัมพันธ์ของข้อมูลที่เก็บอยู่ในฐานข้อมูล (Association rule) เพื่อดูว่าข้อมูลแบบไหนมีโอกาสเกิดร่วมกันบ่อย ๆ
การเลือกข้อมูล (Select attributes)	เป็นฟังก์ชันที่เกี่ยวกับเทคนิคการคัดเลือกข้อมูล เน้นการหาว่าตัวแปรไหนที่สำคัญและตัวแปรไหนที่ไม่สำคัญในชุดข้อมูลบ้าง ซึ่งตัวแปรที่ไม่สำคัญจะถูกกำจัดทิ้งไป ก่อนที่จะวิเคราะห์ด้วยข้อมูลวิธีต่างๆ
การนำเสนอข้อมูลด้วยรูปภาพ (Visualize)	เป็นฟังก์ชันที่เกี่ยวกับเทคนิคการนำเสนอข้อมูลด้วยรูปภาพ เป็นส่วนของการพลอตจุดข้อมูลใน 2 มิติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

โดยหลังจากการประมวลผลเสร็จเรียบร้อย ผลลัพธ์จะแสดงออกมาในรูปแบบตัวอักษร และมีรายละเอียดที่แตกต่างกันออกไป โดยตัวอย่างผลลัพธ์ทีแสดงออกมานั้น ดังรูปที่ 2.24



รูปที่ 2.24 ตัวอย่างผลลัพธ์ที่แสดงในโปรแกรมเวก้า

2.2 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องมี 3 งานวิจัย อันประกอบด้วยงานวิจัยที่เกี่ยวข้องกับการหาปัจจัยที่มีผลกับการรีวิวโค้ด ซึ่งเป็นงานวิจัยของ Daricélio M. Soares และคณะ ได้แก่ งานวิจัย “What Factor Influence The Reviewer Assignment To Pull Requests?” งานวิจัย “Acceptance Factors of Pull Requests in Open-Source Projects” และงานวิจัย “Rejection Factors of Pull Requests Field by Core Team Developers in Software Projects with High Acceptance Rates” มีรายละเอียดดังต่อไปนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2.1 งานวิจัย “What Factor Influence The Reviewer Assignment To Pull Requests?”

งานวิจัยของ Daricélio M. Soares และคณะ [3] นำเสนอปัจจัยที่มีอิทธิพลกับการกำหนดผู้รีวิวไปยังรายการร้องขอการนำเข้าซอร์สโค้ด โดยการใช้หลักการเหมืองข้อมูล วิธีการกฎความสัมพันธ์ (Association Rules) ในการวิเคราะห์ข้อมูลในอดีตของรายการร้องขอการนำเข้าซอร์สโค้ด 22,523 รายการ และพิจารณาภายใต้สามโครงการโอเพนซอร์สบนกิตฮับ รวมถึงมีการประยุกต์ใช้กระบวนการด้านฐานข้อมูล (Knowledge Discovery in Databases: KDD) มาประยุกต์ใช้

ขั้นตอนที่1 การเลือกข้อมูล (Data Selection) ทางผู้จัดทำมีการใช้เครื่องมือ GHTorrent Tool ในการดึงข้อมูลมาจากถังการเก็บซอร์สโค้ดบนกิตฮับ

ขั้นตอนที่2 การเตรียมข้อมูล (Pre-Processing) มีการจัดการข้อมูล เลือกลักษณะของข้อมูล รวมถึงลบข้อมูลที่ไม่ถูกต้อง (Incorrect) และไม่มีความต้องกัน (Inconsistency) ออกจากฐานข้อมูล

ขั้นตอนที่3 การแปลงและการตกแต่งข้อมูล โดยการแปลงข้อมูลบางส่วนให้เป็นข้อมูลหรือลักษณะที่เป็นตัวเลข (Numerical Attributes)

ขั้นตอนที่4 การใช้กฎความสัมพันธ์ (Association Rules Extraction) โดยมีการประยุกต์ใช้กฎความสัมพันธ์และอัลกอริทึมเอไพรออไร ในเครื่องมืออัตโนมัติเวก้า (Weka: Waikato Environment for Knowledge Analysis) ในการวิเคราะห์ข้อมูล

ขั้นตอนที่5 แปลงผลและวิเคราะห์ข้อมูล โดยการใช้เครื่องมือเสริม (WekaPar Plugin) เข้ามาช่วยในการวิเคราะห์และประเมินผล

จากการใช้กฎความสัมพันธ์ในการวิเคราะห์ข้อมูลพบว่า สามารถวิเคราะห์ข้อมูลรายการร้องขอการนำเข้าซอร์สโค้ดเพื่อหาปัจจัยที่มีอิทธิพลได้จริง โดยจากผลสรุปกล่าวว่า ปัจจัยเช่น จำนวนการคอมมิท (Number of Commits) และไฟล์ในรายการร้องขอ มีอิทธิพลกับผู้รีวิว ปัจจัยที่เกี่ยวกับประวัติของผู้ร้องขอมีอิทธิพลกับการจัดสรรของผู้รีวิว ความสัมพันธ์ระหว่างผู้ร้องขอและผู้รีวิว มีอิทธิพลกับผลการประเมิน โดยเมื่อผู้รีวิวรู้จักผู้ร้องขอ โอกาสการประเมินจะมีโอกาสผ่านสูงขึ้น และปัจจัยอื่นๆ เช่นความเป็นเจ้าของ (Ownership) หรือความท้องถิ่นของรายการร้องขอ (Locality of Pull Request) มีความสัมพันธ์กับการคาดเดาของผู้รีวิว

ผู้วิจัยได้แนวคิดการวิเคราะห์ปัจจัย เพื่อใช้ในการหาความสัมพันธ์ของตัวแปรจากงานวิจัยนี้ หากแต่งานวิจัยนี้ยังมีข้อจำกัด เนื่องจากไม่ได้มีการนำเอาปัจจัยต่าง ๆ ที่หาได้มาหาความสัมพันธ์และรวมเป็นหนึ่งสมการที่ระบุถึงทุกปัจจัยที่มีผลกระทบ อีกทั้งปัจจัยที่ถูกระบุในงานวิจัยนี้ยังเป็นปัจจัยที่เกี่ยวข้องกับผู้รีวิวที่เป็นคนให้ผลการรีวิวในรายการร้องขอการนำเข้าซอร์สโค้ด ไม่ใช่การวิเคราะห์โดยตรงจากรายการร้องขอดังกล่าว

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2.2 งานวิจัย “Acceptance Factors of Pull Requests in Open-Source Projects”

งานวิจัยของ Daricélio M. Soares และคณะ [4] นำเสนอปัจจัยการยอมรับรายการร้องขอการนำเข้าซอร์สโค้ดในโครงการโอเพนซอร์สโดยการใช้หลักการเหมืองข้อมูล วิธีการกฎความสัมพันธ์ (Association Rules) ในการวิเคราะห์ข้อมูลที่บันทึกไว้ในอดีตของรายการร้องขอการนำเข้าซอร์สโค้ด อีกทั้งยังเพื่อหารูปแบบ (Patterns) ใหม่และเป็นประโยชน์ของการเกิดปัจจัยดังกล่าว โดยจากผลลัพธ์ของงานวิจัยพบว่า การใช้กฎของความสัมพันธ์สามารถช่วยระบุปัจจัยที่เพิ่มโอกาสในการนำเข้าซอร์สโค้ด การระบุลักษณะประจำที่มีอิทธิพลกับการร้องขอการนำเข้าซอร์สโค้ด รวมถึงเพิ่มความรู้ทางด้านการใช้การนำเข้าซอร์สโค้ด อีกทั้งยังมีการนำมาประยุกต์เพื่อหาว่าปัจจัยอะไรที่ทำให้การนำเข้าซอร์สโค้ดรวดเร็วมากยิ่งขึ้น

ข้อจำกัดของงานวิจัยนี้คือ ไม่ได้มีการนำเอาปัจจัยต่าง ๆ ที่หาได้มาหาความสัมพันธ์ และรวมเป็นหนึ่งสมการที่ระบุถึงทุกปัจจัยที่มีผลกระทบกับการยอมรับการนำเข้าซอร์สโค้ด อีกทั้งการวิเคราะห์ที่ได้ยังเป็นการนำเสนอปัจจัยจำนวนน้อย เพียงไม่กี่ปัจจัยเท่านั้น เนื่องจากข้อมูลที่นำมาพิจารณาไม่ได้มากพอที่จะทำให้หารูปแบบความสัมพันธ์จำนวนมากได้

2.2.3 งานวิจัย “Rejection Factors of Pull Requests Field by Core Team Developers in Software Projects with High Acceptance Rates”

งานวิจัยของ Daricélio M. Soares และคณะ [5] นำเสนอปัจจัยการปฏิเสธรายการร้องขอการนำเข้าซอร์สโค้ดโดยนักพัฒนาหลักในโครงการซอฟต์แวร์ที่มีอัตรายการยอมรับสูง โดยการใช้หลักการเหมืองข้อมูล วิธีการกฎความสัมพันธ์ (Association Rules) ในการวิเคราะห์ข้อมูลในอดีตของรายการร้องขอการนำเข้าซอร์สโค้ด กระบวนการวิจัยเป็นวิธีการเดียวกันกับงานวิจัย [3] และ [4] หากแต่เป็นการพิจารณาปัจจัยที่เกี่ยวข้องกับการปฏิเสธการร้องขอโดยนักพัฒนาหลัก อีกทั้งโครงการตัวอย่างยังเป็นโครงการที่มีอัตรายการยอมรับสูงอีกด้วย

การนำเข้าข้อมูลเข้ามาใช้มีการใช้เครื่องมือ GHTorrent Tool เช่นเดียวกับงานวิจัยก่อนหน้า หากแต่มีการแตกรายละเอียดของรายการร้องขอการนำเข้าซอร์สโค้ด 20,140 รายการโดยนักพัฒนาซอฟต์แวร์หลัก รวมถึงเป็นการรวบรวมข้อมูลจาก 7 โครงการ (Akka, Bitcoin, Brackets, Commcare-hq, IPython Katello และ Kuma) โดยจากข้อมูลทั้ง 7 โครงการ มีอัตรายการยอมรับสูงถึง 93.43% และมีการปฏิเสธเพียงแค่ 6.58% โดยรายการร้องขอการนำเข้าซอร์สโค้ดที่ถูกปฏิเสธเป็นรายการร้องขอการนำเข้าซอร์สโค้ดของนักพัฒนาภายนอกโครงการ (หรือกล่าวคือนักพัฒนาที่ไม่ใช่นักพัฒนาหลัก) ทั้งหมด รายละเอียดดังตารางที่ 2.3

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 2.3 ลักษณะพิเศษของโครงการที่นำมาพิจารณา [5]

Project	Language	Pull requests	Merged (%)	Closed (%)
Akka	Scala	954	86.16	13.84
Bitcoin	C++	1563	83.37	16.63
Brackets	Java Script	3685	93.05	6.95
Commcare-hq	Python	6961	98.85	1.15
IPython	Python	3163	84.71	10.28
Katello	Ruby	1046	94.36	5.64
Kuma	HTML	2768	92.30	7.70

โดยลักษณะประจำที่นำมาพิจารณาในการวิเคราะห์ ได้แก่ รหัสโครงการ (project_id) รายการร้องขอรายการแรก (first_pull) เพื่อใช้ในการพิจารณาว่ารายการร้องขอนี้เป็นรายการร้องขอรายการแรกของผู้ร้องขอหรือไม่ จำนวนการคอมมิทต่อหนึ่งรายการร้องขอ (commits_pull) จำนวนไฟล์ที่มีการเพิ่ม ลบ และแก้ไข (files_changed) ไฟล์ที่มีการเปลี่ยนแปลง (files) ตำแหน่งที่มีการเปลี่ยนแปลง (directory) และค่าสุดท้ายที่นำมาพิจารณาคือค่าการยอมรับการนำเข้าซอร์สโค้ด หรือ Accepted โดยในงานวิจัยนี้จะให้ค่าเป็น Merged หรือการปฏิเสธ หรือ Rejected โดยงานวิจัยนี้จะให้ค่าเป็น Closed

จากบทสรุปงานวิจัย ได้มีการพบว่า ประสบการณ์ของผู้ร้องขอ พื้นที่ และขนาดของการนำเข้า มีอิทธิพลอย่างสูงกับการปฏิเสธการนำเข้า รวมถึงการรวมกันของฟีเจอร์ในรายการร้องขอการนำเข้าเพิ่มโอกาสการถูกปฏิเสธได้ และการระบุโดเมนของโครงการมีผล ไม่ว่าจะเป็นการให้คำแนะนำ รวมถึงการทดสอบ ล้วนนำไปสู่การปฏิเสธทั้งสิ้น

งานวิจัยนี้มีการระบุรายละเอียดในการวิเคราะห์ข้อมูลการร้องขอการนำเข้าซอร์สโค้ดที่ดี มีการใช้กฎความสัมพันธ์ในการวิเคราะห์ความสัมพันธ์ของข้อมูลรายการร้องขอการนำเข้าซอร์สโค้ดในอดีต หากแต่ข้อมูลที่นำมาพิจารณา เป็นเพียงการปฏิเสธจากนักพัฒนาหลักเท่านั้น ไม่ใช่การปฏิเสธทั้งหมดของรายการร้องขอการนำเข้าซอร์สโค้ด อีกทั้งยังเป็นเพียงการหาปัจจัยเพียงอย่างเดียว ไม่มีการนำปัจจัยเหล่านั้นมาหาความสัมพันธ์โดยรวม เพื่อสรุปเป็นผลกระทบของปัจจัยหลาย ๆ ปัจจัยรวมกัน

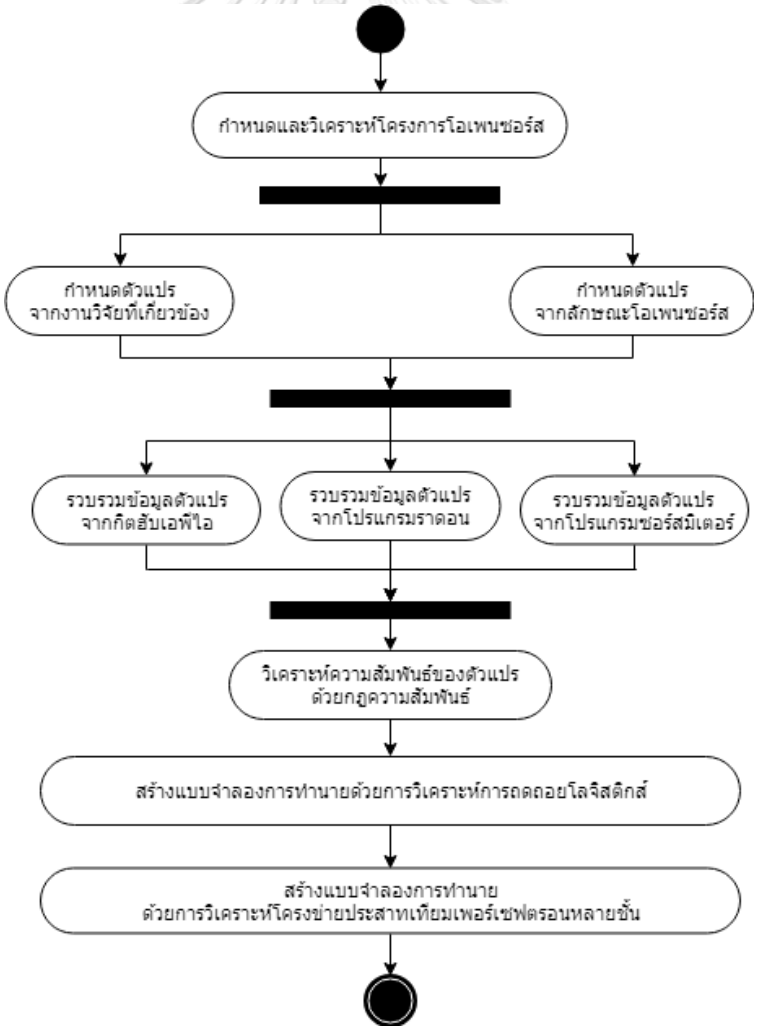
จากการพิจารณางานวิจัยที่เกี่ยวข้องทั้งหมดพบว่า งานวิจัยที่เกี่ยวข้องมีการหาปัจจัยที่มีส่วนเกี่ยวข้องกับการรีวิวโค้ดบนกิตฮับ หากแต่การหาปัจจัยนั้น เป็นเพียงการหาปัจจัยเพื่อนำไปใช้ในการป้องกัน หรือการพัฒนาซอฟต์แวร์ในการร้องขอการนำเข้าซอร์สโค้ดในครั้งถัด ๆ ไป หากแต่ยังขาดในส่วนของการหาความสัมพันธ์ของปัจจัยดังกล่าว เพื่อนำไปใช้ในการทำนายผลการรีวิว รวมถึงการสร้างสมการเพื่อใช้ในการทำนาย งานวิจัยนี้จึงจัดทำขึ้นมาเพื่อนำเอาตัวแปรที่ได้รวบรวมมาทั้งจากงานวิจัยที่อ้างอิง งานวิจัยในอดีต และตัวแปรที่เป็นลักษณะเฉพาะของโครงการโอเพนซอร์สที่นำมาพิจารณา เพื่อนำมาหาความสัมพันธ์และสร้างเป็นสมการทำนายเพื่อใช้ในการทำนายผลการรีวิวโค้ด

บทที่ 3
แบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

บทนี้กล่าวถึงรายละเอียดและขั้นตอนในการสร้างแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับเพื่อใช้ในการทำนายผลการรีวิวโค้ดก่อนนักพัฒนาซอฟต์แวร์ทำการร้องขอการนำซอร์สโค้ดเข้าสู่ระบบ โดยจะกล่าวถึงภาพรวมของวิธีการออกแบบและสร้างแบบจำลอง จากนั้นจึงอธิบายรายละเอียดของแต่ละขั้นตอน โดยมีรายละเอียดดังนี้

3.1 ภาพรวมวิธีการออกแบบการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

งานวิจัยนี้นำเสนอการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ โดยภาพรวมวิธีการออกแบบและการสร้างแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับนั้น ดังรูปที่ 3.1



รูปที่ 3.1 ภาพรวมวิธีการสร้างแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

จากรูปที่ 3.1 ภาพรวมวิธีการสร้างแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ ประกอบไปด้วย 7 ขั้นตอนด้วยกัน ได้แก่

ขั้นตอนที่ 1 กำหนดและวิเคราะห์โครงการโอเพนซอร์สที่นำมาพิจารณาในงานวิจัย มีจุดประสงค์เพื่อกำหนดขอบเขตการวิเคราะห์ข้อมูล และศึกษารายละเอียด รวมถึงลักษณะพิเศษ (Characteristic) ของโครงการโอเพนซอร์สที่แตกต่างกันออกไป โดยรายละเอียดของขั้นตอนนี้ ได้อธิบายไว้ในหัวข้อ 3.2.1

ขั้นตอนที่ 2 กำหนดตัวแปรที่คาดว่าจะมีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ มีจุดประสงค์ในการกำหนดตัวแปรเพื่อใช้ในการวิเคราะห์ข้อมูล โดยวิธีการในการกำหนดตัวแปรทำได้ 2 วิธีด้วยกัน คือการกำหนดตัวแปรจากงานวิจัยที่เกี่ยวข้อง และการกำหนดตัวแปรจากลักษณะของโครงการโอเพนซอร์ส รายละเอียดของการกำหนดตัวแปรอธิบายในหัวข้อ 3.2.2

ขั้นตอนที่ 3 รวบรวมข้อมูลตัวแปรที่มีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ เพื่อให้ได้ข้อมูลหรือค่าของตัวแปรว่าแต่ละตัวแปรมีค่าเท่าใดเพื่อใช้ในการวิเคราะห์ต่อไป โดยวิธีการในการรวบรวมแบ่งเป็น 3 วิธีการด้วยกัน คือการรวบรวมข้อมูลผ่านเอพีไอบนกิตฮับ การรวบรวมข้อมูลจากโปรแกรมราดอน และการรวบรวมข้อมูลผ่านโปรแกรมซอร์สมิเตอร์ รายละเอียดของการรวบรวมข้อมูลตัวแปรอธิบายไว้ในหัวข้อ 3.2.3

ขั้นตอนที่ 4 วิเคราะห์ความสัมพันธ์ของตัวแปรด้วยกฎความสัมพันธ์ เพื่อหารูปแบบของความสัมพันธ์ที่เกิดบ่อยในชุดของข้อมูล โดยรายละเอียดของขั้นตอนนี้ ได้อธิบายไว้ในหัวข้อ 3.2.4

ขั้นตอนที่ 5 สร้างแบบจำลองการทำนายด้วยการวิเคราะห์การถดถอยโลจิสติก ซึ่งเป็นหลักการทางสถิติ โดยรายละเอียดของขั้นตอนนี้ ได้อธิบายไว้ในหัวข้อ 3.2.5

ขั้นตอนที่ 6 สร้างแบบจำลองการทำนายด้วยการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น ซึ่งเป็นหลักการเรียนรู้ของเครื่อง โดยรายละเอียดของขั้นตอนนี้ ได้อธิบายไว้ในหัวข้อ 3.2.6

3.2 วิธีการสร้างแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

3.2.1 กำหนดและวิเคราะห์โครงการโอเพนซอร์สที่นำมาวิเคราะห์ในงานวิจัย

เนื่องจากมีโครงการโอเพนซอร์สจำนวนมากที่ถูกจัดเก็บอยู่ในกิตฮับซึ่งเป็นระบบควบคุมเวอร์ชันแบบกระจายศูนย์ [9] ในรูปแบบเว็บไซต์ที่ใช้ในการเก็บข้อมูลและควบคุมเวอร์ชันของโอเพนซอร์สที่เยอะที่สุดในโลก จึงต้องมีการกำหนดโครงการโอนเพนซอร์สที่จะใช้ในการวิเคราะห์ เพื่อจำกัดโครงการโอเพนซอร์สที่ใช้ในการวิเคราะห์ข้อมูล โดยพิจารณาจากผลการวิเคราะห์การใช้งานกิตฮับในปี พ.ศ. 2560 หรือ ค.ศ. 2017 ที่ผ่านมาจากการวิเคราะห์ในโครงการ The State of The Octoverse 2017 (https://octoverse.github.com/) จากผลการวิเคราะห์ถึง 10 อันดับแรกของโครงการโอเพนซอร์ส พบว่ามีจำนวนการรีวิวโค้ดผ่านกิตฮับรวมกันสูงถึง 5,090 การรีวิว โดย

รายละเอียดของโครงการโอเพนซอร์สที่ติด 10 อันดับแรกที่นำมาพิจารณา รายละเอียดดังตารางที่ 3.1

ตารางที่ 3.1 รายละเอียดโครงการโอเพนซอร์สสิบอันดับที่มีการรีวิวสูงสุดในกิตฮับ

อันดับ	ชื่อโครงการ	ศูนย์ควบคุมที่เป็นที่นิยม	การรีวิว	ศูนย์ควบคุม	ผู้ดูแลหลัก
1	DefinitelyTyped	DefinitelyTyped/DefinitelyTyped	800	15	12
2	Kubernetes	Kubernetes/kubernetes	680	56	280
3	Homebrew	Homebrew/Homebrew-core	580	37	17
4	Ansible	Ansible/Ansible	550	31	62
5	Node.js Foundation	Nodejs/node	480	1,532	306
6	NixOS	NixOs/nixpkgs	480	31	62
7	The Apache Software Foundation	Apache/spark	450	1,532	470
8	The Rust Programming Language	Rust-lang/rust	390	32	62
9	Symfony	Symfony/symfony	340	109	22
10	Tensorflow	Tensorflow/tensorflow	340	41	136

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

โดยหลังจากได้รายการโครงการโอเพนซอร์สสิบอันดับที่มีการรีวิวสูงสุดมาแล้ว ขั้นตอนต่อมาคือการนำเอาโครงการเหล่านั้นมาวิเคราะห์ข้อมูลหรือดูรายละเอียดเพิ่มเติม เพื่อเลือกโครงการโอเพนซอร์สที่จะนำมาวิเคราะห์ โดยรายละเอียดโครงการโอเพนซอร์สสิบอันดับที่มีการรีวิวสูงสุดในกิตฮับเพิ่มเติม ได้แก่ จำนวนการรีวิว (Number of reviews) จำนวนประเด็นปัญหา (Number of issues) จำนวนการร้องขอการนำเข้าซอร์สโค้ด (Number of pull requests) จำนวนโครงการ (Number of repository) จำนวนการคอมมิท (Number of commit) จำนวนสาขา (Number of branch) จำนวนเวอร์ชันที่มีการออกสู่ตลาด (Number of release) และจำนวนผู้สนับสนุน (Number of contributions) รายละเอียดดังตารางที่ 3.2

ตารางที่ 3.2 รายละเอียดโครงการโอเพนซอร์สสิบอันดับที่มีการรีวิวสูงสุดในกิตฮับ

#	ศูนย์ควบคุมที่เป็นที่นิยม	การรีวิว	ประเด็นปัญหา	การร้องขอการนำเข้า	โครงการ	คอมมิท	สาขา	เวอร์ชันที่ออกสู่ตลาด	ผู้สนับสนุน
1	DefinitelyTyped/ DefinitelyTyped	800	1,985	68	1	43,942	59	0	5,952
2	Kubernetes/ kubernetes	680	2,372	1,067	11	64,232	38	373	1,645
3	Homebrew/ homebrew-core	580	11	66	0	112,720	0	0	6,907
4	Ansible/Ansible	550	3,561	1,431	21	36,762	204	204	3,391
5	Nodejs/node	480	597	128	7	21,946	487	487	1,964
6	NixOS/nixpkgs	480	2,358	832	11	137,257	87	87	1,597
7	Apache/spark	450	N/A	516	0	21,737	59	59	1,218
8	Rust-lang/rust	390	3,547	144	1	76,922	53	53	2,056
9	Symfony/ symfony	340	664	159	3	36,372	338	338	1,622
10	Tensorflow/ tensorflow	340	1,316	200	0	31,623	54	54	1,432

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จากตารางที่ 3.2 ซึ่งเป็นตารางการสรุปรายละเอียดดังกล่าว ทางผู้วิจัยได้คัดเลือกโครงการที่มีจำนวนศูนย์ควบคุมเวอร์ชันสูงสุดสามอันดับออก เนื่องจากการวิเคราะห์โครงการโอเพนซอร์สในทุก ๆ ศูนย์ควบคุมเวอร์ชันจะช่วยป้องกันปัญหาการเกิดข้อมูลที่ผิดปกติ หรือสามารถควบคุมช่วงของข้อมูลได้ จึงคัดโครงการโอเพนซอร์สอื่น ๆ ออกจากการวิเคราะห์ โดยโครงการโอเพนซอร์สที่คัดออกคือโครงการ symphony (อันดับ 9) โครงการ nodejs.foundation (อันดับ 5) และโครงการ The Apache software foundation (อันดับ 7) และคัดบางโครงการยังไม่มีการเปิดเผยซอฟต์แวร์สู่การตลาดเลย ถึงแม้จะเป็นโครงการที่เป็นที่รู้จัก หากแต่ไม่ได้รับการพูดคุยในวงกว้าง คือโครงการ Definitely/Typed (อันดับ 1) และ โครงการ Homebrew (อันดับ 3)

จากการคัดโครงการโอเพนซอร์สออกไปแล้วห้าโครงการ หากพิจารณาถึงโครงการที่หลือนั้น พบว่าทุกโครงการมีความน่าสนใจแตกต่างกัน หากแต่ทางผู้วิจัยได้เลือก โครงการ Ansible เป็นโครงการที่ดำเนินการวิจัย เนื่องจากโครงการ Ansible มีอันดับการรีวิวสูงถึงอันดับที่ 4 ในปี ค.ศ. 2017 คือมีจำนวนการรีวิวทั้งหมด 550 การรีวิว อีกทั้งยังมีจำนวนจำนวนศูนย์ควบคุมเวอร์ชันสูงสุดในห้าโครงการที่เหลืออยู่ อีกทั้งหากพิจารณาผลการวิเคราะห์การใช้งานกิตฮับในปีพ.ศ. 2560 หรือ ค.ศ. 2017 เพิ่มเติม พบว่าโครงการ Ansible เป็นโครงการที่มีผู้สนับสนุนสูงสุดเป็นอันดับ 10 โดยมีจำนวนสูงถึงห้าพันกว่าคน และมีจำนวนการปรึกษาหารือกันในโครงการสูงสุดเป็นอันดับ 10 เป็นจำนวนสูงถึงห้าหมื่นกว่าการหารือ อีกทั้งโครงการ Ansible ยังเป็นโครงการโอเพนซอร์สทางซอฟต์แวร์ที่มีชื่อเสียงในกิตฮับ เป็นแพลตฟอร์มที่ใช้ในการอัตโนมัติ ที่จะทำให้แอปพลิเคชันและระบบง่ายในการขึ้นระบบ ซึ่ง Ansible เป็นโครงการโอเพนซอร์สที่มีผู้ใช้กิตฮับให้คะแนนความนิยมถึงสองหมื่นกว่าราย อีกทั้งยังมีจำนวนการนำเข้าซอร์สโค้ดถึงสามหมื่นกว่าครั้ง และมีจำนวนระบบการเก็บข้อมูล สูงถึงแปดสิบโครงการภายใต้โครงการโอเพนซอร์สนี้ ด้วยเหตุผลนี้ทางผู้วิจัยจึงเลือกโครงการ Ansible เป็นโครงการโอเพนซอร์สที่นำมาพิจารณา

หลังจากที่กำหนดโครงการ Ansible เป็นโครงการโอเพนซอร์สที่นำมาวิเคราะห์ข้อมูลแล้ว ขั้นตอนต่อมาคือ การวิเคราะห์โครงการโอเพนซอร์สที่กำหนดเพื่อให้ได้มาซึ่งรายละเอียดของโครงการโอเพนซอร์สนั้น ๆ กระบวนการทำงานของโครงการโอเพนซอร์ส แผ่นป้ายที่แสดงผลลัพธ์ที่ได้จากการรีวิวของรายการร้องขอการนำเข้าซอร์สโค้ดทั้งหมด รายการเอพีไอที่ใช้ในการวิเคราะห์รายละเอียดของโครงการ อีกทั้งยังมีการศึกษาเพิ่มเติมในส่วนของแผ่นป้ายที่ใช้ในการรีวิว และเหตุผลเบื้องหลังในแปะแผ่นป้ายนั้น ๆ โดยวิธีการในการวิเคราะห์โครงการโอเพนซอร์สแบ่งเป็น 3 วิธีการ ได้แก่ การวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านคู่มือการพัฒนา Ansible การวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านกิตฮับ และการวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านโปรแกรมโพสต์แมน มีภาพรวมวิธีการวิเคราะห์โครงการโอเพนซอร์ส แสดงดังตารางที่ 3.3

ตารางที่ 3.3 ภาพรวมวิธีการวิเคราะห์โครงการโอเพนซอร์ส

ลำดับ	รายการกิจกรรม	ข้อมูลนำเข้า	ข้อมูลนำออก
1	การวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านคู่มือการพัฒนา Ansible	- โครงการโอเพนซอร์สที่วิเคราะห์ - คู่มือพัฒนา	- กระบวนการทำงานของโครงการโอเพนซอร์ส

ตารางที่ 3.3 ภาพรวมวิธีการวิเคราะห์โครงการโอเพนซอร์ส (ต่อ)

ลำดับ	รายการกิจกรรม	ข้อมูลนำเข้า	ข้อมูลนำออก
2	การวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านกิตฮับ	- เว็บไซต์กิตฮับที่ระบุรายละเอียดของการพัฒนาโครงการโอเพนซอร์สบนกิตฮับ - เว็บไซต์ของโครงการโอเพนซอร์สที่วิเคราะห์บนกิตฮับ	- รายการเอพีไอ ที่ใช้ในการวิเคราะห์รายละเอียดโครงการโอเพนซอร์ส - โครงสร้างกระบวนการทำงานของโครงการโอเพนซอร์ส - แผ่นป้ายที่ใช้ในโครงการโอเพนซอร์สและความหมายของแผ่นป้าย - แผ่นป้ายที่แสดงผลลัพธ์ที่ได้จากการรีวิวของรายการร้องขอการนำเข้าซอร์สโค้ด
3	การวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านโปรแกรมโพสต์แมน	- โครงการโอเพนซอร์สที่วิเคราะห์ - โปรแกรมโพสต์แมน	- ผลลัพธ์การทดสอบการใช้เอพีไอที่ใช้ในการวิเคราะห์รายละเอียดโครงการโอเพนซอร์ส

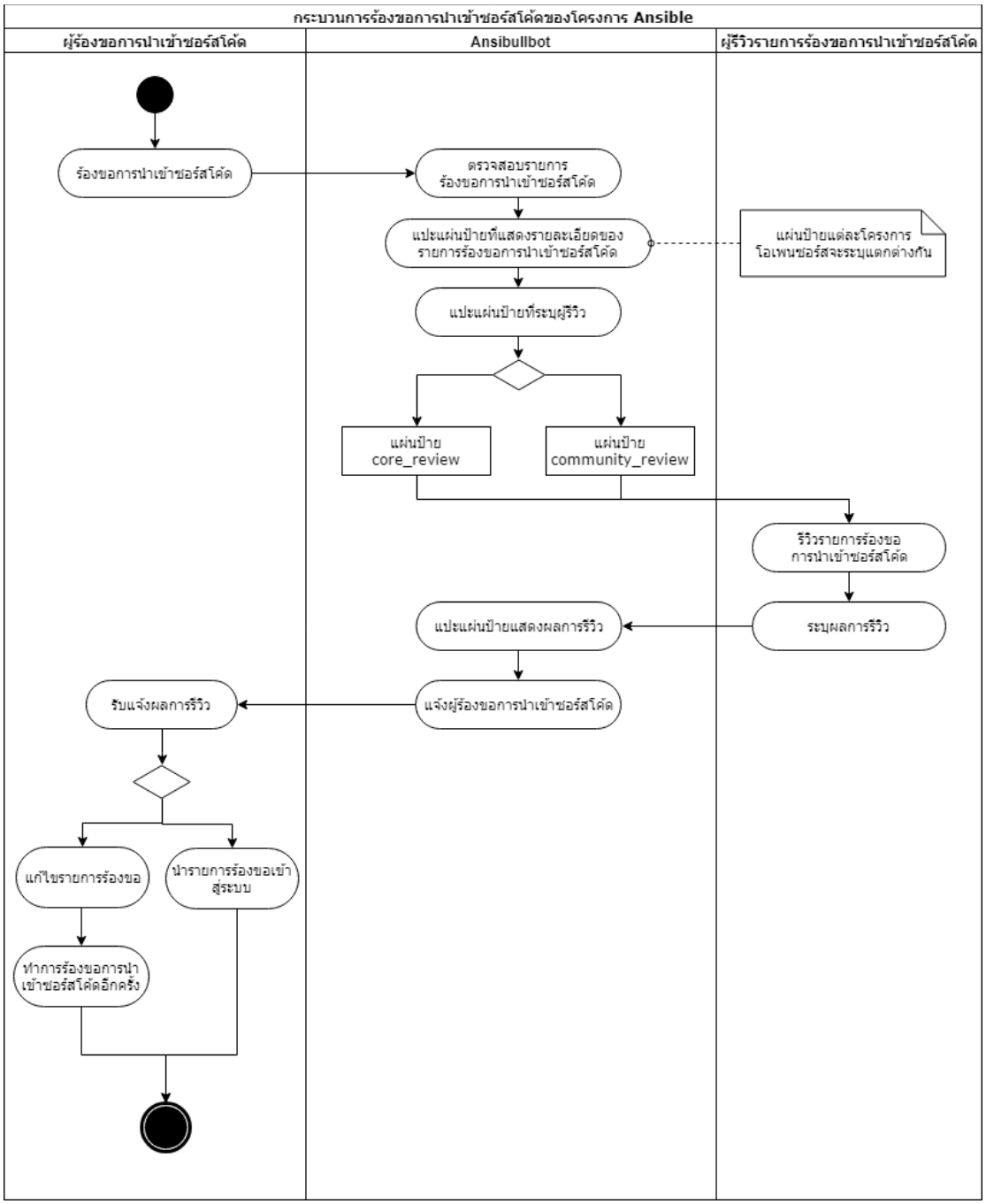
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

โดยวิธีการวิเคราะห์โครงการโอเพนซอร์สมีรายละเอียดดังต่อไปนี้

(1) การวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านคู่มือการพัฒนาของ Ansible

คู่มือการพัฒนาโครงการ Ansible ที่สร้างไว้เพื่อการศึกษาของนักพัฒนาซอฟต์แวร์ มีรายละเอียดเกี่ยวกับโครงการที่เป็นประโยชน์กับงานวิจัยมากมาย ไม่ว่าจะเป็นรายละเอียดทั่วไปของซอฟต์แวร์ Ansible กระบวนการทำงานของโครงการโอเพนซอร์สรายละเอียดของการร้องขอการนำเข้าซอร์สโค้ดผ่านกิตฮับ ภาษาที่ใช้ในการพัฒนา รายละเอียดเกี่ยวกับเครื่องมืออัตโนมัติที่ Ansible ใช้ในการดำเนินงานในกิตฮับ ที่มีชื่อเรียกว่า Ansibullbot อีกทั้งยังมีคำแนะนำทั่วไปสำหรับนักพัฒนาซอฟต์แวร์ที่สนใจในโครงการ ซึ่งจากการวิเคราะห์รายละเอียดโครงการ Ansible ผ่านคู่มือการพัฒนาพบว่าหัวข้อ Pull Requests ที่อยู่ในหัวข้อ The Ansible Development Process ในคู่มือการพัฒนานั้นเป็นหัวข้อที่เกี่ยวกับงานวิจัยนี้โดยตรง โดย Ansible จะยอมรับซอร์สโค้ดที่ทำการ

เปลี่ยนแปลงหรือแก้ไขโดยตรงผ่านเมนู Pull Requests บนกิตฮับ ซึ่งมีตัวย่อว่า PRs หรือชื่อเรียกภาษาไทยว่า รายการร้องขอการนำเข้าซอร์สโค้ด และเนื่องจากโครงการ Ansible มีรายการร้องขอการนำเข้าซอร์สโค้ดเป็นจำนวนมาก จึงต้องมีการจัดการรายการร้องขอการนำเข้าซอร์สโค้ดต่าง ๆ ผ่านเครื่องมืออัตโนมัติที่เรียกว่า Ansibullbot ดังที่กล่าวไปข้างต้น โดย Ansibullbot ช่วยในการดำเนินกระบวนการรีวิวซอร์สโค้ด และการรวมซอร์สโค้ด (Merge) เข้าสู่ระบบ ซึ่งกระบวนการทำงานที่ระบุในคู่มือการพัฒนานั้น มีรายละเอียดดังรูปที่ 3.2



รูปที่ 3.2 กระบวนการร้องขอการนำเข้าซอร์สโค้ดของโครงการ Ansible

รูปที่ 3.2 แสดงกระบวนการร้องขอการนำเข้าซอร์สโค้ดของโครงการ Ansible ซึ่งประกอบด้วย 11 ขั้นตอนด้วยกัน มีรายละเอียดดังนี้

ขั้นตอนที่ 1 นัฒพัฒนาซอฟต์แวร์หรือผู้ร้องขอการนำเข้าซอร์สโค้ด ทำการร้องขอการนำเข้าซอร์สโค้ดที่ได้ทำการแก้ไขหรือเพิ่มเติมจากซอร์สโค้ดที่ดาวน์โหลดมาจากถังเก็บข้อมูลหลัก

ขั้นตอนที่ 2 ระบบอัตโนมัติ Ansibullbot จะดำเนินการอย่างต่อเนื่องในโครงการโอเพนซอร์สโดยนักพัฒนาซอฟต์แวร์จะเห็นการเปลี่ยนแปลงหรือการไหลของกระบวนการร้องขอการนำเข้าซอร์สโค้ดอย่างต่อเนื่อง ภายใน 30 นาทีหลังทำการร้องขอการนำเข้าซอร์สโค้ด โดย Ansibullbot จะทำการตรวจสอบทุก ๆ รายการร้องขอการนำเข้าซอร์สโค้ดที่เปิดขึ้นมา

ขั้นตอนที่ 3 ระบบอัตโนมัติ Ansiblebullbot จะทำการแปะแผ่นป้ายการรีวิวที่แสดงรายละเอียดของรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ๆ เช่นหากรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ๆ เกี่ยวกับระบบคลาวด์ (Cloud) ระบบอัตโนมัติจะแปะแผ่นป้าย Cloud เป็นต้น

ขั้นตอนที่ 4 ระบบอัตโนมัติจะทำการแปะแผ่นป้ายที่ระบุผู้รีวิวโดยพิจารณาโมดูลของผู้ร้องขอการนำเข้าซอร์สโค้ดหากผู้ร้องขอการนำเข้าซอร์สโค้ดไม่ได้อยู่ในทีมงานพัฒนาหลักของโครงการ Ansible ระบบจะย้ายสถานะของรายการร้องขอการนำเข้าซอร์สโค้ดไปยังสถานะการรีวิวโดยประชาคม (community_review) และทำการแปะแผ่นป้าย community_review แต่หากนักพัฒนาซอฟต์แวร์อยู่ในทีมงานพัฒนาหลักของโครงการ Ansible ระบบจะย้ายสถานะของรายการร้องขอการนำเข้าซอร์สโค้ดไปยังสถานะการรีวิวโดยหลัก (core_review) และทำการแปะแผ่นป้าย core_review

ขั้นตอนที่ 5 ผู้รีวิวรายการร้องขอการนำเข้าซอร์สโค้ด จะทำการรีวิวรายการร้องขอการนำเข้าซอร์สโค้ดที่ส่งมาโดยพิจารณาจากซอร์สโค้ดที่นำเข้า และรายละเอียดของรายการร้องขอการนำเข้าซอร์สโค้ดดังกล่าว

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ขั้นตอนที่ 6 ผู้รีวิวจะทำการระบุผลการรีวิวรายการร้องขอการนำเข้าซอร์สโค้ด และทำการระบุผลรีวิวดังกล่าว โดยผลการรีวิวมีอยู่สองผลลัพธ์ ได้แก่ ผลการรีวิวเป็นผ่าน คืออนุญาตให้นำซอร์สโค้ดเข้าสู่ระบบ และผลการรีวิวเป็นไม่ผ่าน คือไม่อนุญาตให้นำซอร์สโค้ดเข้าสู่ระบบ

ขั้นตอนที่ 7 ระบบอัตโนมัติ Ansibullbot จะทำการแปะแผ่นป้ายแสดงผลการรีวิว โดยแผ่นป้ายที่อยู่ในขั้นตอนนี้ ประกอบไปด้วย 12 แผ่นป้าย ซึ่งอธิบายรายละเอียดในตารางที่ 3.5

ขั้นตอนที่ 8 ระบบอัตโนมัติ Ansibullbot จะทำการแจ้งผู้ร้องขอการนำเข้าซอร์สโค้ดถึงผลลัพธ์การรีวิวผ่านแผ่นป้าย

ขั้นตอนที่ 9 ผู้รีวิวรับแจ้งผลการรีวิว ซึ่งประกอบไปด้วย 2 ผลลัพธ์ด้วยกัน คือผลการรีวิวเป็นผ่านและผลการรีวิวเป็นไม่ผ่าน

ขั้นตอนที่ 10 หากรายการร้องขอการนำเข้าซอร์สโค้ดมีผลการรีวิวเป็นผ่าน ผู้ร้องขอการนำเข้าซอร์สโค้ดจะนำรายการร้องขอการนำเข้าซอร์สโค้ดนั้นเข้าสู่ระบบ และหากรายการร้องขอการนำเข้า

ซอร์สโค้ดนั้นไม่ผ่าน ผู้ร้องขอการนำเข้าซอร์สโค้ดจะต้องทำการแก้ไขรายการร้องขอดังกล่าว และทำการร้องขอการนำเข้าซอร์สโค้ดอีกครั้ง โดยการแก้ไขจะเป็นไปตามรายละเอียดในแผ่นป้าย หรือความคิดเห็นที่ผู้รีวิวระบุไว้ ตัวอย่างเช่นหากรายการร้องขอการนำเข้าซอร์สโค้ดใดที่มีการแปะแผ่นป้ายต้องการข้อมูล ผู้ร้องขอการนำเข้าซอร์สโค้ดจะต้องทำการเพิ่มเติมข้อมูลก่อนที่จะทำการร้องขอการนำเข้าซอร์สโค้ดอีกครั้ง

(2) การวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านกิตฮับ

เว็บไซต์กิตฮับ มีคู่มือระบุชัดเจนถึงรายการเอพีไอที่เปิดให้นักพัฒนาซอฟต์แวร์เรียกใช้ เพื่อเป็นการดึงข้อมูลบนกิตฮับไปประมวลผลต่อ โดยรูปแบบการเรียกใช้งานเอพีไอที่ใช้ในงานวิจัยนี้คือเรส (REST) หรือ Representational State Transfer ซึ่งเป็นการเรียกเซอร์วิสผ่านยูอาร์แอล (URL) โดยคำสั่งที่ใช้ในงานวิจัยนี้ทั้งหมด รายละเอียดดังตารางที่ 3.4

ตารางที่ 3.4 รายการเอพีไอที่ใช้ในการวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านกิตฮับ

ลำดับ	คำสั่ง	วัตถุประสงค์ในการเรียกใช้เอพีไอ	ผลลัพธ์ที่ได้
1	GET single pull request	เป็นคำสั่งเอพีไอที่ใช้เพื่อดึงรายละเอียดของรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ๆ	- จำนวนคอมมิท - รายการคอมมิท - ผู้คอมมิท - วันและเวลาที่สร้างรายการร้องขอการนำเข้าซอร์สโค้ด
2	List commit	เป็นคำสั่งเอพีไอที่ใช้เพื่อดึงค่าชาร์ (sha) ซึ่งเป็นค่าเฉพาะของแต่ละคอมมิท และเป็นค่าที่ต้องนำไปใช้ในการร้องขอคำสั่ง Single commit ต่อไป	- ค่าชาร์ หรือค่าเฉพาะของแต่ละคอมมิท
3	Single commit	เป็นคำสั่งเอพีไอที่ใช้เพื่อดึงรายละเอียดของแต่ละคอมมิท	- จำนวนไฟล์ที่มีการเปลี่ยนแปลง - ชื่อไฟล์ที่มีการเปลี่ยนแปลง - จำนวนชังก์ - จำนวนบรรทัดที่มีการเปลี่ยนแปลง

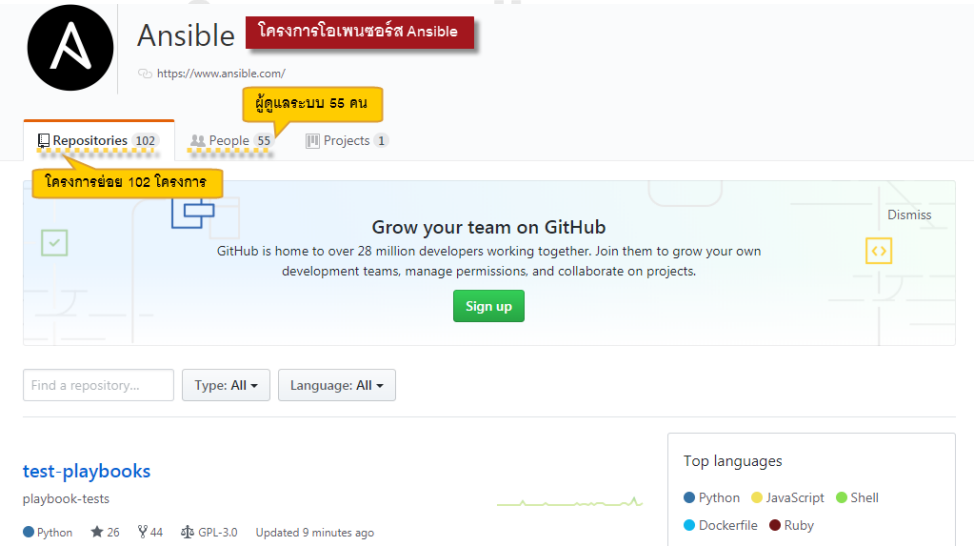
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.4 รายการเอพีไอที่ใช้ในการวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านกิตฮับ (ต่อ)

ลำดับ	คำสั่ง	วัตถุประสงค์ในการเรียกใช้เอพีไอ	ผลลัพธ์ที่ได้
4	Single commit comment	เป็นคำสั่งเอพีไอที่ใช้เพื่อดึงคอมเม้นของแต่ละคอมมิท	- คอมเม้นที่อยู่ในคอมมิท
5	List label	เป็นคำสั่งเอพีไอที่ใช้เพื่อดึงแผ่นป้ายทั้งหมดของรายการร้องขอการนำเข้าซอร์สโค้ด	- แผ่นป้ายของรายการร้องขอการนำเข้าซอร์สโค้ด
6	Issue event	เป็นคำสั่งเอพีไอที่ใช้ในการหาแผ่นป้ายที่ถูกสร้างและลบทิ้งไปแล้ว ซึ่งไม่สามารถหาได้จากการใช้คำสั่งเอพีไอ List label	- แผ่นป้ายของรายการร้องขอการนำเข้าซอร์สโค้ดที่ถูกสร้างและลบทิ้งไปแล้ว

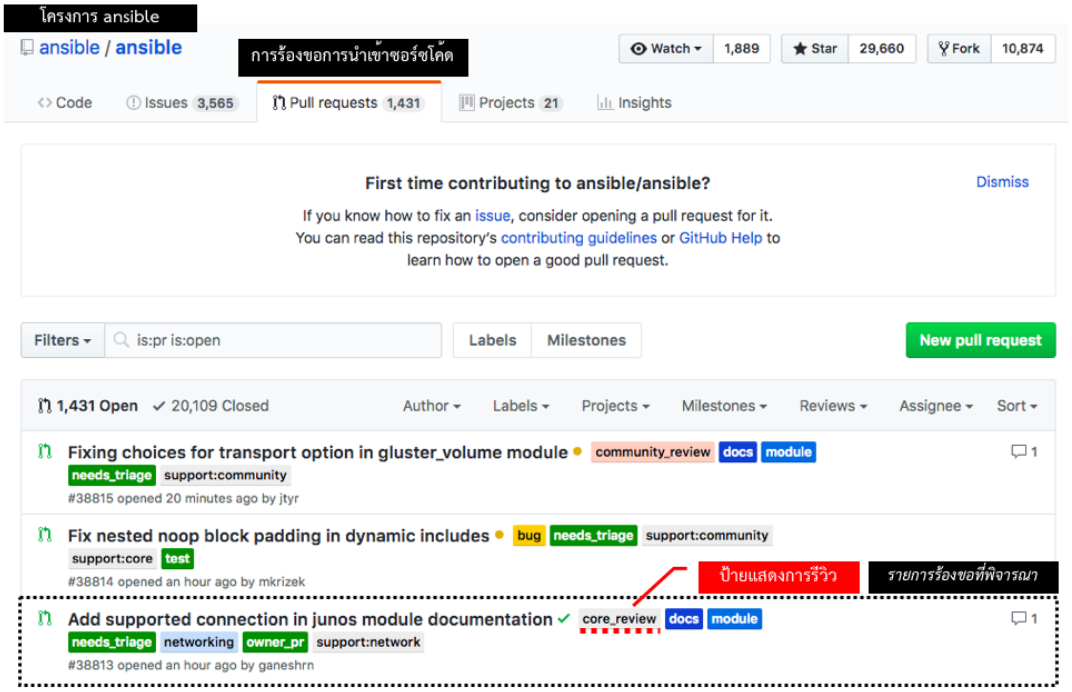
นอกจากเว็บไซต์ของกิตฮับแล้ว เว็บไซต์ของโครงการ Ansible บนกิตฮับนั้น ยังมีรายละเอียดของโครงการระบุอยู่ จากการวิเคราะห์พบว่าโครงการ Ansible มีโครงการย่อยอยู่ทั้งหมด 102 โครงการย่อย และมีดูแลระบบ 55 คนด้วยกัน โดยโครงการย่อยแต่ละโครงการมีรายละเอียดที่แตกต่างกันออกไป ไม่ว่าจะเป็นภาษาหลักที่ใช้ในการพัฒนา หรือจำนวนรายการร้องขอการนำเข้าซอร์สโค้ดที่มีจำนวนมากน้อยต่างกัน หากแต่กระบวนการร้องขอการนำเข้าซอร์สโค้ดของแต่ละโครงการย่อยนั้นดำเนินการเหมือนกัน รายละเอียดที่แสดงจำนวนโครงการย่อยและผู้ดูแลระบบ ดังรูปที่ 3.3

จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3.3 จำนวนโครงการย่อยและผู้ดูแลระบบของโครงการ Ansible

หากพิจารณาถึงรายละเอียดโครงการโอเพนซอร์สดังกล่าว และพิจารณาเมนู “Pull Requests” จะเห็นว่ามีจำนวนการร้องขอการรีวิวโค้ดมากมาย โดยแต่ละรายการที่ได้ร้องขอการเข้าระบบนั้นมีป้ายแสดงสถานะต่าง ๆ โดยตัวอย่างป้ายที่ระบุนั้น ดังรูปที่ 3.4



รูปที่ 3.4 ตัวอย่างแผ่นป้ายในเมนู Pull Requests ของโครงการ Anisible บนกิตฮับ

ซึ่งแผ่นป้ายที่อยู่ในรายการร้องขอการนำเข้าซอร์สโค้ดของโครงการโอเพนซอร์สนั้น ประกอบไปด้วย 102 แผ่นป้ายด้วยกัน รายละเอียดของแผ่นป้ายทุกแผ่นป้ายโครงการ ถูกระบุอยู่ในภาคผนวก ก แผ่นป้ายในรายการร้องขอการนำเข้าซอร์สโค้ดของโครงการ Ansible โดยจากแผ่นป้ายทั้งหมดนั้นมีแผ่นป้าย 12 แผ่นป้ายที่นำมาใช้ในการวิเคราะห์โครงการ แสดงรายละเอียดดังตารางที่ 3.5

ตารางที่ 3.5 รายละเอียดของแผ่นป้ายของโครงการ Ansible

ลำดับ	แผ่นป้าย	ความหมาย
1	shipit	แผ่นป้ายนำส่ง แผ่นป้ายที่ระบุเพื่อบ่งบอกสถานะของรายการร้องขอการนำเข้าซอร์สโค้ด คืออนุญาตให้นำซอร์สโค้ดเข้าสู่ระบบได้ กล่าวอีกนัยยะคือ ผ่านการรีวิว
2	discuss_topic	แผ่นป้ายที่ระบุเพื่อบ่งบอกว่า รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการการหารือเพิ่มเติม อาจจะมีการสร้างการประชุมขึ้นมาเพื่อหารือรายการร้องขอการนำเข้าซอร์สโค้ดดังกล่าว

ตารางที่ 3.5 รายละเอียดของแผ่นป้ายของโครงการ Ansible (ต่อ)

ลำดับ	แผ่นป้าย	ความหมาย
3	needs_ci	แผ่นป้ายที่ระบุเพื่อบ่งบอกว่าต้องการการทดสอบในรูปแบบซีเอ (CI) หรือ Continuous Integration ซึ่งเป็นแนวทางการพัฒนาซอฟต์แวร์ที่อิงกับการรวมโค้ดลงในที่เก็บข้อมูลร่วมกัน เพื่อใช้ในการตรวจสอบซอร์สโค้ดว่า ซอร์สโค้ดแต่ละชุดบิวส์ (Build) ผ่านหรือไม่
4	needs_ci_update	แผ่นป้ายที่ระบุพื่อบ่งบอกข้อมูลว่าผลการรีวิวถูกบล็อก (block) โดยการอัพเดทซีเอ หรือโครงข่ายซีเอ ก่อนที่จะทำการทดสอบซีเอ
5	needs_info	แผ่นป้ายที่ระบุเพื่อร้องขอเอกสารหรือข้อมูลเพิ่มเติม อีกทั้งยังรวมไปถึงการตอบคำถามที่ผู้รีวิวถาม
6	needs_maintainer	แผ่นป้ายที่ระบุเพื่อร้องขอให้ผู้ดูแลระบบเข้ามาตรวจสอบรายการร้องขอการนำเข้าซอร์สโค้ด
7	needs_rebase	แผ่นป้ายที่ระบุเพื่อบ่งบอกว่ารายการร้องขอการนำเข้าซอร์สโค้ดนี้ล้าสมัยแล้ว (out-of-date) เนื่องจากรายการร้องขอการนำเข้าซอร์สโค้ดนี้อาจจะมีข้อขัดแย้งกับการเปลี่ยนแปลงที่ได้นำเข้าระบบไปก่อนหน้า หรือรายการร้องขอการนำเข้าซอร์สนี้ไม่สามารถทดสอบกับระบบอัตโนมัติได้ เนื่องด้วยปัจจุบันมีฟังก์ชันที่เปลี่ยนแปลงไปมากแล้ว
8	needs_repo	แผ่นป้ายที่ระบุว่ารายการร้องขอการนำเข้าซอร์สนี้ไม่มีสาขาที่อยู่ ซึ่งอาจถูกลบไปโดยผู้นำเข้ารายการร้องขอการนำเข้าซอร์สโค้ด
9	needs_revision	แผ่นป้ายที่ระบุบ่งบอกว่ารายการร้องขอการนำเข้าซอร์สโค้ดนั้นต้องการการแก้ไข เนื่องจากไม่ผ่านการทดสอบซีไอ (CI) หรือไม่ได้รับการอนุมัติโดยผู้ดูแลระบบ
10	needs_shippable	แผ่นป้ายที่ระบุเพื่อบ่งบอกว่ารายละเอียดในโครงสร้างการร้องขอการนำเข้าซอร์สโค้ดขาด shippable.yml ในโครงสร้างซอร์สโค้ด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.5 รายละเอียดของแผ่นป้ายของโครงการ Ansible (ต่อ)

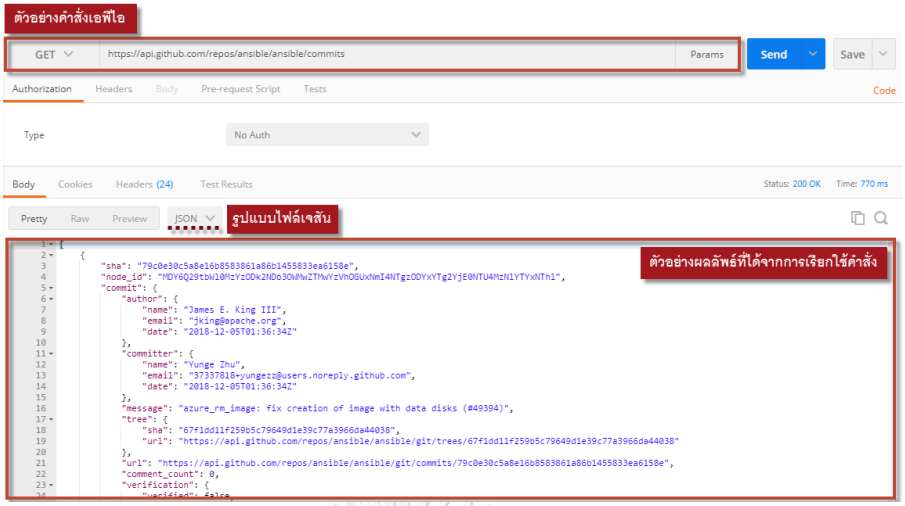
ลำดับ	แผ่นป้าย	ความหมาย
11	needs_template	แผ่นป้ายที่ระบุเพื่อบ่งบอกว่ารายการร้องขอการนำเข้าซอร์สโค้ดขาดรายละเอียดหรือขาดคำอธิบาย ให้ใส่ข้อมูลในแม่แบบ (template) ให้เรียบร้อย
12	needs_tests	แผ่นป้ายที่ระบุเพื่อบ่งบอกว่ารายการร้องขอการนำเข้าซอร์สโค้ดขาดการทดสอบปริพันธ์ (Integration test) หรือการทดสอบย่อย (Unit test) สำหรับซอร์สโค้ด

จากตาราง 3.5 แผ่นป้ายทั้งหมดแสดงรายละเอียดของโครงการโอเพนซอร์ส โดยรายการร้องขอการนำเข้าซอร์สโค้ดใด ๆ ที่มีสถานะการรีวิวโค้ดเป็นผ่านการรีวิวหรือยอมรับ (accepted) จะมีแผ่นป้ายนำส่ง หรือ shipit แปะกำกับไว้ (แผ่นป้ายลำดับที่ 1) และหากรายการร้องขอการนำเข้าซอร์สโค้ดใด ๆ มีผลการรีวิวเป็นไม่ผ่าน หรือปฏิเสธ (rejected) จะมีแผ่นป้าย discuss_topic, needs_ci, needs_ci_update, needs_info, needs_maintainer, needs_rebase, needs_repo, needs_revision, needs_shippable, needs_template และ needs_tests (แผ่นป้ายลำดับที่ 2-12) โดยแผ่นป้ายเป็นสิ่งที่บ่งบอกว่าผลการรีวิวเป็นไม่ผ่านเพราะเหตุใด ยกตัวอย่างเช่น หากรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ไม่ได้มีการระบุข้อมูลการทดสอบซอร์สโค้ด ระบบจะทำการแปะแผ่นป้าย need_tests เป็นต้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

(3) การวิเคราะห์รายละเอียดโครงการโอเพนซอร์สผ่านโปรแกรมโพสต์แมน (Post man)

โปรแกรมโพสต์แมนเป็นโปรแกรมที่ช่วยในการพัฒนาส่วนต่อประสานโปรแกรมประยุกต์ หรือเอพีไอ (Application Programming Interface) ทดสอบเซอร์วิสต่าง ๆ โดยนักพัฒนาซอฟต์แวร์ใช้โปรแกรม Postman ในการทดสอบการเชื่อมต่อกันระหว่างผู้ใช้บริการ (Client) และผู้ให้บริการ (Server) โดยการร้องขอข้อมูลต่าง ๆ ตามที่มาตรฐานเอพีไอได้กำหนดไว้ โดยงานวิจัยนี้จะนำโปรแกรมโพสต์แมนมาช่วยในทดสอบเอพีไอที่ใช้ในการวิเคราะห์ข้อมูล ว่าเอพีไอที่นำมาวิเคราะห์นั้นมีการแสดงผลอย่างไร ซึ่งจากการทดสอบเอพีไอที่กำหนดผ่านโปรแกรมโพสต์แมน พบว่ารายการเอพีไอที่กำหนดสามารถนำไปเขียนโปรแกรมเพื่อใช้ในการวิเคราะห์ข้อมูลได้ โดยตัวอย่างเอพีไอที่ใช้ในงานวิจัย ได้แก่ การนำข้อมูลรายการร้องขอการนำเข้าซอร์สโค้ดออกมา ซึ่งมีคำสั่งที่ใช้ในการดึงข้อมูล คือ GET/repos/:owner/:repo/pulls สามารถแสดงหน้าจอของการใช้โปรแกรมโพสต์แมนทดสอบได้ดังรูปที่ 3.5

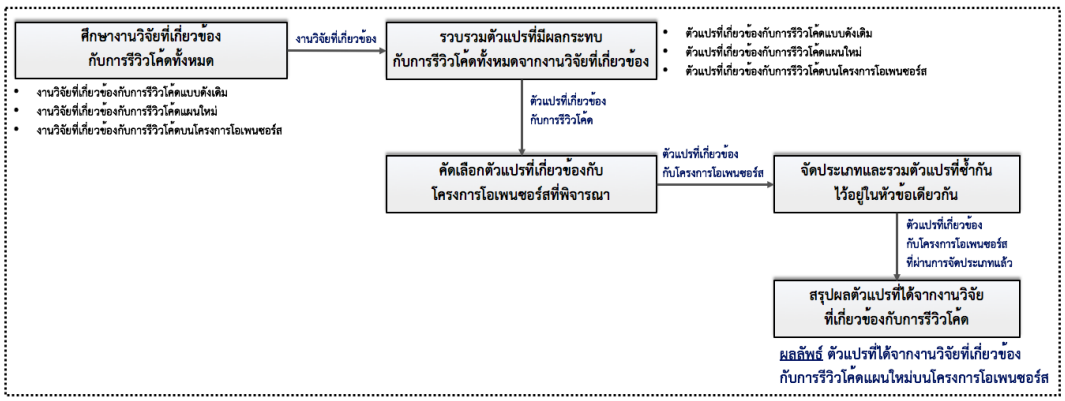


รูปที่ 3.5 ตัวอย่างหน้าจอของโปรแกรม Postman ในการดึงข้อมูลรายการ Pull requests

จากรูปที่ 3.5 คำสั่งที่ใช้คือ GET https://api.github.com/repos/ansible/ansible/commit โดยผลลัพธ์ที่ได้จะออกมาในรูปแบบของไฟล์เจสันที่แสดงรายละเอียดต่าง ๆ ของคอมมิทนั้น

3.2.2 กำหนดตัวแปรที่คาดว่าจะมีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

วิธีการกำหนดตัวแปรที่มีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์ส ทำได้โดยการกำหนดตัวแปรที่มีผลกับการรีวิวโค้ดบนโครงการโอเพนซอร์สทั้งหมดจากงานวิจัยที่เกี่ยวข้อง โดยจากการศึกษารายละเอียดของงานวิจัยที่เกี่ยวข้องกับการรีวิวโค้ด ทั้งการรีวิวโค้ดแบบการตรวจสอบซอร์สโค้ด หรือการรีวิวโค้ดแบบดังเดิม และการรีวิวโค้ด พบว่ามีตัวแปรมากมายที่เกี่ยวข้องกับการรีวิวโค้ด ไม่ว่าจะเป็นตัวแปรทางเทคนิค (Technical) และตัวแปรที่ไม่เกี่ยวข้องกับทางเทคนิค (Non-Technical) อีกทั้งยังมีตัวแปรหลายตัวที่ถูกพิจารณาให้เป็นตัวแปรที่สำคัญในส่งผลกับการรีวิวโค้ด โดยวิธีการหาตัวแปรจากงานวิจัยที่อ้างอิง มีขั้นตอน ดังรูปที่ 3.6



รูปที่ 3.6 วิธีการหาตัวแปรจากงานวิจัยที่เกี่ยวข้องกับการรีวิวโค้ด

ขั้นตอนที่ 1 ศึกษางานวิจัยที่เกี่ยวข้องกับการรีวิวโค้ดทั้งหมด โดยไม่กำหนดประเภทของการรีวิวโค้ด (ทุกประเภท) เพื่อให้มั่นใจว่าได้ศึกษาตัวแปรที่มีผลหรืออาจส่งผลกับการรีวิวโค้ดในทุกรูปแบบ โดยวัตถุประสงค์ของการศึกษามีเพื่อ ทำให้เข้าใจถึงรายละเอียดการรีวิวโค้ดโดยแท้จริง เพื่อให้ได้มาซึ่งตัวแปรที่มีผลกับการรีวิวโค้ดต่อไป

ขั้นตอนที่ 2 รวบรวมตัวแปรที่มีผลกับการรีวิวโค้ดทั้งหมดจากงานวิจัยที่เกี่ยวข้อง โดยทำการระบุประเภทของตัวแปร และวิเคราะห์รายละเอียดของการมีผลกับการรีวิวโค้ดดังกล่าว

ขั้นตอนที่ 3 จัดประเภท และรวมตัวแปรที่ซ้ำกันไว้อยู่ในหัวข้อเดียวกัน โดยการนำตัวแปรนั้น ๆ มาจำแนกตามประเภท และทำการคัดกรองตัวแปรที่ซ้ำกันออกให้เหลือเพียงตัวแปรที่พิจารณาเพียงตัวแปรเดียวต่อหนึ่งหัวข้อ เนื่องด้วยมีบางตัวแปรที่งานวิจัยใช้คำศัพท์หรือคำเรียก ในการเรียกถึงต่างกัน หากแต่มีความหมายเดียวกัน

ขั้นตอนที่ 4 คัดเลือกตัวแปรที่เกี่ยวข้องกับโครงการโอเพนซอร์สที่พิจารณา หรือโครงการโอเพนซอร์ส Ansible มาพิจารณาเพียงเท่านั้น และคัดกรองตัวแปรที่ไม่เกี่ยวข้องออก โดยพิจารณาจากการเกี่ยวข้องกันของข้อมูลโครงการโอเพนซอร์สที่ได้วิเคราะห์มา และพิจารณาถึงขั้นตอนการหาข้อมูลของตัวแปรนั้น ๆ ในโครงการโอเพนซอร์สดังกล่าว

ขั้นตอนที่ 5 สรุปผลตัวแปรที่ได้จากงานวิจัยที่เกี่ยวข้องกับการรีวิวโค้ด โดยสรุปผลตัวแปรที่ได้จากงานวิจัยที่เกี่ยวข้อง บันทึกข้อมูล และเตรียมพร้อมสำหรับการวิเคราะห์ตัวแปรในขั้นตอนต่อไป

โดยตัวแปรที่ได้จากการศึกษางานวิจัยที่เกี่ยวข้อง แสดงรายละเอียดดังตารางที่ 3.6

ตารางที่ 3.6 ตัวแปรที่ได้จากการวิเคราะห์รายละเอียดโครงการ

ลำดับ	ตัวแปร	ประเภท	อ้างอิง	คำอธิบาย
1	Pull request ID	Pull request	-	หมายเลขอ้างอิงรายการร้องขอการนำเข้าซอร์สโค้ด
2	Commit count	Pull request	[3] [4] [5] [9]	จำนวนการคอมมิทในหนึ่งรายการนำเข้าซอร์สโค้ด
3	File changed count	Pull request	[3] [4] [5] [9]	จำนวนไฟล์ที่ได้รับการเปลี่ยนแปลงต่อหนึ่งรายการร้องขอการนำเข้าซอร์สโค้ด
4	Lines of code changed count	Pull request	[23] [24] [25]	จำนวนบรรทัดการเปลี่ยนแปลงของโค้ดต่อหนึ่งรายการร้องขอการนำเข้าซอร์สโค้ด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.6 ตัวแปรที่ได้จากการวิเคราะห์รายละเอียดโครงการ (ต่อ)

ลำดับ	ตัวแปร	ประเภท	อ้างอิง	คำอธิบาย
5	Code chunk	Pull request	[3]	จำนวนของช่วงซอร์สโค้ดต่อหนึ่งรายการร้องขอการนำเข้าซอร์สโค้ด
6	Function	Code quality	[20]	จำนวนฟังก์ชัน
7	Method	Code quality	[20]	จำนวนเมธ็อด
8	Class	Code quality	[20]	จำนวนคลาส
9	Logical lines of code (LLOC)	Code quality	[20] [24]	จำนวนบรรทัดที่ไม่รวมบรรทัดที่ระบุคอมเม้น และไม่รวมบรรทัดที่ไม่ระบุข้อมูลหากแต่ยังไม่รวมฟังก์ชัน Nested
10	Source line of code (SLOC)	Code quality	[20] [24]	จำนวนบรรทัดที่มีเฉพาะซอร์สโค้ด
11	Comments	Code quality	[20]	จำนวนคอมเม้นหรือจำนวนการแสดงความคิดเห็น
12	Multi lines count	Code quality	[19] [24] [25]	จำนวนบรรทัดหลายรูปแบบ
13	Blank lines count	Code quality	[19] [24] [25]	จำนวนบรรทัดที่ไม่ระบุโค้ด
14	Hal:h1	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
15	Hal:h2	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
16	Hal:N1	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
17	Hal:N2	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.6 ตัวแปรที่ได้จากการวิเคราะห์รายละเอียดโครงการ (ต่อ)

ลำดับ	ตัวแปร	ประเภท	อ้างอิง	คำอธิบาย
18	Hal:vocabulary	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
19	Hal:length	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
20	Hal:calculated_length	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
21	Hal:volume	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
22	Hal:difficulty	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
23	Hal:effort	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
24	Hal:time	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
25	Hal:bugs	Code quality	[19] [26]	ค่าความซับซ้อนในรูปแบบ Halstead
26	Clone coverage (CC)	Code quality	[20] [26]	สัดส่วนระหว่างซอร์สโค้ดที่ซ้ำซ้อนกันกับซอร์สโค้ดทั้งหมด
27	Clone classes (CCL)	Code quality	[20] [26]	จำนวนของคลาสที่มีการโคลนที่ประกอบด้วยอย่างน้อยหนึ่งการโคลนในส่วนประกอบของซอร์สโค้ด
28	Clone complexity (CCO)	Code quality	[20] [26]	ความซับซ้อนของการโคลนวัดโดยใช้ McCabe
29	Clone instances (CI)	Code quality	[20] [26]	จำนวนกรณีตัวอย่างในซอร์สโค้ด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.6 ตัวแปรที่ได้จากการวิเคราะห์รายละเอียดโครงการ (ต่อ)

ลำดับ	ตัวแปร	ประเภท	อ้างอิง	คำอธิบาย
30	Clone line coverage (CLC)	Code quality	[20] [26]	สัดส่วนระหว่างบรรทัดซอร์สโค้ดที่ซ้ำซ้อนกับบรรทัดของซอร์สโค้ดทั้งหมด
31	Lines of duplicated code (LDC)	Code quality	[20] [23] [25] [26]	จำนวนบรรทัดซอร์สโค้ดที่มีความซ้ำซ้อน
32	Logical lines of duplicated code (LLDC)	Code quality	[20] [25] [26]	จำนวนบรรทัดซอร์สโค้ดที่มีความซ้ำซ้อน ที่มีการตัดบรรทัดว่างเปล่าและบรรทัดที่แสดงความคิดเห็นออก
33	Lack of Cohesion in Methods 5 (LCOM5)	Code quality	[20] [23] [24] [26]	ค่าการขาดความยึดเหนี่ยวกันและคำนวณจำนวนคลาสที่ควรจะแตกแยกออกจากกัน
34	Nesting level (NL)	Code quality	[20] [23] [26]	ความซับซ้อนของเมท็อดหรือฟังก์ชันที่อธิบายโดยระดับความลึกสูงสุดของการเข้าเงื่อนไข
35	Nesting Level Else-If (NLE)	Code quality	[20] [23] [24] [26]	ความซับซ้อนของเมท็อดหรือฟังก์ชันที่อธิบายโดยระดับความลึกสูงสุดของการเข้าเงื่อนไข โดย if-else-if เป็นโครงสร้างแรกที่พิจารณา
36	Weight Methods per Class (WMC)	Code quality	[20] [26]	ความซับซ้อนของคลาสที่อธิบายจำนวนความเป็นอิสระในกราฟการคอนโทรล (control flow graph) ที่คำนวณจาก McCabe's Cyclomatic Complexity (McCC)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.6 ตัวแปรที่ได้จากการวิเคราะห์รายละเอียดโครงการ (ต่อ)

ลำดับ	ตัวแปร	ประเภท	อ้างอิง	คำอธิบาย
37	Coupling between object classes (CBO)	Code quality	[20] [23] [24] [26]	จำนวนคลาสที่ขึ้นกับคลาสอื่น
38	Coupling between object classes Inverse (CBOI)	Code quality	[20] [23]	จำนวนคลาสที่ขึ้นกับคลาสอื่นที่มีผลกระทบต่อพฤติกรรมของระบบ
39	Number of Incoming Invocations (NII)	Code quality	[20] [24] [26] [26]	จำนวนเมท็อด ฟังก์ชัน ลักษณะประจำ หรือคลาสที่ถูกเรียกโดย เมท็อด ฟังก์ชัน ลักษณะประจำ หรือคลาสอื่น
40	Number of Outgoing Invocations (NOI)	Code quality	[20] [24] [26]	จำนวนเมท็อด ฟังก์ชัน ลักษณะประจำ หรือคลาสที่เรียก เมท็อด ฟังก์ชันลักษณะประจำ หรือคลาสอื่น
41	Response set for class (RFC)	Code quality	[20] [26]	จำนวนเมท็อดเฉพาะที่ในคลาสรวมกับจำนวนตัวเลขที่เกี่ยวข้องกับเมท็อดอื่นด้วยตัวมันเองหรือค่าเริ่มต้นของลักษณะประจำ (attribute initializations: NOI)
42	Comment Density (CD)	Code quality	[20] [26]	ความหนาแน่นของบรรทัดการคอมเม้นต่อบรรทัดคอมเม้นรวมกับบรรทัดที่เป็นโลจิก
43	Command lines of code (CLOC)	Code quality	[20] [25] [26]	จำนวนบรรทัดคอมเม้นรวมกับบรรทัดที่ระบุถึงข้อความ
44	Documentation lines of code (DLOC)	Code quality	[20] [25]	จำนวนบรรทัดที่ระบุถึงข้อความ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.6 ตัวแปรที่ได้จากการวิเคราะห์รายละเอียดโครงการ (ต่อ)

ลำดับ	ตัวแปร	ประเภท	อ้างอิง	คำอธิบาย
45	Total Comment Density (TCD)	Code quality	[20]	ความหนาแน่นของบรรทัดการคอมเม้นต่อบรรทัดคอมเม้นรวมกับบรรทัดที่เป็นโลจิคทั้งหมด
46	Total comment lines of code (TCLOC)	Code quality	[20] [25] [26]	จำนวนบรรทัดคอมเม้นรวมกับบรรทัดที่ระบุถึงข้อความทั้งหมด
47	Depth of inheritance tree (DIT)	Code quality	[20] [26]	จำนวนความยาวของเส้นทางที่นำไปสู่คลาสที่เป็นลูกหลานในต้นไม้ที่ยาวที่สุด
48	Number of ancestors (NOA)	Code quality	[20] [24] [26]	จำนวนคลาสที่ขึ้นตรงหรือถูกถ่ายทอดมา
49	Number of children (NOC)	Code quality	[20] [26]	จำนวนคลาสที่ถูกแบ่งมาจากคลาสอื่นๆ
50	Number of descendants (NOD)	Code quality	[20] [26]	จำนวนคลาสที่ทั้งขึ้นตรงและไม่ขึ้นตรงที่แบ่งมาจากคลาสอื่นๆ
51	Number of parent (NOP)	Code quality	[20] [26]	จำนวนคลาสที่ถูกต่อตรงโดยคลาสอื่นๆ
52	Number of attributes (NA)	Code quality	[20] [26]	จำนวนลักษณะประจำ
53	Number of local attributes (NLA)	Code quality	[20] [26]	จำนวนลักษณะประจำเฉพาะที่
54	Number of local methods (NLM)	Code quality	[20] [26]	จำนวนเมท็อดเฉพาะที่
55	Number of methods (NM)	Code quality	[20] [26]	จำนวนเมท็อด

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 3.6 ตัวแปรที่ได้จากการวิเคราะห์รายละเอียดโครงการ (ต่อ)

ลำดับ	ตัวแปร	ประเภท	อ้างอิง	คำอธิบาย
56	Number of statements (NOS)	Code quality	[20]	จำนวนข้อความสั่ง
57	Total logical lines of code (TLLOC)	Code quality	[20] [25] [26]	จำนวนบรรทัดที่ไม่รวมบรรทัดที่ระบุคอมเม้น และไม่รวมบรรทัดที่ไม่ระบุข้อมูล
58	Total lines of code (TLOC)	Code quality	[20] [25]	จำนวนบรรทัดซอร์สโค้ดทั้งหมด
59	Total number of attribute (TNA)	Code quality	[20]	จำนวนลักษณะประจำทั้งหมด
60	Total number of classes (TNLA)	Code quality	[20]	จำนวนลักษณะประจำเฉพาะที่ทั้งหมด
61	Total number of local methods (TNLM)	Code quality	[20]	จำนวนเมท็อดเฉพาะที่ทั้งหมด
62	Total number of methods (TNM)	Code quality	[20]	จำนวนเมท็อดทั้งหมด
63	Total number of statements (TNOS)	Code quality	[20]	จำนวนข้อความสั่งทั้งหมด
64	Result	Pull request	[3]	ผลการร้องขอการนำเข้าซอร์สโค้ด

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3.2.3 รวบรวมข้อมูลตัวแปรที่มีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

หลังจากที่ได้ตัวแปรที่มีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับมาแล้ว ขั้นตอนต่อไปคือการรวบรวมข้อมูลตัวแปรต่าง ๆ โดยขั้นตอนนี้นั้น ได้นำเอากระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่มาประยุกต์ใช้ โดยมีทั้งหมด 6 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 ทำความเข้าใจโดเมนการประยุกต์ และการเตรียมพร้อมความเข้าใจก่อนการวิเคราะห์ข้อมูล เนื่องด้วยงานวิจัยนี้มีเป้าหมายที่ชัดเจน คือการทำนาย โดยข้อมูลที่ได้มานั้น ได้มาจากกิตฮับ โดยนำซอร์สโค้ดของโครงการโอเพนซอร์สต่าง ๆ บน กิตฮับมาใช้ในการวิเคราะห์ ซึ่งการ

ใช้กระบวนการค้นพบความรู้ในฐานข้อมูลขนาดใหญ่นั้น มีเป้าหมายคือการหาตัวแปรที่มีผลกับการรีวิวโค้ด และการใช้ข้อมูลที่รวบรวมมาในการทำนายผลการรีวิวโค้ดนั่นเอง

ขั้นตอนที่ 2 รวบรวมข้อมูลและคัดเลือกข้อมูล (Data Selection) ที่จะทำการวิเคราะห์ โดยการคัดเลือกข้อมูลนั้น โดยข้อมูลที่นำมาพิจารณานั้นมีข้อกำหนดทั้งหมด 2 ข้อ ดังต่อไปนี้

1. รายการร้องขอการนำเข้าซอร์สโค้ดที่นำมาพิจารณา จะต้องเป็นรายการร้องขอการนำเข้าซอร์สโค้ดที่อยู่ในโครงการ Ansible และเปิดในปี 2017 เท่านั้น เพื่อให้สอดคล้องกับเหตุผลในการเลือกโครงการที่ได้กล่าวไปข้างต้น

2. รายการร้องขอการนำเข้าซอร์สโค้ดที่นำมาพิจารณา จะต้องมีซอร์สโค้ดแนบมากับรายการร้องขอ เนื่องจากการหาค่าตัวแปรต่าง ๆ ที่ได้รวบรวมมานั้น จำเป็นต้องใช้ซอร์สโค้ดในการคำนวณผล เนื่องจาก Ansible ถูกเขียนโดยภาษาไพธอน ดังนั้นการพิจารณารายการร้องขอการนำเข้าซอร์สโค้ดที่มีซอร์สโค้ดนั้น สามารถพิจารณาได้ว่ารายการใดที่มีไฟล์ .py แนบมา ให้ถือว่าเป็นรายการร้องขอการนำเข้าที่มีซอร์สโค้ดระบุอยู่

หลังจากที่กำหนดข้อกำหนดในการคัดเลือกข้อมูลแล้วนั้น ขั้นตอนต่อไปเป็นการรวบรวมข้อมูลตัวแปรที่มีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ เนื่องด้วยจำนวนตัวแปรมีปริมาณมาก ทำให้ไม่สามารถใช้วิธีการเดียวการหาค่าตัวแปรต่าง ๆ ได้ครอบคลุม ดังนั้นวิธีการที่ใช้ในการรวบรวมข้อมูลตัวแปร ต้องทำทั้ง 3 วิธีด้วยกัน ได้แก่ การเขียนสคริปเพื่อดึงข้อมูลผ่านกิตฮับเอพีไอ การคำนวณค่าตัวแปรผ่านเครื่องมือราดอน และการคำนวณค่าตัวแปรผ่านเครื่องมือซอร์สมิเตอร์ โดยรายละเอียดแต่ละวิธีการมีดังนี้

วิธีการที่ 1 การเขียนสคริปเพื่อดึงข้อมูลผ่านกิตฮับเอพีไอ (GitHub API)

เนื่องด้วยโครงการในกิตฮับส่วนมากเป็นโครงการโอเพนซอร์ส กิตฮับจึงมีรายการเอพีไอ (API) หรือ Application Programming Interface ให้นักพัฒนาใช้ประโยชน์อยู่หลาย ๆ ส่วน หนึ่งในนั้นคือการใช้เอพีไอในการดึงข้อมูลโครงการโอเพนซอร์สบนกิตฮับ โดยรูปแบบของเอพีไอที่ใช้ในกิตฮับนั้นมีหลากหลายรูปแบบด้วยกัน หากแต่ในงานวิจัยนี้ใช้รูปแบบของเอพีไอแบบเรส (RESTful) หรือ Representational state transfer คำสั่งเอพีไอที่ใช้ประกอบด้วย 6 คำสั่งคือ GET Single pull request, List commit, Single commit, Single commit comment, List label และ Issue event โดยมีรายละเอียดดังที่อธิบายไปในหัวข้อ 3.4 การเขียนสคริปเพื่อดึงข้อมูลผ่านกิตฮับเอพีไอประกอบไปด้วย 2 ขั้นตอนด้วยกัน

1. เขียนโปรแกรมจาวาเพื่อใช้เรียกคำสั่งเอพีไอดังกล่าว โดยผลลัพธ์ที่ได้จากการเขียนจะออกมาในรูปแบบไฟล์เจสัน (JASON) คำอธิบาย

2. แปลงรูปแบบไฟล์เจสันให้อยู่ในรูปแบบไฟล์ข้อความ นามสกุล .txt เพื่อนำไปใช้ประโยชน์ต่อไป

โดยผลลัพธ์ที่ได้จากการเขียนสคริปเพื่อดึงข้อมูลผ่านกิตฮับเอพีไอ คือไฟล์ข้อความ (.txt) ที่ประกอบไปด้วย หมายเลขรายการร้องขอการนำเข้าซอร์สโค้ด จำนวนการคอมมิท จำนวนไฟล์ที่มีการเปลี่ยนแปลง จำนวนบรรทัดซอร์สโค้ดที่มีการเปลี่ยนแปลง จำนวนชังก์

วิธีการที่ 2 การคำนวณค่าตัวแปรผ่านเครื่องมือราดอน (Radon)

วิธีการคำนวณค่าตัวแปรผ่านเครื่องมือราดอน ทำได้โดยการเขียนสคริป เพื่อใช้ประมวลผลคำสั่งบรรทัดคำสั่ง (Command Line) เรียกเอพีไอ (API) เพื่อไปดึงข้อมูลต่าง ๆ โดยการคำนวณค่าตัวแปรผ่านเครื่องมือราดอนของแต่ละรายการร้องขอการนำเข้าซอร์สโค้ดนั้นประกอบด้วย 3 ขั้นตอนด้วยกัน มีรายละเอียดดังนี้

1. เลือกคำสั่งการประมวลผลผ่านบรรทัดคำสั่งมาแสดง โดยคำสั่งการประมวลผลต้องมีทั้งคำสั่งการประมวลผลของกิตฮับและคำสั่งการประมวลผลของราดอน เนื่องจากเป็นการดึงข้อมูลร่วมกัน โดยคำสั่งบรรทัดคำสั่งที่ใช้ในโปรแกรมราดอนประกอบไปด้วย 3 คำสั่งคือ คำสั่ง cc คำสั่ง raw และคำสั่ง hal และคำสั่งของกิตฮับประกอบด้วย 2 คำสั่งได้แก่ git fetch และ get check out โดยรายละเอียดของแต่ละคำสั่ง รายละเอียดดังตารางที่ 3.7

ตารางที่ 3.7 คำสั่งที่ใช้ในการดึงข้อมูลผ่านโปรแกรมราดอน

ลำดับ	คำสั่ง	วัตถุประสงค์การใช้คำสั่ง
1	คำสั่ง cc	ใช้ในการคำนวณค่าความซับซ้อนไซโคลมาติก (Cyclomatic Complexity)
2	คำสั่ง raw	ใช้ในการคำนวณตัววัดดิบ ได้แก่ จำนวนฟังก์ชัน จำนวนเมธ็อด จำนวนคลาส จำนวนบรรทัดที่ไม่รวมบรรทัดที่ระบุคอมเม้น จำนวนบรรทัดที่ระบุซอร์สโค้ด จำนวนคอมเม้นหรือจำนวนการแสดงความคิดเห็น จำนวนบรรทัดหลายรูปแบบ และจำนวนบรรทัดที่ไม่ระบุโค้ด
3	คำสั่ง hal	ใช้ในการคำนวณค่าความซับซ้อนฮัลสเตด (Halstead Complexity) ซึ่งประกอบด้วย 12 ผลลัพธ์ด้วยกัน คือค่าความซับซ้อนฮัลสเตดในแต่ละรูปแบบ ได้แก่ Hal:h1, Hal:h2, Hal:N1, Hal:N2, Hal:vocabulary, Hal:length, hal:calculated_length, Hal_volume, Hal_difficulty, Hal_time และ Hal_bug

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 3.7 คำสั่งที่ใช้ในการดึงข้อมูลผ่านโปรแกรมราดอน (ต่อ)

ลำดับ	คำสั่ง	วัตถุประสงค์การใช้คำสั่ง
4	คำสั่ง git fetch	ใช้ในการดึงความเปลี่ยนแปลงจากรีโมทมายังเครื่องส่วนตัว (local) ตัวอย่างคำสั่งเช่น git fetch origin pull/21151/head:pr-21551 หมายความว่าให้แปลงจากการรีโมทไปยังเซิร์ฟเวอร์ มาที่เครื่องส่วนตัว หรือหมายความว่าให้ดึงข้อมูลทั้งหมดจากฝั่งรีโมทในรายการร้องขอการนำเข้าซอร์สโค้ดเลขที่ 21151 มายังเครื่องส่วนตัว
5	คำสั่ง git check out	ใช้ในการเปลี่ยนสาขา จากสาขาหลักไปยังสาขาที่ระบุรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ตัวอย่างคำสั่งเช่น git checkout pr-21551 หมายความว่า ให้เปลี่ยนที่อยู่จากสาขาหลัก ไปยังสาขาที่มีรายการร้องขอการนำเข้าซอร์สโค้ดหมายเลข 21551 ระบุอยู่

2. ดาวน์โหลดซอร์สโค้ดทั้งหมดของรายการร้องขอการนำเข้าซอร์สโค้ดที่พิจารณาลงมาบนเครื่องคอมพิวเตอร์ส่วนตัว (Personal Computer) เพื่อใช้ในการคำนวณ โดยวิธีการดาวน์โหลดคือการใช้คำสั่ง Git clone https://github.com/ansible/ansible.git เพื่อใช้ในการดาวน์โหลดข้อมูลทั้งหมดลงมาในพาท (path) ที่เรากำหนด

3. ประมวลผลผ่านบรรทัดคำสั่ง โดยใช้คำสั่งบรรทัดคำสั่ง 5 คำสั่ง ที่ระบุไว้ในตารางข้างต้นในการประมวลผล จากนั้นใช้โปรแกรมจาวาที่เขียนขึ้นมาในการจัดรูปแบบข้อมูลให้อยู่ในรูปแบบไฟล์ข้อความ (.txt) เพื่อให้สามารถนำไปใช้ในการประมวลผลได้

โดยผลลัพธ์จากการใช้โปรแกรมราดอนในการคำนวณค่าตัวแปรได้แก่ ไฟล์ในรูปแบบข้อความ (.txt) จำนวน 3 ไฟล์คือ ไฟล์ที่ประมวลผลโดยคำสั่ง cc ไฟล์ที่ประมวลผลโดยคำสั่ง hal และไฟล์ที่ประมวลผลโดยคำสั่ง mi โดยมีตัวแปรที่เกี่ยวข้องในการดึงข้อมูลผ่านโปรแกรมราดอนอยู่ทั้งหมด 20 ตัวแปรด้วยกัน ได้แก่ Pull request ID, Function, Method, Class, Logical lines of code, Source line of code, Comments, Multi line count, Blank lines count, Hal:h1, Hal:h2, Hal:N1, Hal:N2, Hal:vocabulary, Hal:length, hal:calculated_length, Hal_volume, Hal_difficulty, Hal_time และ Hal_bug

วิธีที่ 3 การคำนวณค่าตัวแปรผ่านเครื่องมือซอร์สมิเตอร์ (SourceMetter)

วิธีการคำนวณค่าตัวแปรผ่านเครื่องมือซอร์สมิเตอร์ ใช้วิธีการเดียวกับเครื่องมือราดอน โดยการเขียนสคริป เพื่อใช้ประมวลผลคำสั่งบรรทัดคำสั่ง (Command Line) เรียกเอพีไอ (API) เพื่อไปดึง

ข้อมูลต่าง ๆ หากแต่ต้องมีบรรทัดคำสั่งที่แตกต่างกัน และมีบรรทัดคำสั่งที่เพิ่มเติม นอกเหนือจากการเขียนสคริปเพื่อดึงข้อมูลแล้ว ยังต้องมีการเขียนสคริปเพื่อลบข้อมูลที่ไม่ใช้ออกด้วย เนื่องจากซอร์สมิเตอร์เป็นเครื่องมือทางการตลาด (Commercial Tool) จึงมีผลการวิเคราะห์ที่แสดงกราฟ และข้อมูลต่าง ๆ ปริมาณมาก การเขียนสคริปเพิ่มเติมขึ้นมา ทำขึ้นมาเพื่อลบข้อมูลที่ไม่ใช้ออก และยังทำให้การดึงข้อมูลประมวลผลเร็วขึ้นอีกด้วย คำสั่งที่ใช้ประกอบไปด้วย 4 คำสั่งด้วยกัน ได้แก่ คำสั่ง call คำสั่ง get fetch คำสั่ง get checkout และคำสั่ง kill โดยคำสั่งที่ใช้มีรายละเอียดดังตารางที่ 3.8

ตารางที่ 3.8 บรรทัดคำสั่งที่ใช้ในการดึงข้อมูลผ่านโปรแกรมซอร์สมิเตอร์

ลำดับ	คำสั่ง	วัตถุประสงค์การใช้คำสั่ง
1	คำสั่ง call	ใช้ในการเรียกใช้โปรแกรมซอร์สมิเตอร์ให้ทำงาน โดยใช้คำสั่ง call ตามด้วยพาทที่โปรแกรมซอร์สมิเตอร์ติดตั้งไว้ในคอมพิวเตอร์ส่วนตัว ตัวอย่างคำสั่งเช่น call C:\devsw\SourceMeter-8.2.0-x64-windows\SourceMeter-8.2.0-x64-windows\Python\ SourceMeterPython.exe เป็นการเรียกใช้โปรแกรมซอร์สมิเตอร์ในพาท (path) ดังกล่าว
2	คำสั่ง git fetch	ใช้ในการดึงความเปลี่ยนแปลงจากรีโมทมายังเครื่องส่วนตัว (local) ตัวอย่างคำสั่งเช่น git fetch origin pull/21151/head:pr-21551 หมายความว่าให้แปลงจากการรีโมทไปยังเซิร์ฟเวอร์ มาที่เครื่องส่วนตัว หรือหมายความว่าให้ดึงข้อมูลทั้งหมดจากฝั่งรีโมทในรายการร้องขอการนำเข้าซอร์สโค้ดเลขที่ 21151 มายังเครื่องส่วนตัว
3	คำสั่ง git check out	ใช้ในการเปลี่ยนสาขา จากสาขาหลักไปยังสาขาที่ระบุรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ตัวอย่างคำสั่งเช่น git checkout pr-21551 หมายความว่า ให้เปลี่ยนที่อยู่จากสาขาหลัก ไปยังสาขาที่มีรายการร้องขอการนำเข้าซอร์สโค้ดหมายเลข 21551 ระบุอยู่
4	คำสั่ง kill	ใช้ในการลบข้อมูลที่ไม่ใช้ทิ้ง ตัวอย่างคำสั่ง kill เช่น :loop taskkill /f /IM graphdump.exe timeout /t 7 goto loop ซึ่งหมายความว่าให้ทำงานเป็นลูป หรือวนคำสั่ง taskkill ไปเรื่อยๆ โดยระบุสิ่งที่จะลบทิ้งคือ graphdump.exe ซึ่งเป็นไฟล์รูปภาพ และกำหนดให้หมดเวลาใน 7 วินาที แล้วเริ่มทำใหม่อีกครั้ง

ผลลัพธ์ที่ได้จากการดึงข้อมูลผ่านโปรแกรมซอร์สมิเตอร์ คือค่าตัวแปรต่าง ๆ ที่ได้กำหนดไว้จำนวน 38 ตัวแปรด้วยกัน ได้แก่ Clone coverage (CC), Clone classes (CCL), Clone complexity (CCO), Clone instances (CI), Clone line coverage (CLC), Lines of duplicated code (LDC), Logical lines of duplicated code (LLDC), Lack of Cohesion in Methods 5 (LCOM5), Nesting level (NL), Nesting Level Else-If (NLE), Weight Methods per Class (WMC), Coupling between object classes (CBO), Coupling between object classes Inverse (CBOI), Number of Incoming Invocations (NII), Number of Outgoing Invocations (NOI), Response set for class (RFC), Comment Density (CD), Command lines of code (CLOC), Documentation lines of code (DLOC), Total Comment Density (TCD), Total comment lines of code (TCLOC), Depth of inheritance tree (DIT), Number of ancestors (NOA), Number of children (NOC), Number of descendants (NOD), Number of parent (NOP), Number of attributes (NA), Number of local attributes (NLA), Number of local methods (NLM), Number of methods (NM), Number of statements (NOS), Total logical lines of code (TLLOC), Total lines of code (TLOC), Total number of attribute (TNA), Total number of classes (TNLA), Total number of local methods (TNLM), Total number of methods (TNM), Total number of statements (TNOS)

ขั้นตอนที่ 3 การประมวลผลข้อมูลก่อน (Data Processing) หรือกล่าวอีกนัยยะคือการทำความสะอาดข้อมูล ทำได้โดยการพิจารณาค่าของตัวแปรทุกตัว และดูความผิดปกติของข้อมูล โดยการประมวลผลข้อมูลก่อนนั้น มีการนำเทคนิคการลบรายการที่มีข้อมูลไม่สมบูรณ์ออกจากการวิเคราะห์และการกำจัดค่าผิดปกติออกจากการวิเคราะห์ โดยแต่ละเทคนิคมีรายละเอียด ดังนี้

1. การลบรายการที่มีข้อมูลไม่สมบูรณ์ออกจากการวิเคราะห์

เนื่องด้วยรายการร้องขอการนำเข้าซอร์สโค้ดมีจำนวนมาก อีกทั้งยังมีวิธีในการดึงข้อมูลถึง 3 วิธีด้วยกัน ทำให้รายการร้องขอการนำเข้าซอร์สโค้ดดังกล่าว มีบางข้อมูลที่ไม่สมบูรณ์ เช่น รายการร้องขอการนำเข้าซอร์สโค้ดนั้น ๆ ไม่สามารถใช้เครื่องมือราดอนในการประมวลผลได้ หรือรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ๆ ไม่สามารถใช้เครื่องมือซอร์สมิเตอร์ในการประมวลผลได้ เป็นต้น

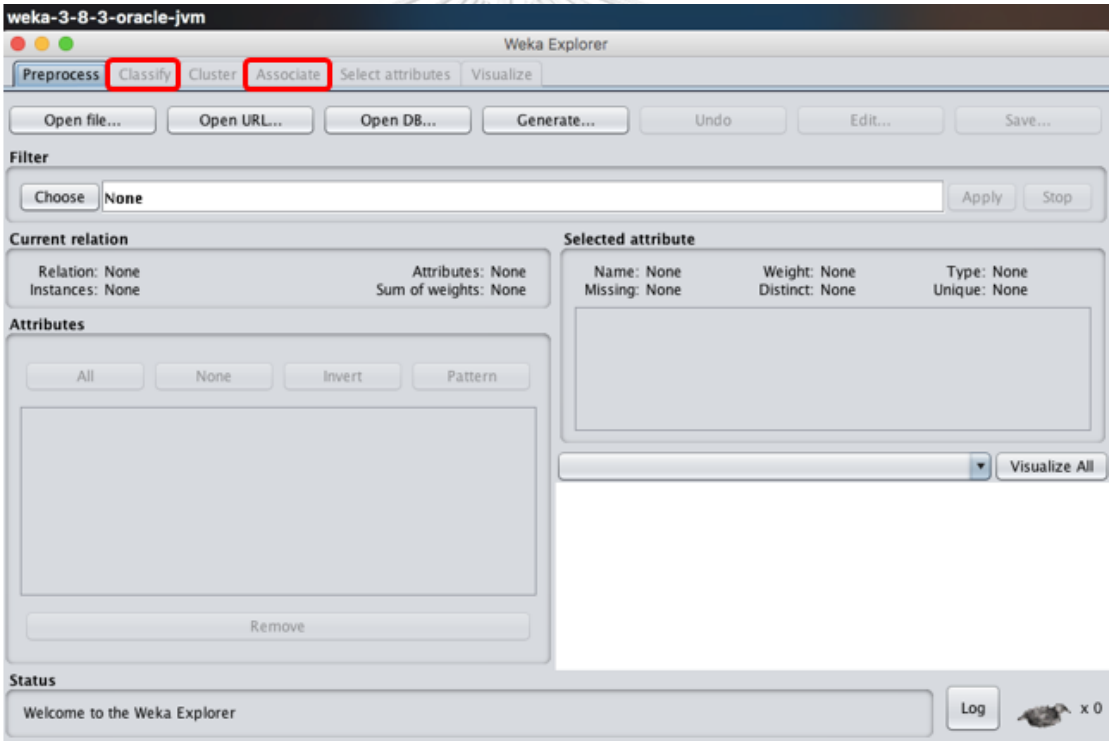
2. การจำกัดค่าผิดปกติออกจากการวิเคราะห์

บางรายการร้องขอการนำเข้าซอร์สโค้ด มีค่าที่ไม่อิงกับรายการส่วนมาก ทำให้ต้องตัดรายการร้องขอการนำเข้าซอร์สโค้ดนั้นออกจากระบบ เพื่อป้องกันไม่ให้ค่าผิดปกติ (Outlier) เข้าสู่การวิเคราะห์ เนื่องด้วยข้อมูลที่มีค่าผิดปกติระบุอยู่ จะทำให้การวิเคราะห์ข้อมูลมีความผิดพลาด ยกตัวอย่างเช่น รายการร้องขอการนำเข้าซอร์สโค้ดเลขที่ 22,878 มีบรรทัดการเปลี่ยนแปลงซอร์สโค้ดสูงถึง 9,333 บรรทัด เนื่องด้วยเป็นรายการร้องขอการนำเข้าซอร์สโค้ดที่ไม่ปกติ เพราะเป็นการเพิ่มโมดูลใหม่ทั้ง

ระบบ แตกต่างกับรายการร้องขอการนำเข้าซอร์สโค้ดทั่วไป ที่เป็นเพียงแค่การแก้ไข หรือการเพิ่มเติมซอร์สโค้ดเท่านั้น รายการร้องขอการนำเข้าซอร์สโค้ดเลขที่ 22,878 จึงไม่ได้ถูกนำมาวิเคราะห์ข้อมูล

ขั้นตอนที่ 4 การแปลงข้อมูล (Data Transformation) จากข้อมูลที่นำมาพิจารณาจะเห็นว่ามีเพียงผลของการรีวิวโค้ดเพียงตัวแปรเดียวที่เป็นข้อมูลในรูปแบบอัตราส่วนแบบนามบัญญัติ ทำให้ต้องแปลงค่าของผลการรีวิวโค้ดก่อนนำไปใช้งาน โดยใช้หลักการตัวแปรหุ่นเข้ามาใช้ แทนค่าผลการรีวิวที่ผ่านการรีวิว (Accepted) ด้วยค่า 1 และแทนค่าผลการรีวิวที่ไม่ผ่านการรีวิว (Rejected) ด้วยค่า 0

ขั้นตอนที่ 5 เลือกประเภทเหมืองข้อมูลมาใช้ในการวิเคราะห์ เนื่องด้วยเป้าหมายหลักของการวิเคาะห์คือการหาความสัมพันธ์ระหว่างตัวแปร และการทำนายผลการรีวิวโค้ด ประเภทเหมืองข้อมุลที่นำมาใช้จึงใช้เป็นสองประเภทด้วยกัน ได้แก่ การหาความสัมพันธ์ของข้อมูล (Association) และการทำนายข้อมูล (Prediction) หากพิจารณาส่วนต่อประสานของเครื่องมือเวก้าแล้วนั้น การวิเคราะห์จะอยู่ในเมนู Associate และ Classify ดังรูปที่ 3.7



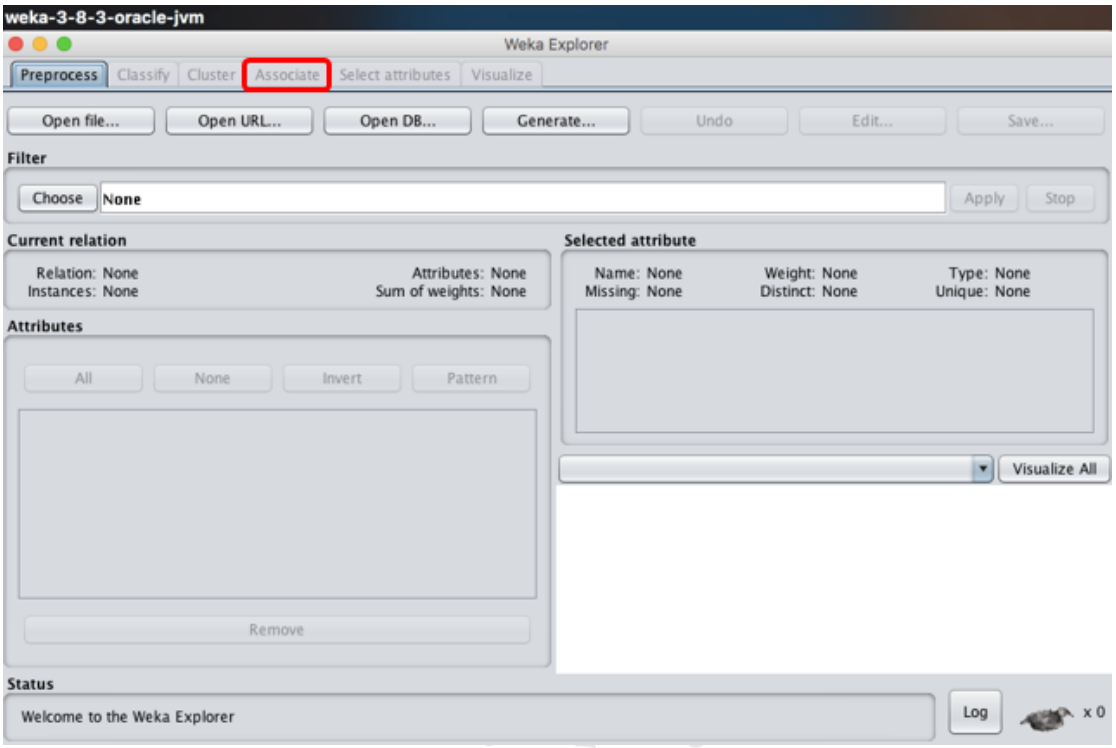
รูปที่ 3.7 ส่วนต่อประสานของเครื่องมือเวก้าที่ใช้

ขั้นตอนที่ 6 เลือกขั้นตอนวิธีการทำเหมืองข้อมูล (Data Mining Alogrithm) โดยขั้นตอนวิธีที่เลือกใช้ในการหาความสัมพันธ์ของตัวแปรที่มีผลกับการรีวิวโค้ดได้แก่ ขั้นตอนวิธีเอไพรออไร และขั้นตอนวิธีที่ใช้ในการทำนายผลการรีวิวโค้ดได้แก่ การวิเคราะห์เพอร์เซปตรอนหลายชั้น (Multi-layer Perceptron)

3.2.4 วิเคราะห์ความสัมพันธ์ระหว่างตัวแปรด้วยกฎความสัมพันธ์

หลังจากที่ได้รวบรวมข้อมูลตัวแปรทั้งหมดมาแล้ว การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรด้วยกฎความสัมพันธ์มีวัตถุประสงค์เพื่อหารูปแบบของความสัมพันธ์ที่เกิดขึ้นบ่อยในชุดข้อมูลที่นำมาวิเคราะห์ โดยมีรายละเอียด ดังนี้

1. วิธีการหาความสัมพันธ์ของตัวแปรด้วยกฎความสัมพันธ์ ด้วยขั้นตอนวิธีเอไพรออไร สามารถทำได้โดยการประมวลผลผ่านเครื่องมือเวก้า เมนู Associate ดังรูปที่ 3.8



รูปที่ 3.8 เมนู Associate ในโปรแกรมเวก้า

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

โดยขั้นตอนในการหากฎความสัมพันธ์ประกอบด้วย 4 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การเตรียมข้อมูลก่อนการวิเคราะห์ (Preprocess)

เนื่องด้วยตัวแปรส่วนมากที่นำมาวิเคราะห์ เป็นตัวแปรที่มีค่าต่อเนื่อง การวิเคราะห์ผ่านเครื่องมือเวก้านั้น ต้องทำการแบ่งค่าต่อเนื่องให้เป็น ช่วง ๆ ก่อนจะนำไปวิเคราะห์ข้อมูลได้ โดยเทคนิคที่เลือกใช้ในการเตรียมข้อมูลก่อนการวิเคราะห์ คือเทคนิคความไม่ต่อเนื่อง (Discretize) ซึ่งเป็นเทคนิคการทำบินนิง (Binning Method) รูปแบบหนึ่งที่แปลงช่วงของค่าตัวเลขในกลุ่มข้อมูลให้เป็นค่านามบัญญัติ ซึ่งจากการใช้เทคนิคความไม่ต่อเนื่องในการประมวลผลนั้น สามารถกำหนดค่าบิน หรือค่าช่วงของข้อมูลในการประมวลผลได้ ในที่นี้มีการกำหนดค่าบินเท่ากับสิบ หรือเรียกได้ว่าเป็นการแบ่งข้อมูลออกเป็น 10 ช่วง ตัวอย่างการแบ่งช่วงข้อมูล รายละเอียดดังตารางที่ 3.9

ตารางที่ 3.9 ตัวอย่างการแบ่งช่วงข้อมูล

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวนตัวแปร
1	จำนวนคอมมิท (Commit count)	1	'(-inf-3.9]'	3128
		2	'(3.9-6.8]'	361
		3	'(6.8-9.7]'	116
		4	'(9.7-12.6]'	73
		5	'(12.6-15.5]'	27
		6	'(15.5-18.4]'	20
		7	'(18.4-21.3]'	11
		8	'(21.3-24.2]'	8
		9	'(24.2-27.1]'	6
		10	'(27.1-inf)'	31

ตารางที่ 3.9 อธิบายการแบ่งช่วงข้อมูลของตัวแปรจำนวนคอมมิท โดยจำนวนคอมมิทที่ได้รวบรวมข้อมูลมา มีค่าตั้งแต่ 1 ถึง 30 สามารถแบ่งช่วงข้อมูลได้ 10 ช่วง ตามการแบ่งช่วงข้อมูลที่ได้ทำการแบ่งโดยโปรแกรมเวก้า ซึ่งรายละเอียดทั้งหมดของการแบ่งช่วงข้อมูลตัวแปร เพื่อใช้ในการหาความสัมพันธ์ด้วยกฎความสัมพันธ์ อธิบายใน ภาคผนวก ค ผลลัพธ์การวิเคราะห์ตัวแปรด้วยกฎความสัมพันธ์โดยละเอียด

ขั้นตอนที่ 2 การตั้งค่ากฎความสัมพันธ์

การกำหนดค่าตัววัดของการประมวลผล ในที่นี้กำหนดให้ตัววัดคือค่าความเชื่อมั่น (Confidence) โดยค่าความเชื่อมั่นต้องมีค่าไม่ต่ำกว่า 0.9 และกำหนดจำนวนกฎที่แสดงให้แสดงกฎความสัมพันธ์ทั้งหมด 10 กฎด้วยกัน

ขั้นตอนที่ 3 การวิเคราะห์ผลลัพธ์การหากฎความสัมพันธ์ด้วยขั้นตอนวิธีเอไพรออไร

ผลการวิเคราะห์ข้อมูลแบ่งออกเป็น 3 ส่วนด้วยกัน ส่วนแรกเป็นรายละเอียดของไฟล์ที่ใช้ในการวิเคราะห์ข้อมูล เค้าร่าง (Schema) แสดงวิธีการที่ใช้ในการวิเคราะห์ข้อมูล และการตั้งค่าต่าง ๆ ในการวิเคราะห์ข้อมูล ความสัมพันธ์(Relation) แสดงไฟล์ที่ใช้ในการวิเคราะห์ข้อมูล และวิธีการนำไฟล์เข้าสู่ระบบ กรณีตัวอย่าง (Instance) แทนกรณีตัวอย่างหรือรายการทรานแซคชันที่นำเข้าสู่ระบบ แอดทริบิวซ์ (Attibutes) คือจำนวนตัวแปรที่ใช้ในการวิเคราะห์ข้อมูล และในส่วนสุดท้ายของส่วนแรก จะเป็นการบอกตัวแปรหรือแอดทริบิวซ์ต่าง ๆ ที่นำเข้าระบบ ดังรูปที่ 3.9

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ข้อมูลแสดงรายละเอียดการวิเคราะห์

```
=== Run information ===

Scheme:       weka.associations.Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:     02-Remove outlier (import data)-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6
Instances:    3781
Attributes:   64
               pull request id
              Count (commit#)
              file_changed
              LOC change
              Code chunk
              function
```

รูปที่ 3.9 รายละเอียดการวิเคราะห์ (ข้อมูลส่วนแรก)

ส่วนที่สองคือค่าตัววัดที่ใช้ในการวิเคราะห์ข้อมูล ประกอบด้วยค่าสนับสนุนน้อยที่สุดที่เกิดจากการวิเคราะห์ ค่าความมั่นใจที่น้อยที่สุดที่เกิดจากการวิเคราะห์ จำนวนรอบการคำนวณด้วยขั้นตอนวิธีเอไพรออไร และจำนวนกลุ่มของไอเทมเซตที่ใหญ่ที่สุด ดังรูปที่ 3.10

รายละเอียดตัววัดในการวิเคราะห์

```
Apriori
=======

Minimum support: 0.85 (3214 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 3

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10

Size of set of large itemsets L(2): 23

Size of set of large itemsets L(3): 8
```

รูปที่ 3.10 รายละเอียดตัววัดในการวิเคราห์ (ข้อมูลส่วนที่สอง)

ส่วนสุดท้ายเป็นส่วนของกฎความสัมพันธ์หรือผลลัพธ์ที่ได้จากการวิเคราะห์ (Best rule found) เป็นกฎที่ดีที่สุดที่ค้นพบ หรือรูปแบบความสัมพันธ์ที่เกิดขึ้นมากในข้อมูลนำเข้าชุดนี้ จำนวนกฎที่แสดงจะตรงกับการตั้งค่า numRules ที่ตั้งไว้ ในที่นี้แสดงผลเป็นกฎความสัมพันธ์ 10 กฎ ดังรูป 3.11

กฎที่ดีที่สุด

```
Best rules found:

 1. hal:time='(-inf-1768.643324]' 3478 ==> hal:effort='(-inf-31835.579233]' 3478    <conf:(1)> lift:(1.09) lev:(0.07) [278] conv:(278.72)
 2. hal:effort='(-inf-31835.579233]' 3478 ==> hal:time='(-inf-1768.643324]' 3478    <conf:(1)> lift:(1.09) lev:(0.07) [278] conv:(278.72)
 3. file_changed='(-inf-17.3]' hal:time='(-inf-1768.643324]' 3297 ==> hal:effort='(-inf-31835.579233]' 3297    <conf:(1)> lift:(1.09) lev:(0.07) [264] conv:(264.21)
 4. file_changed='(-inf-17.3]' hal:effort='(-inf-31835.579233]' 3297 ==> hal:time='(-inf-1768.643324]' 3297    <conf:(1)> lift:(1.09) lev:(0.07) [264] conv:(264.21)
 5. hal:time='(-inf-1768.643324]' Result=reject 3251 ==> hal:effort='(-inf-31835.579233]' 3251    <conf:(1)> lift:(1.09) lev:(0.07) [260] conv:(260.53)
 6. hal:effort='(-inf-31835.579233]' Result=reject 3251 ==> hal:time='(-inf-1768.643324]' 3251    <conf:(1)> lift:(1.09) lev:(0.07) [260] conv:(260.53)
 7. function='(-inf-23.1]' hal:time='(-inf-1768.643324]' 3232 ==> hal:effort='(-inf-31835.579233]' 3232    <conf:(1)> lift:(1.09) lev:(0.07) [259] conv:(259)
 8. function='(-inf-23.1]' hal:effort='(-inf-31835.579233]' 3232 ==> hal:time='(-inf-1768.643324]' 3232    <conf:(1)> lift:(1.09) lev:(0.07) [259] conv:(259)
 9. hal:h1='(-inf-30.4]' hal:time='(-inf-1768.643324]' 3217 ==> hal:effort='(-inf-31835.579233]' 3217    <conf:(1)> lift:(1.09) lev:(0.07) [257] conv:(257.8)
10. hal:h1='(-inf-30.4]' hal:effort='(-inf-31835.579233]' 3217 ==> hal:time='(-inf-1768.643324]' 3217    <conf:(1)> lift:(1.09) lev:(0.07) [257] conv:(257.8)
```

รูปที่ 3.11 กฎความสัมพันธ์ที่ดีที่สุดสิบกฎ (ข้อมูลชุดที่สาม)

โดยสรุปผลลัพธ์ หรือกฎความสัมพันธ์ที่ดีที่สุดสิบกฎความสัมพันธ์ที่ได้จากการวิเคราะห์ข้อมูล ซึ่งบ่งบอกถึงความสัมพันธ์ที่เกิดขึ้นบ่อยในชุดข้อมูลนี้ รายละเอียดดังตารางที่ 3.10

ตารางที่ 3.10 กฎที่ดีที่สุดที่ได้จากการวิเคราะห์ข้อมูล

อันดับ	กฎความสัมพันธ์	ค่าความเชื่อมั่น
1	Halstead time = (-inf-1768.643324] --> Halstead effort = (-inf-31635.579233]	1
2	Halstead effort = (-inf-31635.579233] --> Halstead time = (-inf-1768.643324]	1
3	File changed = (-inf-31685.579233] Halstead time = (-inf-1768.643324] --> Halstead effort = (-inf-31635.579233]	1
4	File changed = (-inf-31685.579233] Halstead effort = (-inf-31635.579233] --> Halstead time = (-inf-1768.643324]	1
5	Halstead time = (-inf-1768.643324] Result = Reject --> Halstead effort = (-inf-31635.579233]	1
6	Halstead effort = (-inf-31635.579233] Result = Reject --> Halstead time = (-inf-1768.643324]	1
7	Function = (-inf-23.1] Halstead time = (-inf-1768.643324] --> Halstead effort = (-inf-31635.579233]	1
8	Function = (-inf-23.1] Halstead effort = (-inf-31635.579233] --> Halstead time = (-inf-1768.643324]	1
9	Halstead H1 = (inf-30.4] Halstead effort = (-inf-31635.579233] --> Halstead time = (-inf-1768.643324]	1
10	Halstead H1 = (inf-30.4] Halstead effort = (-inf-31635.579233] --> Halstead time = (-inf-1768.643324]	1

จากตารางที่ 7 สามารถแปลผลได้ดังนี้

กฎข้อที่ 1 Halstead time = (-inf-1768.643324] --> Halstead effort = (-inf-31635.579233] หมายความว่า ถ้ารายการร้องขอการนำเข้าซอร์สโค้ดใด มีค่า Halstead time มีค่าอยู่ในช่วง–inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 1768.643324 แล้วค่า Halstead effort จะมีค่าอยู่ในช่วง–inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31635.579233

กฎข้อที่ 2 สามารถตีความได้แบบเดียวกันกับกฎข้อที่หนึ่ง หากแต่เป็นการกลับค่ากัน กล่าวคือถ้ารายการร้องขอการนำเข้าซอร์สโค้ดใด มีค่า Halstead effort มีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบ

อินฟินิตี้ (Infinity) ถึง 31635.579233 แล้วค่า Halstead time จะมีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 1768.643324

กฎข้อที่ 3 หมายความว่า ถ้าจำนวนการเปลี่ยนแปลงของไฟล์ (File changed) มีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31685.579233 และ Halstead time มีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 1768.643324 แล้ว Halstead effort จะมีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31635.579233

กฎข้อที่ 4 สามารถตีความได้แบบเดียวกันกับกฎข้อที่สาม หากแต่เป็นการกลับค่ากัน กล่าวคือว่า ถ้าจำนวนการเปลี่ยนแปลงของไฟล์ (File changed) มีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31685.579233 และ Halstead effort มีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31635.579233 แล้ว Halstead time จะมีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 1768.643324 แล้ว

กฎข้อที่ 5 หมายความว่า ถ้า Halstead time มีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 1768.643324 และผลลัพธ์การรีวิวเป็นไม่ตอบรับการรีวิวแล้ว Halstead effort จะมีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31635.579233

กฎข้อที่ 6 สามารถตีความได้แบบเดียวกันกับกฎข้อที่ห้า หากแต่เป็นการกลับค่ากัน กล่าวคือถ้า Halstead effort มีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31635.579233 และผลลัพธ์การรีวิวเป็นไม่ตอบรับการรีวิวแล้ว Halstead time จะมีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 1768.643324

กฎข้อที่ 7 หมายความว่า ถ้าฟังก์ชัน (Function) มีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 23.1 และ Halstead time มีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 1768.643324 แล้ว Halstead effort จะมีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31635.579233

กฎข้อที่ 8 สามารถตีความได้แบบเดียวกันกับกฎข้อที่ห้า หากแต่เป็นการกลับค่ากัน กล่าวคือถ้าฟังก์ชัน (Function) มีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 23.1 และ Halstead effort ค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31635.579233 แล้ว Halstead time มีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 1768.643324

กฎข้อที่ 9 หมายความว่า ถ้า Halstead H1 มีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 30.4 และ Halstead time มีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 1768.643324 แล้ว Halstead effort จะมีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31635.579233

กฎข้อที่ 10 สามารถตีความได้แบบเดียวกันกับกฎข้อที่ห้า หากแต่เป็นการกลับค่ากัน กล่าวคือถ้า Halstead H1 มีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 30.4 และ Halstead effort มีค่าอยู่ในช่วง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 31635.579233 แล้วทำให้ Halstead time มีค่าอยู่ในช่วงระหว่าง –inf หรือ ติดลบอินฟินิตี้ (Infinity) ถึง 1768.643324

จากการวิเคราะห์กฎความสัมพันธ์ทั้งสิบข้อนั้นสามารถสรุปได้ว่า ข้อมูลส่วนมากอยู่ในช่วงเริ่มต้นของข้อมูล จะเห็นได้จากค่าแรกที่พิจารณาจะเป็นค่าที่เริ่มต้นด้วย ติดลบอินฟินิตี้ (-inf) ความสัมพันธ์ของตัวแปรที่พบบ่อยส่วนมาก คือตัวแปร Halstead โดยสังเกตุได้จากการที่ทุกกฎความสัมพันธ์ประกอบไปด้วยค่า Halstead

นอกจากการวิเคราะห์กฎความสัมพันธ์แล้ว ขั้นตอนวิธีเอไพรออไร ยังสามารถแสดงไอเทมเซตที่เกิดขึ้นบ่อยในข้อมูลอีกด้วย เพื่อเป็นการวิเคราะห์แนวโน้ม ว่าถ้าเทียบกับข้อมูลทั้งข้อมูล ตัวแปรทุกตัวแปรแล้ว ตัวแปรไหนที่มีโอกาสการเกิดได้บ่อยหรือเกิดจำนวนมากในฐานข้อมูล โอเท็มเซตที่เกิดขึ้นบ่อย รายละเอียดดังตารางที่ 3.11

ตารางที่ 3.11 ไอเทมเซตที่เกิดขึ้นบ่อย

ไอเทมเซต	ตัวแปรและช่วงของตัวแปร	จำนวน
ไอเทมเซตที่เกิดขึ้นบ่อยใน 1 ตัวแปร	file_changed='(-inf-17.3]'	3593
	LOC change='(-inf-550.9]'	3397
	Code chunk='(-inf-25.7]'	3436
	function='(-inf-23.1]'	3486
	method='(-inf-38.4]'	3320
	class='(-inf-5.7]'	3338
	hal:h1='(-inf-30.4]'	3505
	hal:effort='(-inf-31835.579233]'	3478
	hal:time='(-inf-1768.643324]'	3478
	Result=reject	3538
ไอเทมเซตที่เกิดขึ้นบ่อยใน 2 ตัวแปร	file_changed='(-inf-17.3]' LOC change='(-inf-550.9]'	3332
	file_changed='(-inf-17.3]' Code chunk='(-inf-25.7]'	3420
	file_changed='(-inf-17.3]' function='(-inf-23.1]'	3366
	file_changed='(-inf-17.3]' class='(-inf-5.7]'	3222
	file_changed='(-inf-17.3]' hal:h1='(-inf-30.4]'	3397

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.11 ไอเทมเซตที่เกิดขึ้นบ่อย (ต่อ)

ไอเทมเซต	ตัวแปรและช่วงของตัวแปร	จำนวน
ไอเทมเซตที่เกิดขึ้นบ่อยใน 2 ตัวแปร	file_changed='(-inf-17.3]' hal:effort='(-inf-31835.579233]'	3297
	file_changed='(-inf-17.3]' hal:time='(-inf-1768.643324]'	3297
	file_changed='(-inf-17.3]' Result=reject	3356
	LOC change='(-inf-550.9]' Code chunk='(-inf-25.7]'	3255
	Code chunk='(-inf-25.7]' function='(-inf-23.1]'	3238
	Code chunk='(-inf-25.7]' hal:h1='(-inf-30.4]'	3281
	function='(-inf-23.1]' hal:h1='(-inf-30.4]'	3362
	function='(-inf-23.1]' hal:effort='(-inf-31835.579233]'	3232
	function='(-inf-23.1]' hal:time='(-inf-1768.643324]'	3232
	function='(-inf-23.1]' Result=reject	3254
	method='(-inf-38.4]' hal:h1='(-inf-30.4]'	3216
	class='(-inf-5.7]' hal:h1='(-inf-30.4]'	3224
	hal:h1='(-inf-30.4]' hal:effort='(-inf-31835.579233]'	3217
	hal:h1='(-inf-30.4]' hal:time='(-inf-1768.643324]'	3217
	hal:h1='(-inf-30.4]' Result=reject	3273
	hal:effort='(-inf-31835.579233]' hal:time='(-inf-1768.643324]'	3478
	hal:effort='(-inf-31835.579233]' Result=reject	3251
	hal:time='(-inf-1768.643324]' Result=reject	3251
ไอเทมเซตที่เกิดขึ้นบ่อยใน 3 ตัวแปร	file_changed='(-inf-17.3]' LOC change='(-inf-550.9]' Code chunk='(-inf-25.7]'	3242
	file_changed='(-inf-17.3]' Code chunk='(-inf-25.7]' function='(-inf-23.1]'	3227
	file_changed='(-inf-17.3]' Code chunk='(-inf-25.7]' hal:h1='(-inf-30.4]'	3270

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.11 ไอเทมเซตที่เกิดขึ้นบ่อย (ต่อ)

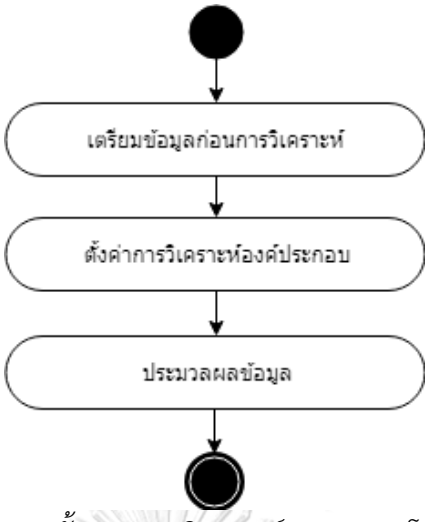
ไอเทมเซต	ตัวแปรและช่วงของตัวแปร	จำนวน
ไอเทมเซตที่เกิดขึ้นบ่อยใน 3 ตัวแปร	file_changed='(-inf-17.3]' Code chunk='(-inf-25.7]' hal:h1='(-inf-30.4]'	3270
	file_changed='(-inf-17.3]' function='(-inf-23.1]' hal:h1='(-inf-30.4]'	3263
	file_changed='(-inf-17.3]' hal:effort='(-inf-31835.579233]' hal:time='(-inf-1768.643324]'	3297
	function='(-inf-23.1]' hal:effort='(-inf-31835.579233]' hal:time='(-inf-1768.643324]'	3232
	hal:h1='(-inf-30.4]' hal:effort='(-inf-31835.579233]' hal:time='(-inf-1768.643324]'	3217
	hal:effort='(-inf-31835.579233]' hal:time='(-inf-1768.643324]' Result=reject	3251

จากตารางที่ 3.11 จะเห็นได้ว่ามีตัวแปรมากมายที่มีความสัมพันธ์กับผลลัพธ์การรีวิวเป็นไม่ผ่านการรีวิว กล่าวคือตัวแปรที่เกิดร่วมกับผลลัพธ์การรีวิวไม่ผ่าน ควรมีการพิจารณาและตรวจทานให้ถี่ถ้วนดีก่อนทำการร้องขอการนำเข้าซอร์สโค้ด โดยตัวแปรที่มีความสัมพันธ์กับผลลัพธ์การรีวิวโค้ดและเป็นไอเทมเซตที่เกิดขึ้นบ่อยได้แก่ file_changed='(-inf-17.3]' function='(-inf-23.1]' hal:h1='(-inf-30.4]' hal:h1='(-inf-30.4]' hal:effort='(-inf-31835.579233]' hal:time='(-inf-1768.643324]' และ hal:effort='(-inf-31835.579233]' ร่วมกับ hal:time='(-inf-1768.643324]'

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.2.5 สร้างแบบจำลองการทำนายด้วยการวิเคราะห์การถดถอยโลจิสติก

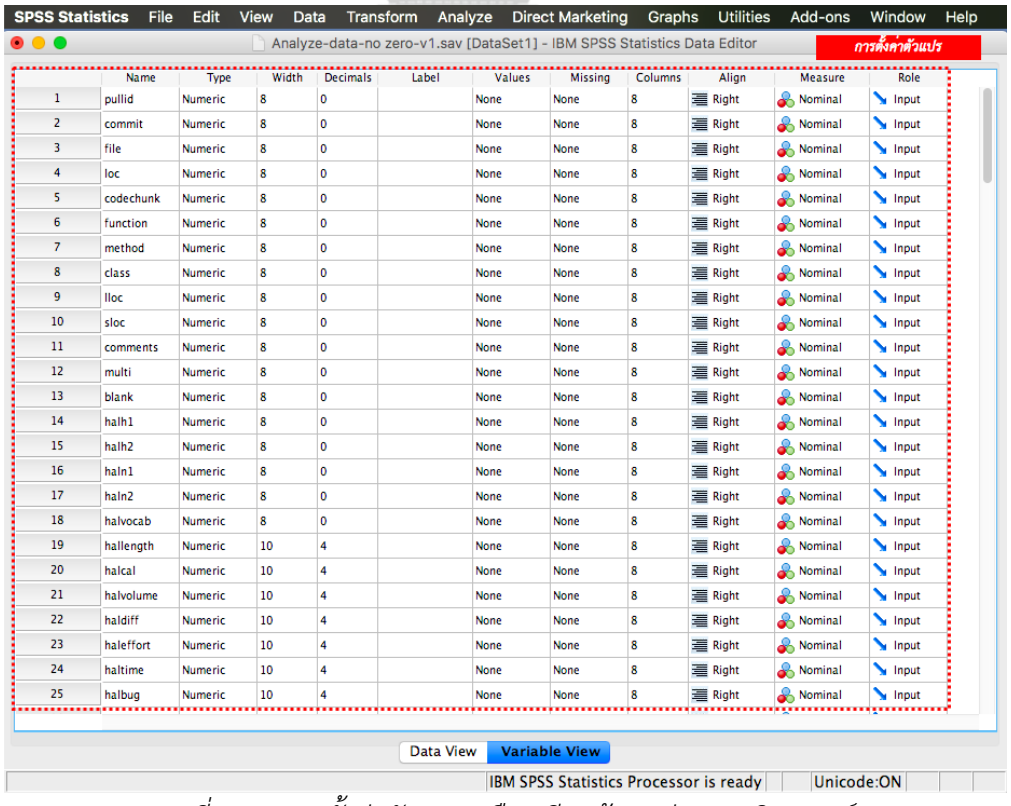
หลังจากที่ได้รวบรวมตัวแปรที่ผ่านการวิเคราะห์ และผ่านการทำความสะอาด (Data cleansing) มาแล้ว ขั้นตอนต่อไปคือการนำรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ๆ มาสร้างแบบจำลองการทำนายด้วยการวิเคราะห์การถดถอยโลจิสติก ซึ่งวิธีการสร้างโมเดลการทำนายด้วยการวิเคราะห์การถดถอยโลจิสติก ทำได้โดยใช้เครื่องมือเอสพีเอสเอส ประกอบด้วย 3 ขั้นตอนด้วยกัน คือ การเตรียมข้อมูลก่อนการวิเคราะห์ การตั้งค่าการวิเคราะห์องค์ประกอบ และการประมวลผลข้อมูล มีภาพรวมดังรูปที่ 3.12



รูปที่ 3.12 ขั้นตอนการวิเคราะห์การถดถอยโลจิสติก

ขั้นตอนที่ 1 การเตรียมข้อมูลก่อนการวิเคราะห์

ทำได้โดยการกำหนดตัวแปรทั้งหมดที่ได้รวบรวมไว้ รวมถึงการตั้งค่าของตัวแปรต่าง ๆ ไม่ว่าจะเป็น ชื่อของตัวแปร ประเภทของตัวแปร ความกว้างของคอลัมน์ จำนวนจุดทศนิยม การกำหนดค่าสำหรับตัวแปรที่เป็นนามบัญญัติ (Ordinal) และการกำหนดตัววัด ว่าตัวแปรนั้น ๆ จัดอยู่ในตัวแปรประเภทใด รายละเอียดการกำหนดตัวแปร ดังรูปที่ 3.13

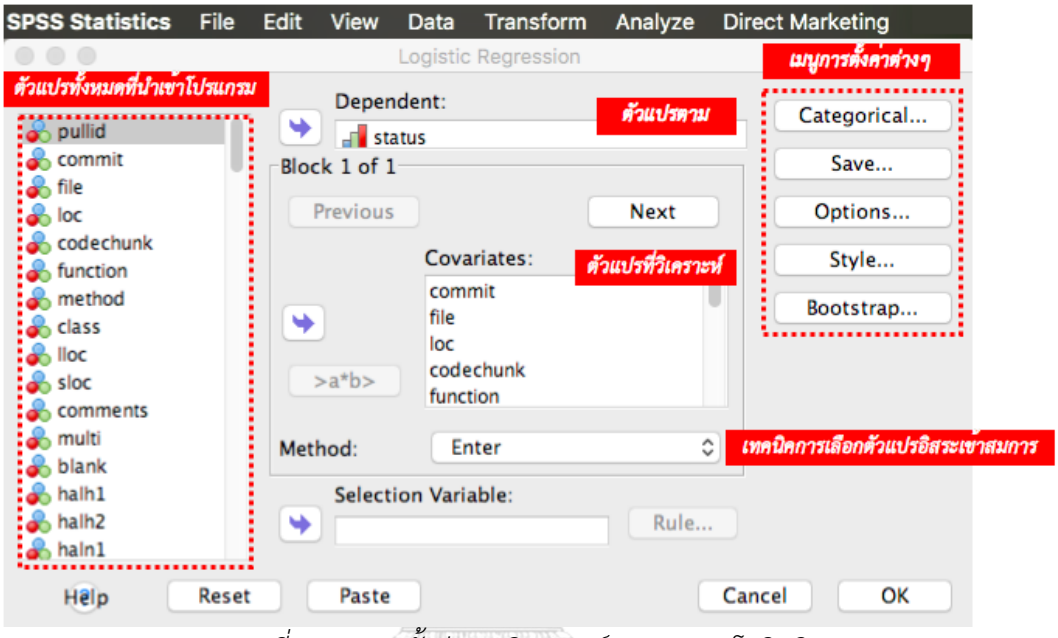


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	pullid	Numeric	8	0		None	None	8	Right	Nominal	Input
2	commit	Numeric	8	0		None	None	8	Right	Nominal	Input
3	file	Numeric	8	0		None	None	8	Right	Nominal	Input
4	loc	Numeric	8	0		None	None	8	Right	Nominal	Input
5	codechunk	Numeric	8	0		None	None	8	Right	Nominal	Input
6	function	Numeric	8	0		None	None	8	Right	Nominal	Input
7	method	Numeric	8	0		None	None	8	Right	Nominal	Input
8	class	Numeric	8	0		None	None	8	Right	Nominal	Input
9	lloc	Numeric	8	0		None	None	8	Right	Nominal	Input
10	sloc	Numeric	8	0		None	None	8	Right	Nominal	Input
11	comments	Numeric	8	0		None	None	8	Right	Nominal	Input
12	multi	Numeric	8	0		None	None	8	Right	Nominal	Input
13	blank	Numeric	8	0		None	None	8	Right	Nominal	Input
14	halh1	Numeric	8	0		None	None	8	Right	Nominal	Input
15	halh2	Numeric	8	0		None	None	8	Right	Nominal	Input
16	haln1	Numeric	8	0		None	None	8	Right	Nominal	Input
17	haln2	Numeric	8	0		None	None	8	Right	Nominal	Input
18	halvocab	Numeric	8	0		None	None	8	Right	Nominal	Input
19	hallength	Numeric	10	4		None	None	8	Right	Nominal	Input
20	halcal	Numeric	10	4		None	None	8	Right	Nominal	Input
21	halvolume	Numeric	10	4		None	None	8	Right	Nominal	Input
22	haldiff	Numeric	10	4		None	None	8	Right	Nominal	Input
23	haleffort	Numeric	10	4		None	None	8	Right	Nominal	Input
24	haltime	Numeric	10	4		None	None	8	Right	Nominal	Input
25	halbug	Numeric	10	4		None	None	8	Right	Nominal	Input

รูปที่ 3.13 การตั้งค่าตัวแปร หรือเตรียมข้อมูลก่อนการวิเคราะห์

ขั้นตอนที่ 2 การตั้งค่าการวิเคราะห์องค์ประกอบ

นำข้อมูลเข้าโปรแกรมการประมวลผลเอสพีเอสเอส กำหนดตัวแปร Status หรือผลลัพธ์การทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับเป็นตัวแปรตาม กำหนดตัวแปรอื่นๆอีก 62 ตัว เป็นตัวแปรที่ใช้ในการวิเคราะห์ หรือตัวแปรควบคุม (Covariates) และทำการตั้งค่าการวิเคราะห์การถดถอยโลจิสติกในขั้นตอนต่าง ๆ ดังรูปที่ 3.14 ซึ่งมีรายละเอียดดังตารางที่ 3.12



รูปที่ 3.14 การตั้งค่าการวิเคราะห์การถดถอยโลจิสติก

ตารางที่ 3.12 รายละเอียดการตั้งค่าการวิเคราะห์การถดถอยโลจิสติก

เมนู	เมนูย่อย	การตั้งค่า	คำอธิบาย
Logistic Regression: Save	Predicted Values	Probabilities	การเลือกตัวแปรที่ต้องการวิเคราะห์องค์ประกอบ
		Group membership	เป็นค่าพยากรณ์ว่าแต่ละเคสจะต้องอยู่กลุ่มใด
	Influence	Cook's	ใช้วัดอิทธิพลของแต่ละเคส
	Residuals	Unstandardized	ไม่มาตรฐาน
		Standardized	มาตรฐาน
	Include the covariance matix	เลือก Include the covariance matix	รวมเมตริกความแปรปรวน

ตารางที่ 3.12 รายละเอียดการตั้งค่าการวิเคราะห์การถดถอยโลจิสติก (ต่อ)

เมนู	เมนูย่อย	การตั้งค่า	คำอธิบาย
Logistic Regression: Options	Statistics and Plots	Classification plots	แสดงกราฟฮิสโตแกรม
		Hosmer-Lemeshow goodness-of-fit	ใช้ในการตรวจสอบความเหะมาสมของแบบจำลอง
		Casewise listing of residuals	แสดงค่า Unstandardized Residual, Predicted probability, Observed group membership และ Predicting group membership
		Outliers outside 2 std. dev.	แสดงเฉพาะค่า
	Display	At each step	ในทุก ๆ ขั้นตอนการทำงาน
	Include constant in model	เลือก Include constant in model	รวมค่าคงที่ในแบบจำลองด้วย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ขั้นตอนที่ 3 การประมวลผลข้อมูล

ประมวลผลข้อมูลหรือทำการวิเคราะห์การถดถอยโลจิสติกหลังจากที่นำข้อมูลเข้าและตั้งค่า ค่าต่าง ๆ ครบถ้วนดีแล้ว โดยหลังจากกำหนดค่าต่าง ๆ และทำการวิเคราะห์การถดถอยโลจิสติกแล้วนั้น ผลลัพธ์ในเครื่องมือเอสพีเอสเอสจะแสดงใน IBM SPSS Statistics Viewer โดยมีรายละเอียด ดังนี้

1. ผลลัพธ์สรุปการเตรียมกรณีตัวอย่าง (Case Processing Summary)

แสดงการเตรียมกรณีตัวอย่าง โดยกรณีที่เลือกมาวิเคราะห์มีจำนวน 1,208 รายการ ซึ่งคือจำนวนรายการร้องขอการนำซอร์สโค้ดทั้งหมดที่ได้จัดเตรียมไว้ โดยที่ทุกรายการมีค่าครบถ้วน ไม่มีกรณีที่ขาดหายไป รายละเอียดดังตารางที่ 3.13

ตารางที่ 3.13 ผลลัพธ์สรุปการเตรียมกรณีตัวอย่าง (Case Processing Summary)

Unweighted Cases		N	Percent
Selected Cases	Included in Analysis	1208	100.0
	Missing Cases	0	.0
	Total	1208	100.0
Unselected Cases		0	.0
Total		1208	100.0

2. ผลลัพธ์การกำหนดค่าตัวแปรตาม (Dependent Variable Encoding)

ผลลัพธ์การกำหนดค่าตัวแปรตามแสดงการกำหนดค่าตัวแปรตาม โดยในที่นี้กำหนดให้ Accept มีค่าเท่ากับ 0 และ Reject มีค่าเท่ากับ 1 รายละเอียดดังตารางที่ 3.14

ตารางที่ 3.14 ผลลัพธ์การกำหนดค่าตัวแปรตาม

Original Value	Internal Value
accept	0
reject	1

3. ผลลัพธ์การวิเคราะห์ Omnibus Tests of Model Coefficients

การวิเคราะห์ Omnibus Tests of Model Coefficients เป็นการวิเคราะห์การทดสอบค่าไคสแควร์ (Chi Square) และการตรวจสอบค่านัยสำคัญ (Siginificant) รายละเอียดดังตารางที่ 3.15

ตารางที่ 3.15 ผลลัพธ์การวิเคราะห์ Omnibus Tests of Model Coefficients

Details		Chi-square	df	Sig.
Step1	Step	150.529	54	.000
	Block	150.529	54	.000
	Model	150.529	54	.000

จากตาราง 3.15 การวิเคราะห์สถิติทดสอบ 3 ค่าคือ สถิติทดสอบของสเตป (Step) สถิติทดสอบของบล็อก (Block) และสถิติทดสอบของโมเดล (Model) มีรายละเอียดดังนี้สถิติทดสอบ ของสเตป (Step) ใช้ทดสอบสมมติฐาน โดย กำหนด H0 คือ โอกาสที่ผลลัพธ์การรีวิวซอร์สโค้ดไม่ขึ้นอยู่กับตัวแปรที่วิเคราะห์ และ H1 คือโอกาสที่ผลลัพธ์การรีวิวซอร์สโค้ดขึ้นอยู่กับตัวแปรที่วิเคราะห์ ซึ่งค่าสถิติทดสอบของสเตปมีค่า 150.529 และค่านัยสำคัญที่ระดับ .000 นั่นคือปฏิเสธ H0 ในส่วนของค่าสถิติ

ทดสอบของสเตปและค่าสถิติทดสอบของบล็อกมีค่าเท่ากับค่าสถิติทดสอบของแบบจำลอง เนื่องจากแบบจำลองมีเฉพาะค่าคงที่ และตัวแปรอิสระทั้ง 63 ตัว

4. ผลลัพธ์การสรุปแบบจำลอง (Model Summary)

แสดงความเหมาะสมของแบบจำลอง (Goodness of fit) นั่นคือ -2 Log likelihood ซึ่งผลลัพธ์แสดงค่า -2 Log likelihood มีมากกว่าค่าคงที่ รายละเอียดดังตาราง 3.16

ตารางที่ 3.16 ผลลัพธ์การสรุปแบบจำลอง

Step	-2 Loglikeilhood	Cox & Snell R Square	Nagelkerke R Square
1	302.828	0.117	0.374

5. ผลลัพธ์การทดสอบความเหมาะสมของโมเดลด้วย Hosmor and Lemeshow Test ของ Block 1

Block 1 หรือข้อมูลที่ผ่านการวิเคราะห์ตัวแปรเข้าสู่สมการแล้ว โดยผลลัพธ์การทดสอบความเหมาะสมของแบบจำลองด้วย Hosmor and Lemeshow Test ใช้ในการทดสอบความเหมาะสมของแบบจำลอง จากค่าสถิติทดสอบ (Chi-square) มีค่าเท่ากับ 12.481 และค่า Significance มีค่าเท่ากับ 0.013 ซึ่งน้อยกว่า 0.05 จึงสรุปได้ว่าแบบจำลองนี้เหมาะสม

6. ผลลัพธ์การตรวจสอบความน่าเชื่อถือของข้อมูล (Classification Table)

ตารางการจำแนกข้อมูลแสดงความน่าเชื่อถือข้อมูล ซึ่งประกอบด้วยผลลัพธ์ของรายการวิเคราะห์ในรูปแบบการแสดงจำนวนรายการที่ทำนายถูกหรือผิด เพื่อบอกว่ามีความแม่นยำเพียงใด รายละเอียดดังตารางที่ 3.17

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.17 ตารางการจำแนกข้อมูลแสดงความน่าเชื่อถือข้อมูล

Observed			Predicted		
			Status		Percent correct
			accept	reject	
Step1	Status	accept	7	49	12.5
		reject	4	1148	99.7
	Overall Percentage				95.6

จากตารางที่ 3.17 การรีวิวโค้ดที่ผลลัพธ์การรีวิวเป็นผ่าน หรือยอมรับ (accept) มีทั้งหมด 56 รายการ มี 7 รายการที่พยากรณ์ผลการรีวิวถูกต้องหรือตรงตามความจริง นั่นคือพยากรณ์ถูกต้อง

12.5% และหากพิจารณาการรีวิวโค้ดที่ผลลัพธ์การรีวิวเป็นไม่ผ่าน หรือปฏิเสธ (reject) มีทั้งหมด 1,152 รายการ มีการพยาการณ์ผลการรีวิวถูกต้องหรือตรงตามความจริง นั่นคือพยากรณ์ถูกต้อง 99.7% โดยรวมผลลัพธ์การพยากรณ์ถูกต้องแล้ว 95.6 เปอร์เซ็นต์

7. ผลลัพธ์ตัวแปรที่อยู่ในสมการ (Variables in the Equation) ของ Block 1 รายละเอียดดังตารางที่ 3.18 เป็นการสรุปว่ามีตัวแปรใดบ้างที่อยู่ในสมการวิเคราะห์การถดถอยโลจิสติก รวมถึงค่าต่าง ๆ โดยค่า B หมายถึงสัมประสิทธิ์โลจิสติก สถิติการทดสอบวอลด์ (Wald Statistic) ค่านัยสำคัญ (Significant) ของการทดลอง และค่า Exp (B) ที่แปลความมาจากค่า B

ตารางที่ 3.18 ตัวแปรและค่าสัมประสิทธิของตัวแปรที่อยู่ในสมการ

Step1	B	S.E.	Wald	df	Sig.	Exp (B)
commit	-.085	.088	.937	1	.333	.918
file	.036	.062	.329	1	.566	1.036
loc	.002	.001	4.261	1	.039	1.002
code chunk	-.031	.023	1.862	1	.172	.969
function	-.052	.027	3.779	1	.052	.949
method	-.052	.018	8.112	1	.004	.949
class	.028	.090	.098	1	.755	1.028
lloc	.016	.003	21.069	1	.000	1.016
sloc	-.005	.001	27.790	1	.000	.995
comments	.007	.007	1.205	1	.272	1.007
multi	.008	.005	2.714	1	.099	1.008
blank	-.005	.005	1.084	1	.298	.995
halh1	-.053	.059	.785	1	.376	.949
halh2	.086	.031	7.722	1	.005	1.090
haln1	-.045	.043	1.105	1	.293	.956
haln2	.021	.022	.966	1	.326	1.021
hallength	-.009	.003	10.876	1	.001	.991
halvolume	-.001	.001	.896	1	.344	.999
haldiff	-.084	.112	.560	1	.454	.920
haleffort	.000	.000	2.632	1	.105	1.000

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.18 ตัวแปรและค่าสัมประสิทธิของตัวแปรที่อยู่ในสมการ (ต่อ)

Step1	B	S.E.	Wald	df	Sig.	Exp (B)
cc	-1255.242	432.556	8.421	1	.004	.000
ccl	.165	.092	3.206	1	.073	1.180
cco	.003	.013	.038	1	.846	1.003
ci	-.045	.080	.312	1	.577	.956
clc	1622.158	518.588	9.785	1	.002	.
ldc	-.502	.602	.694	1	.405	.605
lldc	.002	.011	.036	1	.850	1.002
lcom5	-.008	.014	.362	1	.547	.992
nl	-.038	.025	2.220	1	.136	.963
nle	-.007	.067	.012	1	.912	.993
wmc	.059	.083	.510	1	.475	1.061
cbo	.018	.007	6.083	1	.014	1.018
cboi	.225	.575	.153	1	.696	1.252
nii	-.001	.559	.000	1	.999	.999
noi	.076	.050	2.279	1	.131	1.079
rfc	.040	.050	.625	1	.429	1.040
cd	1.665	1.010	2.719	1	.099	5.286
cloc	.186	.166	1.256	1	.262	1.204
dloc	-.128	.356	.129	1	.720	.880
tcd	-.187	.551	.115	1	.735	.830
dit	.137	.357	.148	1	.700	1.147
noa	.199	3.355	.004	1	.953	1.220
noc	-12.294	1058.907	.000	1	.991	.000
nod	-1.021	.352	8.422	1	.004	.360
nop	12.733	1058.902	.000	1	.990	338807.948
na	-.018	.045	.162	1	.687	.982
nla	.097	.075	1.706	1	.191	1.102
nlm	-.100	.093	1.140	1	.286	.905

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.18 ตัวแปรและค่าสัมประสิทธิของตัวแปรที่อยู่ในสมการ (ต่อ)

Step1	B	S.E.	Wald	df	Sig.	Exp (B)
nm	.067	.030	5.203	1	.023	1.070
nos	-.052	.049	1.136	1	.287	.949
ttloc	-.006	.004	1.973	1	.160	.994
tloc	.008	.005	3.065	1	.080	1.008
tna	.007	.045	.027	1	.870	1.007
tnos	.017	.052	.108	1	.743	1.017
constant	.676	4.478	.023	1	.880	1.966

โดยจากผลการทดลองนั้น สามารถสรุปเป็นสมการวิเคราะห์การถดถอยโลจิสติกได้ ดังนี้

Logistic regression equation (สมการวิเคราะห์การถดถอยโลจิสติก) = 2.417 + commit (-0.085) + file (0.036) + loc (0.002) + codechunk (-0.031) + function (-0.052) + method (-0.052) + class (0.028) + lloc (0.016) + sloc (-0.005) + comments (0.007) + multi (0.008) + blank (-0.005) + halh1 (-0.053) + halh2 (0.086) + haln1 (-0.045) + haln2 (0.021) + hallength (-0.009) + halvolume (-0.001) + haldiff (-0.084) + haleffort (0) + cc (-1255.242) + ccl (0.165) + cco (0.003) + ci (-0.045) + clc (1622.158) + ldc (-0.502) + lldc (0.002) + lcom5 (-0.008) + nl (-0.038) + nle (-0.007) + wmc (0.059) + cbo (0.018) + cboi (0.225) + nii (-0.001) + noi (0.076) + rfc (0.040) + cd (1.665) + cloc (0.186) + dloc (-0.128) + tcd (-0.187) + dit (0.137) + noa (0.199) + noc (-12.294) + nod (-1.021) + nop (12.733) + na (-0.018) + nla (0.097) + nlm (-0.100) + nm (0.067) + nos (-0.052) + tlloc (-0.006) + tloc (0.008) + tna (0.007) + tnos (0.017)

โดยหลังจากที่ได้สมการวิเคราะห์การถดถอยโลจิสติกแล้วนั้น ขั้นตอนต่อมาคือ การสร้างแบบจำลองการทำนายหรือเรียกได้อีกอย่างว่า สมการทำนาย โดยสมการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ สามารถเขียนอยู่ในรูปเศษส่วน เพื่อแสดงถึงผลลัพธ์ ที่มีค่าอยู่ในช่วง 0 ถึง 1 ดังรูปที่ 3.15

$$\text{สมการทำนาย} = \frac{e^{\text{Logistic regression equation}}}{1 + e^{\text{Logistic regression equation}}}$$

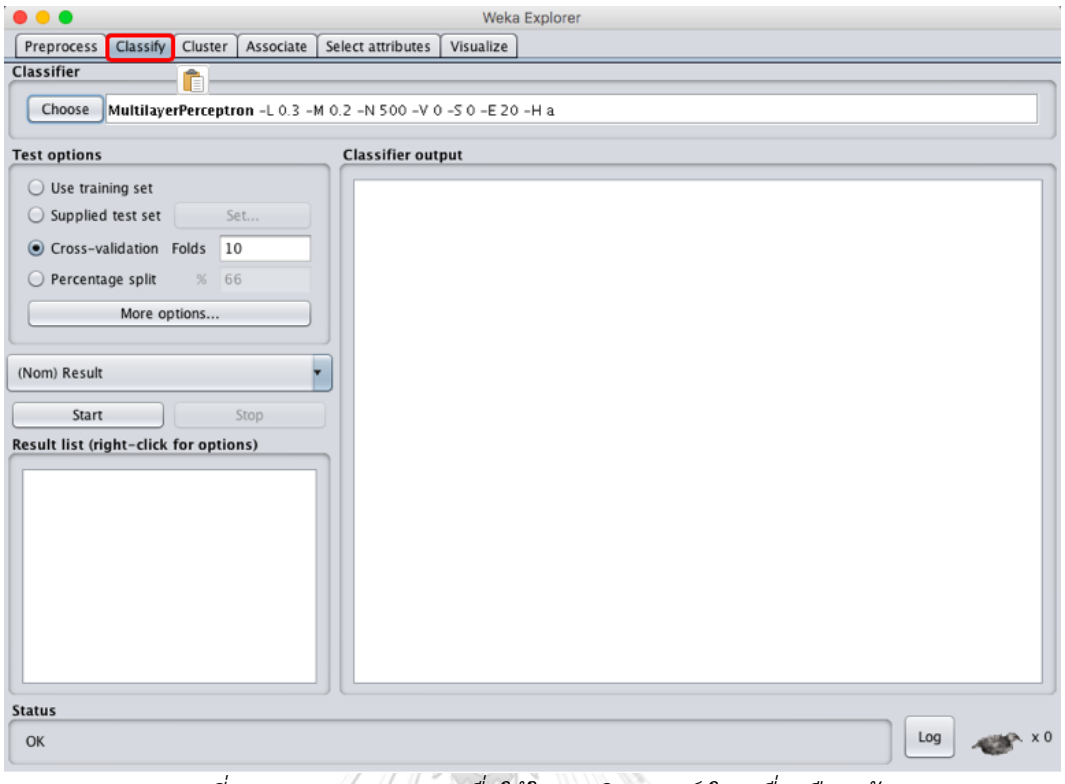
รูปที่ 3.15 สมการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

จากรูปที่ 3.15 สมการทำนาย ประกอบด้วยค่าอี (e) หรือค่าคงตัวทางคณิตศาสตร์ ซึ่งมีค่าเท่ากับ 2.7182818284 .... และ Logistic regression equation คือสมการโลติสจิสติกส์ที่ได้แสดงไปข้างต้น โดยผลลัพธ์ที่ได้จากสมการทำนาย จะอยู่ระหว่างศูนย์ถึงหนึ่ง โดยกำหนดหากค่าที่ได้จากสมการทำนายมีค่าน้อยกว่าหรือเท่ากับ 0.5 ให้พยากรณ์หรือทำนายผลการรีวิวเป็นผ่าน หรือยอมรับ (accepted) แต่หากค่าที่ได้จากสมการทำนายมีค่ามากกว่า 0.5 ให้พยากรณ์หรือทำนายผลการรีวิวเป็นไม่ผ่าน หรือปฏิเสธ (rejected)

3.2.6 สร้างแบบจำลองการทำนายด้วยการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น

แบบจำลองการทำนายด้วยการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอน ใช้รายการร้องขอการนำเข้าซอร์สโค้ดที่ประกอบด้วย 63 ตัวแปรจากการรวบรวมข้อมูลตัวแปรมาวิเคราะห์ข้อมูลในเครื่องมือเวก้า เนื่องด้วยตัวแปร Pull request ID ไม่ได้นำมาใช้ในการวิเคราะห์ข้อมูลในเวก้า เนื่องจากเครื่องมือเวก้า เป็นเครื่องมือที่มีลำดับ 1,2,3 เรียงรายการที่นำเข้าให้อยู่แล้ว หากการวิเคราะห์นั้น นำตัวแปร Pull request ID เข้าไปวิเคราะห์ด้วย เครื่องมือเวก้า จะถือว่า Pull request ID เป็นหนึ่งในตัวแปรที่ต้องวิเคราะห์ และทำการวิเคราะห์ข้อมูลดังกล่าว ซึ่งจะทำให้ข้อมูลมีความผิดพลาด หรือไม่ตรงตามที่ต้องการ เนื่องจาก Pull request ID ไม่ใช่ตัวแปรที่เราสนใจ หรือนำมาวิเคราะห์ หากแต่เป็นตัวแปรที่ใช้ในการแสดงเลขลำดับของรายการร้องขอการนำเข้าซอร์สโค้ด ซึ่งวิธีการการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น ทำได้โดยการประมวลผลผ่านฟีเจอร์ที่มีชื่อว่า MultilayerPerceptron ภายใต้เมนูจำแนกประเภท (Classify) ในเครื่องมือเวก้า ดังรูปที่ 3.16

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

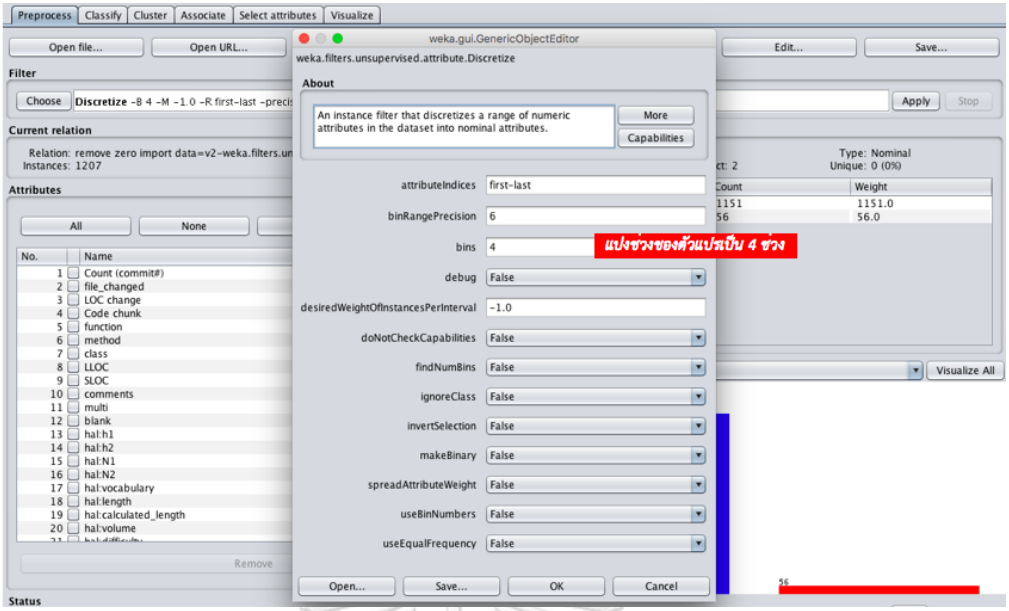


รูปที่ 3.16 เมนู Classify เพื่อใช้ในการวิเคราะห์ ในเครื่องมือเวก้า

วิธีการวิเคราะห์โครงข่ายประสาทเทียมหลายชั้น ประกอบด้วย 3 ขั้นตอนด้วยกัน ได้แก่ การเตรียมข้อมูลก่อนการวิเคราะห์ การตั้งค่าการวิเคราะห์ และการประมวลผลข้อมูล ซึ่งมีรายละเอียดดังนี้

ขั้นตอนที่ 1 การเตรียมข้อมูลก่อนการวิเคราะห์

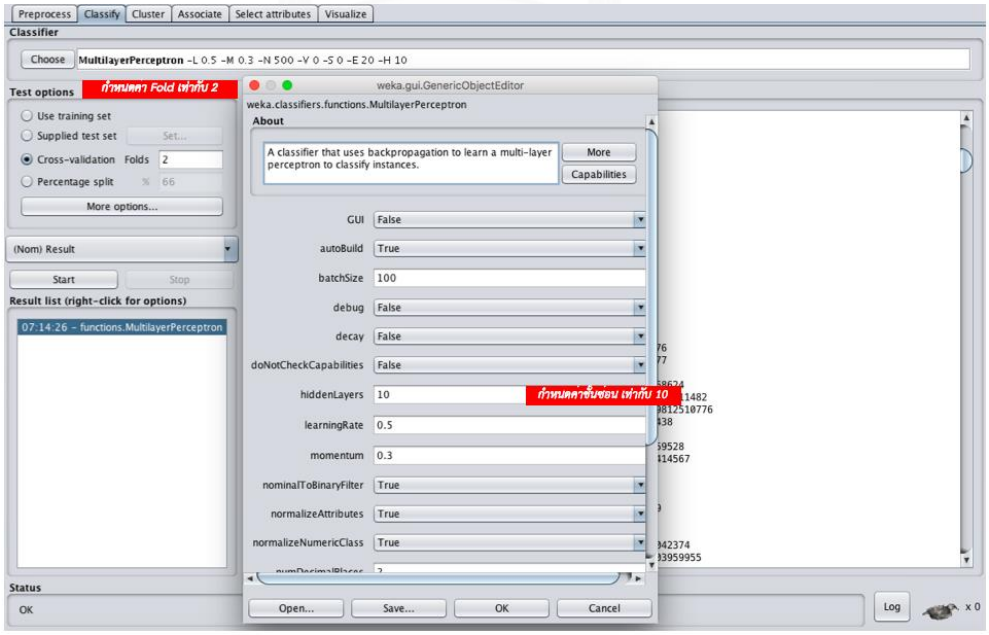
เนื่องด้วยตัวแปรส่วนมากที่นำมาวิเคราะห์ เป็นตัวแปรที่มีค่าต่อเนื่อง การวิเคราะห์ผ่านเครื่องมือเวก้านั้น ต้องทำการแบ่งค่าต่อเนื่องให้เป็น ช่วง ๆ ก่อนจะนำไปวิเคราะห์ข้อมูลได้ โดยเทคนิคที่เลือกใช้ในการเตรียมข้อมูลก่อนการวิเคราะห์ คือเทคนิคความไม่ต่อเนื่อง (Discretize) ซึ่งเป็นเทคนิคการทำบินนิง (Binning Method) รูปแบบหนึ่งที่แปลงช่วงของค่าตัวเลขในกลุ่มข้อมูลให้เป็นค่านามบัญญัติ ซึ่งจากการใช้เทคนิคความไม่ต่อเนื่องในการประมวลผลนั้น สามารถกำหนดค่าบิน หรือค่าช่วงของข้อมูลในการประมวลผลได้ ในที่นี้มีการกำหนดค่าบินเท่ากับสี่ หรือเรียกได้ว่าเป็นการแบ่งข้อมูลออกเป็น 4 ช่วง เนื่องด้วยการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้นเป็นการสร้างโมเดลที่ใช้ในการทำนาย ไม่ใช่การหารูปแบบความสัมพันธ์ของข้อมูลเหมือนการหากฎความสัมพันธ์ การใช้ค่าบินเท่ากับสี่จึงเพียงพอกับการวิเคราะห์ข้อมูล หน้าจอการแบ่งค่าต่อเนื่อง ดังรูปที่ 3.17



รูปที่ 3.17 การแบ่งช่วงของข้อมูลสำหรับการเตรียมข้อมูลก่อนการวิเคราะห์

ขั้นตอนที่ 2 การตั้งค่าการวิเคราะห์

ขั้นตอนวิธีที่ใช้ในการวิเคราะห์คือฟีเจอร์ที่มีชื่อว่า MultilayerPerceptron ซึ่งอยู่ภายใต้การวิเคราะห์ฟังก์ชัน ในเมนูจำแนกประเภท (Classify) ของเครื่องมือเวก้า การวิเคราะห์ขั้นตอนวิธีนี้สามารถตั้งค่าชั้นที่ซ่อน (HiddenLayer) ได้ โดยในที่นี้กำหนดค่าชั้นที่ซ่อนเป็น 10 ชั้น อีกทั้งยังสามารถกำหนดการตั้งค่าตัวเลือกการทดสอบได้ อีกทั้งยังสามารถเลือกวิธีในการทดสอบได้ ในที่นี้ใช้ขั้นตอนวิธี Cross-validation Folds กำหนดค่าให้เท่ากับ 2 รายละเอียดหน้าจอ ดังรูปที่ 3.18

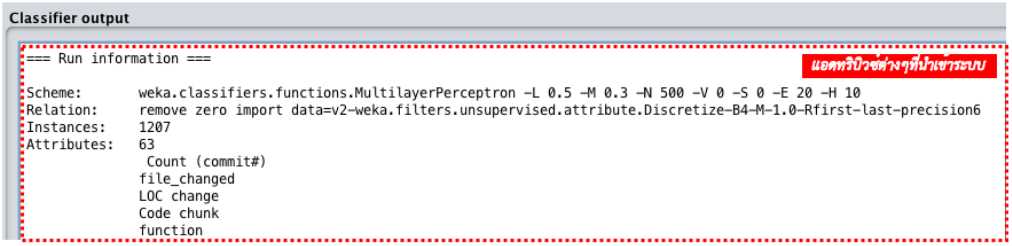


รูปที่ 3.18 หน้าจอการตั้งค่าการวิเคราะห์

ขั้นตอนที่ 3 การประมวลผลข้อมูล

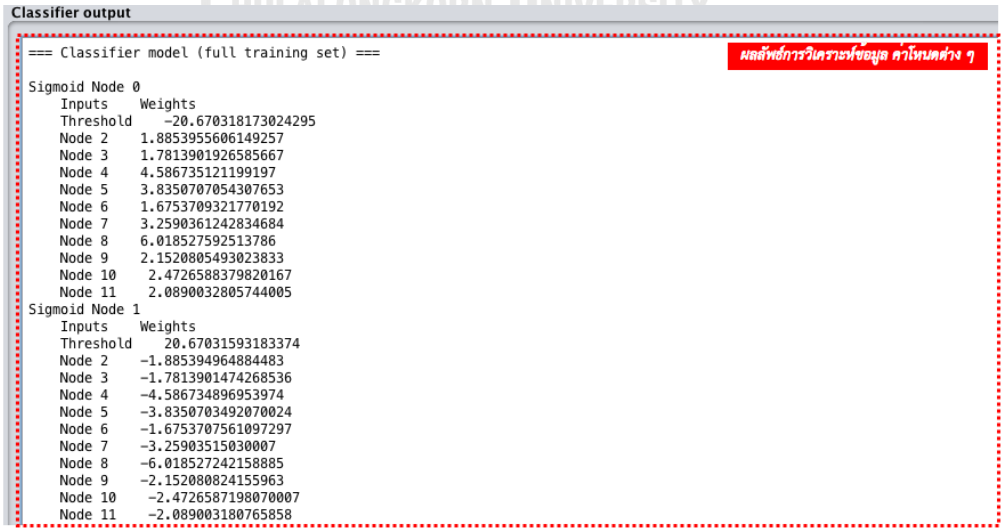
ประมวลผลข้อมูลหรือทำการวิเคราะห์องค์ประกอบหลังจากที่นำข้อมูลเข้าและตั้งค่า ค่าต่าง ๆ ครบถ้วนดีแล้ว ผลลัพธ์การวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้นจะแสดงในรูปแบบข้อความบนกล่อง Classifier output ในโปรแกรมเวก้า ซึ่งผลการวิเคราะห์ข้อมูลแบ่งออกเป็น 4 ส่วนด้วยกัน โดยในแต่ละส่วน มีรายละเอียดดังนี้

ส่วนที่ 1 เป็นรายละเอียดของไฟล์ที่ใช้ในการวิเคราะห์ข้อมูล เค้าร่าง (Schema) แสดงวิธีการที่ใช้ในการวิเคราะห์ข้อมูล และการตั้งค่าต่าง ๆ ในการวิเคราะห์ข้อมูล ความสัมพันธ์(Relation) แสดงไฟล์ที่ใช้ในการวิเคราะห์ข้อมูล และวิธีการนำไฟล์เข้าสู่ระบบ กรณีตัวอย่าง (Instance) แทนกรณีตัวอย่างหรือรายการทรานแซคชันที่นำเข้าสู่ระบบ แอดทริบิวซ์ (Attibutes) คือจำนวนตัวแปรที่ใช้ในการวิเคราะห์ข้อมูล และในส่วนสุดท้ายของส่วนแรก จะเป็นการบอกตัวแปรหรือแอดทริบิวซ์ต่าง ๆ ที่นำเข้าระบบ ดังรูปที่ 3.19



รูปที่ 3.19 รายละเอียดการวิเคราะห์ (ข้อมูลส่วนแรก)

ส่วนที่ 2 คือแบบจำลองที่ได้จากการวิเคราะห์ โดยในตัวแบบจำลองนี้จะระบุ โหนดต่าง ๆ ที่ได้จากการคำนวณ ค่าแอดทริบิวซ์ ค่าช่วงของแต่ละแอดทริบิวซ์ และน้ำหนักของตัวแปร เพื่อใช้ในการคำนวณสูตรในการทำนายต่อไป รายละเอียดหน้าจอดังรูปที่ 3.20



รูปที่ 3.20 หน้าจอผลลัพธ์การวิเคราะห์ข้อมูล (ข้อมูลส่วนที่สอง)

จากรูปที่ 3.20 ในแต่ละโหนดมีรายละเอียดดังนี้

โหนดที่ 0 แสดงค่า Threshold เท่ากับ -20.67031873024295 ซึ่งเป็นการแปลค่าของการผลการรีวิวโค้ดเป็นยอมรับ (Accept) และค่าที่แสดงในโหนดที่ 0 ได้แก่ ตัวแปรโหนด และน้ำหนักของตัวแปรโหนดนั้น ๆ โดยตัวแปรโหนดที่แสดงในโหนดที่ 0 ได้แก่ ตัวแปรโหนดที่ 2, ตัวแปรโหนดที่ 3, ตัวแปรโหนดที่ 4, ตัวแปรโหนดที่ 5, ตัวแปรโหนดที่ 6, ตัวแปรโหนดที่ 7, ตัวแปรโหนดที่ 8 และ ตัวแปรโหนดที่ 9

โหนดที่ 1 แสดงค่า Threshold เท่ากับ 20.67031873024295 ซึ่งเป็นการแปลค่าของการผลการรีวิวโค้ดเป็นปฏิเสธ (Reject) และค่าที่แสดงในโหนดที่ 1 ได้แก่ ตัวแปรโหนด และน้ำหนักของตัวแปรโหนดนั้น ๆ โดยตัวแปรโหนดที่แสดงในโหนดที่ 1 ได้แก่ ตัวแปรโหนดที่ 2, ตัวแปรโหนดที่ 3, ตัวแปรโหนดที่ 4, ตัวแปรโหนดที่ 5, ตัวแปรโหนดที่ 6, ตัวแปรโหนดที่ 7, ตัวแปรโหนดที่ 8 และ ตัวแปรโหนดที่ 9

โหนดที่ 2 ถึงโหนดที่ 9 แสดงรูปแบบผลลัพธ์ลักษณะเหมือนกัน กล่าวคือแสดงค่าช่วงของข้อมูลของตัวแปร (ตามที่ได้กำหนดค่าการแบ่งหรือค่าบินไปข้างต้น) ซึ่งมีทั้งหมด 4 ช่วง และค่าน้ำหนักของตัวแปรในแต่ละช่วง โดยตัวแปรที่แสดงในโหนดที่ 2 ถึงโหนดที่ 9 นั้น ประกอบไปด้วยตัวแปร 63 ตัวแปรที่นำเข้ามาสำหรับการวิเคราะห์

การคำนวณค่าของโหนดแต่ละโหนด ทำได้โดยพิจารณาค่าของตัวแปรแต่ละตัว ว่าอยู่ในช่วงใดและนำค่าน้ำหนักของตัวแปรในช่วงนั้นมาคูณกับค่าของตัวแปรนั้น ๆ และนำซ้ำแบบนี้กับตัวแปรทุกตัว จากนั้นนำผลคูณระหว่างตัวแปรกับน้ำหนักของตัวแปรมาบวกกัน ผลรวมของทุก ๆ ตัวแปร คือค่าของโหนดนั้น ๆ

ตัวอย่างเช่น ในโหนดที่ 2 มีจำนวนคอมมิท (commit) มีค่าเท่ากับ 5 พิจารณาจากช่วงของค่าแล้ว จำนวนคอมมิทจะอยู่ในช่วงที่ 1 หรือ -inf-8 ดังนั้นจึงต้องนำเอาจำนวนคอมมิทไปคูณกับค่าน้ำหนักที่อยู่ในช่วงนั้นในโหนดที่ต้องการพิจารณา หากจะพิจารณาค่าในโหนดที่สอง น้ำหนักจำนวนคอมมิทช่วงแรกคือ -1.7698 ค่าที่คำนวณได้จะออกมาเป็น 5 x -1.7698 ได้เป็น -8.849 ทำแบบนี้ไปจนครบทุกตัวแปรในโหนดที่สอง แล้วเอามาบวกกัน จะได้ค่าของโหนดที่ 2

หลังจากการคำนวณค่าครบทุกโหนดแล้ว จากผลลัพธ์ของโหนดที่ 0 และโหนดที่ 1 สามารถนำมาสร้างเป็นสมการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น ได้มาจากผลรวมของโหนดแต่ละโหนดคูณกับค่าน้ำหนักของโหนดแต่ละตัว ซึ่งสามารถแสดงอยู่ในรูปสมการได้ดังนี้

Multi layer perceptron equation = 20.6703 + (1.8854) Node2 + (1.7814) Node 3 + (4.5867) Node 4 + (3.8351) Node 5 + (1.6754) Node 6 + (3.2590) Node 7 + (6.0185) Node 8 + (2.1521) Node 9 + (2.4727) Node 10 + (2.0890) Node 11

โดยหลังจากคำนวณสมการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้นแล้ว นำค่าที่ได้จากสมการนั้นมาวิเคราะห์ต่อเพื่อสร้างเป็นแบบจำลองการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้นได้ โดยการวิเคราะห์ค่าของผลลัพธ์ หากผลลัพธ์มีค่าเป็นบวก หรือมากกว่า 0 ให้ตีความว่า ผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับเป็นยอมรับ (Accept) และหากผลลัพธ์มีค่าเป็นลบ หรือน้อยกว่า 0 ให้ตีความว่าผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับเป็นปฏิเสธ (Reject)

ส่วนที่ 3 แสดงเมตริกความสับสน (Confusion Matrix) ที่บ่งบอกถึงค่าความระลึก (Recall) และค่าความแม่นยำ (Accuracy) ของแบบจำลอง รายละเอียดดังตารางที่ 3.19

ตารางที่ 3.19 เมตริกความสับสนของการวิเคราะห์เพอร์เซปตรอนหลายชั้น

ผลลัพธ์เป็นไม่ผ่านหรือปฏิเสธ (Reject)	ผลลัพธ์เป็นผ่านหรือยอมรับ (Accept)	<-- Classified as
1125	26	
54	2	

จากตารางที่ 3.19 แสดงค่าความแม่นยำของแบบจำลอง โดยจากข้อมูลที่แสดงพบว่า รายการร้องขอการนำเข้าซอร์สโค้ดที่นำมาวิเคราะห์และมีผลลัพธ์เป็นไม่ผ่านหรือปฏิเสธ (reject) มีรายการที่ตอบตรงตามผลลัพธ์ 1,125 รายการ จากรายการที่มีผลลัพธ์เป็นไม่ผ่านหรือปฏิเสธทั้งหมด 1,179 รายการ คิดเป็นความแม่นยำถูกต้อง 95 เปอร์เซ็นต์ และรายการร้องขอการนำเข้าซอร์สโค้ดที่นำมาวิเคราะห์และมีผลลัพธ์เป็นผ่านหรือยอมรับ (accept) มีรายการที่ตอบตรงตามผลลัพธ์ 2 รายการจาก 28 รายการ รวมเป็นความแม่นยำถูกต้อง 7 เปอร์เซ็นต์ และหากคำนวณเปอร์เซ็นต์ความแม่นยำถูกต้องจากทั้งหมดแล้วพบว่า มีรายการร้องขอการนำเข้าซอร์สโค้ดที่ทำนายหรือพยากรณ์ถูกต้อง (Correctly Classified Instances) 1,127 รายการ และมีรายการร้องขอการนำเข้าซอร์สโค้ดที่ทำนายหรือพยากรณ์ไม่ถูกต้อง (Incorrectly Classified Instances) 80 รายการ สรุปแล้วแบบจำลองมีเปอร์เซ็นต์ความแม่นยำถูกต้อง 93.372%

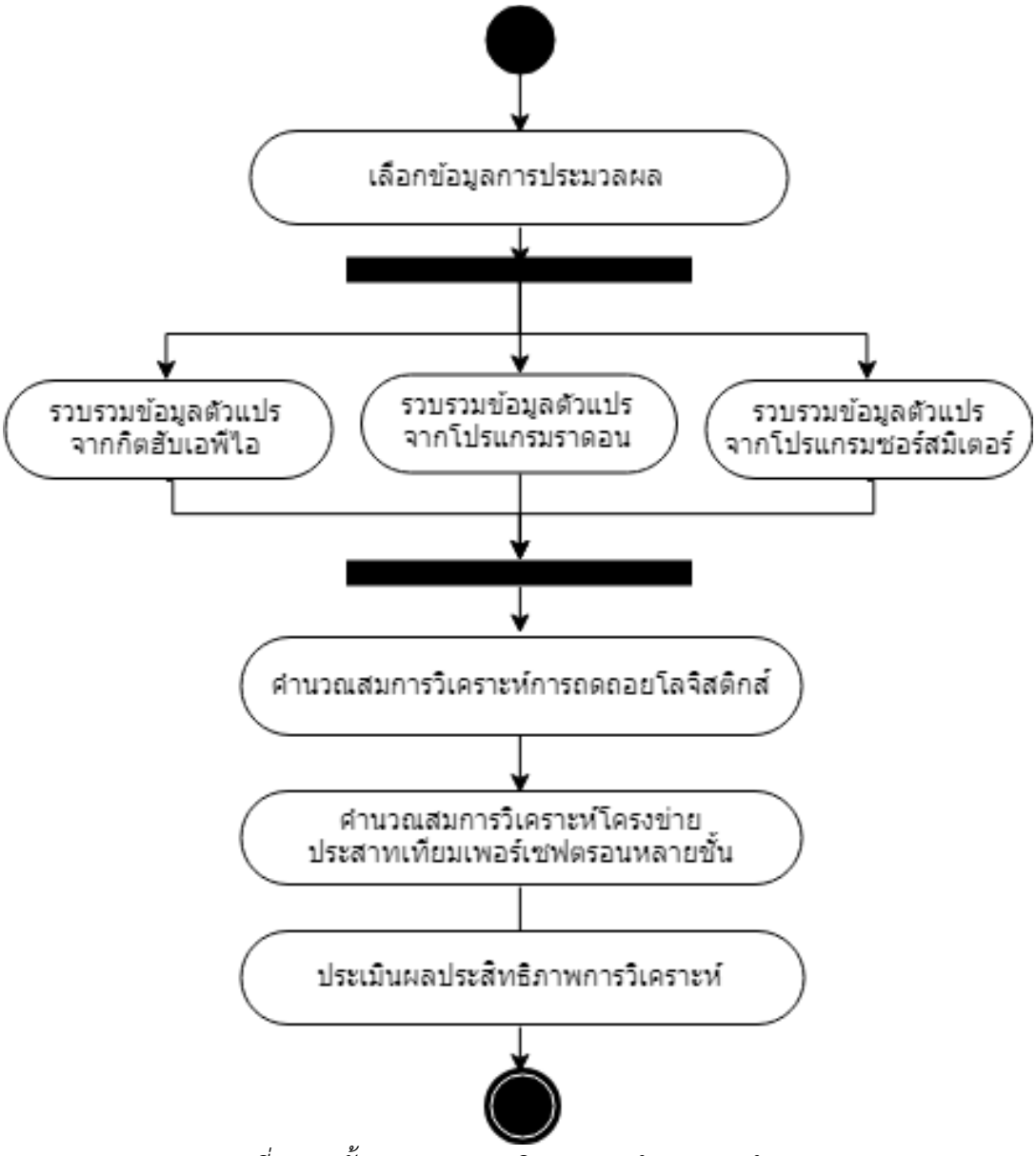
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.3 การประเมินแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

การทดสอบแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับประกอบด้วย 2 หัวข้อด้วยกัน คือ วิธีการทดสอบแบบจำลอง และผลลัพธ์การทดสอบแบบจำลอง โดยแต่ละหัวข้อ มีรายละเอียดดังนี้

3.3.1 วิธีการประเมินแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

เนื่องด้วยการสร้างแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ มี 2 แบบจำลองด้วยกัน ได้แก่การสร้างแบบจำลองด้วยการวิเคราะห์การถดถอยโลจิสติก และการสร้างแบบจำลองด้วยการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น จึงต้องทำการประเมินผลแบบจำลองทั้ง สองแบบจำลองด้วยกัน มีขั้นตอนดังรูป 3.21



รูปที่ 3.21 ขั้นตอนการประเมินผลแบบจำลองการทำนาย

จากรูป 3.21 ขั้นตอนการประเมินผลแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ ประกอบด้วย 5 ขั้นตอนด้วยกัน มีรายละเอียดดังต่อไปนี้

ขั้นตอนที่ 1 เลือกข้อมูลการประเมินผล โดยข้อมูลการทดสอบที่เลือก เลือกมาจากโครงการ Ansible ที่มีรายการร้องขอการนำเข้าซอร์สโค้ดในปี 2018 และโครงการ Tensorflow ที่มีรายการร้องขอการนำเข้าซอร์สโค้ดในปี 2018

ขั้นตอนที่ 2 รวบรวมข้อมูลของชุดข้อมูลที่ได้คัดเลือกจากกิตฮับ เพื่อใช้ในการประมวลผล โดยใช้วิธีการดึงข้อมูลเช่นเดียวกับขั้นตอนการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ จากนั้นทำการตรวจสอบค่าความผิดปกติของชุดข้อมูลทดสอบ และเตรียมข้อมูลในการวิเคราะห์ในขั้นตอนถัดไป

ขั้นตอนที่ 3 คำนวณสมการวิเคราะห์การถดถอยโลจิสติกของข้อมูล โดยใช้สมการทำนายที่ได้สร้างไว้ในการคำนวณผล

ขั้นตอนที่ 4 คำนวณสมการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้นของข้อมูล โดยใช้สมการทำนายที่ได้สร้างไว้ในการคำนวณผล

ขั้นตอนที่ 5 ประเมินผลประสิทธิภาพการวิเคราะห์ ด้วยวิธีการต่าง ๆ

3.3.2 การวิเคราะห์ผลลัพธ์การประเมินผลแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

ผลลัพธ์การประเมินแบบจำลองด้วยการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ ประกอบด้วย 4 ส่วนด้วยกัน ได้แก่ผลลัพธ์การคัดเลือกข้อมูลการประมวลผล ผลลัพธ์การประเมินผลแบบจำลองด้วยการวิเคราะห์การถดถอยโลจิสติก ผลลัพธ์การประเมินแบบจำลองด้วยการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น และผลลัพธ์การวิเคราะห์ประสิทธิภาพด้วยเมตริกความสับสน ซึ่งการวิเคราะห์ผลลัพธ์การประเมินผลแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับมีรายละเอียดดังต่อไปนี้

1. ผลลัพธ์การคัดเลือกข้อมูลการประเมินผล และการดึงข้อมูลที่ใช้ในการประเมินผล

ข้อมูลที่คัดเลือกมาใช้ในการะประเมินผลแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ มีจำนวน 65 รายการ ได้แก่ข้อมูลจากโครงการ Ansible ที่ประกอบด้วยรายการร้องขอการนำเข้าซอร์สโค้ดปีค.ศ. 2018 จำนวน 40 รายการ และข้อมูลจากโครงการ Tensorflow ที่ประกอบด้วยรายการร้องขอการนำเข้าซอร์สโค้ดปีค.ศ. 2018 จำนวน 25 รายการ โดยรายการร้องขอการนำเข้าซอร์สโค้ดทั้งหมดได้ผ่านการตรวจสอบค่าความผิดปกติ (Outlier) มาแล้ว โดยรายละเอียดผลลัพธ์การคัดเลือกข้อมูลการประเมินผล รายละเอียดดังตารางที่ 3.20

ตารางที่ 3.20 รายละเอียดผลลัพธ์การคัดเลือกข้อมูลการประเมินผล

โครงการ	จำนวนทั้งหมด	จำนวนรายการที่ผลลัพธ์เป็นปฏิเสธ (Rejected)	จำนวนรายการที่ผลลัพธ์เป็นยอมรับ (Accepted)
Ansible ปี 2018	40 รายการ	33	7
Tensorflow ปี 2018	25 รายการ	23	2
รวม	65 รายการ	56	9

การดึงข้อมูลของโครงการ Ansible ใช้วิธีการเดียวกับขั้นตอนการดึงข้อมูลที่ระบุในหัวข้อ 3.2.3 รวบรวมข้อมูลตัวแปรที่มีผลกับการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ หากแต่การดึงข้อมูลของโครงการ Tensorflow นั้น ต้องมีการใช้ยูอาร์แอลที่ต่างกัน กล่าวคือใช้วิธีการในการดึงข้อมูลเหมือนกัน หากแต่ต่างกันที่ยูอาร์แอลที่ใช้ระบุที่อยู่ของโครงการและซอร์สโค้ด หากเป็นโครงการ Tensorflow นั้น ต้องใช้ยูอาร์แอล https://github.com/tensorflow/tensorflow.git

2. ผลลัพธ์การคำนวณข้อมูลที่วิเคราะห์ด้วยสมการทำนายการวิเคราะห์การถดถอยโลจิสติก

จากการการคำนวณข้อมูลที่วิเคราะห์ด้วยสมการทำนายการวิเคราะห์การถดถอยโลจิสติก พบว่ามีรายการที่พยากรณ์ผิดพลาด หรือทำนายไม่ถูกต้องอยู่ 7 รายการ จาก 65 รายการ คิดเป็นเปอร์เซ็นต์ความแม่นยำถูกต้องแล้วแบบจำลองการทำนายผลการรีวิวโค้ดจากการวิเคราะห์การถดถอยโลจิสติกมีเปอร์เซ็นต์ความแม่นยำถูกต้องเท่ากับ 89.2308 เปอร์เซ็นต์ โดยรายละเอียดของการวิเคราะห์ด้วยการวิเคราะห์การถดถอยโลจิสติก สามารถแสดงได้ในอีก 2 มุมมอง ได้แก่ มุมมองการพิจารณาผลลัพธ์การวิเคราะห์ด้วยโครงการโอเพนซอร์ส และมุมมองการพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิว ซึ่งแต่ละมุมมอง มีรายละเอียดดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

มุมมองที่ 1 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยโครงการโอเพนซอร์ส

หากพิจารณาผลลัพธ์การวิเคราะห์ด้วยโครงการโอเพนซอร์สแล้ว พบว่ารายการร้องขอการนำเข้าซอร์สโค้ดของโครงการโอเพนซอร์ส Ansible มีรายการที่พยากรณ์ผิดพลาด หรือทำนายไม่ถูกอยู่ 5 รายการจาก 40 รายการ และรายการร้องขอการนำเข้าซอร์สโค้ดของโครงการ Tensorflow มีรายการที่พยากรณ์ผิดพลาด หรือทำนายไม่ถูกอยู่ 2 รายการจาก 25 รายการ รายละเอียดดังตารางที่ 3.21

ตารางที่ 3.21 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยโครงการโอเพนซอร์ส ด้วยวิธีการวิเคราะห์การถดถอยโลจิสติก

โครงการ	พิจารณาสองโครงการ	โครงการ Ansible	โครงการ Tensorflow
จำนวนรายการร้องขอการนำเข้าซอร์สโค้ด	65 รายการ	40	25
จำนวนทำนายถูก	58 รายการ	35	23
จำนวนทำนายผิดพลาด	7 รายการ	5	2

มุมมองที่ 2 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิวโค้ด

ผลการรีวิวของโครงการ Ansible และโครงการ Tensorflow ประกอบไปด้วย 2 ผลลัพธ์การรีวิวด้วยกันได้แก่ ผ่านหรือยอมรับ (Accept) ไม่ผ่านหรือปฏิเสธ (Reject) โดยจากการพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิวโค้ดแล้ว พบว่า รายการร้องขอการนำเข้าซอร์สโค้ดที่มีผลการรีวิวเป็นผ่านหรือยอมรับ (Accept) มีรายการที่พยากรณ์ผิดพลาด หรือทำนายไม่ถูกอยู่ 7 รายการจาก 9 รายการ ในส่วนของรายการร้องขอการนำเข้าซอร์สโค้ดที่มีผลการรีวิวเป็นไม่ผ่านหรือปฏิเสธ ไม่มีรายการพยากรณ์ผิดพลาด หรือกล่าวอีกนัยนึงว่าตอบถูกทุกกรณีทดสอบ รายละเอียดของมุมมองการพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิวโค้ด รายละเอียดดังตารางที่ 3.22

ตารางที่ 3.22 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิวโค้ด ด้วยวิธีการวิเคราะห์การถดถอยโลจิสติก

โครงการ	พิจารณาสองผลลัพธ์	ผลลัพธ์เป็นปฏิเสธ (Rejected)	ผลลัพธ์เป็นยอมรับ (Accepted)
พิจารณาสองโครงการ	65 รายการ	56	9
จำนวนทำนายถูก	58 รายการ	56	2
จำนวนทำนายผิดพลาด	7 รายการ	0	7

3. ผลลัพธ์การคำนวณข้อมูลที่วิเคราะห์ด้วยสมการทำนายการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนกลายชั้น

จากการการคำนวณข้อมูลที่วิเคราะห์ด้วยสมการสมการทำนายการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนกลายชั้น พบว่ามีรายการที่พยากรณ์ผิดพลาด หรือทำนายไม่ถูกต้องอยู่ 6 รายการ จาก 65 รายการ คิดเป็นเปอร์เซ็นต์ความแม่นยำถูกต้องแล้วแบบจำลองการทำนายผลการ

รีวิวโค้ดจากการวิเคราะห์ด้วยสมการทำนายการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนกลายชั้นความแม่นยำถูกต้องเท่ากับ 90.7692 เปอร์เซ็นต์ โดยรายละเอียดของการวิเคราะห์ด้วยการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนกลายชั้นสามารถแสดงได้ในอีก 2 มุมมอง ได้แก่ มุมมองการพิจารณาผลลัพธ์การวิเคราะห์ด้วยโครงการโอเพนซอร์ส และมุมมองการพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิว ซึ่งแต่ละมุมมอง มีรายละเอียดดังนี้

มุมมองที่ 1 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยโครงการโอเพนซอร์ส

หากพิจารณาผลลัพธ์การวิเคราะห์ด้วยโครงการโอเพนซอร์สแล้ว พบว่ารายการร้องขอการนำเข้าซอร์สโค้ดของโครงการโอเพนซอร์ส Ansible มีรายการที่พยากรณ์ผิดพลาด หรือทำนายไม่ถูกอยู่ 4 รายการ จาก 40 รายการ และรายการร้องขอการนำเข้าซอร์สโค้ดของโครงการโอเพนซอร์ส Tensorflow มีรายการที่พยากรณ์ผิดพลาด หรือทำนายไม่ถูกอยู่ 2 รายการจาก 25 รายการ รายละเอียดดังตารางที่ 3.23

ตารางที่ 3.23 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยโครงการโอเพนซอร์ส ด้วยวิธีการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอน

โครงการ	พิจารณาสองโครงการ	โครงการ Ansible	โครงการ Tensorflow
จำนวนรายการร้องขอการนำเข้าซอร์สโค้ด	65 รายการ	40	25
จำนวนทำนายถูก	59 รายการ	36	23
จำนวนทำนายผิดพลาด	6 รายการ	4	2

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

มุมมองที่ 2 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิวโค้ด

ผลการรีวิวของโครงการ Ansible และโครงการ Tensorflow ประกอบไปด้วย 2 ผลลัพธ์การรีวิวด้วยกันได้แก่ ยอมรับ (Accepted) ปฏิเสธ (Rejected) โดยจากการพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิวโค้ดแล้ว พบว่า รายการร้องขอการนำเข้าซอร์สโค้ดที่มีผลการรีวิวเป็นผ่านหรือยอมรับ (Accepted) มีรายการที่พยากรณ์ผิดพลาด หรือทำนายไม่ถูกอยู่ 6 รายการจาก 9 รายการ ในส่วนของรายการร้องขอการนำเข้าซอร์สโค้ดที่มีผลการรีวิวเป็นปฏิเสธ ไม่มีรายการพยากรณ์ผิดพลาด หรือกล่าวอีกนัยนึงว่าตอบถูกทุกกรณีทดสอบ รายละเอียดของมุมมองการพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิวโค้ด รายละเอียดดังตารางที่ 3.24

ตารางที่ 3.24 การพิจารณาผลลัพธ์การวิเคราะห์ด้วยผลการรีวิวโค้ด ด้วยวิธีการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอน

โครงการ	พิจารณาสองผลลัพธ์	ผลลัพธ์เป็นไม่ผ่านหรือปฏิเสธ (Reject)	ผลลัพธ์เป็นผ่านหรือยอมรับ (Accept)
พิจารณาสองโครงการ	65 รายการ	56	9
จำนวนพยากรณ์ถูก (หรือทำนายถูก)	58 รายการ	56	3
จำนวนพยากรณ์ผิดพลาด (หรือทำนายไม่ถูก)	7 รายการ	0	6

4. ผลลัพธ์การวิเคราะห์ประสิทธิภาพด้วยเมตริกส์ความสับสน

หลังจากการคำนวณผลลัพธ์การวิเคราะห์ข้อมูลเรียบร้อยแล้ว นำข้อมูลผลลัพธ์ที่ได้จากการคำนวณมาสร้างเป็นเมตริกความสับสน รายละเอียดดังตารางที่ 3.25 และ 3.26

ตารางที่ 3.25 เมตริกความสับสนของการประเมินผลแบบจำลองการทำนายการวิเคราะห์การถดถอยโลจิสติก

ผลลัพธ์เป็นปฏิเสธ (Rejected)	ผลลัพธ์เป็นยอมรับ (Accepted)	<-- Classified as
56	7	
0	2	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 3.25 แสดงค่าความแม่นยำของแบบจำลอง โดยจากข้อมูลที่แสดงพบว่า รายการร้องขอการนำเข้าซอร์สโค้ดที่นำมาวิเคราะห์และมีผลลัพธ์เป็นปฏิเสธ (rejected) มีรายการที่ตรงตามผลลัพธ์อยู่ 56 รายการ จากรายการที่มีผลลัพธ์เป็นไม่ผ่านหรือปฏิเสธทั้งหมด 56 รายการ คิดเป็นความแม่นยำถูกต้อง 100 เปอร์เซ็นต์ และรายการร้องขอการนำเข้าซอร์สโค้ดที่นำมาวิเคราะห์และมีผลลัพธ์เป็นยอมรับ (accepted) มีรายการที่ตอบตรงตามผลลัพธ์ 2 รายการจาก 9 รายการ รวมเป็นความแม่นยำความถูกต้อง 22.2222 เปอร์เซ็นต์ และหากคำนวณเปอร์เซ็นต์ความถูกต้องจากทั้งหมดแล้วพบว่า มีรายการร้องขอการนำเข้าซอร์สโค้ดที่ทำนายหรือพยากรณ์ถูกต้อง (Correctly Classified Instances) 56 รายการ และมีรายการร้องขอการนำเข้าซอร์สโค้ดที่ทำนายหรือพยากรณ์

ไม่ถูกต้อง (Incorrectly Classified Instances) 7 รายการ สรุปแล้วแบบจำลองมีเปอร์เซ็นต์ความแม่นยำถูกต้อง 89.2308

ตารางที่ 3.26 เมตริกความสับสนของการประเมินผลแบบจำลองการทำนายการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอน

ผลลัพธ์เป็นไม่ผ่านหรือปฏิเสธ (Reject)	ผลลัพธ์เป็นผ่านหรือยอมรับ (Accept)	<-- Classified as
56	6	
0	3	

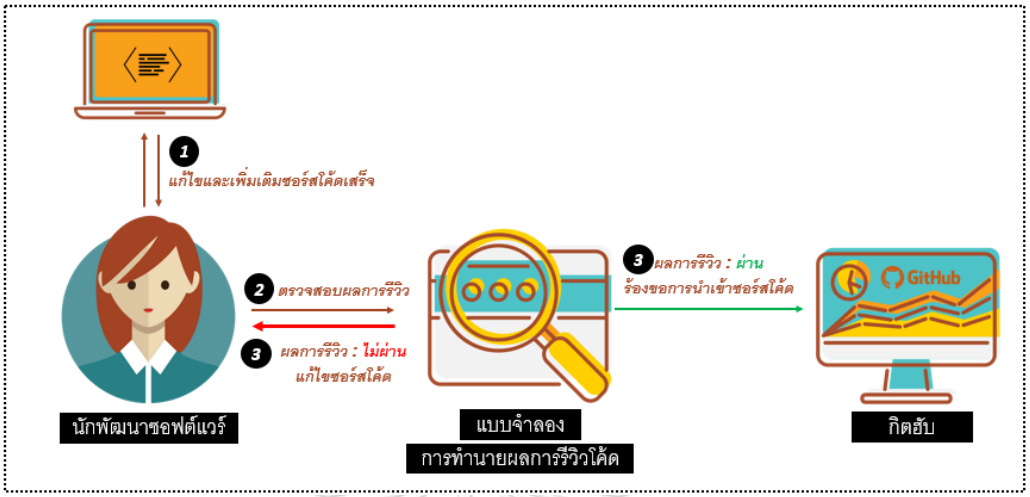
จากตารางที่ 3.26 รายละเอียดค่าความแม่นยำของแบบจำลอง โดยจากข้อมูลที่แสดงพบว่ารายการร้องขอการนำเข้าซอร์สโค้ดที่นำมาวิเคราะห์และมีผลลัพธ์เป็นปฏิเสธ (rejected) มีรายการที่ตรงตามผลลัพธ์อยู่ 56 รายการ จากรายการที่มีผลลัพธ์เป็นไม่ผ่านหรือปฏิเสธทั้งหมด 56 รายการ คิดเป็นความแม่นยำถูกต้อง 100 เปอร์เซ็นต์ และรายการร้องขอการนำเข้าซอร์สโค้ดที่นำมาวิเคราะห์และมีผลลัพธ์เป็นยอมรับ (accepted) มีรายการที่ตอบตรงตามผลลัพธ์ 3 รายการจาก 9 รายการ รวมเป็นความแม่นยำความถูกต้อง 33.3333 เปอร์เซ็นต์ และหากคำนวณเปอร์เซ็นต์ความถูกต้องจากทั้งหมดแล้วพบว่า มีรายการร้องขอการนำเข้าซอร์สโค้ดที่ทำนายหรือพยากรณ์ถูกต้อง (Correctly Classified Instances) 59 รายการ และมีรายการร้องขอการนำเข้าซอร์สโค้ดที่ทำนายหรือพยากรณ์ไม่ถูกต้อง (Incorrectly Classified Instances) 6 รายการ สรุปแล้วแบบจำลองมีเปอร์เซ็นต์ความแม่นยำถูกต้อง 90.7692

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.4 การประยุกต์ใช้แบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

การนำแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับไปใช้ สามารถนำไปใช้ได้ทั้งในโครงการโอเพนซอร์ส และประยุกต์ใช้ในโครงการของบริษัททั่วไป โดยวัตถุประสงค์ในการใช้งานแบบจำลอง เพื่อใช้ในการทำนายผลการรีวิวโค้ดก่อนทำการรีวิวโค้ด เพื่อป้องกันการได้ผลลัพธ์การรีวิวโค้ดเป็นปฏิเสธ และส่งผลกระทบให้เกิดการแก้ไข และการทำงานใหม่อีกครั้ง ซึ่งเป็นการเสียเวลาการทำงานของนักพัฒนาซอฟต์แวร์ อีกทั้งยังสามารถใช้แบบจำลองการทำนายการรีวิวโค้ดในการทดสอบหรือประเมินผลการรีวิวโค้ดโดยการใช้เครื่องมือแทนการใช้คนในการรีวิว ซึ่งอาจจะใช้เครื่องมืออัตโนมัติในการนำการร้องขอการนำเข้าซอร์สโค้ดที่ถูกร้องขอมาส่งข้อมูลที่ใช้ในการวิเคราะห์ต่าง ๆ เพื่อประมวลผลแบบจำลองการทำนายผลการรีวิว และใช้เป็นเครื่องมืออัตโนมัติในการกำหนด

ผลการรีวิวโค้ด ภาพรวมการประยุกต์ใช้แบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ ดังรูปที่ 3.22 และ รูปที่ 3.23



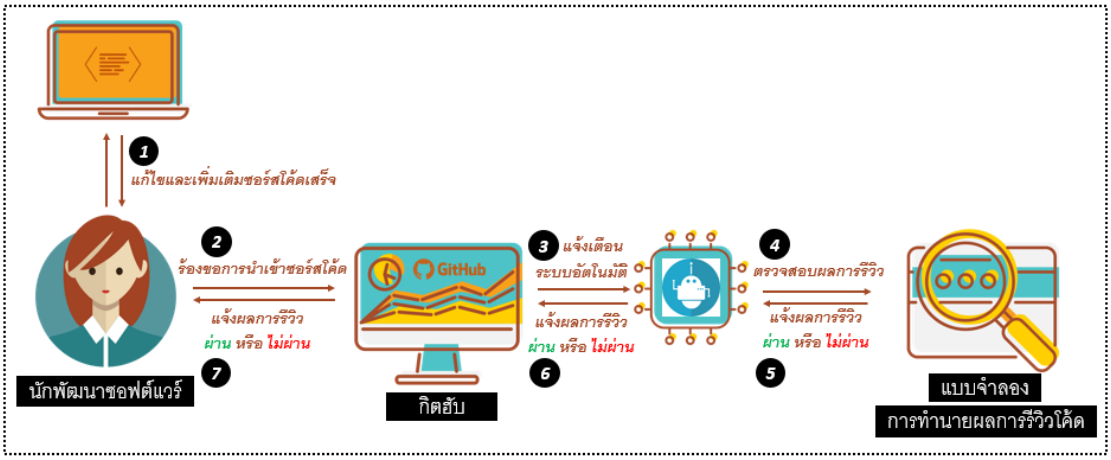
รูปที่ 3.22 ภาพรวมการประยุกต์ใช้แบบจำลองการทำนายก่อนการร้องขอการนำเข้าซอร์สโค้ด

จากรูปที่ 3.22 รายละเอียดภาพรวมการประยุกต์ใช้แบบจำลองการทำนายก่อนการร้องขอการนำเข้าซอร์สโค้ด แบ่งเป็น 3 ขั้นตอน โดยมีรายละเอียดดังนี้

ขั้นตอนที่ 1 นักพัฒนาซอฟต์แวร์ทำการแก้ไขและเพิ่มเติมซอร์สโค้ดที่ดาวน์โหลดซอร์สโค้ดต้นฉบับมาจากกิตฮับเสร็จเรียบร้อยแล้ว

ขั้นตอนที่ 2 นักพัฒนาซอฟต์แวร์ตรวจสอบผลการรีวิวผ่านแบบจำลองการทำนายผลการรีวิวโค้ดเพื่อใช้ในการพยากรณ์หรือประเมินก่อนทำการร้องขอการนำเข้าซอร์สโค้ด

ขั้นตอนที่ 3 แบบจำลองการทำนายการรีวิวแจ้งผลการรีวิวแก่นักพัฒนาซอฟต์แวร์ หากผลการรีวิวเป็นผ่าน แบบจำลองการทำนายผลหรือนักพัฒนาซอฟต์แวร์อาจทำการร้องขอการนำเข้าซอร์สโค้ด แต่หากผลการรีวิวโค้ดเป็นไม่ผ่าน นักพัฒนาซอฟต์แวร์จะต้องทำการแก้ไขซอร์สโค้ดและทำการตรวจสอบผลลัพธ์อีกครั้ง



รูปที่ 3.23 ภาพรวมการประยุกต์ใช้แบบจำลองการทำนายหลังการร้องขอการนำเข้าซอร์สโค้ด

จากรูปที่ 3.23 รายละเอียดภาพรวมการประยุกต์ใช้แบบจำลองการทำนายหลังการร้องขอการนำเข้าซอร์สโค้ด แบ่งเป็น 7 ขั้นตอน โดยมีรายละเอียดดังนี้

ขั้นตอนที่ 1 นักพัฒนาซอฟต์แวร์ทำการแก้ไขและเพิ่มเติมซอร์สโค้ดที่ได้ดาวน์โหลดซอร์สโค้ดต้นฉบับมาจากกิตฮับเสร็จเรียบร้อยแล้ว

ขั้นตอนที่ 2 นักพัฒนาซอฟต์แวร์ทำการร้องขอการนำเข้าซอร์สโค้ดบนกิตฮับหรือระบบใด ๆ

ขั้นตอนที่ 3 กิตฮับทำการแจ้งเตือนระบบอัตโนมัติที่ทางนักพัฒนาซอฟต์แวร์หรือองค์กรพัฒนาไว้ใช้ในการดำเนินการร้องขอการนำเข้าซอร์สโค้ดโดยอัตโนมัติ

ขั้นตอนที่ 4 ตรวจสอบผลการรีวิวผ่านแบบจำลองการทำนายผลการรีวิวโค้ดเพื่อใช้พยากรณ์หรือประเมินก่อนทำการร้องขอการนำเข้าซอร์สโค้ด

ขั้นตอนที่ 5 แบบจำลองการทำนายการรีวิวแจ้งผลการรีวิวแก่ระบบอัตโนมัติ

ขั้นตอนที่ 6 ระบบอัตโนมัติแจ้งผลการรีวิวแก่กิตฮับ จากนั้นกิตฮับจะทำการปรับผลการรีวิวตามที่แบบจำลองการทำนายระบุผลการรีวิวไว้

ขั้นตอนที่ 7 กิตฮับแจ้งผลการรีวิวแก่นักพัฒนาซอฟต์แวร์ หากผลการรีวิวโค้ดเป็นผ่าน ถือเป็นการร้องขอการนำเข้าซอร์สโค้ดที่สมบูรณ์ โค้ดจะถูกนำเข้าระบบ แต่หากผลการรีวิวโค้ดเป็นไม่ผ่าน นักพัฒนาซอฟต์แวร์จะต้องทำการแก้ไขซอร์สโค้ดและทำการตรวจสอบผลลัพธ์อีกครั้ง

บทที่ 4
การออกแบบและการพัฒนาเครื่องมือ

บทนี้กล่าวถึงการออกแบบและการพัฒนาเครื่องมือสนับสนุนการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ โดยมีวัตถุประสงค์ในการนำเสนอการใช้แบบจำลองการทำนายผลการรีวิวโค้ดที่ได้ออกแบบไว้ รายละเอียดในหัวข้อนี้ ประกอบด้วย 4 หัวข้อด้วยกัน ได้แก่ หัวข้อ 1.1 ข้อกำหนดความต้องการ ระบุถึงข้อกำหนดความต้องการในการออกแบบและพัฒนาเครื่องมือ หัวข้อ 1.2 แบบจำลองการวิเคราะห์ ประกอบด้วย แบบจำลองเชิงหน้าที่ แบบจำลองที่ไม่ใช่หน้าที่ และแบบจำลองเชิงพฤติกรรม หัวข้อ 1.3 สถาปัตยกรรมของระบบ อธิบายถึงภาพรวมเชิงสถาปัตยกรรมและแบบจำลองเชิงสถาปัตยกรรม และหัวข้อ 1.4 การออกแบบโครงสร้างการเชื่อมต่อ ซึ่งประกอบด้วยแผนภาพวินโดวน์และส่วนต่อประสานผู้ใช้งาน โดยมีรายละเอียดดังนี้

4.1 ข้อกำหนดความต้องการ

ข้อกำหนดความต้องการของเครื่องมือสนับสนุน ประกอบด้วย 2 หัวข้อด้วยกัน ได้แก่ ข้อกำหนดความต้องการที่เป็นหน้าที่หลัก และข้อกำหนดความต้องการที่ไม่เป็นหน้าที่หลัก มีรายละเอียดดังต่อไปนี้

4.1.1 ข้อกำหนดความต้องการที่เป็นหน้าที่หลัก

รองรับการใช้งานของผู้ใช้งานบนระบบปฏิบัติการวินโดวน์ ประกอบด้วยฟังก์ชันการทำงานต่าง ๆ รายละเอียดดังตารางที่ 4.1

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.1 ข้อกำหนดความต้องการที่เป็นหน้าที่หลัก

รหัสอ้างอิง	ฟังก์ชันการทำงาน	รายการความต้องการ
R-F-01	ฟังก์ชันการนำเข้าข้อมูลซอร์สโค้ด	ผู้ใช้สามารถนำซอร์สโค้ดเข้าสู่ระบบได้ ผ่านเมนูการอัพโหลดซอร์สโค้ด
R-F-02	ฟังก์ชันการเลือกวิธีการทำนาย	ผู้ใช้สามารถระบุวิธีการที่ใช้ในการทำนายผลการรีวิวโค้ดได้
R-F-03	ฟังก์ชันการคำนวณผลทำนาย	ระบบสามารถคำนวณผลการทำนายด้วยวิธีการวิเคราะห์การถดถอยโลจิสติก
R-F-04	ฟังก์ชันการคำนวณผลทำนาย	ระบบสามารถคำนวณผลการทำนายด้วยวิธีการวิเคราะห์เพอร์เซปตรอนหลายชั้น

ตารางที่ 4.1 ข้อกำหนดความต้องการที่เป็นหน้าที่หลัก (ต่อ)

รหัสอ้างอิง	ฟังก์ชันการทำงาน	รายการความต้องการ
R-F-05	ฟังก์ชันการแสดงผลการทำนาย	ระบบสามารถแสดงผลลัพธ์การทำนายได้ 2 ผลลัพธ์ด้วยกัน คือ ผลการรีวิวเป็นยอมรับ (accepted) และผลการรีวิวเป็นปฏิเสธ (rejected)
R-F-06	ฟังก์ชันการออกรายงานผลการทำนาย	ผู้ใช้สามารถดาวน์โหลดรายงานผลการทำนายได้

4.1.2 ข้อกำหนดความต้องการที่ไม่เป็นหน้าที่หลัก

สามารถจำแนกข้อกำหนดความต้องการที่ไม่เป็นหน้าที่หลัก ตามรายการความต้องการเพื่อบรรลุวัตถุประสงค์ของผู้ใช้งาน มีรายละเอียดดังตารางที่ 4.2

ตารางที่ 4.2 ข้อกำหนดความต้องการที่ไม่เป็นหน้าที่หลัก

รหัสอ้างอิง	ประเภทความต้องการ	รายการความต้องการ
R-NF-01	ความต้องการด้านการพัฒนา (Implementation requirements)	ระบบต้องพัฒนาด้วยภาษาดังนี้ - Java - CSS - HTML - JavaScript
R-NF-02	ความต้องการด้านการพัฒนา (Implementation requirements)	เครื่องคอมพิวเตอร์ที่ใช้ในการพัฒนาเครื่องมือสนับสนุน ต้องมีโปรแกรมเหล่านี้ติดตั้งอยู่ - Eclipse - Git - Postman - Tomcat - Radon - SourceMeter

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

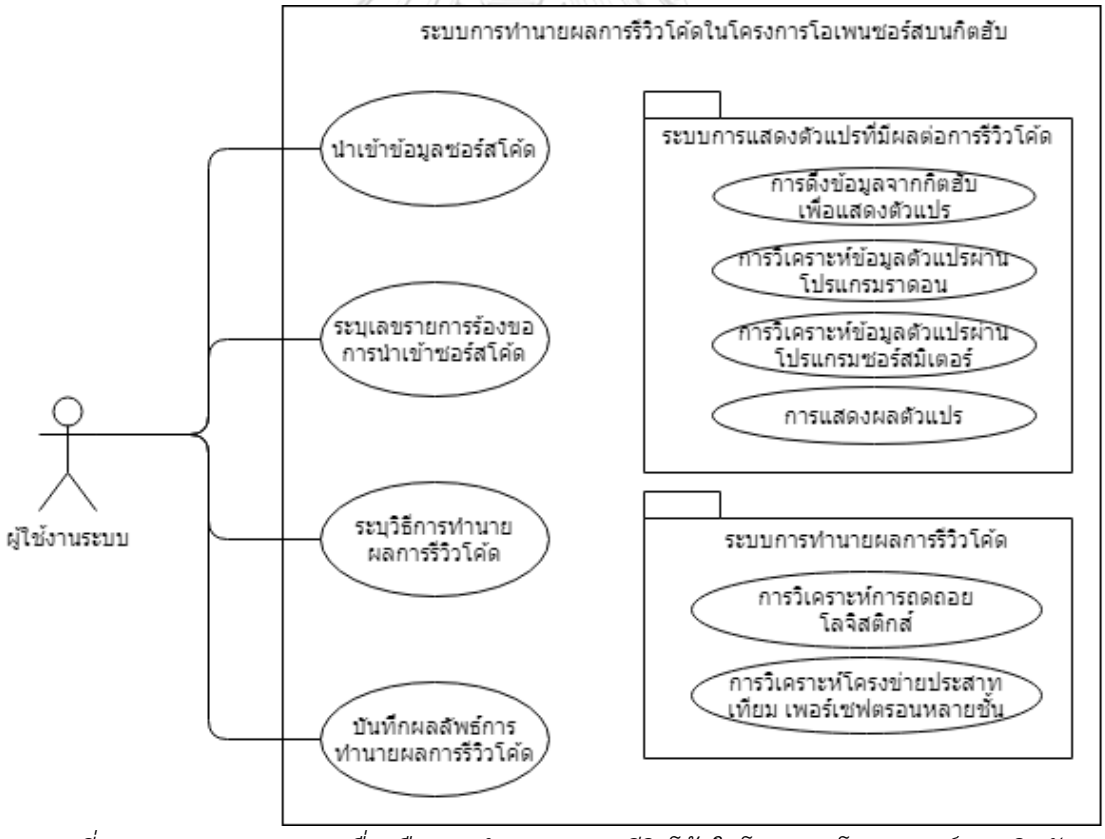
4.2 แบบจำลองการวิเคราะห์

การพัฒนาเครื่องมือที่ใช้ในการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ สามารถนำเสนอการออกแบบเครื่องมือได้หลากหลายรูปแบบ ไม่ว่าจะเป็นแบบจำลองเชิงหน้าที่ (Functional Models) ที่ประกอบไปด้วยแผนภาพยูสเคส (Use Case Diagram) คำอธิบายยูสเคส (Use Case Description) และแผนภาพกิจกรรม (Activity Diagram) แบบจำลองเชิงโครงสร้าง (Structural Models) ซึ่งประกอบไปด้วยแผนภาพคลาส อีกทั้งยังสามารถอธิบายได้ในรูปแบบของแบบจำลองเชิงพฤติกรรม (Behavioral Models) ที่ประกอบไปด้วยแผนภาพลำดับการทำงาน (Sequence Diagram) โดยรายละเอียดของแบบจำลองมีดังต่อไปนี้

4.2.1 แบบจำลองเชิงหน้าที่

1. แผนภาพยูสเคส

ใช้เพื่ออธิบายฟังก์ชันการทำงานพื้นฐานของเครื่องมือการทำนาย โดยแผนภาพยูสเคสของระบบการทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ รายละเอียดดังรูปที่ 4.1



รูปที่ 4.1 แผนภาพยูสเคสเครื่องมือการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

2. คำอธิบายยูสเคส

คำอธิบายยูสเคสของเครื่องมือการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ ประกอบด้วย 4 ตารางด้วยกัน ได้แก่ ตารางที่ 4.3 คำอธิบายยูสเคสนำเข้าข้อมูลซอร์สโค้ด ตารางที่ 4.4 คำอธิบายยูสเคสระบุเลขรายการร้องขอการนำเข้าซอร์สโค้ด ตารางที่ 4.5 คำอธิบายระบุวิธีการทำนายผลการรีวิวโค้ด และตารางที่ 4.6 คำอธิบายยูสเคสบันทึกผลลัพธ์การทำนายผลการรีวิวโค้ด โดยมีรายละเอียดดังนี้

ตารางที่ 4.3 คำอธิบายยูสเคสนำเข้าข้อมูลซอร์สโค้ด

ชื่อยูสเคส: นำเข้าข้อมูลซอร์สโค้ด	รหัส: R-UC-01	ระดับความสำคัญ: มาก
ผู้กระทำหลัก: ผู้ใช้งานระบบ	ประเภทยูสเคส: พื้นฐาน เชิงละเอียด	
ผู้ที่มีส่วนเกี่ยวข้องและการใช้ประโยชน์ : ผู้ใช้งานระบบทำการนำเข้าข้อมูลซอร์สโค้ด		
คำอธิบาย: เป็นยูสเคสที่อธิบายขั้นตอนการนำเข้าข้อมูลซอร์สโค้ด		
สิ่งกระตุ้น: ผู้ใช้งานระบบทำการนำเข้าซอร์สโค้ด ประเภทของสิ่งกระตุ้น: ภายนอก		
ความสัมพันธ์: ความเกี่ยวเนื่อง: ผู้ใช้งานระบบ การรวม: - การขยาย: - การรับทอดคุณสมบัติ: -		
ขั้นตอนการทำงานปกติ: 1. ผู้ใช้งานระบบเข้าใช้งานเครื่องมือการทำนายผลการรีวิวโค้ด 2. ผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูลก่อนทำการร้องขอการนำเข้าซอร์สโค้ด 3. ผู้ใช้งานระบบนำเข้าซอร์สโค้ด 4. ผู้ใช้งานระบบกดปุ่มซับมิทเพื่อทำการวิเคราะห์ข้อมูล 5. ระบบนำซอร์สโค้ดที่ผู้ใช้งานระบบนำเข้าไปวิเคราะห์ผลในขั้นตอนต่อไป		
ขั้นตอนการทำงานย่อย: -		

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.4 คำอธิบายยูสเคสระบุเลขรายการร้องขอการนำเข้าซอร์สโค้ด

ชื่อยูสเคส: ระบุเลขรายการร้องขอการนำเข้าซอร์สโค้ด	รหัส: R-UC-02	ระดับความสำคัญ: มาก
ผู้กระทำหลัก: ผู้ใช้งานระบบ	ประเภทยูสเคส: พื้นฐาน เชิงละเอียด	
ผู้ที่มีส่วนเกี่ยวข้องและการใช้ประโยชน์ : ผู้ใช้งานระบบทำการระบุเลขรายการร้องขอการนำเข้าซอร์สโค้ด		
คำอธิบาย: เป็นยูสเคสที่อธิบายขั้นตอนการระบุเลขรายการร้องขอการนำเข้าซอร์สโค้ด		
สิ่งกระตุ้น: ผู้ใช้งานระบบทำการระบุเลขรายการร้องขอการนำเข้าซอร์สโค้ด ประเภทของสิ่งกระตุ้น: ภายนอก		
ความสัมพันธ์: ความเกี่ยวเนื่อง: ผู้ใช้งานระบบ การรวม: - การขยาย: - การรับทอดคุณสมบัติ: -		
ขั้นตอนการทำงานปกติ: 1. ผู้ใช้งานระบบเข้าใช้งานเครื่องมือการทำนายผลการรีวิวโค้ด 2. ผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูลหลังทำการร้องขอการนำเข้าซอร์สโค้ด 3. ผู้ใช้งานระบบระบุเลขรายการร้องขอการนำเข้าซอร์สโค้ด 4. ผู้ใช้งานระบบกดปุ่มซับมิทเพื่อทำการวิเคราะห์ข้อมูล 5. ระบบทำการวิเคราะห์ข้อมูลผ่านเลขรายการร้องขอการนำเข้าซอร์สโค้ดที่ผู้ใช้งานระบบระบุ		
ขั้นตอนการทำงานย่อย: -		

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.5 คำอธิบายระบุวิธีการทำนายผลการรีวิวโค้ด

ชื่อยูสเคส: ระบุวิธีการทำนายผลการรีวิวโค้ด	รหัส: R-UC-03	ระดับความสำคัญ: มาก
ผู้กระทำหลัก: ผู้ใช้งานระบบ	ประเภทยูสเคส: พื้นฐาน เชิงละเอียด	
ผู้ที่มีส่วนเกี่ยวข้องและการใช้ประโยชน์ : ผู้ใช้งานระบบทำการระบุวิธีการทำนายผลการรีวิวโค้ด		
คำอธิบาย: เป็นยูสเคสที่อธิบายขั้นตอนการระบุวิธีการทำนายผลการรีวิวโค้ด		
สิ่งกระตุ้น: ผู้ใช้งานระบบทำการระบุวิธีการทำนายผลการรีวิวโค้ด ประเภทของสิ่งกระตุ้น: ภายนอก		

ตารางที่ 4.5 คำอธิบายระบุวิธีการทำนายผลการรีวิวโค้ด (ต่อ)

ชื่อยูสเคส: ระบุวิธีการทำนายผลการรีวิวโค้ด	รหัส: R-UC-03	ระดับความสำคัญ: มาก
ความสัมพันธ์: ความเกี่ยวเนื่อง: ผู้ใช้งานระบบ การรวม: - การขยาย: - การรับทอดคุณสมบัติ: -		
ขั้นตอนการทำงานปกติ: 1. ผู้ใช้งานระบบเข้าใช้งานเครื่องมือการทำนายผลการรีวิวโค้ด 2. ผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูล 3. ผู้ใช้งานระบบระบุรายละเอียดการวิเคราะห์ข้อมูล 4. ผู้ใช้งานระบบกดปุ่มซับมิทเพื่อทำการวิเคราะห์ข้อมูล 5. ระบบแสดงตัวแปรที่มีผลกับการรีวิวโค้ด 6. ผู้ใช้งานระบบระบุวิธีการทำนายผลการรีวิวโค้ด 7. ระบบแสดงผลลัพธ์การรีวิวโค้ด		
ขั้นตอนการทำงานย่อย: 2. ผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูล 2.1 ผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ด 2.2 ผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูลหลังการร้องขอการนำเข้าซอร์สโค้ด 3. ผู้ใช้งานระบบระบุรายละเอียดการวิเคราะห์ข้อมูล 3.1 ถ้าผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ด ผู้ใช้งานระบบต้องระบุโครงการ (Repository) ระบุเลขรายการพัฒนาซอฟต์แวร์ และนำการนำเข้าซอร์สโค้ดเพื่อใช้ในการวิเคราะห์ข้อมูล 3.2 ถ้าผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูลหลังการร้องขอการนำเข้าซอร์สโค้ด ผู้ใช้งานระบบต้องระบุโครงการและเลขรายการร้องขอการนำเข้าซอร์สโค้ด 6. ผู้ใช้งานระบบระบุวิธีการทำนายผลการรีวิวโค้ด 6.1 วิธีการทำนายทั้งการวิเคราะห์การถดถอยโลจิสติก และการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น 6.2 วิธีการทำนายการวิเคราะห์การถดถอยโลจิสติก 6.3 วิธีการทำนายการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น		

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.6 คำอธิบายยูสเคสบันทึกผลลัพธ์การทำนายผลการรีวิวโค้ด

ชื่อยูสเคส: บันทึกผลลัพธ์การทำนายผลการรีวิวโค้ด	รหัส: R-UC-04	ระดับความสำคัญ: มาก
ผู้กระทำหลัก: ผู้ใช้งานระบบ	ประเภทยูสเคส: พื้นฐาน เชิงละเอียด	
ผู้ที่มีส่วนเกี่ยวข้องและการใช้ประโยชน์ : ผู้ใช้งานระบบทำการบันทึกผลลัพธ์การทำนายผลการรีวิวโค้ด		
คำอธิบาย: เป็นยูสเคสที่อธิบายขั้นตอนการบันทึกผลลัพธ์การทำนายผลการรีวิวโค้ด		
สิ่งกระตุ้น: ผู้ใช้งานระบบทำการบันทึกผลลัพธ์การทำนายผลการรีวิวโค้ด ประเภทของสิ่งกระตุ้น: ภายนอก		
ความสัมพันธ์: ความเกี่ยวเนื่อง: ผู้ใช้งานระบบ การรวม: - การขยาย: - การรับทอดคุณสมบัติ: -		
ขั้นตอนการทำงานปกติ: 1. ผู้ใช้งานระบบเข้าใช้งานเครื่องมือการทำนายผลการรีวิวโค้ด 2. ผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูล 3. ผู้ใช้งานระบบระบุรายละเอียดการวิเคราะห์ข้อมูล 4. ผู้ใช้งานระบบกดปุ่มซับมิทเพื่อทำการวิเคราะห์ข้อมูล 5. ระบบแสดงตัวแปรที่มีผลกับการรีวิวโค้ด 6. ผู้ใช้งานระบบระบุวิธีการทำนายผลการรีวิวโค้ด 7. ระบบแสดงผลลัพธ์การรีวิวโค้ด 8. ผู้ใช้งานระบบทำการบันทึกผลลัพธ์การทำนายผลการรีวิวโค้ด		
ขั้นตอนการทำงานย่อย: 1. ผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูล 1.1 ผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ด 1.2 ผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูลหลังการร้องขอการนำเข้าซอร์สโค้ด 2. ผู้ใช้งานระบบระบุรายละเอียดการวิเคราะห์ข้อมูล		

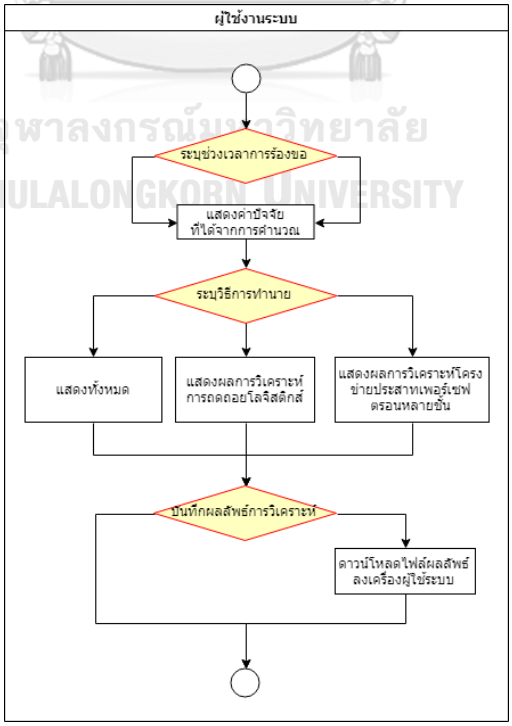
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.6 คำอธิบายยูสเคสบันทึกผลลัพธ์การทำนายผลการรีวิวโค้ด (ต่อ)

ชื่อยูสเคส: บันทึกผลลัพธ์การทำนายผลการรีวิวโค้ด	รหัส: R-UC-04	ระดับความสำคัญ: มาก
2.1 ถ้าผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ด ผู้ใช้งานระบบต้องระบุ โครงการ (Repository) ระบุเลขรายการพัฒนาซอฟต์แวร์ และนำการนำเข้าซอร์สโค้ดเพื่อใช้ในการวิเคราะห์ข้อมูล		
ขั้นตอนการทำงานย่อย (ต่อ): 2. ผู้ใช้งานระบบระบุรายละเอียดการวิเคราะห์ข้อมูล 2.2 ถ้าผู้ใช้งานระบบเลือกเมนูการวิเคราะห์ข้อมูลหลังการร้องขอการนำเข้าซอร์สโค้ด ผู้ใช้งานระบบต้องระบุ โครงการ (Repository) และเลขรายการร้องขอการนำเข้าซอร์สโค้ด 3. ผู้ใช้งานระบบระบุวิธีการทำนายผลการรีวิวโค้ด 3.1 วิธีการทำนายทั้งการวิเคราะห์การถดถอยโลจิสติก และการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น		

3. แผนภาพกิจกรรม

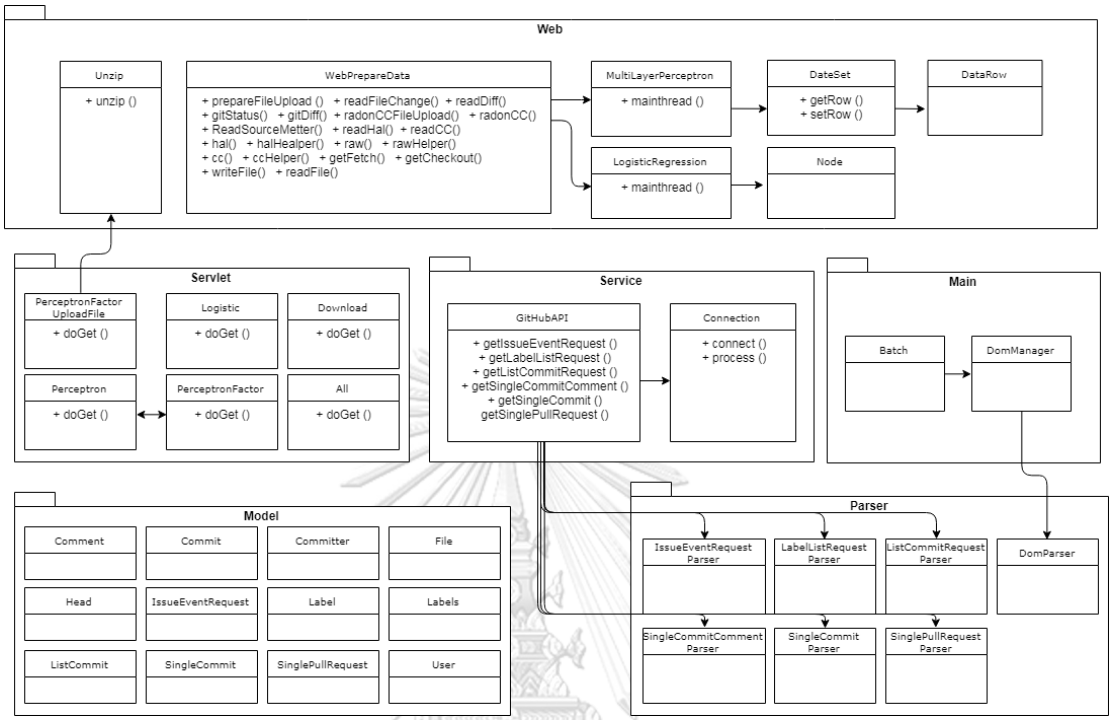
แผนภาพกิจกรรมแสดงถึงลำดับขั้นตอนการทำงานของเครื่องมือการวิเคราะห์ผลการรีวิวโค้ด กิจกรรมที่เกิดขึ้นในการใช้งานเครื่องมือ รวมถึงการทำงานเครื่องมือ โดยมีรายละเอียดดังรูปที่ 4.2



รูปที่ 4.2 แผนภาพกิจกรรมของการทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

4.2.2 แบบจำลองเชิงโครงสร้าง

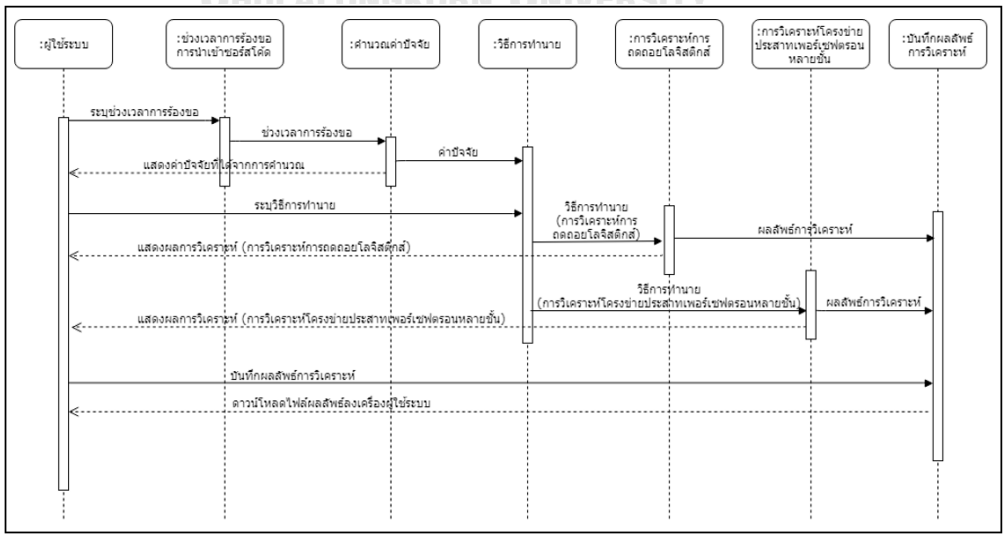
แผนภาพคลาส ใช้ในการแสดงข้อมูลและสิ่งที่เกิดขึ้นในระบบ และแสดงกลุ่มงานที่เกิดขึ้นในแต่ละคลาส สถานะของคลาส รวมถึงความสัมพันธ์ระหว่างคลาส ดังรูปที่ 4.3



รูปที่ 4.3 แผนภาพคลาสของการทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

4.2.3 แบบจำลองเชิงพฤติกรรม

แผนภาพลำดับเป็นแผนกาพที่แสดงให้เห็นถึงการติดต่อกันระหว่างอ็อบเจ็กต์หรือฟังก์ชันการทำงาน รวมถึงการส่งผ่านข้อความระหว่างฟังก์ชัน แผนภาพลำดับของการทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ รายละเอียดดังรูปที่ 42



รูปที่ 4.4 แผนภาพลำดับแบบจำลองการทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

4.3 สถาปัตยกรรมระบบ

สถาปัตยกรรมระบบของการออกแบบเครื่องมือการทำนายผลการรีวิวโค้ด แบ่งเป็น 2 หัวข้อ ได้แก่ ภาพรวมเชิงสถาปัตยกรรม และแบบจำลองเชิงสถาปัตยกรรม มีรายละเอียดดังนี้

4.3.1 ภาพรวมเชิงสถาปัตยกรรม

ภาพรวมของระบบสามารถจำแนกได้ 3 ประเภท ดังนี้

1. Client-Side ใช้โปรโตคอลเฮชทีทีพี (HTTP Protocal) สำหรับการติดต่อสื่อสารกับส่วนประกอบภายนอก เช่นติดต่อกับเซิร์ฟเวอร์ผ่านอินเตอร์เน็ต เป็นต้น

2. Server-Side เป็นส่วนของเซิร์ฟเวอร์ ซึ่งประกอบด้วย 4 ส่วนประกอบด้วยกันได้แก่ View, Service, Controller และ Model

3. GitHub เป็นส่วนที่ใช้ในการติดต่อกับกิตฮับ ประกอบด้วย 3 ส่วนประกอบด้วยกัน ได้แก่ Git Local Repository, Radon และ SourceMeter

สภาพแวดล้อมที่ใช้ในการพัฒนาเครื่องมือ ประกอบด้วย 2 ส่วนด้วยกัน ได้แก่ฮาร์ดแวร์ และซอฟต์แวร์ มีรายละเอียดดังตารางที่ 4.7

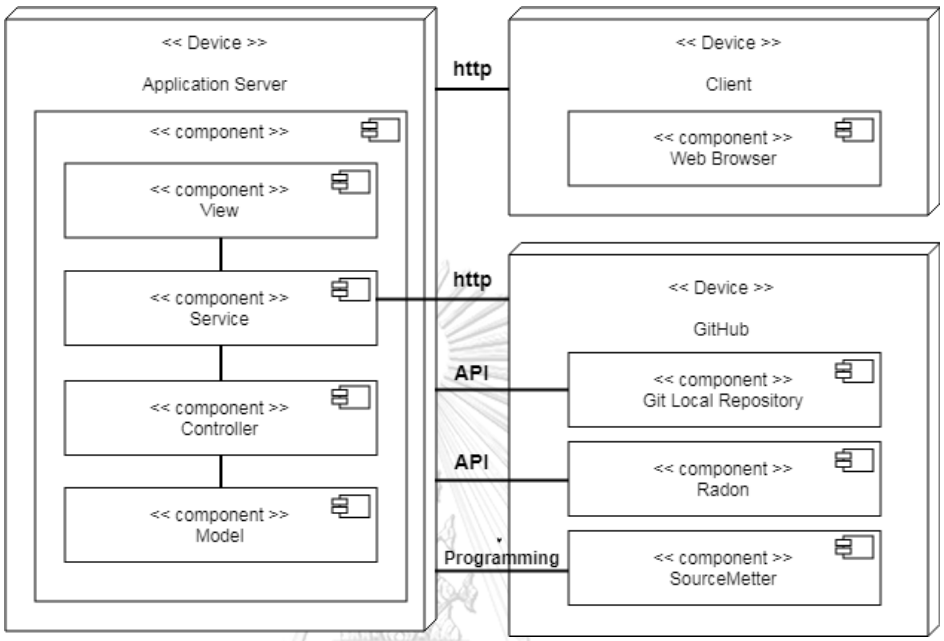
ตารางที่ 4.7 รายละเอียดสภาพแวดล้อมที่ใช้ในการพัฒนาเครื่องมือ

ส่วนที่	ส่วนประกอบ	รายละเอียด
1	ฮาร์ดแวร์ (Hardware)	เครื่องคอมพิวเตอร์แบบพกพา หน่วยการประมวลผลอินเทลคอร์ไอเซเว่น 2.50 กิกะเฮิร์ต (Intel Core i7 2.50 GHz)
		หน่วยความจำหลัก (Ram) 32.0 กิกะไบต์ (512 GB)
2	ซอฟต์แวร์ (Software)	ระบบปฏิบัติการวินโดวน์เท่านั้น เนื่องจากโปรแกรมซอร์สมิเตอร์ไม่รองรับการทำงานบนระบบปฏิบัติการอื่นๆ
		โปรแกรม Eclipse ในการพัฒนาซอร์สโค้ด
		โปรแกรมราดอน สำหรับการคำนวณและวิเคราะห์ค่าตัวแปรที่กำหนด
		โปรแกรมซอร์สมิเตอร์ สำหรับการคำนวณและวิเคราะห์ค่าตัวแปรที่กำหนด
		โปรแกรมไพธอน เนื่องจากโปรแกรมราดอนมีข้อกำหนดให้ลงโปรแกรมไพธอนก่อนการประมวลผล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.3.2 แบบจำลองเชิงสถาปัตยกรรม

เครื่องมือสนับสนุนการทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับพัฒนาในรูปแบบของเว็บเซอร์วิสที่รับค่าผ่าน HTTP เป็น JSON และส่งค่าออกเป็น JSON ไปยังผู้เรียกเซอร์วิส สามารถแสดงในรูปแบบแผนภาพดีพลอยได้ ดังรูปที่ 4.5

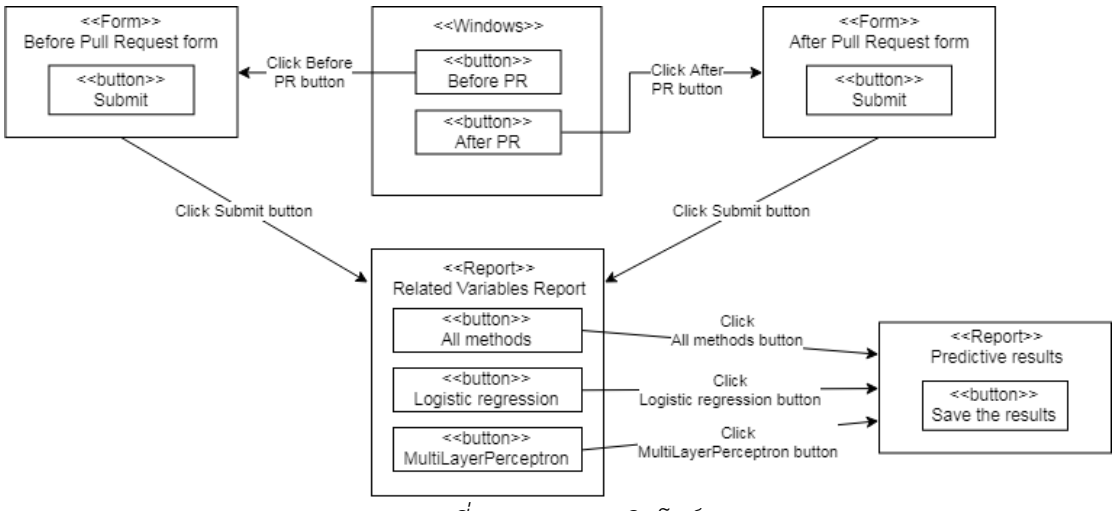


รูปที่ 4.5 แผนภาพดีพลอย

4.4 การออกแบบโครงสร้างการเชื่อมต่อ

4.4.1 แผนภาพวินโดว์

แผนภาพวินโดว์แสดงโครงสร้างการเชื่อมต่อของเครื่องมือการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ รายละเอียดดังรูปที่ 4.6



รูปที่ 4.6 แผนภาพวินโดว์

4.4.2 ส่วนต่อประสานผู้ใช้งาน

ส่วนต่อประสานกับผู้ใช้ของเครื่องมือ ประกอบด้วยทั้งหมด 5 หน้าได้แก่ หน้าแรก หน้าการวิเคราะห์ซอร์สโค้ดก่อนการร้องขอการนำเข้าซอร์สโค้ด หน้าการวิเคราะห์ซอร์สโคดหลังการร้องขอการนำเข้าซอร์สโค้ด หน้าแสดงตัวแปรที่มีผลกระทบกับการรีวิวโค้ด และหน้าการแสดงผลการรีวิวโค้ด แต่ละหน้ามีรายละเอียดดังนี้

1. หน้าแรก หน้าแรกเป็นหน้าที่ให้ผู้ใช้งานระบบเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด โดยหากผู้ใช้งานต้องการวิเคราะห์ซอร์สโค้ดก่อนการร้องขอการนำเข้าซอร์สโค้ด ระบบจะเปิดหน้าการกรอกข้อมูลสำหรับการวิเคราะห์ซอร์สโค้ดก่อนการร้องขอการนำเข้าซอร์สโค้ด หรือหากผู้ใช้งานระบบต้องการวิเคราะห์ซอร์สโค้ดหลังการร้องขอการนำเข้าซอร์สโค้ด ระบบจะเปิดหน้าการกรอกข้อมูลสำหรับการวิเคราะห์ซอร์สโค้ดหลังการร้องขอการนำเข้าซอร์สโค้ด

2. หน้าการวิเคราะห์ซอร์สโค้ดก่อนการร้องขอการนำเข้าซอร์สโค้ด ระบบจะให้ผู้ใช้งานระบุโครงการที่ต้องการตรวจสอบ เลขสาขาในการพัฒนาซอฟต์แวร์ และอัพโหลดซอร์สโค้ดเพื่อการวิเคราะห์ข้อมูล

3. หน้าการวิเคราะห์ซอร์สโค้ดหลังการร้องขอการนำเข้าซอร์สโค้ด ระบบจะให้ผู้ใช้งานระบุโครงการที่ต้องการตรวจสอบ รวมถึงเลขการร้องขอการนำเข้าซอร์สโค้ด หากแต่ไม่ต้องมีการอัพโหลดซอร์สโค้ด เนื่องจากระบบมีความสามารถในการไปดึงข้อมูลซอร์สโค้ดมาได้จากเลขรายการร้องขอการนำเข้าซอร์สโค้ดที่ผู้ใช้งานระบุ

4. หน้าแสดงตัวแปรที่มีผลกระทบกับการรีวิวโค้ด หน้านี้จะเป็นหน้าที่ทั้งแสดงค่าตัวแปรต่าง ๆ ที่มีผลกับการรีวิวโค้ด อีกทั้งยังเป็นหน้าที่ให้ผู้ใช้งานเลือกว่าจะทำการวิเคราะห์แบบใด โดยมี 3 ตัวเลือกให้ผู้ใช้งานเลือกได้แก่ การวิเคราะห์แบบการถดถอยโลจิสติก การวิเคราะห์แบบโครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น และการวิเคราะห์ทั้งสองวิธี

5. หน้าการแสดงผล ไม่ว่าจะเป็นการวิเคราะห์แบบใดแบบหนึ่งในทั้งสามแบบ ผลที่ได้คือผลการรีวิวว่าผ่าน (Accept) หรือไม่ผ่าน (Reject)

4.5 การทดสอบเครื่องมือการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

การทดสอบเครื่องมือสำหรับการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับมีวัตถุประสงค์ในการหาข้อบกพร่องของเครื่องมือที่ผู้วิจัยได้พัฒนา โดยรายละเอียดการทดสอบประกอบด้วย 2 หัวข้อ ได้แก่ สภาพแวดล้อมที่ใช้ในการทดสอบ และการทดสอบเครื่องมือ มีรายละเอียดดังต่อไปนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.5.1 สภาพแวดล้อมที่ใช้ในการทดสอบ

เนื่องด้วยสภาพแวดล้อมที่ใช้ในการทดสอบ มีผลต่อการทดสอบโดยตรง จึงต้องมีการกำหนดสภาพแวดล้อมในการทดสอบก่อนการทำสอบทุกครั้ง ซึ่งจะต้องตรงกับข้อกำหนดความต้องการที่ระบุไว้ โดยสภาพแวดล้อมที่ใช้ในการทดสอบ ประกอบด้วย 2 หัวข้อด้วยกัน ได้แก่ฮาร์ดแวร์ และซอฟต์แวร์ มีรายละเอียดดังต่อไปนี้

1. ฮาร์ดแวร์ (Hardware)

ใช้เครื่องคอมพิวเตอร์แบบพกพา หน่วยการประมวลผลอินเทลคอร์ไอเซเว่น 2.50 กิกะเฮิร์ต (Intel Core i7 2.50 GHz) ซึ่งมีหน่วยความจำหลัก (Ram) 32.0 กิกะไบต์ (512 GB)

2. ซอฟต์แวร์ (Software)

ระบบปฏิบัติการวินโดว์ ที่ลงโปรแกรมราดอน โปรแกรมซอร์สมิเตอร์ และโปรแกรมไพธอนเรียบร้อยแล้ว

4.5.2 การทดสอบเครื่องมือ

ข้อมูลที่ใช้ในการทดสอบคือข้อมูลรายการร้องขอการนำเข้าซอร์สโค้ดของโครงการโอเพนซอร์ส Ansible ทั้งในปีค.ศ. 2017 และในปีค.ศ. 2018 ที่ได้ดึงข้อมูลมาจากกิตฮับ โดยการทดสอบเครื่องมือประกอบด้วย 6 การทดสอบด้วยกัน ภาพรวมการทดสอบรายละเอียดดังตารางที่ 4.8

ตารางที่ 4.8 ภาพรวมการทดสอบเครื่องมือการทำนายผลการรีวิวโค้ด

รหัสกรณีทดสอบ	ชื่อกรณีทดสอบ	ผลการทดสอบ
TC01	การเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด	ผ่าน
TC02	การวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ด	ผ่าน
TC03	การวิเคราะห์ข้อมูลหลังการร้องขอการนำเข้าซอร์สโค้ด	ผ่าน
TC04	การแสดงตัวแปรที่มีผลกับการรีวิวโค้ด	ผ่าน
TC05	การแสดงผลการวิเคราะห์ข้อมูล	ผ่าน
TC06	การบันทึกผลการวิเคราะห์ข้อมูล	ผ่าน

1. การทดสอบที่ 1

การทดสอบการเลือกช่วงเวลาในการร้องขอการนำเข้าซอร์สโค้ด เครื่องมือสนับสนุนนี้ได้นำเสนอการเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด 2 ช่วงเวลา ได้แก่ ก่อนทำการร้องขอการนำเข้าซอร์สโค้ด และหลังจากการร้องขอการนำเข้าซอร์สโค้ด ซึ่งมีรายละเอียดดังตารางที่ 4.9

ตารางที่ 4.9 รายละเอียดการทดสอบที่ 1

รหัสกรณีทดสอบ	TC01
ชื่อกรณีทดสอบ	การเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด
ฟังก์ชันที่ทดสอบ	การเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด ในหน้าแรก (index)
จุดมุ่งหมายในการทดสอบ	เพื่อทำการทดสอบว่า เครื่องมือสามารถเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ดได้ตามที่ผู้ใช้กำหนด
เงื่อนไขการเริ่มทดสอบ	หน้าแรก (index) แสดงผลเรียบร้อยแล้ว
ข้อมูลนำเข้า	ช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด
ข้อมูลนำออก	-
ผลลัพธ์ที่คาดหวัง	1. ผู้ใช้สามารถเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ดได้ 2. หน้าจอหลังจากการเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ดต้องแสดงถูกต้อง กล่าวคือ หากผู้ใช้กดปุ่ม Before pull request ต้องแสดงผลหน้าจอการระบุข้อมูลแบบก่อนการร้องขอการนำเข้าซอร์สโค้ด และหากผู้ใช้กดปุ่ม After pull request ต้องแสดงผลหน้าจอการระบุข้อมูลหลังการร้องขอการนำเข้าซอร์สโค้ด
ลำดับการทดสอบ	1. ผู้ใช้กดปุ่ม Before pull request ในหน้าแรก 2. หน้าจอการระบุข้อมูลแบบก่อนการร้องขอการนำเข้าซอร์สโค้ดแสดงผลอย่างถูกต้อง 3. กลับไปหน้าแรก 1. ผู้ใช้กดปุ่ม After pull request ในหน้าแรก 2. หน้าจอการระบุข้อมูลแบบหลังการร้องขอการนำเข้าซอร์สโค้ดแสดงผลอย่างถูกต้อง
สรุปผลการทดสอบ	ผ่าน

2. การทดสอบที่ 2

การทดสอบการวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ด เครื่องมือสนับสนุนนี้ ได้นำเสนอการเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด 2 ช่วงเวลา ในที่นี้จะทำเป็นการทำสอบการวิเคราะห์ก่อนทำการร้องขอการนำเข้าซอร์สโค้ด ซึ่งมีรายละเอียดดังตารางที่ 4.10

ตารางที่ 4.10 รายละเอียดการทดสอบที่ 2

รหัสกรณีทดสอบ	TC02
ชื่อกรณีทดสอบ	การวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ด
ฟังก์ชันที่ทดสอบ	การวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ด
จุดมุ่งหมายในการทดสอบ	เพื่อทำการทดสอบว่า เครื่องมือสามารถวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ดได้
เงื่อนไขการเริ่มทดสอบ	ผู้ใช้เลือกการวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ด
ข้อมูลนำเข้า	ช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด
ข้อมูลนำออก	รายละเอียดของรายการร้องขอการนำเข้าซอร์สโค้ดที่ผู้ใช้ระบุ
ผลลัพธ์ที่คาดหวัง	1. ผู้ใช้สามารถเลือกโครงการที่ต้องการวิเคราะห์ได้ 2. ผู้ใช้สามารถระบุหมายเลขสาขาการพัฒนาได้ 3. ผู้ใช้สามารถอัพโหลดซอร์สโค้ดที่ต้องการวิเคราะห์ได้
ลำดับการทดสอบ	1. ผู้ใช้เลือกรายละเอียดโครงการที่ต้องการวิเคราะห์ 2. ผู้ใช้ระบุหมายเลขสาขาการพัฒนา 3. ผู้ใช้อัพโหลดซอร์สโค้ดที่ต้องการวิเคราะห์ 4. ผู้ใช้กดปุ่ม Submit เพื่อวิเคราะห์ข้อมูล
สรุปผลการทดสอบ	ผ่าน

3. การทดสอบที่ 3

การทดสอบการวิเคราะห์ข้อมูลก่อนการร้องขอการนำเข้าซอร์สโค้ด เครื่องมือสนับสนุนนี้ ได้นำเสนอการเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด 2 ช่วงเวลา ในที่นี้จะทำเป็นการทำสอบการวิเคราะห์หลังทำการร้องขอการนำเข้าซอร์สโค้ด ซึ่งมีรายละเอียดดังตารางที่ 4.11

ตารางที่ 4.11 รายละเอียดการทดสอบที่ 3

รหัสกรณีทดสอบ	TC03
ชื่อกรณีทดสอบ	การวิเคราะห์ข้อมูลหลังการร้องขอการนำเข้าซอร์สโค้ด
ฟังก์ชันที่ทดสอบ	การวิเคราะห์ข้อมูลหลังการร้องขอการนำเข้าซอร์สโค้ด
จุดมุ่งหมายในการทดสอบ	เพื่อทำการทดสอบว่า เครื่องมือสามารถวิเคราะห์ข้อมูลหลังการร้องขอการนำเข้าซอร์สโค้ดได้
เงื่อนไขการเริ่มทดสอบ	ผู้ใช้เลือกการวิเคราะห์ข้อมูลหลังการร้องขอการนำเข้าซอร์สโค้ด
ข้อมูลนำเข้า	ช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด

ตารางที่ 4.11 รายละเอียดการทดสอบที่ 3 (ต่อ)

รหัสกรณีทดสอบ	TC03
ข้อมูลนำออก	รายละเอียดของรายการร้องขอการนำเข้าซอร์สโค้ดที่ผู้ใช้ระบุ
ผลลัพธ์ที่คาดหวัง	1. ผู้ใช้สามารถเลือกโครงการที่ต้องการวิเคราะห์ได้ 2. ผู้ใช้สามารถระบุหมายเลขรายการร้องขอการนำเข้าซอร์สโค้ดได้
ลำดับการทดสอบ	1. ผู้ใช้เลือกรายละเอียดโครงการที่ต้องการวิเคราะห์ 2. ผู้ใช้ระบุหมายเลขรายการร้องขอการนำเข้าซอร์สโค้ด 3. ผู้ใช้กดปุ่ม Submit เพื่อวิเคราะห์ข้อมูล
สรุปผลการทดสอบ	ผ่าน

4. การทดสอบที่ 4

การทดสอบการแสดงตัวแปรที่มีผลกับการรีวิวโค้ด โดยทำการเปรียบเทียบข้อมูลที่แสดง กับการประมวลผลจริง ๆ ของโปรแกรมที่มีส่วนเกี่ยวข้อง ซึ่งมีรายละเอียดดังตารางที่ 4.12

ตารางที่ 4.12 รายละเอียดการทดสอบที่ 4

รหัสกรณีทดสอบ	TC04
ชื่อกรณีทดสอบ	การแสดงตัวแปรที่มีผลกับการรีวิวโค้ด
ฟังก์ชันที่ทดสอบ	1. การดึงข้อมูลตัวแปรที่มีผลกับการรีวิวโค้ด 2. การแสดงตัวแปรที่มีผลกับการรีวิวโค้ด
จุดมุ่งหมายในการทดสอบ	เพื่อทำการทดสอบว่า เครื่องมือสามารถดึงข้อมูลตัวแปรมาแสดงได้อย่างถูกต้อง และครบถ้วนทุกตัวแปร
เงื่อนไขการเริ่มทดสอบ	ผู้ใช้ระบุรายละเอียดรายการร้องขอการนำเข้าซอร์สโค้ดเรียบร้อยแล้ว
ข้อมูลนำเข้า	รายละเอียดรายการร้องขอการนำเข้าซอร์สโค้ด
ข้อมูลนำออก	ข้อมูลตัวแปร
ผลลัพธ์ที่คาดหวัง	ระบบสามารถดึงข้อมูลตัวแปรมาแสดงได้อย่างถูกต้อง
	1. ตรวจสอบตัวแปรที่แสดงผลในหน้าจอ โดยเปรียบเทียบผลลัพธ์ที่แสดงบนหน้าจอกับรายการร้องขอการนำเข้าซอร์สโค้ดบนกิตฮับ 2. ตรวจสอบตัวแปรที่แสดงผลในหน้าจอ โดยเปรียบเทียบผลลัพธ์ที่แสดงบนหน้าจอกับการคำนวณผลโดยตรงจากโปรแกรมราดอน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.12 รายละเอียดการทดสอบที่ 4 (ต่อ)

รหัสกรณีทดสอบ	TC04
ลำดับการทดสอบ	3. ตรวจสอบตัวแปรที่แสดงผลในหน้าจอ โดยเปรียบเทียบผลลัพธ์ที่แสดงบนหน้าจอกับการคำนวณผลโดยตรงจากโปรแกรมซอร์สมิเตอร์
สรุปผลการทดสอบ	ผ่าน

5. การทดสอบที่ 5

การทดสอบการแสดงผลการวิเคราะห์ข้อมูล มีรายละเอียดดังตารางที่ 4.13

ตารางที่ 4.13 รายละเอียดการทดสอบที่ 5

รหัสกรณีทดสอบ	TC05
ชื่อกรณีทดสอบ	การแสดงผลการวิเคราะห์ข้อมูล
ฟังก์ชันที่ทดสอบ	การแสดงผลการวิเคราะห์ข้อมูล
จุดมุ่งหมายในการทดสอบ	เพื่อทำการทดสอบว่าการแสดงผลการวิเคราะห์ข้อมูลถูกต้องตามที่ผู้ใช้กำหนดวิธีการทำนายผลการรีวิวโค้ด
เงื่อนไขการเริ่มทดสอบ	ผู้ใช้กำหนดวิธีการทำนายผลการรีวิวโค้ดเรียบร้อยแล้ว
ข้อมูลนำเข้า	วิธีการทำนายผลการรีวิวโค้ด
ข้อมูลนำออก	ผลการทำนาย
ผลลัพธ์ที่คาดหวัง	1. ระบบสามารถแสดงผลการทำนายได้ในกรณีที่ผู้ใช้เลือกวิธีการทำนายผลแบบการวิเคราะห์การถดถอยโลจิสติก 2. ระบบสามารถแสดงผลการทำนายได้ในกรณีที่ผู้ใช้เลือกวิธีการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอน 3. ระบบสามารถแสดงผลการทำนายได้ในกรณีที่ผู้ใช้เลือกวิธีการทำนายผลแบบทั้งการวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอน
ลำดับการทดสอบ	1. ผู้ใช้กดปุ่มเลือกวิธีการทำนายผลการรีวิวโค้ด โดยแบ่งเป็น 3 ทางเลือก ได้แก่ - ทางเลือกที่ 1 การแสดงผลทั้งการวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอน - ทางเลือกที่ 2 การแสดงผลการวิเคราะห์การถดถอยโลจิสติก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.13 รายละเอียดการทดสอบที่ 5 (ต่อ)

รหัสกรณีทดสอบ	TC05
ลำดับการทดสอบ	- ทางเลือกที่ 3 การแสดงผลการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอน 2. ระบบแสดงผลการทำนายการรีวิวโค้ด 3. ตรวจสอบผลการทำนายการรีวิวโค้ดกับรายการร้องขอการนำเข้าซอร์สโค้ดในโครงการโอเพนซอร์สบนกิตฮับ ผลลัพธ์จะต้องตรงกัน
สรุปผลการทดสอบ	ผ่าน

6. การทดสอบที่ 6

การทดสอบการเลือกช่วงเวลาในการร้องขอการนำเข้าซอร์สโค้ด เครื่องมือสนับสนุนนี้ ได้นำเสนอการเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด 2 ช่วงเวลา ได้แก่ ก่อนทำการร้องขอการนำเข้าซอร์สโค้ด และหลังจากการร้องขอการนำเข้าซอร์สโค้ด มีรายละเอียดดังตารางที่ 4.14

ตารางที่ 4.14 รายละเอียดการทดสอบที่ 6

รหัสกรณีทดสอบ	TC06
ชื่อกรณีทดสอบ	การบันทึกผลการวิเคราะห์ข้อมูล
ฟังก์ชันที่ทดสอบ	การบันทึกผลการวิเคราะห์ข้อมูล
จุดมุ่งหมายในการทดสอบ	เพื่อทำการทดสอบว่า เครื่องมืออนุญาตให้ผู้ใช้บันทึกผลการวิเคราะห์ข้อมูลได้
เงื่อนไขการเริ่มทดสอบ	ผู้ใช้ทำการระบุวิธีการทำนายผลการรีวิวโค้ดเรียบร้อยแล้ว
ข้อมูลนำเข้า	วิธีการทำนายผลการรีวิวโค้ด
ข้อมูลนำออก	-
ผลลัพธ์ที่คาดหวัง	ผู้ใช้สามารถบันทึกผลการทำนายการรีวิวโค้ดและข้อมูลตัวแปรได้
ลำดับการทดสอบ	1. ผู้ใช้กดปุ่ม Save 2. ตรวจสอบผลของไฟล์ที่บันทึกลงในเครื่องคอมพิวเตอร์
สรุปผลการทดสอบ	ผ่าน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

วิทยานิพนธ์นี้นำเสนอ การออกแบบการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ โดยโครงการโอเพนซอร์สนำมาพิจารณาเป็นกรณีศึกษานั้น ได้แก่ โครงการโอเพนซอร์ส Ansible ซึ่งเป็นโครงการโอเพนซอร์สที่ได้รับการนิยมมากในโลกอุตสาหกรรมซอฟต์แวร์ และติดอันดับโครงการที่มีการรีวิวสูงเป็นอันดับที่ 4 ของกิตฮับในปี 2017 อีกทั้งยังมีการใช้เทคโนโลยีที่ทันสมัยในการดำเนินงานกระบวนการบนกิตฮับ หลังจากที่ได้เลือกโครงการโอเพนซอร์สแล้ว ขั้นตอนต่อมาคือการวิเคราะห์โครงการโอเพนซอร์ส รวมถึงการตรวจสอบลักษณะพิเศษของโครงการโอเพนซอร์สนั้น ๆ เพื่อให้ได้มาซึ่งลักษณะการทำงานของโครงการโอเพนซอร์ส และตัวแปรที่อยู่เบื้องหลังของโครงการโอเพนซอร์ส สำหรับโครงการโอเพนซอร์ส Ansible นั้น แผ่นป้าย (Label) ในรายการร้องขอการนำเข้าซอร์สโค้ดคือตัวแปรหนึ่งในการวิเคราะห์ข้อมูล เนื่องจากแผ่นป้ายในรายการร้องขอการนำเข้าซอร์สโค้ดนั้น เป็นตัวระบุผลการรีวิวโค้ด ว่ารายการร้องขอการนำเข้าซอร์สโค้ดนั้นมีผลการรีวิวโค้ดเป็นผ่านหรือยอมรับ (Accept) และไม่ผ่านหรือปฏิเสธ (Reject) นอกจากตัวแปรที่ได้จากลักษณะพิเศษของโครงการแล้ว งานวิจัยนี้ยังได้มีการกำหนดตัวแปรที่คาดว่าจะมีผลต่อการรีวิวโค้ดจากงานวิจัยที่เกี่ยวข้องอีกด้วย

หลังจากกำหนดตัวแปรที่คาดว่าจะมีผลในการรีวิวโค้ดแล้ว ขั้นตอนต่อมาได้แก่การดึงข้อมูลจากกิตฮับมาประมวลผล โดยวิธีในการดึงข้อมูลนั้นประกอบไปด้วย 3 วิธีด้วยกันคือ การเขียนโปรแกรมเพื่อใช้ในการเรียกเอพีไอที่ทางกิตฮับเปิดให้ใช้งานไปดึงข้อมูลมา อีกทั้งยังมีการใช้เครื่องมือมาช่วยในการประมวลผล ซึ่งประกอบด้วยเครื่องมือราดอนและเครื่องมือซอร์สมิเตอร์ จากนั้นนำข้อมูลที่ได้มาวิเคราะห์ตามหลักการเหมืองข้อมูล มีการทำความสะอาดข้อมูล และตัดข้อมูลบางส่วนที่เป็นความผิดปกติออก อีกทั้งยังมีการจัดรูปแบบข้อมูลใหม่เพื่อให้การวิเคราะห์ที่ได้มีประสิทธิภาพที่ดีมากยิ่งขึ้น จากนั้นทำการวิเคราะห์ความสัมพันธ์ของข้อมูลด้วยกฎความสัมพันธ์ (Association rules) ผลลัพธ์ที่ได้แสดงกฎความสัมพันธ์ 10 กฎ ซึ่งเป็นรูปแบบความสัมพันธ์ที่เกิดขึ้นบ่อยที่สุดในชุดข้อมูลที่ทำการวิเคราะห์ อีกทั้งยังเป็นการแสดงตัวแปรที่เกิดขึ้นบ่อยในชุดข้อมูล และสร้างแบบจำลองการทำนายการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับทั้งจากวิธีการทางสถิติและหลักการเรียนรู้ด้วยเครื่อง ซึ่งปะกอบด้วยการวิเคราะห์การถดถอยโลจิสติก และการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น โดยผลลัพธ์จากการสร้างแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับคือ สมการทำนายสองสมการที่ใช้สำหรับการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์

สบนกิตฮับของทั้งการวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น

และขั้นตอนสุดท้ายคือการประเมินผลแบบจำลองโดยการวิเคราะห์ประสิทธิภาพของแบบจำลองของทั้งสองวิธีการวิเคราะห์ ทำการเปรียบเทียบค่าที่ได้จากเมตริกความสับสน และการนำเอาชุดข้อมูลอื่น ๆ มาทดสอบการทำนายผลของแบบจำลอง จากการประเมินผลพบว่า แบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับที่ได้จากการวิเคราะห์การถดถอยโลจิสติกมีเปอร์เซ็นต์ความแม่นยำในการทำนาย 89.2307% และแบบจำลองการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับที่ได้จากการวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้นมีเปอร์เซ็นต์ความแม่นยำในการทำนาย 90.7692%

5.2 ข้อจำกัดของงานวิจัย

1. การดึงข้อมูลจากกิตฮับใช้เวลานานมาก เนื่องด้วยข้อมูลในกิตฮับมีปริมาณมาก ทำให้บางครั้งการดึงข้อมูลปริมาณมากใช้เวลาดึงข้อมูลนาน ระยะเวลาในการดึงข้อมูลหนึ่งรายการร้องขอการนำเข้าซอร์สโค้ด ใช้เวลาตั้งแต่ 10 นาทีจนถึง 30 นาทีขึ้นอยู่กับปริมาณข้อมูลในรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ๆ

2. ข้อมูลรายการร้องขอการนำเข้าซอร์สโค้ดในกิตฮับมีความหลากหลาย ทำให้จำนวนข้อมูลที่ดึงมา กับจำนวนข้อมูลที่สามารถนำมาวิเคราะห์จริง มีความแตกต่างกัน กล่าวคือ อาจมีบางรายการร้องขอการนำเข้าซอร์สโค้ดที่มีข้อมูลไม่สมบูรณ์และไม่สามารถนำไปใช้ในการวิเคราะห์ข้อมูลได้

3. ข้อมูลรายการร้องขอการนำเข้าซอร์สโค้ดที่ได้จากกิตฮับส่วนมาก เป็นรายการร้องขอการนำเข้าซอร์สโค้ดที่มีผลการรีวิวเป็นปฏิเสธ ทำให้เปอร์เซ็นต์การวิเคราะห์ข้อมูลของจำนวนข้อมูลนำเข้ามีความแตกต่างกันมาก

4. ตัวแปรที่มีผลกับผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ มีปริมาณมากไม่เพียงแต่ตัวแปรที่งานวิจัยนี้กำหนดเท่านั้น

5. ลักษณะของโครงการโอเพนซอร์สมีความสำคัญในการทำนายผลการรีวิว การศึกษาโครงการโอเพนซอร์สที่นำมาใช้ในการวิเคราะห์จึงเป็นส่วนสำคัญที่ต้องทำก่อนการวิเคราะห์ข้อมูล

5.3 แนวทางการพัฒนางานวิจัยต่อ

1. หาตัวแปรที่มีผลกับผลการรีวิวโค้ดมาใช้ในวิเคราะห์ข้อมูลเพิ่มเติม

2. ประยุกต์ใช้แนวทางการออกแบบการทำนายผลการรีวิวโค้ด ไปใช้ในการวิเคราะห์โครงการโอเพนซอร์สอื่น ๆ หรือโครงการที่อยู่ในระดับองค์กร ไม่ใช่โครงการโอเพนซอร์สได้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3. ใช้วิธีการทางสถิติวิธีการอื่น ในการวิเคราะห์ข้อมูล ยกตัวอย่างเช่น บางโครงการอาจมีผลการรีวิวมากกว่า 2 ผลการรีวิว เช่น ผลการรีวิวเป็นผ่าน ผลการรีวิวเป็นไม่ผ่าน และผลการรีวิวเป็นรอพิจารณาเพิ่มเติม อาจใช้การวิเคราะห์จำแนกกลุ่มด้วยเทคนิค Discriminant Analysis

4. ใช้หลักการเหมืองข้อมูลวิธีการอื่นมาประยุกต์ใช้ในการวิเคราะห์ข้อมูล อาจเป็นการแบ่งกลุ่มข้อมูล (Data clustering) เพื่อใช้ในการจำแนกกลุ่มรายการร้องขอการนำเข้าซอร์สโค้ด พิจารณาว่าสามารถแบ่งกลุ่มรายการร้องขอการนำเข้าซอร์สโค้ดเป็นกี่กลุ่ม แต่ละกลุ่มมีข้อมูลอะไรบ้าง เป็นต้น

5. ใช้หลักการเรียนรู้ด้วยเครื่องวิธีการอื่นมาประยุกต์ใช้ในการวิเคราะห์ข้อมูล อาจเป็นการจำแนกข้อมูล (Classification) ด้วยวิธีการอื่น เช่นการเรียนรู้ของต้นไม้ตัดสินใจ (Decision Tree) เป็นต้น

6. ข้อมูลของโครงการโอเพนซอร์สบนกิตฮับที่รวบรวมมา สามารถนำไปประยุกต์ใช้ได้หลากหลาย อาจจะเป็นการวิเคราะห์การแสดงความคิดเห็นที่เกิดขึ้นในรายการร้องขอการนำเข้าซอร์สโค้ดใดๆ หรือจำนวนการคอมมิทที่เกิดขึ้นบ่อยในรายการร้องขอการนำเข้าซอร์สโค้ด เป็นต้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บรรณานุกรม

1. Chacon, S., *Pro Git. [electronic resource]*. 2009: Berkeley, CA : Apress, 2009.
2. Fitzgerald, B., *Adopting Open Source Software : A Practical Guide*. 2011, Cambridge, Mass: The MIT Press.
3. Soares, D.M., et al., *What factors influence the reviewer assignment to pull requests?* Information and Software Technology, 2018. **98**: p. 32-43.
4. Soares, D.M., et al., *Acceptance factors of pull requests in open-source projects.* Symposium on Applied Computing, 2015: p. 1541.
5. Soares, D.M., et al., *Rejection Factors of Pull Requests Filed by Core Team Developers in Software Projects with High Acceptance Rates.* 2015 IEEE 14th International Conference on Machine Learning & Applications (ICMLA), 2015: p. 960.
6. Stark, J., *Peer Reviews as a Quality Management Technique in Open-Source Software Development Projects.* Software Quality - Ecsq 2002, 2006: p. 340.
7. Gousios, G., M.-A. Storey, and A. Bacchelli, *Work Practices and Challenges in Pull-Based Development: The Contributor's Perspective.* ICSE: International Conference on Software Engineering, 2016: p. 285.
8. Gajda, W., *Git Recipes : A Problem-Solution Approach*. The Expert's Voice in Open Source. 2013, [Berkeley, CA]: Apress.
9. Georgios Gousios, M.P., and Arie van Deursen. *An Exploratory Study of the Pull-based Software Development Model*. in *36th International Conference on Software Engineering*. 2014.
10. Li, Z.-X., et al., *What Are They Talking About? Analyzing Code Reviews in Pull-Based Development Model.* Journal of Computer Science & Technology (10009000), 2017. **32**(6): p. 1060.
11. Aggarwal, C.C., *Data Mining. [electronic resource] : The Textbook*. 2015: Cham : Springer International Publishing : Imprint: Springer, 2015.
12. Zhang, C. and S. Zhang, *Association rule mining : models and algorithms*. Lecture notes in artificial intelligence: 2307. 2002: Berlin : Springer, c2002.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

13. Theeramunkong, T., *Introduction to concepts and techniques in data mining and application to text mining*. Second Edition ed.: Sirindhom International Institute of Technology Thammasat University.
14. Dowdy, S.M., D.M. Chilko, and S. Wearden, *Statistics for Research*. Wiley Series in Probability and Statistics. Vol. 3rd ed. 2004, Hoboken, N.J.: Wiley-Interscience.
15. Fenton, N.E. and S.L. Pfleeger, *Software metrics : a rigorous and practical approach*. 1997: Boston : PWS Pub., 1997. 2nd ed., rev. printing.
16. Cureton, E.E. and R.B. D'Agostino, *Factor analysis : an applied approach*. 1983: Hillsdale, N.J. : Lawrence Erlbaum Associates, 1983.
17. Hosmer, D.W. and S. Lemeshow, *Applied logistic regression*. 2000: New York : Wiley, c2000. 2nd ed.
18. Mitchell, T.M., *Machine learning*. 1997: Boston : WCB/McGraw-Hill, c1997.
19. Lacchia, M., *Radon Documentation*. Jan 14, 2018.
20. SourceMetter.
21. McCormick, K., et al., *SPSS Statistics for Data Analysis and Visualization*. 2017, Indianapolis, IN: Wiley.
22. Witten, I.H. and M.A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Vol. 3rd ed. Ian H. Witten, Frank Eibe, Mark A. Hall. 2011, Burlington, MA: Morgan Kaufmann.
23. Hagemeister, P.O.J. *Metrics for assessing a software system's maintainability*. 9-12 Nov. 1992. IEEE.
24. Mens, T., A. Serebrenik, and A. Cleve, *Evolving Software Systems. [electronic resource]*. 2014: Berlin, Heidelberg : Springer Berlin Heidelberg : Imprint: Springer, 2014.
25. Faragó, C., P. Hegedũs, and R. Ferenc, *The Impact of Version Control Operations on the Quality Change of the Source Code.* Computational Science & Its Applications - ICCSA 2014: Part V, 2014: p. 353.
26. Bakota, T.H., P. Kortvelyesi, P. Ferenc, R. Gyimothy, T., *A probabilistic software*

Chulalongkorn University

quality model. 2011. p. 243.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ก

แผ่นป้ายที่แสดงในรายการร้องขอการนำเข้าซอร์สโค้ดของโครงการ Ansible

เนื่องด้วยโครงการ Ansible เป็นโครงการที่ใช้ระบบอัตโนมัติ Ansibullbot เป็นเครื่องมือในการจัดการรายการร้องขอการนำเข้าซอร์สโค้ดซึ่งเป็นการจัดการแบบอัตโนมัติ อีกทั้งในรายการร้องขอการนำเข้าซอร์สโค้ดของโครงการ Ansible นั้นยังมีแผ่นป้ายที่ระบุรายละเอียดของรายการร้องขอการนำเข้าซอร์สโค้ดนั้น ๆ แผ่นป้ายที่ใช้ในการระบุลำดับการทำงาน และการไหลของการทำงาน (Work flow) ที่น่าสนใจ โดยแผ่นป้ายที่แสดงในรายการร้องขอการนำเข้าซอร์สโค้ดในโครงการ Ansible แสดงรายละเอียดดังตารางที่ ก.1

ตารางที่ ก.1 รายละเอียดของแผ่นป้ายของโครงการ Ansible

ลำดับ	แผ่นป้าย	ความหมายของแผ่นป้าย
1	aci	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวกับ Cisco ACI Community
2	affects_0.9	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v0.9
3	affects_1.3	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v1.3
4	affects_1.5	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v1.5
5	affects_1.6	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v1.6
6	affects_1.7	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v1.7
7	affects_1.8	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v1.8
8	affects_1.9	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v1.9
9	affects_2.0	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v2.0
10	affects_2.1	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v2.1
11	affects_2.2	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v2.2
12	affects_2.3	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v2.3
13	affects_2.4	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v2.4
14	affects_2.5	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v2.5
15	affects_2.6	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v2.6
16	affects_2.7	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v2.7
17	affects_2.8	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ส่งผลกระทบกับ Ansible v2.8
18	automerge	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ถูกรวมอัตโนมัติโดยเครื่องมืออัตโนมัติ Ansibullbot
19	avi	แผ่นป้ายที่แสดงประชาคม (Community) ชนิดหนึ่ง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ลำดับ	แผ่นป้าย	ความหมายของแผ่นป้าย
20	aws	แผ่นป้ายที่แสดงประชาคม (Community) ชนิดหนึ่ง
21	azure	แผ่นป้ายที่แสดงประชาคม (Community) ชนิดหนึ่ง
22	backport	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ไม่ได้อยู่ในสาขาการพัฒนา (Devel branch)
23	bot_broken	บอท (Bot) มีพฤติกรรมที่ผิดพลาด และกำลังมีบุคคลตรวจสอบอยู่
24	bug	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวกับข้อผิดพลาด (bug)
25	c:	แผ่นป้าย c: ประกอบไปด้วย 103 แผ่นป้าย ซึ่งเป็นแผ่นป้ายที่ใช้ในการระบุตำแหน่งหรือหัวเรื่องของซอร์สโค้ด ได้แก่แผ่นป้ายดังต่อไปนี้ c:cli/adhoc c:cli/connection c:cli/console c:cli/doc c:cli/galaxy c:cli/playbook c:cli/pull c:cli/vault c:cli/ c:compat/six c:compat/test c:constants c:crypto c:errors/ c:executor/module_common c:executor/play_iterator c:executor/playbook_executor c:executor/playbook_executor c:executor/process/worker c:executor/stats c:executor/task_executor c:executor/task_queue_manager c:executor/task_result c:galaxy/ c:inventory/contrib_script c:inventory/dir c:inventory/expand_hosts c:inventory/group c:inventory/host c:inventory/ini c:inventory/inventory c:inventory/script c:inventory/yaml c:inventory/azure_rm_common c:module_utils/basic c:module_utils/facts c:module_utils/powershell c:module_utils/urls c:module_utils/vca c:module_utils/vmware c:module_utils/ c:parsing/data_loader c:parsing/mod_args c:parsing/quoting c:parsing/splitter c:parsing/utils/ c:parsing/vault/ c:parsing/yaml c:playbook/attribute c:playbook/base c:playbook/become c:playbook/block c:playbook/conditional c:playbook/handler c:playbook/helpers c:playbook/included_file c:playbook/loop_control

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ลำดับ	แผ่นป้าย	ความหมายของแผ่นป้าย
25 (ต่อ)	c:	c:playbook/play_context c:playbook/play c:playbook/role_include c:playbook/role/* c:playbook/taggable c:playbook/task_include c:playbook/task c:playbook/action c:plugins/cache c:plugins/callback c:plugins/connection/chroot c:plugins/connection/docker c:plugins/connection/local c:plugins/connection/network_cli c:plugins/connection/paramiko_ssh c:plugins/connection/parasistent c:plugins/connection/ssh c:plugins/connection/winrm c:plugins/connection/ c:plugins/filter c:plugins/lookup c:plugins/plugin_loader c:plugins/shell c:plugins/strategy c:plugins/terminal c:plugins/test c:plugins/vars c:template/other c:template/safe_eval c:template/templar c:/utils/cmd_functions c:utils/color c:utils/display c:utils/encypt c:utils/hashing c:utils/helpers c:utils/listfy c:utils/path c:utils/shlex c:utils/ssh_functins c:utils/vars c:vars/hostvars c:vars/unsafe_proxy c:vars/variable_manager
26	candidate_to_close	ระบบอัตโนมัติทำการตรวจสอบและพบว่าควรปิดรายการร้องขอการนำเข้าซอร์สโค้ดนี้
27	cherrypick_candidate	รายการร้องขอการนำเข้าซอร์สโค้ดนี้อยู่ในตัวเลือกการย้ายคอมมิทไปยังสาขาอื่น
28	ci_verified	ทำการแก้ไขไปยังรายการร้องขอการนำเข้าซอร์สโค้ดเพื่อสั่งประมวลผลระบบซีเอ (CI)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ลำดับ	แผ่นป้าย	ความหมายของแผ่นป้าย
29	cli/	แผ่นป้าย cli/ ประกอบด้วย 8 แผ่นป้ายด้วยกัน ได้แก่ cli/ cli/adhoc cli/console cli/doc cli/galaxy cli/playbook cli/pull cli/vault
30	cloud	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ cloud
31	cloudstack	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ cloudstack
32	code_cleanup	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ code_cleanup
33	committer_review	แผ่นป้ายที่ระบุกระบวนการทำงาน โดยระบุว่ารายการร้องขอการนำเข้าซอร์สโค้ดนี้ถูกรีวิวโดย committer
34	compat/six	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ compat/six
35	constants	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ constants
36	contrib_follow_up	ระบบอัตโนมัติได้ทำการติดตามนักพัฒนาซอฟต์แวร์ หรือกล่าวอีกนัยคือ contributor
37	core_review	แผ่นป้ายที่ระบุกระบวนการทำงาน โดยระบุว่ารายการร้องขอการนำเข้าซอร์สโค้ดนี้ถูกรีวิวโดย core หรือทีมงานหลัก
38	deprecated	รายการร้องขอการนำเข้าซอร์สโค้ดนี้มีโมดูลที่เลิกใช้แล้ว
39	digital_ocean	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ digital_ocean
40	discussion_topic	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการปรึกษาหารือเพิ่มเติม อาจทำการนัดประชุมมาเพื่อหารือรายการร้องขอการนำเข้าซอร์สโค้ดดังกล่าว
41	docker	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ docker
42	docs	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับเอกสาร
43	docsite_pr	รายการร้องขอการนำเข้าซอร์สโค้ดนี้สร้างมาสำหรับการใช้เอกสารในการแก้ไขลิ้งค์บนกิตฮับ
44	docsite	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ docsite
45	easyfix	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ง่ายต่อการแก้ไข

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ลำดับ	แผ่นป้าย	ความหมายของแผ่นป้าย
46	enhancement	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับฟีเจอร์ใหม่ที่เพิ่มเติมเข้ามา
47	errors/	มีข้อผิดพลาดเกิดขึ้นกับรายการร้องขอการนำเข้าซอร์สโค้ดนี้
48	executor/	แผ่นป้าย executor/ ประกอบด้วย 7 แผ่นป้ายด้วยกัน ได้แก่ executor/module_common executor/play_iterator executor/task_executor executor/process/worker executor/task_queue_manager executor/task_result executor/playbook_executor
49	f5	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ F5
50	feature	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับฟีเจอร์
51	filament	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ filament
52	galaxy/	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ galaxy/
53	gce	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ gce
54	hang	Ansible หยุดทำงาน
55	has_pr	รายการร้องขอการนำเข้าซอร์สโค้ดนี้มีรายการร้องขอการนำเข้าซอร์สโค้ดอื่น ๆ ที่เกี่ยวข้องด้วย
56	in progress	แผ่นป้ายระบุกระบวนการทำงาน โดยกล่าวว่ารายการร้องขอการนำเข้าซอร์สโค้ดนี้กำลังดำเนินการ
57	include_role	รายการร้องขอการนำเข้าซอร์สโค้ดนี้รวมถึง role
58	infoblox	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ infoblox
59	inventory/	แผ่นป้าย inventory/ ประกอบด้วย 5 แผ่นป้ายด้วยกัน ได้แก่ inventory/dir inventory/group inventory/host inventory/ini inventory/yaml
60	inventory	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ inventory
61	jboss	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ jboss
62	k8s	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ k8s
63	linode	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ linode
64	m:unarchive	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับโมดูล Unarchive

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ลำดับ	แผ่นป้าย	ความหมายของแผ่นป้าย
65	m:xml	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับโมดูล xml
66	meraki	แผ่นป้ายระบุประชาคม Cisco Meraki community
67	merge_commit	รายการร้องขอการนำเข้าซอร์สโค้ดมีการรวม (Merge) มากกว่าหนึ่งรายการ หรือหนึ่งคอมมิท
68	module_util	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ module_util
69	module_utils/	แผ่นป้าย module_utils/ ประกอบด้วย 8 แผ่นป้ายด้วยกัน ได้แก่ module_utils/ module_utils/basic module_utils/azure_rm_common module_utils/facts module_utils/powershell module_utils/urls module_utils/vmware
70	module	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับโมดูล
71	native_jinja	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ native_jinja
72	needs_ci_update	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ถูกกีดกันโดยการอัพเดทซีไอเพื่อให้ผ่านการทดสอบบนระบบซีไอก่อน
73	needs_ci	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องผ่านการทดสอบในระบบซีไอก่อน ให้ปิดและทดสอบ จากนั้นจึงร้องขอการนำเข้าซอร์สโค้ดใหม่
74	needs_info	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการข้อมูลเพิ่มเติม
75	needs_maintainer	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการผู้ดูแล
76	needs_rebase	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการเปลี่ยนฐาน
77	needs_repo	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการ repo
78	needs_revision	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการเปลี่ยนเวอร์ชัน หรือต้องการแก้ไข
79	needs_shippable	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการส่งออก
80	needs_template	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการแม่แบบ
81	needs_tests	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการการทดสอบ
82	needs_triage	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการให้ผู้ดูแลมาตรวจสอบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ลำดับ	แผ่นป้าย	ความหมายแผ่นป้าย
83	needs_verified	รายการร้องขอการนำเข้าซอร์สโค้ดนี้ต้องการการตรวจสอบโดยผู้ดูแล
84	net_tools	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ net_tools
85	netapp	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ netapp
86	networking	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ networking
87	new_contributor	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ new_contributor
88	new_inventory	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ new_inventory
89	new_module	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ new_module
90	new_plugin	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ new_plugin
91	nxos	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ nxos
92	openstack	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ openstack
93	ovirt	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ ovirt
94	owner_pr	รายการร้องขอการนำเข้าซอร์สโค้ดนี้จัดทำโดยผู้ดูแลโมดูล
95	P	แผ่นป้าย P หมายถึง Priority หรือลำดับความสำคัญ โดยแบ่งเป็น 3 แผ่นป้ายด้วยกัน ได้แก่ P1 หรือ Priority1 หมายถึงต้องการการใส่ใจอย่างเร่งด่วน P2 หรือ Priority2 หมายถึงต้องการการใส่ใจก่อนทำการปล่อยออกสู่ตลาด (release) และ P3 หรือ Priority3 หมายถึงอนุมัติ (approved) ไม่จำกัดเวลา
96	parsing/	แผ่นป้าย parsing/ ประกอบด้วย 4 แผ่นป้ายด้วยกัน ได้แก่ parsing/mod_args parsing/splitter parsing/vault/ parsing/yaml/
97	pending_action	รายการร้องขอการนำเข้าซอร์สโค้ดนี้มีรายการที่ยังรอการประมวลผลอยู่
98	Pep8	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ Pep8
99	performance	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ performance

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ลำดับ	แผ่นป้าย	ความหมายแผ่นป้าย
100	playbook/	แผ่นป้าย playbook/ ประกอบด้วย 13 แผ่นป้ายด้วยกัน ได้แก่ playbook/block playbook/base playbook/become playbook/conditional playbook/helpers playbook/included_file playbook/play_context playbook/playbook_include playbook/role_include playbook/taggable playbook/task_include playbook/task
101	plugin	แผ่นป้าย plugin ประกอบด้วย 20 แผ่นป้ายด้วยกัน plugin plugins/action plugins/connection/chroot plugins/cache plugins/connection/docker plugins/connection/local plugins/callback plugins/connection/network_cli plugins/connection/paramiko_ssh plugins/connection/persistent plugins/connection/ssh plugins/connection/winrm plugins/connection/ plugin/filter plugins/lookup plugins/shell plugins/strategy plugins/terminal plugins/test plugins/vars
102	pure_storage	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ pure_storage
103	python3	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ python3
104	question	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับคำถาม
105	remote_management	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ remote_management
106	roadmap_candidate	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ roadmap_candidate
107	scaleway	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ scaleway
108	seilnux	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ seilnux
109	shipit	รายการร้องขอการนำเข้าซอร์สโค้ดนี้พร้อมให้นำเข้าแล้ว
110	small_patch	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เป็นแพทช์ขนาดเล็ก
111	stale_ci	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ stale_ci
112	stale_review	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ stale_review

ลำดับ	แผ่นป้าย	ความหมายแผ่นป้าย
113	storage	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ storage
114	suggestion_given	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ suggestion_given
115	support:	แผ่นป้าย support: ประกอบด้วย 6 แผ่นป้ายด้วยกัน ได้แก่ support:certified support:community support:core support:curated support:network support:None
116	template/safe_eval	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ template/safe_eval
117	test_pull_requests	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับการทดสอบ รายการร้องขอการนำเข้าซอร์สโค้ด
118	test	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับการทดสอบ
119	tower_blocker	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ tower_blocker
120	traceback	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ traceback
121	ucs	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ ucs
122	unicode	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ unicode
123	utils/	แผ่นป้าย utils/ ประกอบด้วย 6 แผ่นป้ายด้วยกัน ได้แก่ utils/display utils/encrypt utils/helper utils/path utils/ssh_functions utils/vars
124	vars_hostvars	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ vars_hostvars
125	verified	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ verified
126	vmware	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับ vmware
127	waiting_on_	แผ่นป้ายที่ขึ้นต้นด้วย waiting_on_ ประกอบด้วย 3 แผ่นป้ายด้วยกัน ได้แก่ waiting_on_contributor waiting_on_maintainer waiting_on_vendor
128	windows	รายการร้องขอการนำเข้าซอร์สโค้ดนี้เกี่ยวข้องกับวินโดว์
129	WIP	รายการร้องขอการนำเข้าซอร์สโค้ดนี้อยู่ระหว่างการดำเนินงาน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ข
การคัดเลือกวิธีการวิเคราะห์ข้อมูล

เนื่องด้วยการวิเคราะห์ข้อมูลที่ใช้ในงานวิจัย คัดเลือกวิธีการหาความสัมพันธ์ของข้อมูลด้วยกฎความสัมพันธ์ และคัดเลือกการวิเคราะห์การถดถอยโลจิสติกเป็นการวิเคราะห์ข้อมูลทางสถิติ การวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนเป็นการวิเคราะห์ข้อมูลทางเอไอ (AI) หรือใช้หลักการเรียนรู้ด้วยเครื่อง ภาคผนวก ข จึงเป็นหัวข้อที่อธิบายเหตุผลในการเลือกวิธีนี้ในการวิเคราะห์ข้อมูลโดยการเปรียบเทียบกับวิธีการต่าง ๆ ที่มีอยู่ในปัจจุบัน ซึ่งประกอบด้วย 3 หัวข้อด้วยกัน ได้แก่ การเปรียบเทียบวิธีการหาความสัมพันธ์ของข้อมูล การเปรียบเทียบวิธีการสร้างสมการทำนายโดยวิธีการทางสถิติ และการเปรียบเทียบวิธีการสร้างสมการทำนายโดยหลักการเรียนรู้ด้วยเครื่อง ซึ่งแต่ละหัวข้อมีรายละเอียดดังต่อไปนี้

1. การเปรียบเทียบวิธีการหาความสัมพันธ์ของข้อมูล

ปัจจุบันมีวิธีการหาความสัมพันธ์ของข้อมูลมากมายในตลาด ไม่ว่าจะเป็น แต่ละวิธีการมีความสามารถในการหาความสัมพันธ์ที่แตกต่างกันออกไป โดยรายละเอียดความแตกต่างของแต่ละวิธีรายละเอียดดังตารางที่ ข.1

ตารางที่ ข.1 เปรียบเทียบความแตกต่างของวิธีการหาความสัมพันธ์ของข้อมูล

ลำดับ	วิธีการ	วัตถุประสงค์	เงื่อนไข
1	วิเคราะห์ความสัมพันธ์ด้วยวิธีการทางสถิติ การวิเคราะห์สหสัมพันธ์ (Correlation)	ใช้ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรสองตัว หรือมากกว่าสองตัว เพื่อใช้ในการหาค่าสัมประสิทธิสหสัมพันธ์	1. ค่าสัมประสิทธิเป็นการวัดความสัมพันธ์เชิงเส้น 2. ค่าสัมประสิทธิอยู่ระหว่าง -1 ถึง 1 บ่งบอกว่ามีความสัมพันธ์กันมากหรือน้อย ในทิศทางใด
2	วิเคราะห์ความสัมพันธ์ด้วยกฎความสัมพันธ์ หลักการเหมืองข้อมูล (Association rules)	ใช้ในการหารูปแบบความสัมพันธ์ของข้อมูลขนาดใหญ่ รวมถึงรูปแบบความสัมพันธ์ที่เกิดขึ้นใหม่ของข้อมูล	1. ใช้ในการหาความสัมพันธ์ของข้อมูลสองชุดหรือมากกว่าสองชุดขึ้นไป 2. วิธีการในการหาความสัมพันธ์ขึ้นอยู่กับอัลกอริทึมที่เลือกใช้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จากตาราง ข.1 งานวิจัยนี้ได้เลือกการวิเคราะห์ความสัมพันธ์ของข้อมูลโดยกฎความสัมพันธ์ ของหลักการเหมืองข้อมูล เนื่องด้วยมีวัตถุประสงค์ไม่ใช่เพียงแค่การหาความสัมพันธ์ของตัวแปรแต่ละตัวแปรเท่านั้น หากแต่เป็นการค้นหารูปแบบที่เกิดขึ้นในชุดข้อมูล และรูปแบบที่เกิดขึ้นบ่อยในชุดข้อมูล

2. การเปรียบเทียบวิธีการสร้างสมการทำนายโดยวิธีการทางสถิติ

วิธีการทางสถิติมีหลากหลาย แต่ละวิธีมีวัตถุประสงค์และข้อกำหนดแตกต่างกันออกไป โดยรายละเอียดความแตกต่างของวิธีการ รายละเอียดดังตารางที่ ข.2

ตารางที่ ข.2 เปรียบเทียบความแตกต่างของวิธีการสร้างสมการทำนายโดยวิธีการทางสถิติ

ลำดับ	วิธีการ	วัตถุประสงค์	เงื่อนไข
1	วิเคราะห์การถดถอยพหุคูณ (Multiple regression)	1. เพื่อศึกษาว่าตัวแปรต้นใดบ้างที่มีอิทธิพลกับตัวแปรตาม 2. เพื่อทำนายการเปลี่ยนแปลงของตัวแปรตาม เนื่องจากอิทธิพลของตัวแปรต้น หรือใช้ในการสร้างสมการทำนายในกรณีที่ตัวแปรตามเป็นค่าต่อเนื่อง (Interval)	1. ตัวแปรมีระดับการวัดเป็นแบบช่วงหรือแบบอัตราส่วน 2. ตัวแปรตามเป็นค่าต่อเนื่อง 3. ตัวแปรต้นหรือตัวแปรอิสระต้องไม่มีความสัมพันธ์กันเอง
2	วิเคราะห์การถดถอยโลจิสติก (Logistic regression)	1. เพื่อศึกษาว่าตัวแปรต้นใดบ้างที่มีอิทธิพลกับตัวแปรตาม 2. เพื่อทำนายการเปลี่ยนแปลงของตัวแปรตาม เนื่องจากอิทธิพลของตัวแปรต้น หรือใช้ในการสร้างสมการทำนายในกรณีที่ตัวแปรตามมีเพียง 2 ค่า	1. ตัวแปรตามมีเพียง 2 ค่า (เช่น ใช่หรือไม่ใช่) 2. ตัวแปรต้นหรือตัวแปรอิสระต้องไม่มีความสัมพันธ์กันเอง
3	การจำแนกกลุ่ม (Discriminant Analysis)	1. พยากรณ์กลุ่มให้กับเคสโดยการทราบค่าตัวแปรอิสระ 2. ศึกษาความสำคัญของตัวแปรอิสระว่าตัวแปรอิสระใดทำให้กลุ่มต่างกัน	1. ตัวแปรตามเป็นตัวแปรเชิงกลุ่ม 2. ตัวแปรอิสระควรเป็นตัวแปรเชิงปริมาณหรืออาจมีตัวแปรกลุ่มผสมอยู่ได้ 3. ตัวแปรอิสระต้องมีการแจกแจงแบบปกติเท่านั้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตาราง ข.2 จะเห็นว่าวิธีการที่เหมาะสมที่สุดในการวิเคราะห์ข้อมูล คือวิธีการวิเคราะห์การถดถอยโลจิสติก เนื่องด้วยตัวแปรตามคือผลการรีวิวซอร์สโค้ด ซึ่งมีเพียง 2 ค่าคือ ผลการรีวิวเป็นผ่านหรือยอมรับ และผลการรีวิวเป็นไม่ผ่านหรือปฏิเสธ และตัวแปรอิสระเป็นตัวแปรที่ใช้ในการวัดคุณภาพซอร์สโค้ดต่าง ซึ่งเป็นตัวแปรปริมาณ อีกทั้งวัตถุประสงค์คือการสร้างสมการทำนาย โดยผลลัพธ์การทำนายหรือตัวแปรตามมีเพียง 2 ค่า การวิเคราะห์การถดถอยโลจิสติกจึงเป็นวิธีที่เลือกใช้ในงานวิจัยนี้

2. การเปรียบเทียบวิธีการสร้างสมการทำนายโดยหลักการเรียนรู้ด้วยเครื่อง

วิธีการทางหลักการเรียนรู้ด้วยเครื่องที่ใช้ในการสร้างสมการทำนายมีหลากหลาย แต่ละวิธีมีวัตถุประสงค์และข้อกำหนดแตกต่างกันออกไป โดยรายละเอียดความแตกต่างของวิธีการ สามารถรายละเอียดได้ดังตารางที่ ข.3

ตารางที่ ข.3 เปรียบเทียบความแตกต่างของวิธีการสร้างสมการทำนายโดยหลักการเรียนรู้ด้วยเครื่อง

ลำดับ	วิธีการ	วัตถุประสงค์	ข้อดีและข้อเสีย
1	วิธีต้นไม้ตัดสินใจ (Decision Tree)	เป็นการนำข้อมูลมาใช้ในการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised learning)	ข้อดี 1. คำนวณรวดเร็ว 2. ง่ายต่อการเข้าใจ ข้อเสีย 1. แบบจำลองที่ซับซ้อนยากต่อการแปลความ 2. อาจเกิดการมีกิ่งต้นไม้ซ้ำได้หากแบบจำลองมีความซับซ้อน
2	วิธีความใกล้เคียงกันมากที่สุด (K-nearest neighbor)	ใช้ในการจำแนกกลุ่มของข้อมูลโดยพิจารณาจากค่าที่มีความใกล้เคียงกันมากที่สุดให้อยู่ในกลุ่มเดียวกัน	ข้อดี 1. ทำได้โดยง่าย และได้ผลลัพธ์ที่ดี 2. ไม่มีการใช้การเรียนรู้ข้อมูลแบบมีผู้สอน ข้อเสีย 1. ใช้เวลานานในการทำนายข้อมูลใหม่ที่นำเข้า 2. ไม่เหมาะกับชุดข้อมูลที่มีหลายมิติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ลำดับ	วิธีการ	วัตถุประสงค์	ข้อดีและข้อเสีย
3	วิธีโครงข่ายประสาทเทียม (Artificial neural network)	ใช้ในการพยากรณ์ โดยพิจารณาจากข้อมูลเข้า ชั้นที่ซ่อน และข้อมูลออก โดยวิธีโครงข่ายประสาทเทียมที่นิยมได้แก่โครงข่ายประสาทเทียมแบบเพอร์เซปตรอน ซึ่งเป็นอัลกอริทึมในการค้นหาน้ำหนักของแต่ละตัวแปร	ข้อดี 1. ได้ผลลัพธ์ที่มีประสิทธิภาพดีมาก 2. ใช้ได้กับแบบจำลองที่มีความซับซ้อน ข้อเสีย 1. ใช้เวลานานในการประมวลผลข้อมูล 2. บางครั้งแบบจำลองไม่สามารถอ่านค่าได้

การคัดเลือกวิธีการสร้างสมการทำนายด้วยหลักการเรียนรู้ด้วยเครื่องนั้น มีเป็นวิธีการที่ยาก เนื่องจากแต่ละวิธีมีข้อดีข้อเสียแตกต่างกันออกไป และทุกวิธีสามารถทำการวิเคราะห์ข้อมูลได้หมด หากแต่อาจได้ผลลัพธ์ที่มีประสิทธิภาพมากมายแตกต่างกันออกไป งานวิจัยนี้ใช้วิธีโครงข่ายประสาทเทียม หรือการวิเคราะห์โครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้นในการวิเคราะห์ข้อมูล เนื่องด้วยเป็นวิธีการที่ได้ผลลัพธ์ที่มีประสิทธิภาพดี และใช้ได้กับแบบจำลองหรือชุดข้อมูลที่มีความซับซ้อน หากแต่ใช้ระยะเวลาในการประมวลผลข้อมูลนานกว่าแบบจำลองอื่น ๆ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ค

ผลลัพธ์การวิเคราะห์ตัวแปรด้วยกฎความสัมพันธ์โดยละเอียด

ผลลัพธ์การวิเคราะห์ตัวแปรด้วยกฎความสัมพันธ์ สามารถแสดงออกเป็น 4 ส่วนด้วยกัน ได้แก่ การเตรียมข้อมูลก่อนการประมวลผล รายละเอียดของไฟล์ที่ใช้ในการวิเคราะห์ข้อมูล ตัววัดที่ใช้ในการวิเคราะห์ข้อมูล และกฎที่ดีที่สุด มีรายละเอียดดังนี้

1. การเตรียมข้อมูลก่อนการประมวลผล ทำได้โดยการแบ่งช่วงค่า (เทคนิคการใช้บินนิง Binning) รายละเอียดดังตารางที่ ค.1

ตารางที่ ค.1 การแบ่งช่วงค่าในการวิเคราะห์ความสัมพันธ์ด้วยกฎความสัมพันธ์

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
1	Commit count	1	'(-inf-3.9]'	3128
		2	'(3.9-6.8]'	361
		3	'(6.8-9.7]'	116
		4	'(9.7-12.6]'	73
		5	'(12.6-15.5]'	27
		6	'(15.5-18.4]'	20
		7	'(18.4-21.3]'	11
		8	'(21.3-24.2]'	8
		9	'(24.2-27.1]'	6
		10	'(27.1-inf)'	31
2	File changes count	1	'(-inf-17.3]'	3593
		2	'(17.3-33.6]'	121
		3	'(33.6-49.9]'	40
		4	'(49.9-66.2]'	13
		5	'(66.2-82.5]'	8
		6	'(82.5-98.8]'	4
		7	'(98.8-115.1]'	1
		8	'(115.1-131.4]'	0
		9	'(131.4-147.7]'	0
		10	'(147.7-inf)'	1

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
3	Lines of code changes	1	'(-inf-550.9]'	3397
		2	'(550.9-1100.8]'	228
		3	'(1100.8-1650.7]'	80
		4	'(1650.7-2200.6]'	31
		5	'(2200.6-2750.5]'	26
		6	'(2750.5-3300.4]'	9
		7	'(3300.4-3850.3]'	3
		8	'(3850.3-4400.2]'	3
		9	'(4400.2-4950.1]'	2
		10	'(4950.1-inf)'	2
4	Code chunk	1	'(-inf-25.7]'	3436
		2	'(25.7-50.4]'	212
		3	'(50.4-75.1]'	67
		4	'(75.1-99.8]'	30
		5	'(99.8-124.5]'	21
		6	'(124.5-149.2]'	8
		7	'(149.2-173.9]'	2
		8	'(173.9-198.6]'	2
		9	'(198.6-223.3]'	0
		10	'(223.3-inf)'	3
5	Function	1	'(-inf-23.1]'	3486
		2	'(23.1-46.2]'	214
		3	'(46.2-69.3]'	48
		4	'(69.3-92.4]'	18
		5	'(92.4-115.5]'	7
		6	'(115.5-138.6]'	2
		7	'(138.6-161.7]'	3
		8	'(161.7-184.8]'	1
		9	'(184.8-207.9]'	0

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
5	Function (ต่อ)	10	'(207.9-inf)'	2
6	Method	1	'(-inf-38.4]'	3320
		2	'(38.4-76.8]'	308
		3	'(76.8-115.2]'	90
		4	'(115.2-153.6]'	43
		5	'(153.6-192]'	11
		6	'(192-230.4]'	7
		7	'(230.4-268.8]'	0
		8	'(268.8-307.2]'	0
		9	'(307.2-345.6]'	1
		10	'(345.6-inf)'	1
7	Class	1	'(-inf-5.7]'	3338
		2	'(5.7-11.4]'	292
		3	'(11.4-17.1]'	70
		4	'(17.1-22.8]'	25
		5	'(22.8-28.5]'	18
		6	'(28.5-34.2]'	7
		7	'(34.2-39.9]'	19
		8	'(39.9-45.6]'	7
		9	'(45.6-51.3]'	3
		10	'(51.3-inf)'	2
8	LLOC	1	'(-inf-511.7]'	2969
		2	'(511.7-1013.4]'	553
		3	'(1013.4-1515.1]'	117
		4	'(1515.1-2016.8]'	68
		5	'(2016.8-2518.5]'	21
		6	'(2518.5-3020.2]'	30
		7	'(3020.2-3521.9]'	10
		8	'(3521.9-4023.6]'	5

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
8	LLOC (ต่อ)	9	'(4023.6-4525.3]'	6
		10	(4525.3-inf)	2
9	SLOC	1	'(-inf-963.8]'	3158
		2	'(963.8-1917.6]'	453
		3	'(1917.6-2871.4]'	75
		4	'(2871.4-3825.2]'	49
		5	'(3825.2-4779]'	16
		6	'(4779-5732.8]'	8
		7	'(5732.8-6686.6]'	6
		8	'(6686.6-7640.4]'	7
		9	'(7640.4-8594.2]'	6
		10	'(8594.2-inf)'	3
10	Multi lines count	1	'(-inf-91.9]'	3104
		2	'(91.9-183.8]'	429
		3	'(183.8-275.7]'	124
		4	'(275.7-367.6]'	57
		5	'(367.6-459.5]'	31
		6	'(459.5-551.4]'	16
		7	'(551.4-643.3]'	3
		8	'(643.3-735.2]'	4
		9	'(735.2-827.1]'	8
		10	'(827.1-inf)'	5
11	Blank lines count	1	'(-inf-184.6]'	3148
		2	'(184.6-366.2]'	407
		3	'(366.2-547.8]'	128
		4	'(547.8-729.4]'	42
		5	'(729.4-911]'	24
		6	'(911-1092.6]'	13
		7	'(1092.6-1274.2]'	9

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
11	Blank lines count (ต่อ)	8	'(1274.2-1455.8]'	3
		9	'(1455.8-1637.4]'	2
		10	'(1637.4-inf)'	5
12	Halstead : H1	1	'(-inf-30.4]'	3505
		2	'(30.4-59.8]'	162
		3	'(59.8-89.2]'	54
		4	'(89.2-118.6]'	32
		5	'(118.6-148]'	8
		6	'(148-177.4]'	6
		7	'(177.4-206.8]'	6
		8	'(206.8-236.2]'	4
		9	'(236.2-265.6]'	3
		10	'(265.6-inf)'	1
13	Halstead : H2	1	'(-inf-178.7]'	2922
		2	'(178.7-356.4]'	525
		3	'(356.4-534.1]'	136
		4	'(534.1-711.8]'	95
		5	'(711.8-889.5]'	40
		6	'(889.5-1067.2]'	22
		7	'(1067.2-1244.9]'	21
		8	'(1244.9-1422.6]'	8
		9	'(1422.6-1600.3]'	9
		10	'(1600.3-inf)'	3
14	Halstead : N1	1	'(-inf-107.5]'	2886
		2	'(107.5-214]'	531
		3	'(214-320.5]'	153
		4	'(320.5-427]'	78
		5	'(427-533.5]'	64
		6	'(533.5-640]'	21

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
14	Halstead : N1 (ต่อ)	7	'(640-746.5]'	28
		8	'(746.5-853]'	5
		9	'(853-959.5]'	9
		10	'(959.5-inf)'	6
15	Halstead : N2	1	'(-inf-206.6]'	2908
		2	'(206.6-412.2]'	540
		3	'(412.2-617.8]'	126
		4	'(617.8-823.4]'	81
		5	'(823.4-1029]'	61
		6	'(1029-1234.6]'	14
		7	'(1234.6-1440.2]'	30
		8	'(1440.2-1645.8]'	9
		9	'(1645.8-1851.4]'	8
		10	'(1851.4-inf)'	4
16	Halstead : Vocab	1	'(-inf-202.3]'	2948
		2	'(202.3-402.6]'	520
		3	'(402.6-602.9]'	123
		4	'(602.9-803.2]'	101
		5	'(803.2-1003.5]'	31
		6	'(1003.5-1203.8]'	33
		7	'(1203.8-1404.1]'	6
		8	'(1404.1-1604.4]'	8
		9	'(1604.4-1804.7]'	6
		10	'(1804.7-inf)'	5
17	Halstead : Length	1	'(-inf-1584.4382]'	2904
		2	'(1584.4382-3166.8764]'	472
		3	'(3166.8764-4749.3146]'	184
		4	'(4749.3146-6331.7528]'	66
		5	'(6331.7528-7914.191]'	43

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
17	Halstead : Length (ต่อ)	6	'(7914.191-9496.6292]'	55
		7	'(9496.6292-11079.0674]'	14
		8	'(11079.0674-12661.5056]'	12
		9	'(12661.5056-14243.9438]'	25
		10	'(14243.9438-inf)'	6
18	Halstead : Calculated Length	1	'(-inf-1285.6382]'	2912
		2	'(1285.6382-2571.2764]'	470
		3	'(2571.2764-3856.9146]'	181
		4	'(3856.9146-5142.5528]'	59
		5	'(5142.5528-6428.191]'	43
		6	'(6428.191-7713.8292]'	50
		7	'(7713.8292-8999.4674]'	22
		8	'(8999.4674-10285.1056]'	14
		9	'(10285.1056-11570.7438]'	24
		10	'(11570.7438-inf)'	6
19	Halstead : Volume	1	'(-inf-2188.552066]'	3198
		2	'(2188.552066-4375.51917]'	330
		3	'(4375.51917-6562.486274]'	133
		4	'(6562.486274-8749.453377]'	23
		5	'(8749.453377-10936.420481]'	28
		6	'(10936.420481-13123.387585]'	25
		7	'(13123.387585-15310.354689]'	26
		8	'(15310.354689-17497.321793]'	0
		9	'(17497.321793-19684.288896]'	4
		10	'(19684.288896-inf)'	14

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
20	Halstead: Difficulty	1	'(-inf-1.655929]'	323
		2	'(1.655929-3.145191]'	593
		3	'(3.145191-4.634453]'	818
		4	'(4.634453-6.123716]'	783
		5	'(6.123716-7.612978]'	579
		6	'(7.612978-9.10224]'	387
		7	'(9.10224-10.591502]'	182
		8	'(10.591502-12.080765]'	55
		9	'(12.080765-13.570027]'	41
		10	'(13.570027-inf)'	20
21	Halstead : effort	1	'(-inf-31835.579233]'	3478
		2	'(31835.579233-63670.365985]'	178
		3	'(63670.365985-95505.152737]'	31
		4	'(95505.152737-127339.939489]'	29
		5	'(127339.939489-159174.726241]'	11
		6	'(159174.726241-191009.512992]'	35
		7	'(191009.512992-222844.299744]'	4
		8	'(222844.299744-254679.086496]'	1
		9	'(254679.086496-286513.873248]'	0
		10	'(286513.873248-inf)'	14

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
22	Halstead : time	1	'(-inf-1768.643324]'	3478
		2	'(1768.643324-3537.242621]'	178
		3	'(3537.242621-5305.841919]'	31
		4	'(5305.841919-7074.441216]'	29
		5	'(7074.441216-8843.040513]'	11
		6	'(8843.040513-10611.639811]'	35
		7	'(10611.639811-12380.239108]'	4
		8	'(12380.239108-14148.838405]'	1
		9	'(14148.838405-15917.437703]'	0
		10	'(15917.437703-inf)'	14
23	Halstead : Bugs	1	'(-inf-0.729517]'	3198
		2	'(0.729517-1.458506]'	330
		3	'(1.458506-2.187495]'	133
		4	'(2.187495-2.916484]'	23
		5	'(2.916484-3.645473]'	28
		6	'(3.645473-4.374462]'	25
		7	'(4.374462-5.103451]'	26
		8	'(5.103451-5.83244]'	0
		9	'(5.83244-6.561429]'	4
		10	'(6.561429-inf)'	14
24	CC	1	'(-inf-0.011395]'	231
		2	'(0.011395-0.021064]'	219
		3	'(0.021064-0.030734]'	997
		4	'(0.030734-0.040403]'	463
		5	'(0.040403-0.050073]'	742
		6	'(0.050073-0.059742]'	197

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
24	CC (ต่อ)	7	'(0.059742-0.069412]'	351
		8	'(0.069412-0.079081]'	177
		9	'(0.079081-0.088751]'	338
		10	'(0.088751-inf)'	66
25	CCL	1	'(-inf-65.5]'	2528
		2	'(65.5-130]'	246
		3	'(130-194.5]'	317
		4	'(194.5-259]'	131
		5	'(259-323.5]'	148
		6	'(323.5-388]'	21
		7	'(388-452.5]'	48
		8	'(452.5-517]'	267
		9	'(517-581.5]'	49
		10	'(581.5-inf)'	26
26	CCO	1	'(-inf-300.6]'	2304
		2	'(300.6-587.2]'	465
		3	'(587.2-873.8]'	73
		4	'(873.8-1160.4]'	320
		5	'(1160.4-1447]'	85
		6	'(1447-1733.6]'	134
		7	'(1733.6-2020.2]'	16
		8	'(2020.2-2306.8]'	174
		9	'(2306.8-2593.4]'	179
		10	'(2593.4-inf)'	31
27	CI	1	'(-inf-78.7]'	2371
		2	'(78.7-155.4]'	377
		3	'(155.4-232.1]'	124
		4	'(232.1-308.8]'	291
		5	'(308.8-385.5]'	188

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
27	CI (ต่อ)	6	'(385.5-462.2]'	36
		7	'(462.2-538.9]'	18
		8	'(538.9-615.6]'	154
		9	'(615.6-692.3]'	187
		10	'(692.3-inf)'	35
28	CLC	1	'(-inf-0.009928]'	232
		2	'(0.009928-0.018614]'	301
		3	'(0.018614-0.0273]'	900
		4	'(0.0273-0.035985]'	704
		5	'(0.035985-0.044671]'	511
		6	'(0.044671-0.053357]'	208
		7	'(0.053357-0.062042]'	424
		8	'(0.062042-0.070728]'	104
		9	'(0.070728-0.079414]'	331
		10	'(0.079414-inf)'	66
29	LDC	1	'(-inf-6.513695]'	1567
		2	'(6.513695-12.982946]'	1061
		3	'(12.982946-19.452196]'	343
		4	'(19.452196-25.921447]'	125
		5	'(25.921447-32.390697]'	210
		6	'(32.390697-38.859948]'	76
		7	'(38.859948-45.329198]'	23
		8	'(45.329198-51.798449]'	162
		9	'(51.798449-58.267699]'	179
		10	'(58.267699-inf)'	35
30	LLDC	1	'(-inf-1037.7]'	2458
		2	'(1037.7-2045.4]'	189
		3	'(2045.4-3053.1]'	372
		4	'(3053.1-4060.8]'	273

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
30	LLDC (ต่อ)	5	'(4060.8-5068.5]'	75
		6	'(5068.5-6076.2]'	28
		7	'(6076.2-7083.9]'	168
		8	'(7083.9-8091.6]'	183
		9	'(8091.6-9099.3]'	11
		10	'(9099.3-inf)'	24
31	LCOM5	1	'(-inf-868]'	2415
		2	'(868-1710]'	241
		3	'(1710-2552]'	392
		4	'(2552-3394]'	260
		5	'(3394-4236]'	68
		6	'(4236-5078]'	23
		7	'(5078-5920]'	161
		8	'(5920-6762]'	186
		9	'(6762-7604]'	9
		10	'(7604-inf)'	26
32	NL	1	'(-inf-215.6]'	266
		2	'(215.6-393.2]'	326
		3	'(393.2-570.8]'	539
		4	'(570.8-748.4]'	1174
		5	'(748.4-926]'	609
		6	'(926-1103.6]'	256
		7	'(1103.6-1281.2]'	180
		8	'(1281.2-1458.8]'	62
		9	'(1458.8-1636.4]'	159
		10	'(1636.4-inf)'	210
33	NLE	1	'(-inf-192.8]'	343
		2	'(192.8-365.6]'	442
		3	'(365.6-538.4]'	951

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
33	NLE (ต่อ)	4	'(538.4-711.2]'	996
		5	'(711.2-884]'	298
		6	'(884-1056.8]'	170
		7	'(1056.8-1229.6]'	171
		8	'(1229.6-1402.4]'	43
		9	'(1402.4-1575.2]'	181
		10	'(1575.2-inf)'	186
34	WMC	1	'(-inf-160.8]'	356
		2	'(160.8-307.6]'	446
		3	'(307.6-454.4]'	1006
		4	'(454.4-601.2]'	994
		5	'(601.2-748]'	271
		6	'(748-894.8]'	148
		7	'(894.8-1041.6]'	153
		8	'(1041.6-1188.4]'	34
		9	'(1188.4-1335.2]'	163
		10	'(1335.2-inf)'	210
35	CBO	1	'(-inf-1941.7]'	240
		2	'(1941.7-3596.4]'	337
		3	'(3596.4-5251.1]'	1059
		4	'(5251.1-6905.8]'	1133
		5	'(6905.8-8560.5]'	336
		6	'(8560.5-10215.2]'	200
		7	'(10215.2-11869.9]'	82
		8	'(11869.9-13524.6]'	25
		9	'(13524.6-15179.3]'	164
		10	'(15179.3-inf)'	205

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
36	CBOI	1	'(-inf-32.1]'	442
		2	'(32.1-61.2]'	538
		3	'(61.2-90.3]'	774
		4	'(90.3-119.4]'	899
		5	'(119.4-148.5]'	315
		6	'(148.5-177.6]'	39
		7	'(177.6-206.7]'	31
		8	'(206.7-235.8]'	142
		9	'(235.8-264.9]'	505
		10	'(264.9-inf)'	96
37	NII	1	'(-inf-32]'	441
		2	'(32-61]'	535
		3	'(61-90]'	752
		4	'(90-119]'	944
		5	'(119-148]'	313
		6	'(148-177]'	16
		7	'(177-206]'	34
		8	'(206-235]'	145
		9	'(235-264]'	505
		10	'(264-inf)'	96
38	NOI	1	'(-inf-34.3]'	2567
		2	'(34.3-68.6]'	75
		3	'(68.6-102.9]'	309
		4	'(102.9-137.2]'	98
		5	'(137.2-171.5]'	41
		6	'(171.5-205.8]'	91
		7	'(205.8-240.1]'	173
		8	'(240.1-274.4]'	58
		9	'(274.4-308.7]'	190

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
38	NOI (ต่อ)	10	'(308.7-inf)'	179
39	RFC	1	'(-inf-32.5]'	328
		2	'(32.5-65]'	322
		3	'(65-97.5]'	1063
		4	'(97.5-130]'	953
		5	'(130-162.5]'	227
		6	'(162.5-195]'	194
		7	'(195-227.5]'	66
		8	'(227.5-260]'	334
		9	'(260-292.5]'	230
		10	'(292.5-inf)'	64
40	CD	1	'(-inf-5.294444]'	333
		2	'(5.294444-6.088889]'	688
		3	'(6.088889-6.883333]'	1735
		4	'(6.883333-7.677778]'	726
		5	'(7.677778-8.472223]'	161
		6	'(8.472223-9.266667]'	49
		7	'(9.266667-10.061111]'	50
		8	'(10.061111-10.855556]'	18
		9	'(10.855556-11.650001]'	13
		10	'(11.650001-inf)'	8
41	CLOC	1	'(-inf-9.480625]'	168
		2	'(9.480625-17.904988]'	305
		3	'(17.904988-26.32935]'	323
		4	'(26.32935-34.753713]'	828
		5	'(34.753713-43.178076]'	381
		6	'(43.178076-51.602439]'	832
		7	'(51.602439-60.026802]'	227
		8	'(60.026802-68.451164]'	221

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
41	CLOC (ต่อ)	9	'(68.451164-76.875527]'	425
		10	'(76.875527-inf)'	71
42	DLOC	1	'(-inf-1128.5]'	161
		2	'(1128.5-2061]'	249
		3	'(2061-2993.5]'	482
		4	'(2993.5-3926]'	1061
		5	'(3926-4858.5]'	850
		6	'(4858.5-5791]'	300
		7	'(5791-6723.5]'	189
		8	'(6723.5-7656]'	105
		9	'(7656-8588.5]'	275
		10	'(8588.5-inf)'	109
43	TCD	1	'(-inf-8.156086]'	1023
		2	'(8.156086-9.472076]'	1182
		3	'(9.472076-10.788067]'	674
		4	'(10.788067-12.104057]'	516
		5	'(12.104057-13.420048]'	251
		6	'(13.420048-14.736038]'	80
		7	'(14.736038-16.052029]'	42
		8	'(16.052029-17.368019]'	8
		9	'(17.368019-18.68401]'	1
		10	'(18.68401-inf)'	4
44	TCLOC	1	'(-inf-9.483208]'	169
		2	'(9.483208-17.910154]'	304
		3	'(17.910154-26.337099]'	323
		4	'(26.337099-34.764045]'	828
		5	'(34.764045-43.190991]'	381
		6	'(43.190991-51.617937]'	832
		7	'(51.617937-60.044883]'	227

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
44	TCLOC (ต่อ)	8	'(60.044883-68.471828]'	221
		9	'(68.471828-76.898774]'	425
		10	'(76.898774-inf)'	71
45	DIT	1	'(-inf-1129.1]'	161
		2	'(1129.1-2062.2]'	250
		3	'(2062.2-2995.3]'	480
		4	'(2995.3-3928.4]'	1056
		5	'(3928.4-4861.5]'	854
		6	'(4861.5-5794.6]'	300
		7	'(5794.6-6727.7]'	191
		8	'(6727.7-7660.8]'	105
		9	'(7660.8-8593.9]'	275
		10	'(8593.9-inf)'	109
46	NOA	1	'(-inf-22.4]'	128
		2	'(22.4-43.8]'	927
		3	'(43.8-65.2]'	776
		4	'(65.2-86.6]'	135
		5	'(86.6-108]'	432
		6	'(108-129.4]'	675
		7	'(129.4-150.8]'	519
		8	'(150.8-172.2]'	48
		9	'(172.2-193.6]'	57
		10	'(193.6-inf)'	84
47	NOC	1	'(-inf-22.6]'	128
		2	'(22.6-44.2]'	961
		3	'(44.2-65.8]'	742
		4	'(65.8-87.4]'	135
		5	'(87.4-109]'	421
		6	'(109-130.6]'	702

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
47	NOC (ต่อ)	7	'(130.6-152.2]'	509
		8	'(152.2-173.8]'	42
		9	'(173.8-195.4]'	57
		10	'(195.4-inf)'	84
48	NOD	1	'(-inf-17.7]'	404
		2	'(17.7-34.4]'	721
		3	'(34.4-51.1]'	686
		4	'(51.1-67.8]'	581
		5	'(67.8-84.5]'	590
		6	'(84.5-101.2]'	31
		7	'(101.2-117.9]'	98
		8	'(117.9-134.6]'	437
		9	'(134.6-151.3]'	127
		10	'(151.3-inf)'	106
49	NOP	1	'(-inf-22.6]'	122
		2	'(22.6-44.2]'	964
		3	'(44.2-65.8]'	742
		4	'(65.8-87.4]'	136
		5	'(87.4-109]'	418
		6	'(109-130.6]'	706
		7	'(130.6-152.2]'	509
		8	'(152.2-173.8]'	40
		9	'(173.8-195.4]'	60
		10	'(195.4-inf)'	84
50	NA	1	'(-inf-8945.3]'	239
		2	'(8945.3-16491.6]'	374
		3	'(16491.6-24037.9]'	1337
		4	'(24037.9-31584.2]'	881
		5	'(31584.2-39130.5]'	303

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
50	NA (ต่อ)	6	'(39130.5-46676.8]'	182
		7	'(46676.8-54223.1]'	73
		8	'(54223.1-61769.4]'	23
		9	'(61769.4-69315.7]'	185
		10	'(69315.7-inf)'	184
51	NLA	1	'(-inf-94.3]'	726
		2	'(94.3-175.6]'	693
		3	'(175.6-256.9]'	79
		4	'(256.9-338.2]'	542
		5	'(338.2-419.5]'	951
		6	'(419.5-500.8]'	29
		7	'(500.8-582.1]'	33
		8	'(582.1-663.4]'	83
		9	'(663.4-744.7]'	554
		10	'(744.7-inf)'	91
52	NLM	1	'(-inf-72.5]'	614
		2	'(72.5-132]'	246
		3	'(132-191.5]'	795
		4	'(191.5-251]'	849
		5	'(251-310.5]'	94
		6	'(310.5-370]'	417
		7	'(370-429.5]'	134
		8	'(429.5-489]'	61
		9	'(489-548.5]'	509
		10	'(548.5-inf)'	62
53	NM	1	'(-inf-477]'	266
		2	'(477-890]'	450
		3	'(890-1303]'	1386
		4	'(1303-1716]'	729

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
53	NM (ต่อ)	5	'(1716-2129]'	284
		6	'(2129-2542]'	199
		7	'(2542-2955]'	79
		8	'(2955-3368]'	38
		9	'(3368-3781]'	148
		10	'(3781-inf)'	202
54	NOS	1	'(-inf-625.3]'	376
		2	'(625.3-1186.6]'	504
		3	'(1186.6-1747.9]'	916
		4	'(1747.9-2309.2]'	590
		5	'(2309.2-2870.5]'	648
		6	'(2870.5-3431.8]'	182
		7	'(3431.8-3993.1]'	152
		8	'(3993.1-4554.4]'	149
		9	'(4554.4-5115.7]'	241
		10	'(5115.7-inf)'	23
55	TLLOC	1	'(-inf-4556.1]'	236
		2	'(4556.1-8421.2]'	345
		3	'(8421.2-12286.3]'	1314
		4	'(12286.3-16151.4]'	873
		5	'(16151.4-20016.5]'	336
		6	'(20016.5-23881.6]'	199
		7	'(23881.6-27746.7]'	86
		8	'(27746.7-31611.8]'	20
		9	'(31611.8-35476.9]'	163
		10	'(35476.9-inf)'	209
56	TLOC	1	'(-inf-7059.5]'	246
		2	'(7059.5-13032]'	428
		3	'(13032-19004.5]'	1580

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
56	TLOC (ต่อ)	4	'(19004.5-24977]'	654
		5	'(24977-30949.5]'	256
		6	'(30949.5-36922]'	177
		7	'(36922-42894.5]'	53
		8	'(42894.5-48867]'	26
		9	'(48867-54839.5]'	174
		10	'(54839.5-inf)'	187
57	TNA	1	'(-inf-8945.3]'	239
		2	'(8945.3-16491.6]'	374
		3	'(16491.6-24037.9]'	1337
		4	'(24037.9-31584.2]'	881
		5	'(31584.2-39130.5]'	303
		6	'(39130.5-46676.8]'	182
		7	'(46676.8-54223.1]'	73
		8	'(54223.1-61769.4]'	23
		9	'(61769.4-69315.7]'	185
		10	'(69315.7-inf)'	184
58	TNLA	1	'(-inf-94.4]'	726
		2	'(94.4-175.8]'	693
		3	'(175.8-257.2]'	79
		4	'(257.2-338.6]'	542
		5	'(338.6-420]'	951
		6	'(420-501.4]'	29
		7	'(501.4-582.8]'	33
		8	'(582.8-664.2]'	90
		9	'(664.2-745.6]'	547
		10	'(745.6-inf)'	91

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
59	TNLM	1	'(-inf-72.6]'	614
		2	'(72.6-132.2]'	246
		3	'(132.2-191.8]'	795
		4	'(191.8-251.4]'	849
		5	'(251.4-311]'	95
		6	'(311-370.6]'	406
		7	'(370.6-430.2]'	144
		8	'(430.2-489.8]'	61
		9	'(489.8-549.4]'	509
		10	'(549.4-inf)'	62
60	TNM	1	'(-inf-477.1]'	266
		2	'(477.1-890.2]'	450
		3	'(890.2-1303.3]'	1380
		4	'(1303.3-1716.4]'	735
		5	'(1716.4-2129.5]'	284
		6	'(2129.5-2542.6]'	199
		7	'(2542.6-2955.7]'	79
		8	'(2955.7-3368.8]'	38
		9	'(3368.8-3781.9]'	148
		10	'(3781.9-inf)'	202
61	TNOS	1	'(-inf-629.3]'	365
		2	'(629.3-1193.6]'	478
		3	'(1193.6-1757.9]'	932
		4	'(1757.9-2322.2]'	539
		5	'(2322.2-2886.5]'	715
		6	'(2886.5-3450.8]'	159
		7	'(3450.8-4015.1]'	169
		8	'(4015.1-4579.4]'	140
		9	'(4579.4-5143.7]'	261

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

อันดับ	ตัวแปร	ช่วงที่	ค่าของช่วง	จำนวน
61	TNOS (ต่อ)	10	'(5143.7-inf)'	23

2. รายละเอียดของไฟล์ที่ใช้ในการวิเคราะห์ข้อมูล

Scheme (สคีมา):
weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation (ความสัมพันธ์):
02-Remove outlier-weka.filters.unsupervised. attribute. Discretize-B10-M-1.0-Rfirst-last-precision6
Instances (กรณีตัวอย่าง): 3781
Attributes (ลักษณะประจำ): มีจำนวน 64 ตัว ดังนี้ pull request id, Count (commit#), file_changed, LOC change, Code chunk, function, method, class, LLOC, SLOC, comments, multi, blank, hal:h1, hal:h2, hal:N1, hal:N2, hal:vocabulary, hal:length, hal:calculated_length, hal:volume, hal:difficulty, hal:effort, hal:time, hal:bugs, CC, CCL, CCO, CI, CLC, LDC, LLDC, LCOM5, NL, NLE, WMC, CBO, CBOI, NII, NOI, RFC, CD, CLOC, DLOC, TCD, TCLOC, DIT, NOA, NOC, NOD, NOP, NA, NLA, NLM, NM, NOS, TLLOC, TLOC, TNA, TNLA, TNLM, TNM, TNOS, Result

3. ตัววัดที่ใช้ในการวิเคราะห์ข้อมูล

Minimum support (ค่าสนับสนุนที่น้อยที่สุด): 0.85 (3214 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 3
Generated sets of large itemsets:
Size of set of large itemsets L(1): 10
Size of set of large itemsets L(2): 23
Size of set of large itemsets L(3): 8

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4. กฎที่ดีที่สุด

Best rules found:

1. hal:time='(-inf-1768.643324]' 3478 ==> hal:effort='(-inf-31835.579233]' 3478 <conf:(1)> lift:(1.09) lev:(0.07) [278] conv:(278.72)

2. hal:effort='(-inf-31835.579233]' 3478 ==> hal:time='(-inf-1768.643324]' 3478 <conf:(1)> lift:(1.09) lev:(0.07) [278] conv:(278.72)

3. file_changed='(-inf-17.3]' hal:time='(-inf-1768.643324]' 3297 ==> hal:effort='(-inf-31835.579233]' 3297 <conf:(1)> lift:(1.09) lev:(0.07) [264] conv:(264.21)

4. file_changed='(-inf-17.3]' hal:effort='(-inf-31835.579233]' 3297 ==> hal:time='(-inf-1768.643324]' 3297 <conf:(1)> lift:(1.09) lev:(0.07) [264] conv:(264.21)

5. hal:time='(-inf-1768.643324]' Result=reject 3251 ==> hal:effort='(-inf-31835.579233]' 3251 <conf:(1)> lift:(1.09) lev:(0.07) [260] conv:(260.53)

6. hal:effort='(-inf-31835.579233]' Result=reject 3251 ==> hal:time='(-inf-1768.643324]' 3251 <conf:(1)> lift:(1.09) lev:(0.07) [260] conv:(260.53)

7. function='(-inf-23.1]' hal:time='(-inf-1768.643324]' 3232 ==> hal:effort='(-inf-31835.579233]' 3232 <conf:(1)> lift:(1.09) lev:(0.07) [259] conv:(259)

8. function='(-inf-23.1]' hal:effort='(-inf-31835.579233]' 3232 ==> hal:time='(-inf-1768.643324]' 3232 <conf:(1)> lift:(1.09) lev:(0.07) [259] conv:(259)

9. hal:h1='(-inf-30.4]' hal:time='(-inf-1768.643324]' 3217 ==> hal:effort='(-inf-31835.579233]' 3217 <conf:(1)> lift:(1.09) lev:(0.07) [257] conv:(257.8)

10. hal:h1='(-inf-30.4]' hal:effort='(-inf-31835.579233]' 3217 ==> hal:time='(-inf-1768.643324]' 3217 <conf:(1)> lift:(1.09) lev:(0.07) [257] conv:(257.8)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ค

วิธีการวิเคราะห์การถดถอยโลจิสติกโดยละเอียด

หลังจากที่ทำการวิเคราะห์ข้อมูลผ่านเครื่องมือเอสพีเอสเอสแล้ว ผลลัพธ์การวิเคราะห์ตัวแปรด้วยการวิเคราะห์การถดถอยโลจิสติก สามารถอธิบายเป็น 7 ตารางด้วยกัน มีรายละเอียดดังต่อไปนี้

1. ผลลัพธ์สรุปการเตรียมกรณีตัวอย่าง (Case Processing Summary)

แสดงการเตรียมกรณีตัวอย่าง โดยกรณีที่เลือกมาวิเคราะห์มีจำนวน 1,208 รายการ ซึ่งคือจำนวนรายการร้องขอการนำซอร์สโค้ดทั้งหมดที่ได้จัดเตรียมไว้ โดยที่ทุกรายการมีค่าครบถ้วน

ตาราง ค.1 ผลลัพธ์สรุปการเตรียมกรณีตัวอย่าง (Case Processing Summary)

Unweighted Cases		N	Percent
Selected Cases	Included in Analysis	1208	100.0
	Missing Cases	0	.0
	Total	1208	100.0
Unselected Cases		0	.0
Total		1208	100.0

2. ผลลัพธ์การกำหนดค่าตัวแปรตาม (Dependent Variable Encoding)

ผลลัพธ์การกำหนดค่าตัวแปรตามแสดงการกำหนดค่าตัวแปรตาม โดยในที่นี้กำหนดให้ Accept มีค่าเท่ากับ 0 และ Reject มีค่าเท่ากับ 1 แสดงดังตารางที่ ค.2

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง ค.2 ผลลัพธ์การกำหนดค่าตัวแปรตาม

Original Value	Internal Value
accept	0
reject	1

3. ผลลัพธ์การวิเคราะห์ Omnibus Tests of Model Coefficients

การวิเคราะห์ Omnibus Tests of Model Coefficients เป็นการวิเคราะห์การทดสอบค่าไคสแควร์ (Chi Square) และการตรวจสอบค่านัยสำคัญ (Siginificant) รายละเอียดดังตารางที่ ค.3

ตารางที่ ค.3 ผลลัพธ์การวิเคราะห์ Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step1	Step	150.529	54	.000
	Block	150.529	54	.000
	Model	150.529	54	.000

จากตาราง ค.3 การวิเคราะห์สถิติทดสอบ 3 ค่าคือ สถิติทดสอบของสเตป (Step) สถิติทดสอบของบล็อก (Block) และสถิติทดสอบของโมเดล (Model) มีรายละเอียดดังนี้สถิติทดสอบ ของสเตป (Step) ใช้ทดสอบสมมติฐาน โดย กำหนด H0 คือ โอกาสที่ผลลัพธ์การรีวิวซอร์สโค้ดไม่ขึ้นอยู่กับตัวแปรที่วิเคราะห์ และ H1 คือโอกาสที่ผลลัพธ์การรีวิวซอร์สโค้ดขึ้นอยู่กับตัวแปรที่วิเคราะห์ ซึ่งค่าสถิติทดสอบของสเตปมีค่า 150.529 และค่านัยสำคัญที่ระดับ .000 นั่นคือปฏิเสธ H0 ในส่วนของค่าสถิติทดสอบของสเตปและค่าสถิติทดสอบของบล็อกมีค่าเท่ากับค่าสถิติทดสอบของโมเดล เนื่องจากแบบจำลองมีเฉพาะค่าคงที่ และตัวแปรอิสระทั้ง 63 ตัว

4. ผลลัพธ์การสรุปแบบจำลอง (Model Summary)

แสดงความเหมาะสมของแบบจำลอง (Goodness of fit) นั่นคือ -2 Log likelihood ซึ่งผลลัพธ์แสดงค่า -2 Log likelihood มีมากกว่าค่าคงที่ รายละเอียดดังตาราง ค.4

จุฬาลงกรณ์มหาวิทยาลัย

ตาราง ค.4 ผลลัพธ์การสรุปแบบจำลอง

Step	-2 Loglikeilhood	Cox & Snell R Square	Nagelkerke R Square
1	302.828	0.117	0.374

5. ผลลัพธ์การทดสอบความเหมาะสมของโมเดลด้วย Hosmor and Lemeshow Test ของ Block 1

Block 1 หรือข้อมูลที่ผ่านการวิเคราะห์ตัวแปรเข้าสู่สมการแล้ว โดยผลลัพธ์การทดสอบความเหมาะสมของแบบจำลองด้วย Hosmor and Lemeshow Test ใช้ในการทดสอบความเหมาะสมของแบบจำลอง จากค่าสถิติทดสอบ (Chi-square) มีค่าเท่ากับ 12.481 และค่า Significance มีค่าเท่ากับ 0.013 ซึ่งน้อยกว่า 0.05 จึงสรุปได้ว่าแบบจำลองนี้เหมาะสม

6. ผลลัพธ์การตรวจสอบความน่าเชื่อถือของข้อมูล (Classification Table)

ตาราง ค.5 ตารางการจำแนกข้อมูลแสดงความน่าเชื่อถือข้อมูล

Observed			Predicted		
			Status		Percent correct
			accept	reject	
Step1	Status	accept	7	49	12.5
		reject	4	1148	99.7
	Overall Percentage				95.6

จากตารางที่ ค.5 การรีวิวโค้ดที่ผลลัพธ์การรีวิวเป็นผ่าน หรือยอมรับ (accept) มีทั้งหมด 56 รายการ มี 7 รายการที่พยากรณ์ผลการรีวิวถูกต้องหรือตรงตามความจริง นั่นคือพยากรณ์ถูกต้อง 12.5% และหากพิจารณาการรีวิวโค้ดที่ผลลัพธ์การรีวิวเป็นไม่ผ่าน หรือปฏิเสธ (reject) มีทั้งหมด 1,152 รายการ มีการพยาการณ์ผลการรีวิวถูกต้องหรือตรงตามความจริง นั่นคือพยากรณ์ถูกต้อง 99.7% โดยรวมผลลัพธ์การพยากรณ์ถูกต้องแล้ว 95.6 เปอร์เซ็นต์

7. ผลลัพธ์ตัวแปรที่อยู่ในสมการ (Variables in the Equation) ของ Block 1 รายละเอียดดังตารางที่ ค.6 เป็นการสรุปว่ามีตัวแปรใดบ้างที่อยู่ในสมการวิเคราะห์การถดถอยโลจิสติก รวมถึงค่าต่าง ๆ โดยค่า B หมายถึงสัมประสิทธิ์โลจิสติก สถิติการทดสอบวอลด์ (Wald Statistic) ค่านัยสำคัญ (Significant) ของการทดลอง และค่า Exp (B) ที่แปลความมาจากค่า B

ตารางที่ ค.6 ตัวแปรในสมการและสัมประสิทธิ์ตัวแปร

Step1	B	S.E.	Wald	df	Sig.	Exp (B)
commit	-.085	.088	.937	1	.333	.918
file	.036	.062	.329	1	.566	1.036
loc	.002	.001	4.261	1	.039	1.002
code chunk	-.031	.023	1.862	1	.172	.969
function	-.052	.027	3.779	1	.052	.949
method	-.052	.018	8.112	1	.004	.949
class	.028	.090	.098	1	.755	1.028
lloc	.016	.003	21.069	1	.000	1.016
sloc	-.005	.001	27.790	1	.000	.995

Step1	B	S.E.	Wald	df	Sig.	Exp (B)
comments	.007	.007	1.205	1	.272	1.007
multi	.008	.005	2.714	1	.099	1.008
blank	-.005	.005	1.084	1	.298	.995
halh1	-.053	.059	.785	1	.376	.949
halh2	.086	.031	7.722	1	.005	1.090
haln1	-.045	.043	1.105	1	.293	.956
haln2	.021	.022	.966	1	.326	1.021
hallength	-.009	.003	10.876	1	.001	.991
halvolume	-.001	.001	.896	1	.344	.999
haldiff	-.084	.112	.560	1	.454	.920
haleffort	.000	.000	2.632	1	.105	1.000
cc	-1255.242	432.556	8.421	1	.004	.000
ccl	.165	.092	3.206	1	.073	1.180
cco	.003	.013	.038	1	.846	1.003
ci	-.045	.080	.312	1	.577	.956
clc	1622.158	518.588	9.785	1	.002	.
ldc	-.502	.602	.694	1	.405	.605
lldc	.002	.011	.036	1	.850	1.002
lcom5	-.008	.014	.362	1	.547	.992
nl	-.038	.025	2.220	1	.136	.963
nle	-.007	.067	.012	1	.912	.993
wmc	.059	.083	.510	1	.475	1.061
cbo	.018	.007	6.083	1	.014	1.018
cboi	.225	.575	.153	1	.696	1.252
nii	-.001	.559	.000	1	.999	.999
noi	.076	.050	2.279	1	.131	1.079
rfc	.040	.050	.625	1	.429	1.040
cd	1.665	1.010	2.719	1	.099	5.286
cloc	.186	.166	1.256	1	.262	1.204

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Step1	B	S.E.	Wald	df	Sig.	Exp (B)
dloc	-.128	.356	.129	1	.720	.880
tcd	-.187	.551	.115	1	.735	.830
dit	.137	.357	.148	1	.700	1.147
noa	.199	3.355	.004	1	.953	1.220
noc	-12.294	1058.907	.000	1	.991	.000
nod	-1.021	.352	8.422	1	.004	.360
nop	12.733	1058.902	.000	1	.990	338807.948
na	-.018	.045	.162	1	.687	.982
nla	.097	.075	1.706	1	.191	1.102
nlm	-.100	.093	1.140	1	.286	.905
nm	.067	.030	5.203	1	.023	1.070
nos	-.052	.049	1.136	1	.287	.949
ttloc	-.006	.004	1.973	1	.160	.994
tloc	.008	.005	3.065	1	.080	1.008
tna	.007	.045	.027	1	.870	1.007
tnos	.017	.052	.108	1	.743	1.017
constant	.676	4.478	.023	1	.880	1.966

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ง

การวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้นโดยละเอียด

การวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น ทำได้ผ่านเครื่องมือเวก้า ซึ่งเป็นเครื่องมือในการวิเคราะห์ข้อมูลด้วยหลักการเหมืองข้อมูล และหลักการเรียนรู้ด้วยเครื่อง โดยการวิเคราะห์ผ่านคู่มือเวก้า โดยผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูล ประกอบด้วย 5 ส่วนด้วยกัน มีรายละเอียดดังต่อไปนี้

1. การแสดงข้อมูลการวิเคราะห์และการตั้งค่าการวิเคราะห์ รายละเอียดดังตารางที่ ง.1

ตารางที่ ง.1 ข้อมูลการวิเคราะห์และการตั้งค่าการวิเคราะห์

Scheme:	weka.classifiers.functions.MultilayerPerceptron -L 0.5 -M 0.3 -N 500 -V 0 -S 0 -E 20 -H 10
Relation:	remove zero import data=v2-weka.filters.unsupervised.attribute.Discretize-B4-M-1.0-R first-last-precision6
Instances:	1207
Attributes:	มีตัวแปรในการวิเคราะห์ 63 ตัวแปร ดังนี้ Count (commit#), File_changed, LOC change, Code chunk, Function, Method, Class, LLOC, SLOC, Comments, multi, blank, hal:h1, hal:h2, hal:N1, hal:N2, hal:vocabulary, hal:length, hal:calculated_length, hal:volume, hal:difficulty, hal:effort, hal:time, hal:bugs, CC, CCL, CCO, CI, CLC, LDC, LLDC, LCOM5, NL, NLE, WMC, CBO, CBOI, NII, NOI, RFC, CD, CLOC, DLOC, TCD, TCLOC, DIT, NOA, NOC, NOD, NOP, NA, NLA, NLM, NM, NOS, TLLOC, TLOC, TNA, TNLA, TNLM, TNM, TNOS, Result

2. การแสดงข้อมูลผลลัพธ์การวิเคราะห์ในโหนดที่ 0 และโหนดที่ 1 เพื่อใช้ในการสร้างสมการทำนายผลการรีวิวโค้ด รายละเอียดดังตารางที่ ง.2

ตารางที่ ง.2 ผลลัพธ์การวิเคราะห์โหนดที่ 0 และโหนดที่ 1

โหนดที่	การคำนวณค่าโหนด	ค่าน้ำหนักโหนด
Sigmoid Node 0	Threshold	-20.670318173024295
	Node 2	1.885396
	Node 3	1.78139
	Node 4	4.586735
	Node 5	3.835071
	Node 6	1.675371
	Node 7	3.259036
	Node 8	6.018528
	Node 9	2.152081
	Node 10	2.472659
	Node 11	2.089003
Sigmoid Node 1	Threshold	20.670318173024295
	Node 2	-1.885396
	Node 3	-1.78139
	Node 4	-4.586735
	Node 5	-3.835071
	Node 6	-1.675371
	Node 7	-3.259036
	Node 8	-6.018528
	Node 9	-2.152081
	Node 10	-2.472659
	Node 11	-2.089003

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3. การแสดงข้อมูลผลลัพธ์การวิเคราะห์ในโหนดที่ 2 และโหนดที่ 11 เพื่อใช้การคำนวณค่าในแต่ละโหนด และใช้ในการแทนค่าในสมการทำนาย รายละเอียดแต่ละโหนดรายละเอียดดังตารางที่ ง.3

ตารางที่ ง.3 ผลลัพธ์การวิเคราะห์โหนดที่ 2 และโหนดที่ 6

ค่าที่ใช้ในการวิเคราะห์	โหนด 2	โหนด 3	โหนด 4	โหนด 5	โหนด 6
Threshold	0.2980	0.4673	0.5470	0.2901	0.6692
commit='(-inf-8.25]'	-0.0916	0.7912	-2.5134	0.0548	1.4554
commit='(8.25-15.5]'	0.0360	1.0536	1.8564	-0.0843	-0.3667
commit='(15.5-22.75]'	-0.1926	-0.3820	-0.4710	-0.2433	-0.6036
commit='(22.75-inf)'	-0.2331	-2.4246	-0.0415	-0.3316	-1.8492
file='(-inf-17.25]'	-0.0735	0.5825	-2.4433	0.0526	0.5574
file='(17.25-33.5]'	-0.0248	1.0971	1.8657	-0.1397	0.1137
file='(33.5-49.75]'	-0.2564	-2.7302	-0.3261	-0.2964	-2.2901
file='(49.75-inf)'	-0.2187	0.0632	-0.1973	-0.2782	0.2683
loc='(-inf-722]'	0.5654	-5.7235	-1.8659	0.5427	-3.1233
loc='(722-1443]'	-1.4185	3.1048	-0.1553	-1.1978	1.2408
loc='(1443-2164]'	0.2739	0.0834	0.7328	0.2196	-0.1517
loc='(2164-inf)'	0.0125	1.5180	0.1566	-0.1836	0.7843
codechunk='(-inf-31.25]'	1.0926	2.7918	-0.3983	0.4021	2.2604
codechunk='(31.25-61.5]'	-1.2085	-0.5137	-0.6332	-0.5749	0.1458
codechunk='(61.5-91.75]'	-0.0902	-3.4672	-0.1873	-0.1858	-4.1740
codechunk='(91.75-inf)'	-0.2916	0.2496	0.0209	-0.3347	0.4940
function='(-inf-28.5]'	0.7506	1.3353	-2.0406	0.5574	-0.2128
function='(28.5-56]'	-0.7889	-2.0251	1.0780	-0.6230	-0.2108
function='(56-83.5]'	-0.1907	0.1579	0.3377	-0.2570	-0.2690
function='(83.5-inf)'	-0.2482	-0.4682	-0.4185	-0.3554	-0.6204
method='(-inf-41.5]'	2.1512	-2.9616	1.0469	1.6054	-0.7795
method='(41.5-82]'	-1.3391	1.5788	-1.4580	-0.9750	-0.1760
method='(82-122.5]'	-1.4334	-0.0301	-0.8782	-1.0771	-0.3166
method='(122.5-inf)'	0.1195	0.3830	0.1972	-0.1483	-0.0311
class='(-inf-12.25]'	-0.1809	-1.6947	-1.0412	-0.1058	-0.7313
class='(12.25-23.5]'	0.1514	1.0281	0.2645	0.0431	0.1271
class='(23.5-34.75]'	-0.2388	-0.3364	-0.2255	-0.2953	-0.4003

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 2	โหนด 3	โหนด 4	โหนด 5	โหนด 6
class='(34.75-inf)'	-0.2776	-0.0682	-0.2100	-0.2821	-0.2394
lloc='(-inf-760]'	0.8365	-2.9934	-2.8974	1.0551	-2.2225
lloc='(760-1509]'	-1.3701	1.9739	1.1192	-1.4005	1.2431
lloc='(1509-2258]'	0.1705	0.0725	0.7088	-0.0372	-0.0521
lloc='(2258-inf)'	-0.1856	-0.0616	-0.0985	-0.3285	-0.2071
sloc='(-inf-1333.5]'	0.7077	1.1778	-0.6561	1.7493	3.3177
sloc='(1333.5-2656]'	-0.5431	-2.3331	-0.7399	-1.5273	-4.1032
sloc='(2656-3978.5]'	-0.4294	0.2137	0.5420	-0.5207	-0.0685
sloc='(3978.5-inf)'	-0.2496	-0.0673	-0.2957	-0.2939	-0.4672
comments='(-inf-174]'	-0.9298	-2.3038	-3.0870	-0.9128	0.2435
comments='(174-347]'	0.8304	1.2740	2.5292	0.8478	-0.7466
comments='(347-520]'	-0.1361	0.4684	0.0052	-0.2558	-0.0909
comments='(520-inf)'	-0.3069	-0.5132	-0.5214	-0.3584	-0.6428
multi='(-inf-112.25]'	-0.4150	-4.4746	-4.3986	-0.4211	-3.1442
multi='(112.25-222.5]'	0.1415	3.5635	2.7685	0.0600	2.0625
multi='(222.5-332.75]'	-0.0272	0.0829	0.8617	-0.0290	0.2712
multi='(332.75-inf)'	-0.2048	-0.1874	-0.2774	-0.3274	-0.5349
blank='(-inf-250.5]'	0.1237	0.2034	-1.7530	0.1797	-0.1548
blank='(250.5-490]'	-0.0799	-0.1108	1.0055	-0.0601	-0.1467
blank='(490-729.5]'	-0.4795	-0.7756	-0.1410	-0.4478	-0.4774
blank='(729.5-inf)'	-0.2570	-0.3609	-0.2887	-0.3172	-0.4335
halh1='(-inf-31.25]'	0.7899	0.3795	-1.2624	0.5225	0.0126
halh1='(31.25-61.5]'	0.6719	-1.2353	0.9679	0.6213	-1.3457
halh1='(61.5-91.75]'	-1.7824	-1.1600	-0.9013	-1.5258	0.1076
halh1='(91.75-inf)'	-0.2118	1.0192	0.0443	-0.3261	-0.0318
halh2='(-inf-292]'	0.2112	0.0813	-0.1594	-0.0320	-0.9032
halh2='(292-582]'	-1.2436	-1.3962	-1.6462	-1.0744	0.2008
halh2='(582-872]'	0.7256	0.1606	0.6956	0.7509	-0.3716
halh2='(872-inf)'	-0.2103	0.2641	0.0824	-0.2760	-0.1502

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 2	โหนด 3	โหนด 4	โหนด 5	โหนด 6
haln1='(-inf-184.75]'	0.2273	0.1629	-0.2617	-0.0073	-0.9389
haln1='(184.75-368.5]'	-0.6634	-0.4552	-1.4681	-0.6620	0.3502
haln1='(368.5-552.25]'	0.1975	-0.9767	0.5639	0.3473	-0.6056
haln1='(552.25-inf)'	-0.2033	0.2555	-0.0092	-0.3110	-0.1492
haln2='(-inf-352]'	0.1866	0.0627	-0.2153	0.0160	-0.8902
haln2='(352-702]'	-1.2898	-1.2837	-1.6632	-1.0762	0.1826
haln2='(702-1052]'	0.7711	-0.0216	0.6315	0.6891	-0.3723
haln2='(1052-inf)'	-0.2459	0.1824	0.0135	-0.2762	-0.1699
halvocab='(-inf-307]'	0.1367	-0.0312	-0.8396	-0.0814	-1.0171
halvocab='(307-611]'	-0.8325	-0.8355	-1.2201	-0.3433	0.1324
halvocab='(611-915]'	0.3598	-0.3075	0.8464	0.0528	-0.2558
halvocab='(915-inf)'	-0.2595	0.2640	0.0263	-0.2774	-0.0864
hallength='(-inf-3408.65325]'	-0.1290	0.4828	-0.2848	0.3091	-0.5470
hallength='(3408.65325-6812.3065]'	-0.0275	0.0907	-0.3269	-0.5458	0.8841
hallength='(6812.3065-10215.95975]'	-0.0785	-1.3910	-0.4236	-0.1674	-1.2728
hallength='(10215.95975-inf)'	-0.2204	-0.1432	-0.2120	-0.2488	-0.3068
halcal='(-inf-2871.903]'	-0.0420	0.8981	0.0537	0.3973	0.2243
halcal='(2871.903-5741.806]'	-0.2322	-0.2849	-0.4368	-0.5751	0.1717
halcal='(5741.806-8611.709]'	-0.0609	-1.3969	-0.5493	-0.1780	-1.3314
halcal='(8611.709-inf)'	-0.2502	-0.2293	-0.2045	-0.3079	-0.3080
halvolume='(-inf-5353.774932]'	0.0267	2.4496	0.4627	0.2434	-0.2651
halvolume='(5353.774932-10703.745955]'	-0.1952	-1.9422	-0.6042	-0.3983	1.1643
halvolume='(10703.745955-16053.716978]'	-0.1065	-1.0948	-0.5846	-0.2611	-1.5723

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ค่าที่ใช้ในการวิเคราะห์	โหนด 2	โหนด 3	โหนด 4	โหนด 5	โหนด 6
halvolume='(16053.716978-inf)'	-0.2825	-0.4334	-0.4378	-0.3531	-0.6039
haldiff='(-inf-4.064822]'	-1.5206	5.0820	-2.4844	-1.2196	3.3071
haldiff='(4.064822-7.729645]'	0.8185	-6.4512	1.1211	1.1669	-6.0763
haldiff='(7.729645-11.394467]'	0.1367	-0.0957	-1.0834	-0.4400	0.0463
haldiff='(11.394467-inf)'	-0.1212	0.4618	1.3645	-0.1491	1.4179
haleffort='(-inf-79588.591466]'	0.4562	-0.0149	-0.2483	-0.1037	0.1397
haleffort='(79588.591466-159175.280977]'	-0.6103	-0.7338	-1.5534	0.0136	-2.0825
haleffort='(159175.280977-238761.970489]'	-0.0805	0.2325	1.1278	-0.2086	1.2724
haleffort='(238761.970489-inf)'	-0.2295	-0.4102	-0.4540	-0.3385	-0.6098
haltime='(-inf-4421.588498]'	0.4534	0.0173	-0.2441	-0.1421	0.1952
haltime='(4421.588498-8843.071332]'	-0.5874	-0.7493	-1.5845	-0.0641	-2.0928
haltime='(8843.071332-13264.554166]'	-0.1370	0.2258	1.1037	-0.2411	1.2875
haltime='(13264.554166-inf)'	-0.2744	-0.4573	-0.5005	-0.3196	-0.6382
halbug='(-inf-1.784592]'	0.0259	2.4154	0.5376	0.2228	-0.3408
halbug='(1.784592-3.567915]'	-0.2322	-1.9210	-0.5995	-0.3792	1.0899
halbug='(3.567915-5.351239]'	-0.1467	-1.0963	-0.5862	-0.2235	-1.5419
halbug='(5.351239-inf)'	-0.2560	-0.4269	-0.5053	-0.3449	-0.5547
cc='(-inf-0.025899]'	0.3851	-1.5479	-0.1095	0.1377	1.9926
cc='(0.025899-0.050073]'	-0.4402	0.5120	-1.2622	-0.5309	-3.0591
cc='(0.050073-0.074247]'	-0.7587	-0.8301	0.1468	-0.0910	-0.5565
cc='(0.074247-inf)'	0.2108	0.9660	0.1597	-0.1622	0.3542

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 2	โหนด 3	โหนด 4	โหนด 5	โหนด 6
ccl='(-inf-160.75]'	0.8001	3.5183	0.5990	0.6095	1.7970
ccl='(160.75-320.5]'	-0.9917	-4.2519	-0.7683	-0.5011	-2.2809
ccl='(320.5-480.25]'	-0.2341	0.0598	-0.2007	-0.2123	-0.3707
ccl='(480.25-inf)'	-0.1710	-0.3527	-0.7841	-0.5790	-0.3912
cco='(-inf-730.5]'	0.2726	1.5360	0.5483	-0.0771	0.3297
cco='(730.5-1447]'	-0.9178	-0.3997	1.1009	-0.2005	0.7018
cco='(1447-2163.5]'	0.2567	-2.1273	-2.1545	0.0212	-1.9609
cco='(2163.5-inf)'	-0.0671	-0.0422	-0.5507	-0.5103	-0.3127
ci='(-inf-193.75]'	0.2758	1.6203	0.5587	-0.0128	0.3398
ci='(193.75-385.5]'	-0.3701	-1.0248	0.2114	0.1160	-1.3549
ci='(385.5-577.25]'	-0.2591	-1.4559	-1.3285	-0.2976	0.0553
ci='(577.25-inf)'	-0.1016	-0.0596	-0.6252	-0.4555	-0.2415
clc='(-inf-0.022957]'	-0.9056	-2.2074	-1.3143	-1.1943	-1.3100
clc='(0.022957-0.044671]'	0.7633	1.1469	-0.0539	0.7931	0.1261
clc='(0.044671-0.066385]'	-0.5604	-0.0992	0.6135	-0.0806	-0.0841
clc='(0.066385-inf)'	0.1521	0.1936	-0.3252	-0.2145	0.0384
ldc='(-inf-16.153206]'	0.3922	-0.4474	-0.6678	0.0033	-0.3627
ldc='(16.153206-32.261967]'	-0.6272	0.0059	-0.0938	-0.0972	-1.3878
ldc='(32.261967-48.370729]'	-0.1855	-0.3825	0.3328	-0.0932	0.7905
ldc='(48.370729-inf)'	-0.1036	-0.1167	-0.6598	-0.5164	-0.3147
lldc='(-inf-2426.25]'	0.2556	-0.4707	-0.7584	-0.0398	-0.4197
lldc='(2426.25-4822.5]'	-0.3678	0.7491	1.0172	0.2051	-0.8202
lldc='(4822.5-7218.75]'	-0.1937	-0.2420	-0.3286	-0.2025	0.5456
lldc='(7218.75-inf)'	-0.2125	-1.0614	-1.0593	-0.6011	-0.5773
lcom5='(-inf-2054]'	0.2365	0.5792	0.0540	-0.0204	0.0480
lcom5='(2054-4082]'	-0.4072	-1.7149	-0.8126	0.1610	-0.8546
lcom5='(4082-6110]'	-0.0942	1.6513	0.7152	-0.0997	0.1267
lcom5='(6110-inf)'	-0.2171	-1.4686	-1.0276	-0.7036	-0.6409
nl='(-inf-484.25]'	0.2816	-3.0556	-1.5622	0.4710	-3.6509

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ค่าที่ใช้ในการวิเคราะห์	โหนด 2	โหนด 3	โหนด 4	โหนด 5	โหนด 6
nl='(484.25-927.5]'	-0.7973	4.2851	0.5590	-1.0303	2.7870
nl='(927.5-1370.75]'	0.0639	-2.2116	0.3135	0.3136	-0.1462
nl='(1370.75-inf)'	-0.0879	0.0100	-0.5096	-0.4852	-0.2553
nle='(-inf-449.25]'	-0.3439	-0.6802	-0.1659	-0.6438	-0.1123
nle='(449.25-870.5]'	0.2196	-0.1289	-1.9890	0.2615	-0.7605
nle='(870.5-1291.75]'	-0.2788	-0.1192	1.5923	0.1363	-0.2314
nle='(1291.75-inf)'	-0.1011	-0.0020	-0.4994	-0.4911	-0.2316
wmc='(-inf-384.75]'	-0.3895	-0.1507	0.4543	-0.6871	0.5635
wmc='(384.75-750.5]'	0.7591	0.5228	-1.1475	0.9151	-0.7031
wmc='(750.5-1116.25]'	-0.7982	-1.3093	0.1015	-0.5080	-0.9068
wmc='(1116.25-inf)'	-0.1215	-0.0569	-0.5020	-0.4149	-0.2829
cbo='(-inf-4510.75]'	0.7159	-1.6319	1.0010	0.7273	-1.5047
cbo='(4510.75-8618.5]'	-1.1682	1.5641	-1.9980	-1.2124	1.6555
cbo='(8618.5-12726.25]'	-0.0256	-0.8178	0.3568	0.2457	-1.0684
cbo='(12726.25-inf)'	-0.0838	-0.1173	-0.5662	-0.4852	-0.2714
cboi='(-inf-73.75]'	0.0640	-3.2538	-2.6769	-0.1775	-1.7596
cboi='(73.75-144.5]'	-0.2514	3.1365	-0.2416	-0.1605	0.2353
cboi='(144.5-215.25]'	0.3871	-0.8043	2.0802	0.3784	0.8391
cboi='(215.25-inf)'	-0.6788	-0.1402	-0.4007	-0.6728	-0.5683
nii='(-inf-73.5]'	0.1371	-3.1541	-2.6241	-0.1831	-1.7483
nii='(73.5-144]'	-0.4289	2.9328	-0.4334	-0.3733	0.0244
nii='(144-214.5]'	0.5101	-0.6896	2.2591	0.6017	0.8669
nii='(214.5-inf)'	-0.6705	-0.1702	-0.3612	-0.7034	-0.4836
noi='(-inf-85.75]'	0.2521	0.3838	-0.2285	-0.0751	-0.0007
noi='(85.75-171.5]'	-0.6081	0.3546	-0.2402	-0.4258	-0.5986
noi='(171.5-257.25]'	-0.0653	-1.7255	-0.2119	0.2131	-0.4098
noi='(257.25-inf)'	-0.1045	0.0227	-0.5305	-0.4842	-0.2263
rfc='(-inf-81.75]'	-0.4201	0.8697	1.2671	-0.8049	-1.5259
rfc='(81.75-162.5]'	-0.0294	-1.0463	-3.5031	0.1377	0.0825

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 2	โหนด 3	โหนด 4	โหนด 5	โหนด 6
rfc='(162.5-243.25]'	0.1112	0.7645	3.6208	0.0422	1.8914
rfc='(243.25-inf)'	-0.1313	-1.5392	-2.4777	-0.0755	-1.6948
cd='(-inf-6.423913]'	0.0053	-2.0086	-0.9386	0.8116	-1.1689
cd='(6.423913-8.282609]'	-2.2800	2.6437	2.0689	-2.2955	0.0086
cd='(8.282609-10.141304]'	1.9577	-1.2792	-1.9314	1.1417	0.3941
cd='(10.141304-inf)'	-0.2542	-0.2741	-0.2709	-0.3195	-0.5365
cloc='(-inf-21.475371]'	-0.4332	-0.2139	2.0935	-0.2537	0.9345
cloc='(21.475371-41.861888]'	0.2714	0.2833	-2.1442	-0.2402	-1.4931
cloc='(41.861888-62.248404]'	0.2835	-0.4749	0.1597	0.3173	0.3737
cloc='(62.248404-inf)'	-0.7308	-0.6390	-1.2211	-0.5108	-1.0641
dloc='(-inf-2549]'	-0.1035	-0.4048	0.9737	0.0732	1.3199
dloc='(2549-4873]'	0.3501	-0.6766	-1.1420	0.0127	-1.9575
dloc='(4873-7197]'	-0.6096	-0.5506	-1.1810	-0.1834	-0.7689
dloc='(7197-inf)'	-0.0614	0.5862	0.2309	-0.4161	0.0861
tcd='(-inf-9.895292]'	0.2760	2.9105	-0.6923	0.3834	3.3739
tcd='(9.895292-12.930195]'	-1.8936	-3.3178	-0.3723	-1.2321	0.2740
tcd='(12.930195-15.965097]'	1.3057	-0.1306	0.3286	0.5171	-4.2597
tcd='(15.965097-inf)'	-0.2941	-0.4419	-0.4491	-0.2880	-0.6384
tcloc='(-inf-21.481224]'	-0.3909	-0.2276	2.0886	-0.2812	0.9825
tcloc='(21.481224-41.873593]'	0.2783	0.2819	-2.1422	-0.1818	-1.4687
tcloc='(41.873593-62.265961]'	0.3385	-0.5113	0.2296	0.3551	0.3534
tcloc='(62.265961-inf)'	-0.7334	-0.5745	-1.2168	-0.5849	-1.0531
dit='(-inf-2550.5]'	-0.0825	-0.3602	0.9782	0.0306	1.3839
dit='(2550.5-4876]'	0.3091	-0.7524	-1.2055	0.0110	-2.0303
dit='(4876-7201.5]'	-0.6509	-0.5474	-1.1954	-0.2067	-0.7934
dit='(7201.5-inf)'	-0.0894	0.5805	0.2401	-0.4375	0.0624
noa='(-inf-54.5]'	0.1409	1.0113	-1.1240	-0.1719	0.0968
noa='(54.5-108]'	-0.1261	-1.5739	0.0339	0.0824	-1.7784
noa='(108-161.5]'	-0.0674	-1.0439	-1.1232	0.3108	-0.0720

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ค่าที่ใช้ในการวิเคราะห์	โหนด 2	โหนด 3	โหนด 4	โหนด 5	โหนด 6
noa='(161.5-inf)'	-0.5208	0.5569	1.0207	-0.8794	0.4460
noc='(-inf-55]'	0.1436	1.0017	-1.1137	-0.2057	0.0229
noc='(55-109]'	-0.1903	-1.5918	0.0313	0.0984	-1.7459
noc='(109-163]'	-0.0568	-0.9993	-1.0638	0.2897	-0.0139
noc='(163-inf)'	-0.5116	0.6293	1.0188	-0.9181	0.4087
nod='(-inf-42.75]'	0.1127	0.6494	-1.0313	-0.2231	0.1900
nod='(42.75-84.5]'	0.1012	-0.6184	-0.8702	0.2847	-1.5741
nod='(84.5-126.25]'	-0.6299	0.8401	2.4321	-0.4173	1.4787
nod='(126.25-inf)'	-0.1468	-1.9356	-1.6029	-0.3935	-1.4669
nop='(-inf-55]'	-0.0085	0.9731	-1.3018	-0.2064	0.0298
nop='(55-109]'	-0.1518	-1.5449	0.1713	0.1512	-1.7245
nop='(109-163]'	-0.1299	-1.1153	-1.2555	0.1445	-0.0984
nop='(163-inf)'	-0.2810	0.6928	1.2297	-0.7380	0.4867
na='(-inf-20595.5]'	-0.5271	-0.3930	0.6023	-0.3947	0.5101
na='(20595.5-39351]'	0.3116	1.1473	-0.1165	-0.0179	0.1428
na='(39351-58106.5]'	-0.2158	-1.7912	-0.9693	0.1943	-1.5732
na='(58106.5-inf)'	-0.0625	-0.0702	-0.5306	-0.5048	-0.2750
nla='(-inf-211]'	0.0947	1.0968	-1.0981	-0.2350	-0.3149
nla='(211-409]'	0.1747	-1.0432	-0.7815	0.3534	-0.9140
nla='(409-607]'	-0.9350	-1.8070	0.5078	-0.9844	0.4557
nla='(607-inf)'	0.1603	0.7060	0.1793	0.1781	-0.4667
nlm='(-inf-160.5]'	-0.7072	-2.0371	0.6323	-1.2438	2.4947
nlm='(160.5-308]'	1.0827	4.6845	-0.3769	1.2430	-0.5034
nlm='(308-455.5]'	-0.5720	-3.4720	-0.8852	-0.7849	-2.3375
nlm='(455.5-inf)'	-0.3940	-0.0962	-0.4613	0.1468	-0.9138
nm='(-inf-1105.5]'	-0.1716	0.8177	-0.0950	0.1626	-0.1899
nm='(1105.5-2135]'	-0.1346	-1.0386	-0.3580	-0.4688	0.3963
nm='(2135-3164.5]'	-0.2099	-0.6898	-0.1506	0.1542	-1.1793
nm='(3164.5-inf)'	-0.1071	-0.0775	-0.5250	-0.4972	-0.3149

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 2	โหนด 3	โหนด 4	โหนด 5	โหนด 6
nos='(-inf-1476.25]'	0.0647	-0.1139	-1.0031	-0.1868	-0.1433
nos='(1476.25-2876.5]'	-0.7075	-1.2808	0.0270	-0.4907	-0.1124
nos='(2876.5-4276.75]'	0.1576	0.8398	0.5265	0.4667	-0.5906
nos='(4276.75-inf)'	-0.1600	-0.4303	-0.5799	-0.4803	-0.3934
tlloc='(-inf-10568.25]'	-0.4930	-0.1468	0.5097	-0.2721	0.1261
tlloc='(10568.25-20159.5]'	0.2414	0.7962	-0.5703	-0.1063	0.3725
tlloc='(20159.5-29750.75]'	-0.1970	-1.4983	-0.5065	0.2259	-1.4534
tlloc='(29750.75-inf)'	-0.1206	-0.0221	-0.5538	-0.5087	-0.2767
tloc='(-inf-16271.75]'	0.1184	0.8632	0.1456	0.4162	0.7459
tloc='(16271.75-31118.5]'	-1.6264	-1.1760	-0.6215	-1.4563	-0.6559
tloc='(31118.5-45965.25]'	1.0555	-0.6087	-0.0707	0.8524	-1.0919
tloc='(45965.25-inf)'	-0.0650	-0.0887	-0.5861	-0.5028	-0.3066
tna='(-inf-20595.5]'	-0.6052	-0.3945	0.5579	-0.3450	0.4179
tna='(20595.5-39351]'	0.3111	1.1666	-0.1328	-0.0492	0.1486
tna='(39351-58106.5]'	-0.1836	-1.7430	-1.0488	0.2026	-1.6413
tna='(58106.5-inf)'	-0.1403	-0.0227	-0.5705	-0.4241	-0.2774
tnla='(-inf-211.25]'	0.1106	1.1521	-1.1096	-0.2017	-0.3375
tnla='(211.25-409.5]'	0.1629	-1.0234	-0.7827	0.3167	-1.0068
tnla='(409.5-607.75]'	-0.9423	-1.8212	0.4187	-1.0003	0.4408
tnla='(607.75-inf)'	0.1986	0.7885	0.2566	0.0901	-0.4329
tnlm='(-inf-160.75]'	0.3245	-1.9692	0.6032	1.0318	2.1382
tnlm='(160.75-308.5]'	-0.0018	4.5855	-0.3595	-1.0611	-0.1297
tnlm='(308.5-456.25]'	-0.5332	-3.4586	-0.9291	-0.7753	-2.3675
tnlm='(456.25-inf)'	-0.3928	-0.0844	-0.4730	0.1577	-0.9357
tnm='(-inf-1105.75]'	-0.1163	0.8298	-0.1160	0.1818	-0.1442
tnm='(1105.75-2135.5]'	-0.2335	-1.0514	-0.6790	-0.6330	0.2320
tnm='(2135.5-3165.25]'	-0.1186	-0.7137	0.2880	0.2304	-1.1031
tnm='(3165.25-inf)'	-0.0707	-0.0810	-0.5509	-0.4675	-0.2556
tnos='(-inf-1484.75]'	0.0655	-0.3051	-1.0620	-0.2014	-0.2322

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 2	โหนด 3	โหนด 4	โหนด 5	โหนด 6
tnos='(1484.75-2892.5]'	-0.6773	-1.1744	-0.2322	-0.5031	-0.1240
tnos='(2892.5-4300.25]'	0.1782	0.8674	0.7746	0.4044	-0.5990
tnos='(4300.25-inf)'	-0.0468	-0.4058	-0.6157	-0.4590	-0.3613

ตารางที่ ง.4 ผลลัพธ์การวิเคราะห์โหนดที่ 7 และโหนดที่ 11

ค่าที่ใช้ในการวิเคราะห์	โหนด 7	โหนด 8	โหนด 9	โหนด 10	โหนด 11
Threshold	0.5448	0.6012	0.3783	0.1572	0.0537
commit='(-inf-8.25]'	0.2666	-1.0774	0.2269	-0.1078	-0.6356
commit='(8.25-15.5]'	-0.3768	0.5292	-0.2544	0.0171	0.6391
commit='(15.5-22.75]'	-0.4747	-0.4278	-0.3561	-0.0236	0.0083
commit='(22.75-inf)'	-0.5470	-0.3444	-0.3739	-0.0530	-0.2182
file='(-inf-17.25]'	0.0010	-1.0490	0.2492	-0.3318	-0.8830
file='(17.25-33.5]'	-0.1520	0.4601	-0.2633	0.2511	0.8279
file='(33.5-49.75]'	-0.5517	-0.1798	-0.3376	-0.0704	-0.3834
file='(49.75-inf)'	-0.5997	-0.4880	-0.3434	-0.1205	0.1924
loc='(-inf-722]'	-0.5069	-3.1087	-0.5292	0.1176	-4.8198
loc='(722-1443]'	-0.0097	1.7365	-0.2892	-1.1063	3.1722
loc='(1443-2164]'	-0.3728	0.2100	0.2060	0.7245	0.6076
loc='(2164-inf)'	-0.3587	-0.0425	-0.0292	-0.0464	0.9187
codechunk='(-inf-31.25]'	1.1078	0.1177	-0.6985	1.1621	-0.5332
codechunk='(31.25-61.5]'	-1.2611	-0.6962	0.6450	-1.2207	1.1575
codechunk='(61.5-91.75]'	-0.4331	-0.3975	-0.3168	-0.1275	-1.2860
codechunk='(91.75-inf)'	-0.5779	-0.3166	-0.3148	-0.0622	0.4757
function='(-inf-28.5]'	1.7271	3.1532	-0.0767	0.1810	-0.3064
function='(28.5-56]'	-1.9868	-4.0093	-0.0625	-0.3985	-0.1247
function='(56-83.5]'	-0.2409	0.1057	-0.3347	-0.0188	0.2781
function='(83.5-inf)'	-0.5646	-0.5632	-0.3804	-0.1317	0.0214
method='(-inf-41.5]'	-0.5892	-0.9500	1.1876	1.7316	-0.2933
method='(41.5-82]'	0.1195	-0.1882	-1.6742	-1.3306	-1.1451

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 7	โหนด 8	โหนด 9	โหนด 10	โหนด 11
method='(82-122.5]'	-0.4725	-0.4141	-0.3269	-0.8672	0.4266
method='(122.5-inf)'	-0.2662	0.3251	0.0495	0.3342	0.9417
class='(-inf-12.25]'	-1.1181	-1.6433	0.0260	-0.6909	-1.5718
class='(12.25-23.5]'	0.7970	1.3418	-0.1822	0.3542	1.3119
class='(23.5-34.75]'	-0.4118	-0.4179	-0.3854	-0.0276	0.0909
class='(34.75-inf)'	-0.4786	-0.3983	-0.3020	0.1226	0.0918
lloc='(-inf-760]'	-3.0111	-3.2977	-1.6954	0.7120	-1.6942
lloc='(760-1509]'	2.0979	3.9786	1.2177	-1.6409	1.3716
lloc='(1509-2258]'	0.1841	-1.9049	0.0573	0.5159	-0.0485
lloc='(2258-inf)'	-0.3876	-0.0262	-0.2499	0.1939	0.2707
sloc='(-inf-1333.5]'	-0.7942	0.3016	-0.5284	1.5756	3.8034
sloc='(1333.5-2656]'	0.9661	-0.9497	1.2458	-1.1928	-4.6954
sloc='(2656-3978.5]'	-0.9161	-0.3998	-1.0247	-0.6094	0.7033
sloc='(3978.5-inf)'	-0.5128	-0.2184	-0.3389	0.0042	0.0519
comments='(-inf-174]'	0.0174	0.4984	-0.8547	-2.0660	-0.5884
comments='(174-347]'	-0.4723	-1.1726	0.7710	1.6373	0.0957
comments='(347-520]'	-0.1379	0.0256	-0.2672	0.2408	0.4571
comments='(520-inf)'	-0.5865	-0.6135	-0.3895	-0.1299	-0.0515
multi='(-inf-112.25]'	-1.4950	-3.6140	-0.1632	-2.1736	-2.8328
multi='(112.25-222.5]'	0.5435	1.8526	-0.1155	1.5436	1.9536
multi='(222.5-332.75]'	0.1483	0.8703	-0.1344	0.4007	0.6591
multi='(332.75-inf)'	-0.3745	-0.2994	-0.3181	0.0093	0.0341
blank='(-inf-250.5]'	0.7717	1.5158	-0.2732	1.4578	0.8008
blank='(250.5-490]'	-0.9876	-0.2538	0.8195	-0.4178	-1.2409
blank='(490-729.5]'	-0.4714	-2.0674	-0.9298	-1.3286	0.0945
blank='(729.5-inf)'	-0.5071	-0.4255	-0.2930	-0.0339	0.1261
halh1='(-inf-31.25]'	0.8539	1.6268	0.5109	0.8930	0.0384
halh1='(31.25-61.5]'	-1.1130	-2.9309	0.1679	0.6798	-1.2364
halh1='(61.5-91.75]'	-0.9810	-0.6401	-1.0942	-1.9569	0.4306

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 7	โหนด 8	โหนด 9	โหนด 10	โหนด 11
halh1='(91.75-inf)'	0.0992	0.7780	-0.3260	0.0527	0.5383
halh2='(-inf-292]'	-0.3112	0.2222	0.1330	-0.0667	-0.9208
halh2='(292-582]'	-0.9765	-0.7200	-1.2493	-0.7134	0.8394
halh2='(582-872]'	0.3112	-1.1828	0.6700	0.3009	-0.3264
halh2='(872-inf)'	-0.2161	0.4686	-0.2670	0.2555	0.3988
haln1='(-inf-184.75]'	-0.3145	0.1993	0.0949	-0.0414	-0.9363
haln1='(184.75-368.5]'	-0.7109	-0.5464	-0.5620	0.0791	1.0283
haln1='(368.5-552.25]'	0.0476	-1.1611	0.1141	-0.5075	-0.5680
haln1='(552.25-inf)'	-0.2392	0.2699	-0.2816	0.2268	0.3721
haln2='(-inf-352]'	-0.2638	0.2964	0.1369	-0.0280	-1.0079
haln2='(352-702]'	-1.0213	-0.8441	-1.1585	-0.7145	0.8004
haln2='(702-1052]'	0.2967	-0.9346	0.6711	0.2967	-0.1876
haln2='(1052-inf)'	-0.1947	0.2552	-0.2316	0.1918	0.1893
halvocab='(-inf-307]'	-0.4008	0.0566	0.0309	-0.1701	-1.1096
halvocab='(307-611]'	-1.2889	-1.2174	-0.7109	0.3112	1.1131
halvocab='(611-915]'	0.7208	-0.5289	0.1599	-0.5923	-0.5140
halvocab='(915-inf)'	-0.2435	0.4834	-0.2023	0.1817	0.3490
hallength='(-inf-3408.65325]'	1.3607	1.6987	1.8101	0.2532	0.3650
hallength='(3408.65325-6812.3065]'	-1.9613	-0.9213	-2.0395	0.6333	0.8041
hallength='(6812.3065-10215.95975]'	-0.1637	-1.8099	-0.2844	-1.2528	-1.4475
hallength='(10215.95975-inf)'	-0.3597	-0.2727	-0.2609	0.0966	0.1885
halcal='(-inf-2871.903]'	1.5192	1.9950	1.8190	0.5435	0.6396
halcal='(2871.903-5741.806]'	-2.2185	-1.1368	-2.0067	0.2840	0.5750

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 7	โหนด 8	โหนด 9	โหนด 10	โหนด 11
halcal='(5741.806-8611.709]'	-0.2690	-1.8885	-0.2268	-1.2809	-1.5287
halcal='(8611.709-inf)'	-0.3586	-0.2396	-0.2933	0.1243	0.1969
halvolume='(-inf-5353.774932]'	1.8248	1.7840	1.0234	-0.3346	-0.8597
halvolume='(5353.774932-10703.745955]'	-2.1556	-2.9468	-1.1152	1.7200	1.5206
halvolume='(10703.745955-16053.716978]'	-0.3365	0.3922	-0.3049	-1.5880	-0.7253
halvolume='(16053.716978-inf)'	-0.4502	-0.6136	-0.3078	-0.0667	0.0306
haldiff='(-inf-4.064822]'	-1.2872	-0.4013	-1.1520	-1.2182	4.3875
haldiff='(4.064822-7.729645]'	-0.2765	-2.1259	0.7184	0.4304	-5.4659
haldiff='(7.729645-11.394467]'	0.6107	0.5145	-0.0237	-1.2704	-0.1863
haldiff='(11.394467-inf)'	-0.2812	0.7765	-0.1629	1.8041	1.0883
haleffort='(-inf-79588.591466]'	-0.9339	-0.9559	-0.8993	-0.0073	-0.4120
haleffort='(79588.591466-159175.280977]'	0.5607	-0.4229	0.7213	-1.8270	-0.7131
haleffort='(159175.280977-238761.970489]'	-0.3234	0.6621	-0.2779	1.6158	1.0438
haleffort='(238761.970489-inf)'	-0.5291	-0.5567	-0.3110	-0.1276	-0.0600
haltime='(-inf-4421.588498]'	-0.8456	-0.9247	-0.8169	0.0053	-0.3909
haltime='(4421.588498-8843.071332]'	0.5956	-0.3555	0.7174	-1.8347	-0.7359

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 7	โหนด 8	โหนด 9	โหนด 10	โหนด 11
haltime='(8843.071332-13264.554166]'	-0.3525	0.6628	-0.2343	1.6254	1.0299
haltime='(13264.554166-inf)'	-0.4836	-0.5709	-0.3176	-0.1335	-0.0369
halbug='(-inf-1.784592]'	1.7803	1.8319	1.0219	-0.3550	-0.8464
halbug='(1.784592-3.567915]'	-2.1036	-2.9403	-1.1400	1.7591	1.4535
halbug='(3.567915-5.351239]'	-0.3437	0.4114	-0.3405	-1.6091	-0.7054
halbug='(5.351239-inf)'	-0.5452	-0.5800	-0.3396	-0.0738	-0.0585
cc='(-inf-0.025899]'	-0.4465	0.0484	-0.5397	-1.3403	-0.5175
cc='(0.025899-0.050073]'	-0.5976	-0.7814	0.3568	0.8012	0.8355
cc='(0.050073-0.074247]'	-0.4674	-1.1737	-0.7799	-0.0235	-1.1960
cc='(0.074247-inf)'	0.3720	0.6933	0.2761	0.2465	0.8325
ccl='(-inf-160.75]'	0.6246	2.0342	0.8719	0.4464	0.3196
ccl='(160.75-320.5]'	-0.9863	-2.7087	-1.1638	-0.6035	-0.8163
ccl='(320.5-480.25]'	-0.3202	-0.2771	-0.2366	-0.0577	0.7058
ccl='(480.25-inf)'	-0.4246	-0.3270	-0.2029	-0.0428	-0.3594
cco='(-inf-730.5]'	0.0418	0.5381	0.2765	-0.2915	0.6884
cco='(730.5-1447]'	-0.8667	-0.7328	-1.0243	-0.3050	-1.0734
cco='(1447-2163.5]'	-0.1918	-1.0674	0.0681	0.2781	0.1635
cco='(2163.5-inf)'	-0.1917	-0.0306	-0.0570	0.0302	0.1463
ci='(-inf-193.75]'	0.0317	0.5392	0.2686	-0.2142	0.6955
ci='(193.75-385.5]'	-0.4408	-1.3394	-0.6141	0.1137	-0.3293
ci='(385.5-577.25]'	-0.5431	-0.4368	-0.3891	-0.0895	-0.5822
ci='(577.25-inf)'	-0.2381	-0.0491	-0.1184	0.0461	0.1416
clc='(-inf-0.022957]'	-0.8314	-0.8072	-1.0012	-2.6452	-0.6799
clc='(0.022957-0.044671]'	-0.2708	0.0410	0.6818	2.1186	0.9396
clc='(0.044671-0.066385]'	-0.1091	-0.6930	-0.4842	0.0160	-0.9903

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 7	โหนด 8	โหนด 9	โหนด 10	โหนด 11
clc='(0.066385-inf)'	-0.0255	0.2055	0.0361	0.2414	0.5176
ldc='(-inf-16.153206]'	-0.1843	0.1828	0.4214	-0.1659	0.4010
ldc='(16.153206-32.261967]'	-0.3683	-1.9599	-0.7194	-0.4827	-0.8052
ldc='(32.261967-48.370729]'	-0.4114	0.4986	-0.2616	0.3110	0.1843
ldc='(48.370729-inf)'	-0.1466	0.0399	-0.0964	-0.0321	0.1243
lldc='(-inf-2426.25]'	-0.4249	-0.0815	0.1478	-0.3587	0.3666
lldc='(2426.25-4822.5]'	-0.0057	-0.7073	-0.3678	0.1915	-0.0984
lldc='(4822.5-7218.75]'	-0.2919	0.1812	-0.1895	0.1015	0.5080
lldc='(7218.75-inf)'	-0.4753	-0.5720	-0.2185	-0.2155	-0.8446
lcom5='(-inf-2054]'	-0.1114	0.3092	0.2927	-0.3137	0.5591
lcom5='(2054-4082]'	-0.3671	-1.0513	-0.4930	0.1776	-1.2552
lcom5='(4082-6110]'	-0.1228	0.2190	-0.1839	0.1824	1.6529
lcom5='(6110-inf)'	-0.6109	-0.7501	-0.3081	-0.3161	-1.1555
nl='(-inf-484.25]'	-0.7660	-1.9823	-0.2464	-0.6210	-1.5313
nl='(484.25-927.5]'	-0.1479	1.3243	-0.6994	-0.6234	1.4179
nl='(927.5-1370.75]'	-0.0646	-0.6157	0.2728	0.9310	-0.2706
nl='(1370.75-inf)'	-0.1356	-0.0103	-0.1032	0.0333	0.2472
nle='(-inf-449.25]'	-0.7737	-0.3135	-0.6228	-1.7261	0.9195
nle='(449.25-870.5]'	-1.1787	-2.1743	-0.0229	0.6753	-1.1350
nle='(870.5-1291.75]'	1.0289	1.3209	-0.0478	0.8011	-0.2025
nle='(1291.75-inf)'	-0.1738	0.0134	-0.0300	-0.0024	0.2534
wmc='(-inf-384.75]'	-0.7904	-0.1605	-0.6527	-1.9653	1.3769
wmc='(384.75-750.5]'	-0.3442	-1.0766	0.4539	2.0629	-1.2443
wmc='(750.5-1116.25]'	0.1564	0.0386	-0.5275	-0.4040	-0.5149
wmc='(1116.25-inf)'	-0.2243	-0.0447	-0.0790	-0.0113	0.2444
cbo='(-inf-4510.75]'	-0.6844	-1.0703	-0.0737	0.5658	-0.6453
cbo='(4510.75-8618.5]'	-0.4111	-0.8356	-0.6764	-1.2668	-0.2588
cbo='(8618.5-12726.25]'	0.1185	0.7199	0.0649	0.4058	0.5923
cbo='(12726.25-inf)'	-0.2288	0.0278	-0.0849	-0.0069	0.2901

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 7	โหนด 8	โหนด 9	โหนด 10	โหนด 11
cboi='(-inf-73.75]'	-0.2404	-1.5586	-0.2582	-2.7591	-0.8718
cboi='(73.75-144.5]'	-0.5334	0.4901	-0.3841	1.6073	0.6843
cboi='(144.5-215.25]'	-0.3989	-0.2783	0.4959	1.1158	0.4115
cboi='(215.25-inf)'	-0.0702	0.1552	-0.6194	-0.3033	-0.4543
nii='(-inf-73.5]'	-0.2701	-1.5974	-0.2415	-2.6064	-0.7509
nii='(73.5-144]'	-0.5881	0.2774	-0.6240	1.3191	0.5454
nii='(144-214.5]'	-0.2638	-0.1316	0.7693	1.2650	0.5246
nii='(214.5-inf)'	-0.0984	0.2198	-0.6631	-0.3024	-0.3895
noi='(-inf-85.75]'	-0.1278	0.2486	0.2177	-0.3267	0.5585
noi='(85.75-171.5]'	0.7653	-0.0805	-0.5295	-0.2944	-1.0103
noi='(171.5-257.25]'	-1.6912	-1.4323	-0.3054	0.3822	0.0319
noi='(257.25-inf)'	-0.2153	-0.0128	-0.0709	0.0200	0.3232
rfc='(-inf-81.75]'	-0.8511	-0.6011	-0.6581	-1.8069	0.4955
rfc='(81.75-162.5]'	-0.2445	-0.5537	-0.4766	0.3517	-0.7807
rfc='(162.5-243.25]'	0.2687	0.6626	0.8274	0.8431	0.5425
rfc='(243.25-inf)'	-0.3681	-0.6028	-0.4631	0.3910	-0.3387
cd='(-inf-6.423913]'	-4.0928	-3.7638	-1.1846	1.4786	-0.8479
cd='(6.423913-8.282609]'	2.5685	2.9402	-2.9936	-2.3426	1.2258
cd='(8.282609-10.141304]'	1.0031	0.1193	3.8940	0.7246	-0.5568
cd='(10.141304-inf)'	-0.5622	-0.4723	-0.3188	-0.0895	0.0745
cloc='(-inf-21.475371]'	-1.3994	-0.5759	-1.4965	-0.3251	0.0649
cloc='(21.475371-41.861888]'	-0.6235	-1.8332	1.0913	0.1711	0.0841
cloc='(41.861888-62.248404]'	1.1233	0.5779	0.5600	0.5266	-0.1948
cloc='(62.248404-inf)'	-0.3070	0.5143	-0.8348	-0.6746	-0.1219
dloc='(-inf-2549]'	-1.0677	-0.0367	-0.5732	0.6854	0.9724
dloc='(2549-4873]'	1.5755	0.5331	0.5007	-0.6141	-0.8738
dloc='(4873-7197]'	-1.4498	-1.8248	-0.5765	-0.3404	-0.9783

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ค่าที่ใช้ในการวิเคราะห์	โหนด 7	โหนด 8	โหนด 9	โหนด 10	โหนด 11
dloc='(7197-inf)'	-0.1544	0.1533	-0.1097	0.0228	0.6410
tcd='(-inf-9.895292]'	-1.7629	-0.7287	-0.3899	1.4406	0.6834
tcd='(9.895292-12.930195]'	1.0236	-0.1912	-1.6320	-1.8103	1.0285
tcd='(12.930195-15.965097]'	0.2003	0.2154	1.6486	0.2623	-1.8439
tcd='(15.965097-inf)'	-0.5908	-0.5432	-0.3309	-0.1292	0.0030
tcloc='(-inf-21.481224]'	-1.4641	-0.5538	-1.4389	-0.3118	0.0847
tcloc='(21.481224-41.873593]'	-0.6466	-1.8407	1.0754	0.2172	0.1318
tcloc='(41.873593-62.265961]'	1.1551	0.5463	0.4847	0.5584	-0.2339
tcloc='(62.265961-inf)'	-0.2318	0.5456	-0.8617	-0.6583	-0.0920
dit='(-inf-2550.5]'	-1.0412	-0.0881	-0.5860	0.6112	1.0120
dit='(2550.5-4876]'	1.5791	0.5297	0.4729	-0.6124	-0.8576
dit='(4876-7201.5]'	-1.4921	-1.9034	-0.5917	-0.2728	-0.8978
dit='(7201.5-inf)'	-0.1817	0.1492	-0.0359	0.0360	0.6728
noa='(-inf-54.5]'	-0.0461	-0.1733	-0.1698	-0.3435	1.2738
noa='(54.5-108]'	1.5034	1.1908	0.7151	-0.6137	-0.3863
noa='(108-161.5]'	-1.9293	-2.6587	-0.4489	1.3639	-0.4968
noa='(161.5-inf)'	-0.7189	0.5076	-0.7804	-0.6284	-0.5214
noc='(-inf-55]'	-0.0715	-0.2226	-0.1823	-0.4096	1.3234
noc='(55-109]'	1.5200	1.2016	0.6571	-0.6078	-0.4501
noc='(109-163]'	-1.9145	-2.6739	-0.4299	1.4087	-0.4941
noc='(163-inf)'	-0.7489	0.4284	-0.7530	-0.6748	-0.5745
nod='(-inf-42.75]'	-0.1618	-0.8458	-0.1804	-0.4520	1.4567
nod='(42.75-84.5]'	0.4816	0.5710	0.2240	0.4075	-1.2930
nod='(84.5-126.25]'	-1.0208	-0.3924	-0.3437	-0.4822	0.9946
nod='(126.25-inf)'	-0.5030	-0.5807	-0.3905	0.2122	-1.3342
nop='(-inf-55]'	-0.1057	-0.2899	-0.2214	-0.5347	1.2700

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 7	โหนด 8	โหนด 9	โหนด 10	โหนด 11
nop='(55-109]'	1.5610	1.1317	0.6925	-0.4770	-0.3934
nop='(109-163]'	-2.2022	-2.8804	-0.6284	1.3039	-0.5636
nop='(163-inf)'	-0.4733	0.7289	-0.5268	-0.5639	-0.3822
na='(-inf-20595.5]'	-0.8003	-0.4959	-0.1041	-0.7177	0.9031
na='(20595.5-39351]'	-0.1109	-0.8399	-0.6182	0.1855	-1.5087
na='(39351-58106.5]'	-0.0763	0.1119	0.0796	0.2698	0.2460
na='(58106.5-inf)'	-0.1325	-0.0271	-0.0965	0.0150	0.2560
nla='(-inf-211]'	-0.2777	-0.7542	-0.2122	-0.9534	0.1022
nla='(211-409]'	0.5881	0.6377	0.2206	0.9412	0.1608
nla='(409-607]'	-1.7358	-2.2221	-0.7318	-0.7676	-1.0738
nla='(607-inf)'	0.2864	1.1743	-0.0188	0.5642	0.6521
nlm='(-inf-160.5]'	-1.9476	1.6160	-2.2080	-2.3024	1.7053
nlm='(160.5-308]'	0.6808	-2.3602	1.5304	3.1938	-0.8548
nlm='(308-455.5]'	-0.0222	-0.2362	-0.0557	-2.2383	-1.8159
nlm='(455.5-inf)'	0.2059	-0.2765	0.0354	1.0772	0.8072
nm='(-inf-1105.5]'	-0.6928	-0.7534	-0.0100	-1.4100	-0.7724
nm='(1105.5-2135]'	-0.2983	-0.7804	-0.1940	1.1793	0.2778
nm='(2135-3164.5]'	-0.0533	0.2875	-0.3874	-0.0325	0.1159
nm='(3164.5-inf)'	-0.1386	0.0341	-0.0958	-0.0079	0.2331
nos='(-inf-1476.25]'	-0.1857	-0.8736	-0.2226	-0.3851	1.4368
nos='(1476.25-2876.5]'	-0.9945	-1.3249	-0.3866	-0.4247	-2.2456
nos='(2876.5-4276.75]'	0.4349	1.1321	0.1157	0.4895	0.6045
nos='(4276.75-inf)'	-0.4217	-0.2640	-0.1382	-0.0051	0.1598
tlloc='(-inf-10568.25]'	-0.7228	-0.6539	-0.0942	-0.5996	-0.6349
tlloc='(10568.25-20159.5]'	-0.1921	-0.8374	-0.6297	0.0259	-0.0436
tlloc='(20159.5-29750.75]'	-0.0865	0.2841	0.0329	0.2782	0.2852
tlloc='(29750.75-inf)'	-0.2036	0.0143	-0.0974	-0.0061	0.2831
tloc='(-inf-16271.75]'	-0.6125	0.5241	0.1461	-0.6837	0.0872
tloc='(16271.75-31118.5]'	-1.0179	-2.0315	-1.6145	-1.0095	-1.5686

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าที่ใช้ในการวิเคราะห์	โหนด 7	โหนด 8	โหนด 9	โหนด 10	โหนด 11
tloc='(31118.5-45965.25]'	0.6101	0.2011	0.8054	1.4119	1.1141
tloc='(45965.25-inf)'	-0.1590	0.0151	-0.0491	0.0682	0.2928
tna='(-inf-20595.5]'	-0.8276	-0.5395	-0.0166	-0.7255	0.8983
tna='(20595.5-39351]'	-0.0987	-0.7869	-0.6252	0.1969	-1.5274
tna='(39351-58106.5]'	-0.0852	0.0825	0.0047	0.2189	0.1852
tna='(58106.5-inf)'	-0.2266	-0.0194	-0.0738	0.0309	0.2519
tnla='(-inf-211.25]'	-0.1986	-0.7952	-0.2386	-0.9668	0.1145
tnla='(211.25-409.5]'	0.6002	0.6643	0.2416	0.9178	0.1264
tnla='(409.5-607.75]'	-1.8021	-2.2773	-0.7783	-0.7928	-1.0027
tnla='(607.75-inf)'	0.2155	1.2288	0.0528	0.5639	0.6731
tnlm='(-inf-160.75]'	-0.0155	1.9184	0.4018	-0.6521	1.0741
tnlm='(160.75-308.5]'	-1.2905	-2.6184	-1.1037	1.5438	-0.2143
tnlm='(308.5-456.25]'	-0.0420	-0.2282	-0.0373	-2.2230	-1.7422
tnlm='(456.25-inf)'	0.1898	-0.2901	-0.0082	1.1119	0.7609
tnm='(-inf-1105.75]'	-0.6354	-0.8142	0.0366	-1.4037	-0.7408
tnm='(1105.75-2135.5]'	-0.6337	-1.1890	-0.6644	1.0880	-0.0866
tnm='(2135.5-3165.25]'	0.3852	0.7181	-0.0063	0.0263	0.4672
tnm='(3165.25-inf)'	-0.1737	0.0049	-0.0280	0.0482	0.2103
tnos='(-inf-1484.75]'	-0.1542	-1.0011	-0.1839	-1.2454	1.1584
tnos='(1484.75-2892.5]'	-1.0485	-1.2766	-0.3600	0.4820	-2.0333
tnos='(2892.5-4300.25]'	0.3092	1.0620	-0.0148	0.4903	0.5834
tnos='(4300.25-inf)'	-0.1765	-0.0561	-0.0875	0.0162	0.1649

4. การแสดงผลลัพธ์การวิเคราะห์ข้อมูล ความแม่นยำ และจำนวนรายการร้องขอการนำเข้าซอร์สโค้ดที่พยากรณ์หรือทำนายถูกต้อง รายละเอียดดังตารางที่ ง.5

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ง.5 แสดงผลลัพธ์การวิเคราะห์ข้อมูล

รายละเอียด	จำนวน
จำนวนรายการร้องขอการนำเข้าซอร์สโค้ดที่พยากรณ์หรือทำนายถูกต้อง (Correctly Classified Instances)	1127 รายการ คิดเป็น 93.372%
จำนวนรายการร้องขอการนำเข้าซอร์สโค้ดที่พยากรณ์หรือทำนายผิดพลาด (Incorrectly Classified Instances)	80 รายการ คิดเป็น 6.628%
สถิติ Kappa	0.0172
ค่าความผิดพลาด Mean absolute error	0.0808
ค่าความผิดพลาด Root mean squared error	0.2392
ค่าความผิดพลาด Relative absolute error	89.9441%
ค่าความผิดพลาด Root relative squared error	113.7274%
จำนวนตัวเลขการวิเคราะห์	1207

5. ผลลัพธ์การตรวจสอบความน่าเชื่อถือของข้อมูล (Classification Table) ดังตารางที่ ง.6

ตารางที่ ง.6 ผลลัพธ์การตรวจสอบความน่าเชื่อถือของข้อมูล (Classification Table)

รายการร้องขอที่มีผลลัพธ์เป็นไม่ผ่านหรือปฏิเสธ	รายการร้องขอที่มีผลลัพธ์เป็นผ่านหรือยอมรับ	<-- Classified as
1125	26	
54	2	

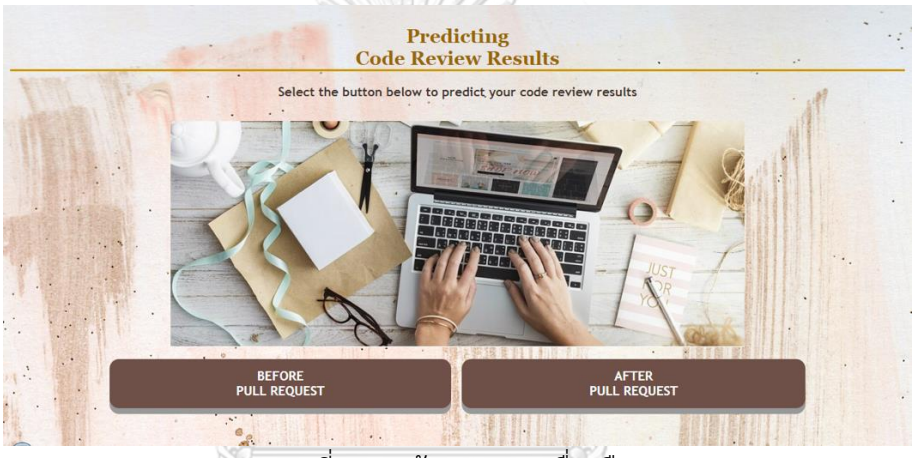
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก จ

คู่มือการใช้งานเครื่องมือการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ

เครื่องมือการทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ ประกอบด้วย 5 หน้าด้วยกัน มีรายละเอียด ดังนี้

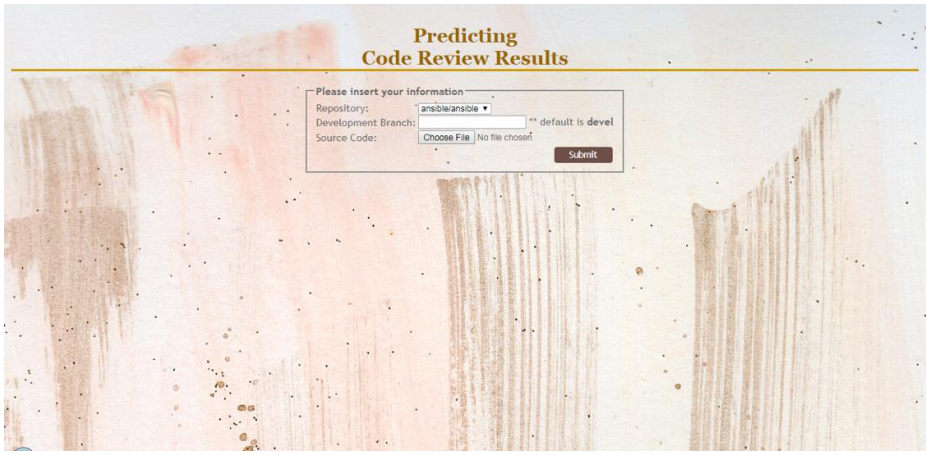
1. หน้าแรก หน้าแรกเป็นหน้าที่ให้ผู้ใช้งานระบบเลือกช่วงเวลาการร้องขอการนำเข้าซอร์สโค้ด โดยหากผู้ใช้งานต้องการวิเคราะห์ซอร์สโค้ดก่อนการร้องขอการนำเข้าซอร์สโค้ด ระบบจะเปิดหน้าการกรอกข้อมูลสำหรับการวิเคราะห์ซอร์สโค้ดก่อนการร้องขอการนำเข้าซอร์สโค้ด หรือหากผู้ใช้งานระบบต้องการวิเคราะห์ซอร์สโค้ดหลังการร้องขอการนำเข้าซอร์สโค้ด ระบบจะเปิดหน้าการกรอกข้อมูลสำหรับการวิเคราะห์ซอร์สโค้ดหลังการร้องขอการนำเข้าซอร์สโค้ด



รูปที่ จ.1 หน้าแรกของเครื่องมือ

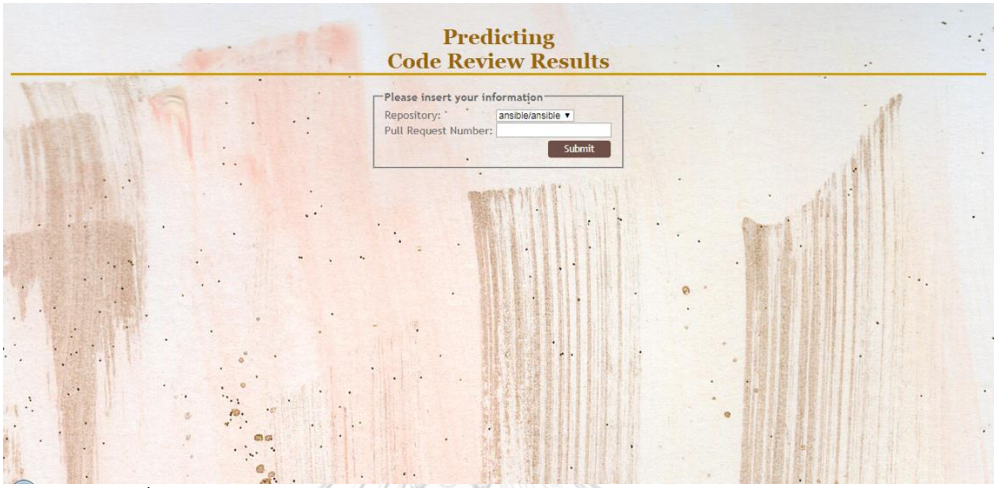
2. หน้าการวิเคราะห์ซอร์สโค้ดก่อนการร้องขอการนำเข้าซอร์สโค้ด ระบบจะให้ผู้ใช้งานระบุโครงการที่ต้องการตรวจสอบ เลขสาขาในการพัฒนาซอฟต์แวร์ และอัพโหลดซอร์สโค้ดเพื่อการวิเคราะห์ข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



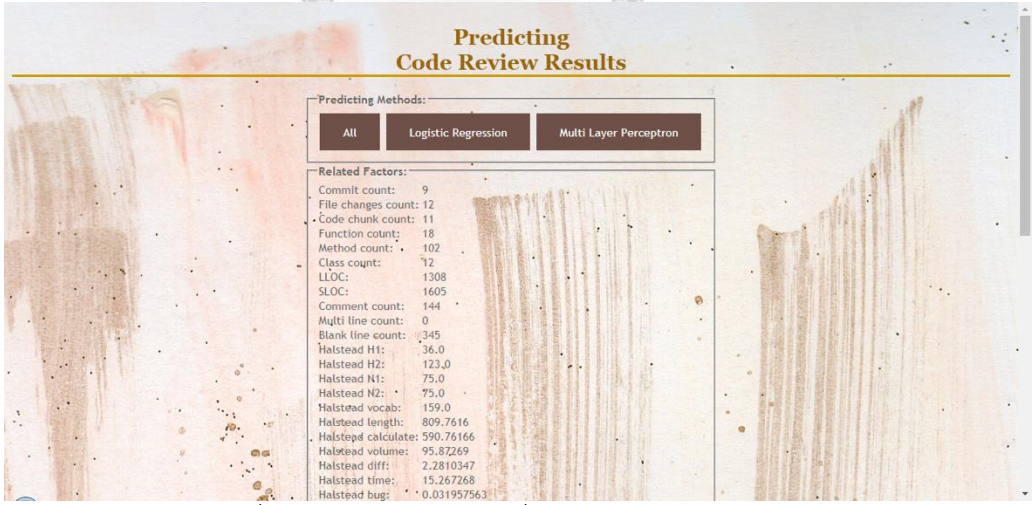
รูปที่ จ.2 หน้าการวิเคราะห์ซอร์สโค้ดก่อนการร้องขอการนำเข้าซอร์สโค้ด

3. หน้าการวิเคราะห์ซอร์สโค้ดหลังการร้องขอการนำเข้าซอร์สโค้ด ระบบจะให้ผู้ใช้งานระบุโครงการที่ต้องการตรวจสอบ รวมถึงเลขการร้องขอการนำเข้าซอร์สโค้ด หากแต่ไม่ต้องมีการอัพโหลดซอร์สโค้ด เนื่องจากระบบมีความสามารถในการไปดึงข้อมูลซอร์สโค้ดมาได้จากเลขรายการร้องขอการนำเข้าซอร์สโค้ดที่ผู้ใช้งานระบุ



รูปที่ จ.3 หน้าการวิเคราะห์ซอร์สโค้ดหลังการร้องขอการนำเข้าซอร์สโค้ด

4. หน้าแสดงตัวแปรที่มีผลกระทบกับการรีวิวโค้ด หน้านี้จะเป็นหน้าที่ทั้งแสดงค่าตัวแปรต่าง ๆ ที่มีผลกับการรีวิวโค้ด อีกทั้งยังเป็นหน้าที่ให้ผู้ใช้งานเลือกว่าจะทำการวิเคราะห์แบบใด โดยมี 3 ตัวเลือกให้ผู้ใช้งานเลือกได้แก่ การวิเคราะห์แบบการถดถอยโลจิสติก การวิเคราะห์แบบโครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น



รูปที่ จ.4 หน้าแสดงตัวแปรที่มีผลกระทบกับการรีวิวโค้ด

5. หน้าการแสดงผล ไม่ว่าจะเป็นการวิเคราะห์แบบใดแบบหนึ่งในทั้งสามแบบ ผลที่ได้คือผลการรีวิวว่าผ่าน (Accept) หรือไม่ผ่าน (Reject)

รูปที่ จ.5 หน้าการแสดงผลของเครื่องมือ (1)

รูปที่ จ.5 หน้าการแสดงผลของเครื่องมือ (2)

ประวัติผู้เขียน

ชื่อ-สกุล	ปานทิพย์ พู่พุฒ
วัน เดือน ปี เกิด	24 กันยายน 2534
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	จบการศึกษาปริญญาตรีจาก สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง คณะวิศวกรรมศาสตร์สารสนเทศ
ที่อยู่ปัจจุบัน	164/19 หมู่บ้านชวนชื่นซิตี้เรสซิเด้นท์ปาร์ค 5 ถนนรามอินทรา 71 เขตบางเขน แขวงท่าแร้ง กทม. 10220
ผลงานตีพิมพ์	งานวิจัยนี้ได้รับการคัดเลือกและตีพิมพ์เป็นบทความวิชาการเรื่อง Finding Impact Factors for Rejection of Pull Requests on GitHub โดยปานทิพย์ พู่พุฒและพรศิริ หมื่นไชยศรี ในการประชุมวิชาการ “2018 VII International Conference on Network, Communication and Computing (ICNCC 2018)” ระหว่างวันที่ 14-16 ธันวาคม 2561 ณ Horward Civil Service International House เมืองไทเป ประเทศไต้หวัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University