การจำแนกพฤติกรรมการนอนหลับตามการจัดกลุ่มของคุณภาพการนอนหลับ

น.ส.ปวรรัตน์ ขุมเงิน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Sleep Behavior Classification based on Clusters of Sleep Quality

Miss Pawonrat Khumngoen

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2022

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การจำแนกพฤติกรรมการนอนหลับตามการจัดกลุ่มของคุณภาพการนอนหลับ
โดย	น.ส.ปวรรัตน์ ขุมเงิน
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

---------- คณบดีคณะวิศวกรรมศาสตร์

(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

---------- ประธานกรรมการ

(รองศาสตราจารย์ ดร.ณัฐพงศ์ ชินธเนศ)

---------- อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

---------- กรรมการภายนอกมหาวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร.เด่นดวง ประดับสุวรรณ)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ปวรรัตน์ ขุมเงิน : การจำแนกพฤติกรรมการนอนหลับตามการจัดกลุ่มของคุณภาพการนอนหลับ. (Sleep Behavior Classification based on Clusters of Sleep Quality) อ.ที่ปรึกษาหลัก : ผศ. ดร.สุกรี สินธุภิญโญ

การนอนหลับเป็นกิจกรรมสำคัญที่มีอิทธิพลต่อการดำรงชีวิต ซึ่งความสำคัญของการนอนหลับคือการฟื้นฟู ซ่อมแซมเซลล์ภายในร่างกายและเตรียมพร้อมพลังงานสำหรับการเริ่มต้นในวันถัดไป การนอนหลับที่ดีสามารถสื่อได้ถึงสุขภาพกายและสุขภาพจิตที่แข็งแรงโดยสามารถวัดได้จากคุณภาพการนอนหลับ การศึกษาทั่วไปจะนิยมใช้ข้อมูลทั้งหมดในการสอนแบบจำลองแต่ในการศึกษานี้เชื่อว่ารูปแบบการนอนหลับในแต่ละบุคคลจะมีความแตกต่างกัน ดังนั้นวิทยานิพนธ์ฉบับนี้ขอเสนอการจำแนกพฤติกรรมการนอนหลับตามการจัดกลุ่มของคุณภาพการนอนหลับ ขั้นตอนแรกคือการจัดกลุ่มคนที่มีรูปแบบการนอนหลับที่คล้ายคลึงกันโดยใช้เทคนิคการวิเคราะห์องค์ประกอบหลักและอัลกอริทึมเคมีน ต่อมาใช้อัลกอริทึมการถดถอยโลจิสติกและการป่าแบบสุ่มในการจำแนกพฤติกรรมการนอนหลับ หลังจากนั้นทดสอบประสิทธิภาพการวิเคราะห์ของแบบจำลองด้วยการตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัว จากผลการทดลองพบว่าค่าความแม่นยำแบบจำลองของอัลกอริทึมการป่าแบบสุ่มในทุกกลุ่มให้ผลลัพธ์ที่ดีกว่าแบบจำลองของการถดถอยโลจิสติกระหว่าง 2.1% ถึง 7.6%

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่อนิสิต .....

ปีการศึกษา 2565 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6372072521 : MAJOR COMPUTER SCIENCE

KEYWORD: Machine Learning, Classification, Sleep behavior, Sleep quality

Pawonrat Khumngoen : Sleep Behavior Classification based on Clusters of Sleep Quality. Advisor: Asst. Prof. SUKREE SINTHUPINYO, Ph.D.

Sleep is a significant activity that can influence livelihoods. The critical part of sleep is recovery, repairing cells physically, and preparing energy for the beginning of the next living days. Good sleep can refer to strong health and mental health which is capably measured by sleep quality. Normally, many works used the whole dataset to train models. But we believe that each person has a different sleeping pattern. So, in this paper, we presented a classification of sleep behavior based on a cluster of sleep quality. We first clustered people who have similar sleep patterns using the Principal Component Analysis technique and K-means algorithm. Then, we used Logistic Regression and Random Forest algorithm to classify sleep behavior. We performed models from the analysis with Leave-one-out cross-validation. The results showed that the accuracy given by Random Forest algorithm models in every group was better than Logistic Regression models between 2.1% and 7.6%.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Field of Study: Computer Science Student's Signature .....

Academic Year: 2022 Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ได้รับความอนุเคราะห์จากที่ปรึกษาวิทยานิพนธ์คือ ผู้ช่วยศาสตราจารย์ ดร. สุกรี สินธุภิญโญ สำหรับคำแนะนำในเรื่องแนวทางการทำวิทยานิพนธ์ การตั้งคำถามข้อสงสัย วิธีการหาคำตอบ การตีพิมพ์ผลงานทางวิชาการ รวมถึงการตรวจทานแก้ไขวิทยานิพนธ์ฉบับนี้ให้สามารถสำเร็จลุล่วงไปได้เป็นอย่างดี ผู้วิจัยขอขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณ รองศาสตราจารย์ ดร. ณัฐพงศ์ ชินธเนศ ประธานกรรมการสอบวิทยานิพนธ์ และ ผู้ช่วยศาสตราจารย์ ดร. เด่นดวง ประดับสุวรรณ กรรมการสอบวิทยานิพนธ์ที่ได้สละเวลาอันมีค่าในการชี้แนะและแสดงความคิดเห็นที่เป็นประโยชน์ต่อการทำวิทยานิพนธ์

สุดท้ายนี้ขอขอบคุณสมาชิกครอบครัวสำหรับกำลังใจและคอยสนับสนุนทำให้วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดี

ปวรรัตน์ ขุมเงิน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ค

บทคัดย่อภาษาอังกฤษ ..... ง

กิตติกรรมประกาศ ..... จ

สารบัญ ..... ฉ

สารบัญตาราง ..... ซ

สารบัญรูปภาพ ..... ฌ

บทที่ 1 บทนำ 1

1.1. ความเป็นมาและความสำคัญของปัญหา 1

1.2. วัตถุประสงค์ 4

1.3. ขอบเขตการดำเนินงาน 4

1.4. ขั้นตอนการวิจัย 4

1.5. ประโยชน์ที่คาดว่าจะได้รับ 5

1.6. โครงสร้างเนื้อหาในโครงร่างวิทยานิพนธ์ 5

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 6

2.1. ทฤษฎีที่เกี่ยวข้อง 6

2.2. งานวิจัยที่เกี่ยวข้อง 17

บทที่ 3 แนวคิดและวิธีดำเนินงาน 19

3.1. สภาพแวดล้อมและเครื่องมือที่ใช้ 19

3.2. แนวคิดและการออกแบบการทดลอง 19

3.3. การจัดการข้อมูลดิบ 24

3.4. การวัดประสิทธิภาพแบบจำลองการจำแนกพฤติกรรมการนอนหลับ 25

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4 การทดลองและผลการทดลอง....28
4.1. การทดลอง....28
บทที่ 5 สรุปผลการทดลอง....42
5.1. สรุปผลการทดลอง....42
5.2. งานวิจัยในอนาคต....43
บรรณานุกรม....44
ประวัติผู้เขียน....47

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญตาราง

หน้า

ตารางที่ 1 ตารางแสดงสัดส่วนของกลุ่มตามการติดป้ายกำกับจากอัลกอริทึมเคมีนตามรายบุคคล....31

ตารางที่ 2 ตารางแสดงจำนวนคนที่ถูกจัดกลุ่มตามสัดส่วนของกลุ่มสูงสุดตามรายบุคคล 37

ตารางที่ 3 ตารางแสดงค่าความถูกต้องจากแบบจำลองการจำแนกด้วยข้อมูลทั้งหมด.....38

ตารางที่ 4 ตารางแสดงค่าความถูกต้องจากแบบจำลองการจำแนกด้วยข้อมูลรายบุคคล 38

ตารางที่ 5 ตารางแสดงค่าความถูกต้องจากแบบจำลองการจำแนกด้วยข้อมูลรายกลุ่ม.....40

ตารางที่ 6 ตารางแสดงการเปรียบเทียบค่าความถูกต้องระหว่างข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่มด้วยการถดถอยโลจิสติก 40

ตารางที่ 7 ตารางแสดงการเปรียบเทียบค่าความถูกต้องระหว่างข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่มด้วยการป่าแบบสุ่ม 40

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญรูปภาพ

หน้า

รูปภาพที่ 1 ตัวอย่างภาพการจัดกลุ่มข้อมูลด้วยอัลกอริทึมเคมีน โดย k=3 10

รูปภาพที่ 2 ตัวอย่างภาพกราฟแสดงค่า WCSS เทียบจำนวนกลุ่ม ด้วยวิธี Elbow โดยค่า k ที่เหมาะสมของกราฟนี้ คือ k=3 11

รูปภาพที่ 3 ตัวอย่างการป่าแบบสุ่ม 13

รูปภาพที่ 4 ตัวอย่างภาพการแบ่งชุดข้อมูลการสอนและทดสอบของการตรวจสอบไขว้แบบแบ่งออกเป็นส่วน 16

รูปภาพที่ 5 รูปภาพแสดงการเปรียบเทียบค่าประสิทธิภาพการนอนหลับระหว่างของวันถัดไปกับวันก่อนหน้าของรายบุคคลและการติดป้ายกำกับการจำแนก 20

รูปภาพที่ 6 ตัวอย่างรูปภาพการแบ่งข้อมูลรายบุคคลจากข้อมูลทั้งหมด 60 คน 21

รูปภาพที่ 7 ตัวอย่างรูปภาพการคำนวณสัดส่วนความน่าจะเป็นของกลุ่มในข้อมูลรายบุคคล 22

รูปภาพที่ 8 รูปภาพไดอะแกรมขั้นตอนการทดลอง 23

รูปภาพที่ 9 รูปภาพแสดงกราฟเปรียบเทียบระหว่างค่าอธิบายความแปรปรวนกับจำนวนองค์ประกอบหลัก 29

รูปภาพที่ 10 รูปภาพแสดงกราฟเปรียบเทียบระหว่างค่าอธิบายความแปรปรวนสะสมกับจำนวนองค์ประกอบหลัก 29

รูปภาพที่ 11 รูปภาพแสดงกราฟเทียบระหว่างค่า WCSS กับจำนวนกลุ่ม (k) 31

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1. ความเป็นมาและความสำคัญของปัญหา

การนอนหลับถือเป็นกิจกรรมอย่างหนึ่งที่สำคัญต่อการดำรงชีวิตของสิ่งมีชีวิตที่นอกเหนือจากการรับประทานอาหาร หรือการออกกำลังกาย การนอนหลับมีบทบาทหน้าที่สำคัญต่อการทำงานของระบบต่าง ๆ ภายในร่างกาย เช่น ระบบประสาท ระบบความจำ การควบคุมหรือแสดงออกทางด้านอารมณ์ ระบบเผาผลาญ ระบบกำจัดของเสีย และระบบภูมิคุ้มกัน เป็นต้น [1] ซึ่งหน้าที่หลักนี้ คือ การฟื้นฟู ซ่อมแซมเซลล์ในร่างกาย รวมถึงการเตรียมความพร้อมของร่างกายสำหรับการทำงานและดำรงชีวิตในวันถัดไป เมื่อร่างกายมีระบบการทำงานที่ดีและมีประสิทธิภาพแล้วจะส่งผลให้ผู้คนมีสุขภาพกายและสุขภาพจิตที่แข็งแรง ด้วยเหตุผลสำคัญนี้ผู้คนจึงควรมีการนอนหลับที่เพียงพอเพื่อให้ร่างกายสามารถฟื้นฟูในขณะที่นอนหลับได้อย่างเต็มที่ โดยปกติคนทั่วไปจะใช้เวลากับการนอนหลับเป็นเปอร์เซ็นต์ระหว่าง 20 ถึง 40 เปอร์เซ็นต์ต่อวัน หรือประมาณ 4.8 ถึง 9.6 ชั่วโมงต่อวัน [2] ทั้งนี้ระยะเวลานอนหลับที่เหมาะสมและเพียงพอต่อการพักผ่อนขึ้นอยู่ตามแต่ละช่วงอายุ ได้แก่ 1.เด็กแรกเกิด คือ 14-17 ชั่วโมง 2.เด็กทารก คือ 12-15 ชั่วโมง 3.เด็กวัยหัดเดิน คือ 11-14 ชั่วโมง 4.เด็กวัยก่อนเข้าเรียน 10-13 ชั่วโมง 5.เด็กวัยเรียน คือ 9-11 ชั่วโมง 6.วัยรุ่น คือ 8-10 ชั่วโมง 7.วัยผู้ใหญ่ คือ 7-9 ชั่วโมง 8.วัยผู้สูงอายุ คือ 7-8 ชั่วโมง [3]

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การมีระยะเวลานอนหลับที่เพียงพอสามารถสื่อได้ถึงการมีการนอนหลับที่ดี นอกจากระยะเวลานอนหลับที่เหมาะสม กล่าวคือมีระยะนอนหลับที่ไม่ยาวหรือสั้นจนเกินไปตามที่ได้กล่าวไว้ข้างต้นแล้วยังมีปัจจัยอื่น ๆ ที่ส่งผลต่อการนอนหลับ ตัวอย่างเช่น การบริโภคอาหาร การทำกิจกรรมระหว่างวัน ความเครียดสะสมที่ได้รับ พฤติกรรมการใช้อุปกรณ์อิเล็กทรอนิกส์ก่อนนอนหรือหลังตื่นนอน เวลาที่เข้านอน เวลาตื่นนอน และระยะเวลาทั้งหมดที่นอนหลับ เป็นต้น

การศึกษานี้มุ่งเน้นศึกษาปัจจัยสำคัญที่ส่งผลต่อการนอนหลับโดยเลือกศึกษาเกี่ยวกับตัวแปรพฤติกรรมการนอนหลับ เช่น ระยะเวลาที่ใช้อุปกรณ์อิเล็กทรอนิกส์ก่อนนอน เวลาตื่นนอน เวลาที่เข้านอน จำนวนครั้งที่ตื่นระหว่างคืน และระยะเวลาทั้งหมดที่นอนหลับ

เป็นต้น ซึ่งช่วงเวลานอนหลับที่ดีที่สุด คือ ช่วงเวลา 2 ทุ่มถึง 4 ทุ่ม เนื่องจากเป็นช่วงเวลาที่มีการหลั่งฮอร์โมนที่สำคัญต่อร่างกายออกมา และฮอร์โมนที่สำคัญนี้จะมีการหลั่งเฉพาะช่วงเวลากลางคืนหรือช่วงที่ร่างกายกำลังนอนหลับอยู่เท่านั้น ซึ่งได้แก่ โกรทฮอร์โมน และฮอร์โมนเมลาโทนิน

โกรทฮอร์โมนมีหน้าที่สำคัญในการควบคุมการเจริญเติบโตของร่างกาย รวมถึงควบคุมระบบการทำงานของร่างกายและการสร้างเซลล์ใหม่ ๆ ขึ้นมาแทนที่เซลล์ที่มีสภาพเสื่อมโทรมลง ซึ่งโกรทฮอร์โมนจะมีการหลั่งออกมาเฉพาะช่วงเวลากลางคืนเท่านั้น การได้รับโกรทฮอร์โมนที่น้อยและไม่เพียงพอต่อร่างกาย ในวัยเด็กจะส่งผลให้ร่างกายมีการเจริญเติบโตที่ช้า มีส่วนสูงน้อยและมีความหนาแน่นของมวลกระดูกที่ต่ำเมื่อเทียบกับเด็กปกติทั่วไป ส่วนในวัยผู้ใหญ่จะส่งผลให้ร่างกายมีการสะสมไขมันเพิ่มขึ้น กล้ามเนื้อลดลง มีความเครียด และอารมณ์ทางเพศที่ลดลง ในขณะเดียวกันฮอร์โมนเมลาโทนินจะมีช่วงเวลาการหลั่งเหมือนกันกับโกรทฮอร์โมน คือ จะหลั่งเฉพาะช่วงเวลากลางคืนเท่านั้น เพียงแต่ฮอร์โมนเมลาโทนินจะมีเงื่อนไขเพิ่มเติมจากโกรทฮอร์โมน โดยฮอร์โมนนี้จะสามารถหลั่งออกมาได้เฉพาะในพื้นที่มืดเท่านั้น หน้าที่ของฮอร์โมนชนิดนี้ คือ การช่วยให้ร่างกายสามารถนอนหลับได้ดีขึ้นและหลับได้อย่างสบายมากขึ้น การขาดฮอร์โมนเมลาโทนินสามารถก่อให้เกิดการนอนหลับที่ยากขึ้นได้ อีกทั้งการใช้อุปกรณ์อิเล็กทรอนิกส์ก่อนนอนสามารถกระตุ้นให้ร่างกายมีการตื่นตัวและนำมาซึ่งการนอนหลับยากได้เช่นกัน ดังนั้นก่อนนอนควรหลีกเลี่ยงการใช้อุปกรณ์อิเล็กทรอนิกส์เพื่อให้ร่างกายได้มีการเตรียมความพร้อมสำหรับการนอนหลับและทำให้มีการนอนหลับที่ดีขึ้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตัวอย่างที่ได้กล่าวไว้ข้างต้นทำให้เห็นว่าพฤติกรรมการนอนหลับสามารถก่อให้เกิดการนอนหลับที่ดีและไม่ดีตามมา ซึ่งในทั่วไปผู้คนสามารถรับรู้ได้ว่าการที่มีการนอนหลับที่ดีนั้นมีความสำคัญและเป็นประโยชน์ต่อการดำรงชีวิตมากกว่าการนอนหลับที่ไม่ดี การนอนหลับที่ไม่ดีอาจได้รับอิทธิพลมาจากการมีระยะเวลานอนหลับที่ไม่เพียงพอ มีเวลาการเข้านอนที่ไม่ปกติ รวมถึงมีตารางเวลาการดำรงชีวิตที่ผิดแปลกไปจากคนทั่วไป เช่น การทำงานกะกลางคืนซึ่งส่งผลให้ร่างกายพลาดการได้รับฮอร์โมนที่สำคัญในช่วงเวลากลางคืน เป็นต้น ซึ่งสาเหตุของการนอนหลับไม่ดีส่วนใหญ่แล้วมักมาจากความผิดปกติด้านการนอนหลับ (Sleep disorders) โดยจะส่งผลทางตรงต่อพฤติกรรมการนอนหลับ [4] ยกตัวอย่างความผิดปกติด้านการนอนหลับ เช่น ภาวะหยุดหายใจขณะนอนหลับ (Sleep apnea) คือ ภาวะการหยุดหายใจที่ผิดปกติ ซึ่งมีสาเหตุมาจากที่ร่างกายขาดออกซิเจนไหลเวียนในร่างกายทำให้ร่างกายมีการกระตุ้นให้ร่างกายตื่นขึ้นจากสภาวะนอนหลับเพื่อให้สามารถหายใจเอาออกซิเจนเข้าสู่ร่างกาย

ได้อย่างเพียงพอ โรคนอนไม่หลับ (Insomnia) คือ สภาวะที่ผู้ป่วยมีการนอนหลับยากหรือใช้เวลาในการเข้าสู่สภาวะนอนหลับนานกว่าปกติ และโรคลมหลับ (Narcolepsy) คือ ความผิดปกติด้านการนอนหลับที่มีลักษณะง่วงนอนตลอดเวลา แม้จะนอนหลับเพียงพอแล้วก็ตาม และโรคนี้อาจทำให้เกิดอุบัติเหตุทางยานพาหนะได้ เป็นต้น โดยผลเสียของการนอนหลับไม่ดีในระยะสั้นจะส่งผลให้ความสนใจหรือความใจจดใจจ่อกับเรื่องบางอย่างลดลง ส่งผลให้มีประสิทธิภาพการทำงานลดลง และมีคุณภาพชีวิตที่แย่ลง ในขณะที่ระยะยาวสามารถส่งผลอันตรายถึงชีวิตได้

การศึกษานี้มุ่งเน้นศึกษาพฤติกรรมการนอนหลับที่ส่งผลให้เกิดการนอนหลับที่ดีขึ้น โดยใช้ตัวแปรเกี่ยวกับพฤติกรรมการนอนหลับที่ได้กล่าวไว้ข้างต้น และใช้ตัวแปรคุณภาพการนอนหลับที่เป็นตัวแปรสำคัญในการวัดและระบุถึงการนอนหลับที่ดีและไม่ดี ซึ่งคุณภาพการนอนหลับได้ถูกนำมาประยุกต์ใช้กับงานวิจัยต่าง ๆ ได้แก่ Shu Hui Cheng และคณะ [5] และ Aarti Sathyanarayana และคณะ [6] ที่นำตัวแปรนี้มาศึกษาและวิเคราะห์เกี่ยวกับปัจจัยที่มีอิทธิพลต่อการนอนหลับ รวมถึงรูปแบบพฤติกรรมต่าง ๆ ที่ส่งผลต่อการนอนหลับด้วย เช่น ไลฟ์สไตล์ การรับประทานอาหาร พฤติกรรมการใช้อินเทอร์เน็ต กิจกรรมระหว่างวัน และรูปแบบการนอนหลับ เป็นต้น จากงานวิจัยของ Shu Hui Cheng และคณะใช้คุณภาพการนอนหลับในการศึกษาและวิเคราะห์ถึงอิทธิพลของปัจจัยที่ส่งผลต่อการนอนหลับ โดยส่วนมากจะใช้ตัวแปรต้นเป็นแบบข้อมูลเชิงปริมาณที่ไม่มีความต่อเนื่องโดยเก็บข้อมูลมาจากการตอบแบบสอบถาม เช่น ตัวแปรเพศ ระดับการศึกษา ค่าดัชนีมวลกาย การดื่มแอลกอฮอล์ กิจกรรมระหว่างวัน และการสูบบุหรี่ เป็นต้น ซึ่งได้ใช้เทคนิควิเคราะห์ทางสถิติ คือ การทดสอบทีและการถดถอยโลจิสติกพหุคูณในการจำแนกประเภทการนอนหลับที่ดีและไม่ดี โดยมีการติดป้ายกำกับการจำแนกโดยใช้ค่าแบบสอบถามคุณภาพการนอนหลับของพิตต์สเบิร์ก (PSQI) ซึ่งเป็นการให้คะแนนด้วยตนเองตามคุณภาพการนอนหลับที่เกิดขึ้นในช่วงเดือนนั้น ส่วนงานวิจัยของ Aarti Sathyanarayana และคณะได้ใช้เทคนิคการเรียนรู้ของโปรแกรมด้วยตัวเองแบบไม่เชิงลึกและเชิงลึกเพื่อให้เหมาะสมกับลักษณะข้อมูลที่ใช้ ซึ่งงานวิจัยนี้ศึกษาเกี่ยวกับการพยากรณ์การนอนหลับที่ดีและไม่ดีด้วยคุณภาพการนอนหลับ โดยใช้ข้อมูลกิจกรรมทางกายและรูปแบบการนอนหลับมาวิเคราะห์

ดังนั้นในการศึกษานี้ผู้วิจัยมุ่งเน้นที่การจำแนกพฤติกรรมการนอนหลับด้วยคุณภาพการนอนหลับโดยใช้เทคนิคเรียนรู้ของโปรแกรมด้วยตัวเองและตัวแปรของรูปแบบการนอนหลับสำหรับการจัดกลุ่มและการจำแนกประเภท ผู้วิจัยเชื่อว่ารูปแบบการนอนหลับมีการ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เปลี่ยนแปลงไปในแต่ละวัน ดังนั้นผู้วิจัยจึงเปรียบเทียบคุณภาพการนอนหลับที่เปลี่ยนแปลงไปจากวันนี้และวันก่อนหน้าสำหรับการติดป้ายกำกับ หากคุณภาพการนอนหลับของวันนี้มีค่าเพิ่มขึ้นจากวันก่อนหน้าจะถือว่าพฤติกรรมการนอนหลับของวันนั้นมีความเหมาะสมขึ้นและส่งผลให้เกิดการนอนหลับที่ดีขึ้น ซึ่งจะแตกต่างจากงานวิจัยของ Shu Hui Cheng และคณะ และ Aarti Sathyanarayana และคณะที่ใช้คุณภาพการนอนหลับแบบแบ่งช่วงค่าเฉพาะในการติดป้ายกำกับ

1.2. วัตถุประสงค์

1.2.1.เพื่อพัฒนาการสร้างแบบจำลองสำหรับการจำแนกพฤติกรรมการนอนหลับ

1.2.2. เพื่อเปรียบเทียบประสิทธิภาพและวิธีการของการสร้างแบบจำลองสำหรับการจำแนกพฤติกรรมการนอนหลับ

1.3. ขอบเขตการดำเนินงาน

1.3.1. ข้อมูลเกี่ยวกับพฤติกรรมการนอนหลับและค่าตัวแปรสำคัญทางด้านการนอนหลับที่ถูกคำนวณมาอย่างถูกต้อง

1.3.2. หลักการการสร้างแบบจำลองสำหรับการจำแนกพฤติกรรมการนอนหลับและการตรวจสอบประสิทธิภาพของแบบจำลอง ได้แก่ การวิเคราะห์องค์ประกอบหลัก อัลกอริทึมเคมีน การถดถอยโลจิสติกส์ การป่าแบบสุ่ม และการตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัว

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.4. ขั้นตอนการวิจัย

1.4.1.ศึกษาและค้นคว้าเกี่ยวกับทฤษฎี งานวิจัยและหลักการเครื่องมือที่เกี่ยวข้อง

1.4.2.จัดรูปแบบข้อมูลและทำความสะอาดข้อมูล

1.4.3.ออกแบบการทดลอง และการวิเคราะห์องค์ประกอบหลัก

1.4.4.วิเคราะห์องค์ประกอบหลัก และสร้างแบบจำลอง

1.4.5.ทดสอบประสิทธิภาพแบบจำลอง และเปรียบเทียบประสิทธิภาพระหว่างแบบจำลอง

1.4.6.วิเคราะห์และสรุปผลการทดลอง

1.4.7.เรียบเรียงและจัดทำเล่มวิทยานิพนธ์

1.5. ประโยชน์ที่คาดว่าจะได้รับ

1.5.1.สามารถวิเคราะห์องค์ประกอบหลัก และสร้างแบบจำลองสำหรับการจำแนกพฤติกรรมการนอนหลับได้

1.5.2. สามารถเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองสำหรับการจำแนกพฤติกรรมการนอนหลับได้

1.6. โครงสร้างเนื้อหาในโครงร่างวิทยานิพนธ์

บทที่ 1 กล่าวถึงความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์ ขอบเขตการดำเนินงาน ขั้นตอนการวิจัย ประโยชน์ที่คาดว่าจะได้รับ และโครงสร้างเนื้อหาในวิทยานิพนธ์

บทที่ 2 กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้อง

บทที่ 3 กล่าวถึงสภาพแวดล้อมและเครื่องมือที่ใช้ แนวคิดและการออกแบบการทดลอง การจัดการข้อมูลดิบ และการวัดประสิทธิภาพแบบจำลองการจำแนกพฤติกรรมการนอนหลับ

บทที่ 4 กล่าวถึงการทดลองและผลการทดลอง

บทที่ 5 กล่าวถึงสรุปผลการทดลอง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้องและงานวิจัยที่เกี่ยวข้องที่นำมาประยุกต์ใช้การศึกษานี้

2.1. ทฤษฎีที่เกี่ยวข้อง

2.1.1. คุณภาพการนอนหลับ (Sleep quality)

คุณภาพการนอนหลับเป็นตัวแปรสำคัญสำหรับการบ่งชี้ถึงการนอนหลับที่ดีและไม่ดี โดยสามารถวัดได้ 2 ด้าน คือ ด้านอัตนัย (Subjective aspect) และด้านปรนัย (Objective aspect) การวัดจากด้านอัตนัยเป็นการวัดโดยใช้ความรู้สึกและอารมณ์ที่เกิดขึ้นหลังจากตื่นนอนหรือหลังจากที่ร่างกายได้รับการพักผ่อนแล้ว ซึ่งความรู้สึกนี้จะแตกต่างกันตามแต่ละบุคคล ส่วนการวัดจากด้านปรนัยเป็นการวัดโดยใช้จอแสดงผลการตรวจทางสรีรวิทยาไฟฟ้าหัวใจที่ประกอบด้วย ระยะเวลานอนหลับทั้งหมด (Total sleep time หรือ TST) การนอนหลับแบบแฝง (Sleep latency หรือ SL) ประสิทธิภาพการนอนหลับ (Sleep efficiency หรือ SE) และการตื่นในช่วงเวลานอนหลับ (Wake after sleep onset หรือ WASO) ของแต่ละบุคคล [7]

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.1.1.1. ระยะเวลานอนหลับทั้งหมด

ระยะเวลานอนหลับทั้งหมด คือ จำนวนเวลานอนหลับทั้งหมดระหว่างคืนที่เกิดขึ้นภายใน 24 ชั่วโมง ซึ่งจำนวนเวลาการนอนหลับนี้ขึ้นอยู่กับกิจวัตรประจำวันและพฤติกรรมของแต่ละบุคคล โดยการมีค่าระยะเวลานอนหลับทั้งหมดที่สูงสามารถบ่งบอกได้ถึงการนอนหลับที่ดี

2.1.1.2. การนอนหลับแบบแฝง

การนอนหลับแบบแฝง คือ ระยะเวลาทั้งหมดที่ใช้ในการเปลี่ยนสถานะจากสภาวะตื่นนอนเข้าสู่สภาวะการนอนหลับ โดยการที่มีค่าการนอนหลับแบบแฝงมากหมายถึงใช้เวลาในการนอนหลับนานหรือมีสภาวะนอนหลับยาก

2.1.1.3. ประสิทธิภาพการนอนหลับ

ประสิทธิภาพการนอนหลับ คือ สัดส่วนที่ได้จากระยะเวลานอนหลับทั้งหมดส่วนด้วยเวลาทั้งหมดที่อยู่บนเตียง มีหน่วยเป็นเปอร์เซ็นต์ ซึ่งค่าประสิทธิภาพการนอนหลับที่สูงสามารถสื่อถึงการนอนหลับที่เพียงพอ

2.1.1.4. การตื่นในช่วงเวลานอนหลับ

การตื่นในช่วงเวลานอนหลับ คือ จำนวนครั้งของการตื่นหลังจากผ่านช่วงการนอนหลับแบบแฝงจนกระทั่งตื่นนอน หรือกล่าวคือจำนวนครั้งที่ตื่นระหว่างคืน หากมีจำนวนการตื่นนอนระหว่างคืนที่มากบ่งบอกถึงการนอนหลับที่แย่ลง

2.1.2. การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis หรือ PCA)

การวิเคราะห์องค์ประกอบหลักเป็นเทคนิคที่นิยมใช้สำหรับการลดมิติข้อมูล โดยมีแนวคิดที่ถูกพัฒนามาจากเทคนิคทางสถิติและมีรากฐานมาจากพีชคณิตเชิงเส้น (Linear algebra) การวิเคราะห์องค์ประกอบหลักมีความสามารถในการคัดกรองและเปิดเผยองค์ประกอบสำคัญที่ซ่อนอยู่ภายในข้อมูลเพื่อใช้สำหรับการวิเคราะห์ อีกทั้งยังสามารถลดความซับซ้อนของข้อมูลที่มีปริมาณตัวแปรต้นเยอะ โดยจะวิเคราะห์และเลือกสนใจเฉพาะตัวแปรที่มีค่าความสำคัญที่มาก [8] ซึ่งจะส่งผลให้เมื่อนำตัวแปรเหล่านี้ที่ได้มาหลังจากผ่านการวิเคราะห์องค์ประกอบหลักและนำเข้าสู่กระบวนการเรียนรู้ของโปรแกรมด้วยตัวเองจะทำให้ประสิทธิภาพของแบบจำลองและระยะเวลาการประมวลผลของโปรแกรมดีขึ้น

หลักการของการวิเคราะห์องค์ประกอบหลัก คือ การเปลี่ยนตัวแปรจำนวนมากสู่ตัวแปรใหม่ที่มีขนาดเล็กลงและเป็นอิสระต่อกัน โดยตัวแปรใหม่ที่ได้มานี้จะถือว่าเป็นข้อมูลที่มีความสำคัญและเพียงต่อการนำไปวิเคราะห์ข้อมูลในกระบวนการถัดไป ซึ่งตัวแปรนี้จะถูกเรียกว่า องค์ประกอบหลัก (principal components หรือ PCs) โดยจะสามารถคำนวณได้จากค่าเฉพาะเจาะจง (Eigenvalue) หรือเวกเตอร์เฉพาะเจาะจง (Eigenvector) จากความสัมพันธ์ความแปรปรวนร่วมเกี่ยว (Covariance) และเนื่องจากรากฐานของการวิเคราะห์องค์ประกอบหลักมาจากพีชคณิตเชิงเส้นซึ่งจะศึกษาเกี่ยวกับเวกเตอร์ พื้นที่ของเวกเตอร์ เมทริกซ์ การหมุนแกน รวมถึงการแปลงเชิงเส้น ดังนั้นการวิเคราะห์องค์ประกอบหลักจึงมีการใช้

ผลรวมเชิงเส้นของเวกเตอร์ข้อมูลในการวิเคราะห์หาตัวแปรใหม่ โดยมีสมการดังต่อไปนี้

$$XP = Y \tag{1}$$

เมื่อ X คือ เมทริกซ์ข้อมูลดั้งเดิม ขนาด m x n

Y คือ เมทริกซ์ขนาด m x n

P คือ เมทริกซ์แถวที่ประกอบด้วยองค์ประกอบหลักของ X

m คือ จำนวนประเภทการวัดตัวแปร

n คือ จำนวนตัวอย่าง

ซึ่งในการหาเวกเตอร์เฉพาะเจาะจงจากสมการข้างต้นนั้นจะต้องนำสมการเมทริกซ์ความแปรปรวนร่วมเกี่ยว หรือ C_Y เข้ามาช่วยในการแก้ไขสมการเพิ่มเติม โดยจะนำสมการ C_Y มาเขียนให้อยู่ในรูปแบบที่มีเมทริกซ์ P ที่ซึ่งประกอบด้วยองค์ประกอบหลักอยู่ จะได้สมการดังด้านล่างนี้

$$C_Y = \frac{1}{n-1} YY^T$$

$$C_Y = \frac{1}{n-1} (PX)(PX)^T$$

$$C_Y = \frac{1}{n-1} P(XX^T)P^T$$

$$C_Y = \frac{1}{n-1} PAP^T \tag{2}$$

เมื่อ A คือ XX^T หรือเรียกว่าเมทริกซ์สมมาตร

ขั้นตอนต่อมาคือการนำเมทริกซ์สมมาตรที่ได้จากสมการข้างบนมาแปลงให้อยู่ในรูปของเมทริกซ์รูปทแยง โดยใช้เมทริกซ์เชิงตั้งฉาก (Orthogonal matrix) ของเวกเตอร์เฉพาะเจาะจงเข้ามาช่วย ซึ่งจะได้สมการสรุปดังนี้

$$A = EDE^T \tag{3}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เมื่อ D คือ เมทริกซ์รูปทแยง

E คือ เมทริกซ์ของเวกเตอร์เฉพาะเจาะจง

ต่อมากำหนดให้เมทริกซ์ P เป็นเมทริกซ์ของเวกเตอร์เฉพาะเจาะจงของ XX^T ดังนั้นจึงได้ว่า $P \equiv E^T$ และนำความสัมพันธ์นี้แทนค่าเข้าไปในสมการที่ 3 จะได้ว่า $A = P^T DP$ ถัดไปนำสมการที่ได้นี้เข้าแทนค่าในสมการที่ 2 จะได้สมการดังด้านล่างนี้ [9]

$$
\begin{aligned}
C_Y &= \frac{1}{n-1} PAP^T \\
C_Y &= \frac{1}{n-1} P(P^T DP)P^T \\
C_Y &= \frac{1}{n-1} (PP^T)D(PP^T) \\
C_Y &= \frac{1}{n-1} (PP^{-1})D(PP^{-1}) \\
C_Y &= \frac{1}{n-1} D
\end{aligned}
\tag{4}
$$

เมื่อ $P^T = P^{-1}$

จากสมการที่ 4 จะแสดงให้เห็นว่าเมทริกซ์ P สามารถแปลง C_Y ให้อยู่ในรูปแบบเมทริกซ์เส้นทแยงมุมได้ อีกทั้งค่าเมทริกซ์ C_Y ของแต่ละตำแหน่งบนเส้นทแยง คือ ค่าความแปรปรวนของ X

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

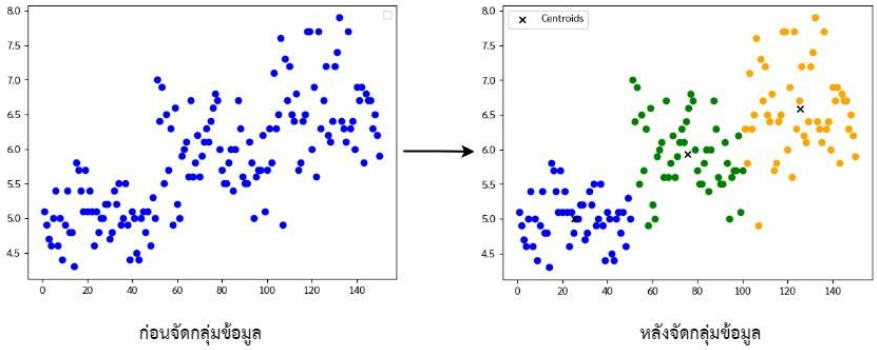
2.1.3. อัลกอริทึมเคมีน (K-means alogorithm)

อัลกอริทึมเคมีนเป็นการจัดกลุ่มหรือระบุรูปแบบกลุ่มของข้อมูลที่มีลักษณะคล้ายกันหรือแตกต่างกัน อัลกอริทึมเคมีนเป็นอัลกอริทึมการจัดกลุ่มประเภทแบบแบ่งส่วน (Partitional clustering algorithms) ซึ่งการจัดกลุ่มในลักษณะนี้จะเป็นการแบ่งข้อมูลของทั้งหมดหนึ่งกลุ่มออกเป็นกลุ่มย่อยที่ไม่เชื่อมต่อกัน และจะมีการกำหนดจำนวนกลุ่มที่ต้องการแบ่งออกอย่างชัดเจน โดยอัลกอริทึมจะมีการประมวลผลหลายครั้งในการคำนวณหาเกณฑ์ค่าความคลาดเคลื่อนกำลังสองที่น้อย

ที่สุดกับการกำหนดจุดใหม่ ๆ สำหรับจัดกลุ่มข้อมูล เพื่อให้สามารถค้นหาการจัดกลุ่มที่ดีที่สุดของข้อมูล [10]

ดังนั้นหลักการของอัลกอริทึมเคมีน คือ จะสุ่มจุดขึ้นมาและกำหนดจุดนั้นให้เป็นจุดศูนย์กลางชั่วคราว หลังจากนั้นจะคำนวณระยะห่างระหว่างจุดที่อยู่รอบจุดศูนย์กลางกับจุดศูนย์กลางหรือคำนวณค่าความคลาดเคลื่อนกำลังสองที่น้อยที่สุด ดังสมการด้านล่าง เพื่อจัดกลุ่มให้กับจุดที่มีระยะห่างใกล้ที่สุดหรือมีค่าคลาดเคลื่อนน้อยที่สุดกับจุดศูนย์กลางนั้น และทำวนซ้ำจนกระทั่งไม่มีการเปลี่ยนแปลงกลุ่ม ตัวอย่างการจัดกลุ่มด้วยอัลกอริทึมเคมีน ดังรูปภาพที่ 1

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} \left| x_j - \mu_i \right|^2 \quad (5)$$



รูปภาพที่ 1 ตัวอย่างภาพการจัดกลุ่มข้อมูลด้วยอัลกอริทึมเคมีน โดย k=3

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ขั้นตอนการจัดกลุ่มอัลกอริทึมเคมีน มีดังนี้

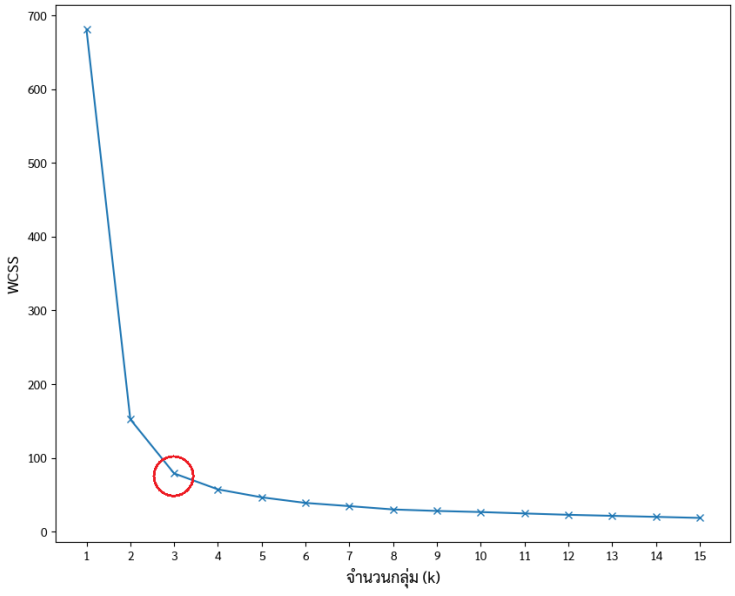
1. เลือกค่า k สำหรับกำหนดจำนวนกลุ่ม
2. สุ่มจุดศูนย์กลาง k จุด
3. คำนวณระยะห่างระหว่างจุดศูนย์กลางกับจุดทั้งหมด หรือคำนวณค่าสมการที่ 5 ของจุดศูนย์กลางกับจุดทั้งหมด
4. จัดกลุ่มให้กับจุดที่มีระยะห่างใกล้ที่สุดกับจุดศูนย์กลางนั้น
5. คำนวณค่าเฉลี่ยระยะห่างภายในกลุ่มเพื่อกำหนดจุดศูนย์กลางใหม่
6. ทำขั้นตอนที่ 3, 4 และ 5 ซ้ำจนกระทั่งไม่มีการเปลี่ยนแปลงของสมาชิกกลุ่ม

นอกจากนี้การระบุค่า k ที่เหมาะสมสำหรับการจัดกลุ่มสามารถระบุได้ด้วยวิธี Elbow วิธี Elbow คือ การวาดกราฟที่ประกอบด้วยค่า Sum square error หรือค่า Within cluster sum square error (WCSS) ที่สามารถคำนวณได้จากสมการที่ 6 เทียบกับจำนวนกลุ่ม โดยการระบุจำนวนกลุ่ม หรือ ค่า k ที่เหมาะสมนั้นสามารถระบุได้จากลักษณะเส้นของจุดที่มีลักษณะหักศอกหรือกล่าวอีกอย่าง คือ การที่ค่า WCSS จากจุดก่อนหน้าเมื่อเทียบกับจุดที่สนใจแล้ว ค่า WCSS มีค่าที่ลดลงมากจนกระทั่งเส้นมีลักษณะคล้ายกับการหักศอก [11] ดังรูปภาพที่ 2

$$WCSS = \sum_{i=1}^{k} \sum_{i \in S_k} \sum_{j=1}^{p} (x_{ij} - \bar{x}_{kj})^2 \quad (6)$$

เมื่อ S_k คือ เซตข้อมูลของกลุ่มที่ k

$\bar{x}_{kj}$ คือ ค่าเฉลี่ยของจุดศูนย์กลางที่ j สำหรับกลุ่มที่ k



รูปภาพที่ 2 ตัวอย่างภาพกราฟแสดงค่า WCSS เทียบจำนวนกลุ่ม ด้วยวิธี Elbow โดยค่า k ที่เหมาะสมของกราฟนี้ คือ k=3

2.1.4. การถดถอยโลจิสติก (Logistic regression)

การถดถอยโลจิสติกเป็นส่วนหนึ่งของวิธีการวิเคราะห์พหุตัวแปรทางสถิติที่มุ่งเน้นการพยากรณ์ค่าตัวแปรตามจากตัวแปรอิสระและสามารถอธิบายถึงความสัมพันธ์การแจกแจงของตัวแปรอิสระที่มีต่อตัวแปรตามได้จากค่าสัมประสิทธิ์ตัวแปร โดยจะวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระที่มากกว่าหนึ่งตัวกับตัวแปรตามแบบทวิภาค [12] ซึ่งตัวแปรทวิภาคนี้จะเป็นตัวแปรที่มีค่าเพียง 2 ค่า ตัวอย่างเช่น เพศ (ชาย-หญิง) หรือการตัดสินใจซื้อสินค้า (ซื้อ-ไม่ซื้อ) เป็นต้น ซึ่งเรียกว่า การถดถอยโลจิสติกแบบสองทาง (Binary logistic regression) หรือกับตัวแปรตามที่มีค่ามากกว่า 2 ทางขึ้นไป ที่ถูกเรียกว่า การถดถอยโลจิสติกแบบพหุกลุ่ม (Multinomial logistic regression) อีกทั้งการถดถอยโลจิสติกสามารถใช้กับข้อมูลประเภทต่อเนื่อง (Continuous data) กับแบบกลุ่มได้ (Categorical data) ซึ่งในการวิเคราะห์จะใช้วิธีการคำนวณสัดส่วนค่าความน่าจะเป็นของเหตุการณ์ที่สนใจที่จะเกิดขึ้นและไม่เกิดขึ้นที่เรียกว่า Odds ซึ่งสามารถคำนวณได้ดังสมการที่ 7

$$Odds = \frac{p}{1-p} \tag{7}$$

เมื่อ p คือ ค่าความน่าจะเป็นของเหตุการณ์ที่สนใจจะเกิดขึ้น

$1-p$ คือ ค่าความน่าจะเป็นของเหตุการณ์ที่สนใจจะไม่เกิดขึ้น

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ซึ่งการถดถอยโลจิสติกสามารถใช้ค่า Odds ในการอธิบายความน่าจะเป็นของเหตุการณ์ที่สนใจจะเกิดขึ้นและไม่เกิดขึ้นได้ แต่ Odds ยังมีข้อจำกัดในกรณีที่ตัวแปรอิสระไม่ได้มีค่าอยู่ในช่วง 0 ถึง 1 จึงทำให้การใช้ Odds ในการถดถอยโลจิสติกยังไม่ดีพอ ดังนั้นวิธีต่อมา คือ การปรับค่าตัวแปรอิสระให้มีค่าอยู่ระหว่าง 0 ถึง 1 เพื่อให้สามารถอธิบายออกมาในรูปแบบของการถดถอยโลจิสติกได้ด้วยลอการิทึมธรรมชาติ ดังสมการที่ 8

$$logit(y) = \ln(Odds) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \tag{8}$$

เมื่อ β คือ สัมประสิทธิ์ของ X

X คือ ตัวแปรอิสระจำนวน k ตัว

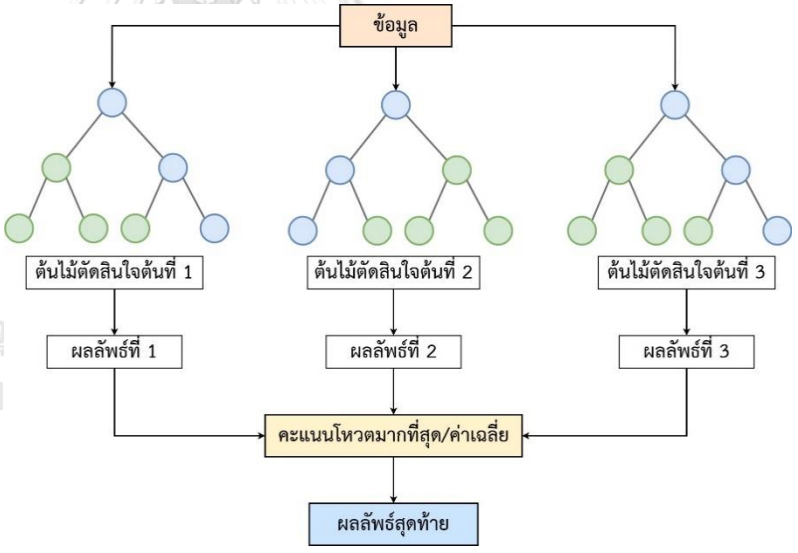
p คือ ค่าความน่าจะเป็นของเหตุการณ์ที่สนใจจะเกิดขึ้น

ซึ่งสามารถคำนวณได้ดังสมการที่ 9 นี้

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_k X_k)}} \quad (9)$$

2.1.5. การป่าแบบสุ่ม (Random forest)

การป่าแบบสุ่มเป็นการเรียนรู้แบบรวมกลุ่ม (Ensemble learning) สามารถใช้ในการจำแนกประเภทและการถดถอยข้อมูลสำหรับข้อมูลแบบต่อเนื่องและแบบกลุ่มได้ ซึ่งมีเทคนิค คือ การรวบรวมผลลัพธ์การตัดสินใจจากต้นไม้ตัดสินใจหลาย ๆ ต้นเข้าด้วยกัน โดยสามารถใช้คะแนนโหวตมากสุดในการจำแนกหรือใช้ค่าเฉลี่ยในการถดถอยสำหรับการตัดสินใจสุดท้าย อีกทั้งต้นไม้ตัดสินใจแต่ละต้นนั้นจะมีการเรียนรู้ที่เป็นอิสระต่อกัน ซึ่งตัวอย่างการป่าแบบสุ่มดังรูปภาพที่ 3



รูปภาพที่ 3 ตัวอย่างการป่าแบบสุ่ม

การป่าแบบสุ่มถูกนำเสนอโดย Leo Breiman ที่พัฒนาแนวคิดมาจาก Amit และ German ที่มีการเขียนต้นไม้หนึ่งต้นในระดับตื้นด้วยมือที่มีการจำแนกลักษณะโดยใช้การเลือกสุ่มตัวแปรจากตัวแปรปริมาณมากสำหรับการแตกกิ่งของแต่ละโหนด ซึ่ง Leo Breiman ได้นำ Bootstap aggregating หรือ Bagging สำหรับลดความแปรปรวนของข้อมูลมาใช้ โดยจะสุ่มเลือกข้อมูลตัวอย่างจากชุดข้อมูลการ

สอนแบบคืนที่ ซึ่งชุดข้อมูลย่อยที่ถูกสุ่มเลือกมานี้จะถูกเรียกว่า Bagging โดยจะใช้สำหรับสร้างแบบจำลองต้นไม้ตัดสินใจ ส่วนข้อมูลที่เหลือจากการสุ่มจะถูกเรียกว่า Out-of-bag หรือข้อมูลสำหรับวัดประสิทธิภาพแบบจำลองโดยจะถูกใช้ทดสอบกับแบบจำลองต้นไม้ตัดสินใจที่ไม่มีข้อมูลชุดนี้อยู่ในการสร้างแบบจำลอง และเหตุผลสำหรับการนำ Bagging มาใช้ คือ เพื่อเพิ่มความแม่นยำของการสุ่มตัวแปรในการสร้างแบบจำลองและใช้ในการประเมินค่า Generalization error (PE) ที่ได้จากการรวมต้นไม้ตัดสินใจหลายต้นเข้าด้วยกัน [13] โดย PE คือ ค่า error ที่ได้มาจากการนำแบบจำลองที่ถูกสร้างจากชุดข้อมูลการสอนมาวัดประสิทธิภาพกับชุดข้อมูลทดสอบ

อัลกอริทึมการป่าแบบสุ่ม [14] มีขั้นตอน ดังนี้

กำหนดให้ $D = \{(x_1, y_1), \ldots, (x_N, y_N)\}$ เป็นชุดข้อมูลการสอน และ $x_i = (x_{i,1}, \ldots, x_{i,p})^T$

1. แบ่งข้อมูลออกเป็นตัวอย่างย่อย D_i ของจำนวน N จาก D
2. ใช้ตัวอย่างย่อย D_i เป็นชุดข้อมูลสอน สำหรับสร้างต้นไม้โดยใช้การแบ่งข้อมูลออกทีละ 2 ส่วน (Binary recursive partitioning)
    2.1 เริ่มจากใช้ตัวอย่างย่อยทั้งหมดในโหนดแรกเริ่ม
    2.2 ทำซ้ำสำหรับโหนดที่ยังไม่มีการแตกกิ่งจนกระทั่งสิ้นสุดการแตกกิ่ง
        2.1.1 สุ่มเลือกตัวทำนาย (Predictor) m ตัว จากตัวทำนายทั้งหมด p ตัว
        2.1.2 ค้นหาการแบ่งออกเป็น 2 ส่วนที่ดีที่สุด (Binary split) สำหรับ m ตัวทำนาย จากขั้นตอนที่ 2.1.1
        2.1.3 แบ่งโหนดออกเป็นโหนดลูกจำนวน 2 โหนดโดยใช้เกณฑ์การแบ่งโหนดจากขั้นตอนที่ 2.1.2

โดยการทำนาย ณ โหนดใหม่ x สามารถคำนวณได้จาก

- $\hat{f}(x) = \frac{1}{J}\sum_{j=1}^{J} \widehat{h_j}(x)$ สำหรับการถดถอย
- $\hat{f}(x) = argmax_y \sum_{j=1}^{J} I(\widehat{h_j}(x) = y)$ สำหรับการจำแนกประเภท

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.1.6. การตรวจสอบไขว้ (Cross-validation)

การตรวจสอบไขว้เป็นกระบวนการสุ่มข้อมูลเพื่อใช้ทดสอบประสิทธิภาพการพยากรณ์ของแบบจำลองและใช้เพื่อป้องกันการเกิด overfitting ในแบบจำลอง [15] การตรวจสอบไขว้มีที่มาจากวิธี Monte Carlo ซึ่งเป็นวิธีการใช้การวิเคราะห์ข้อมูลในอดีตมาพยากรณ์หรือคาดการณ์เหตุการณ์ที่จะเกิดขึ้นในอนาคต

หลักการของการตรวจสอบไขว้ คือ การแบ่งข้อมูลชุดการสอนออกเป็น 2 ส่วนสำหรับสร้างและประเมินประสิทธิภาพแบบจำลอง ได้แก่ ข้อมูลสำหรับการสอนและทดสอบ โดยการตรวจสอบไขว้จะมีการสุ่มนำชุดข้อมูลการสอนและทดสอบที่แบ่งออกมาทำสลับซ้ำ ๆ กันในการตรวจสอบประสิทธิภาพของแบบจำลอง หลังจากวัดประสิทธิภาพของแบบจำลองจากหลาย ๆ ชุดข้อมูลแล้วจะนำผลลัพธ์การวัดประสิทธิภาพแบบจำลองมาคำนวณค่าความแม่นยำเฉลี่ยซึ่งจะถูกกำหนดให้เป็นค่าประมาณของประสิทธิภาพแบบจำลองของแบบจำลองในท้ายที่สุด

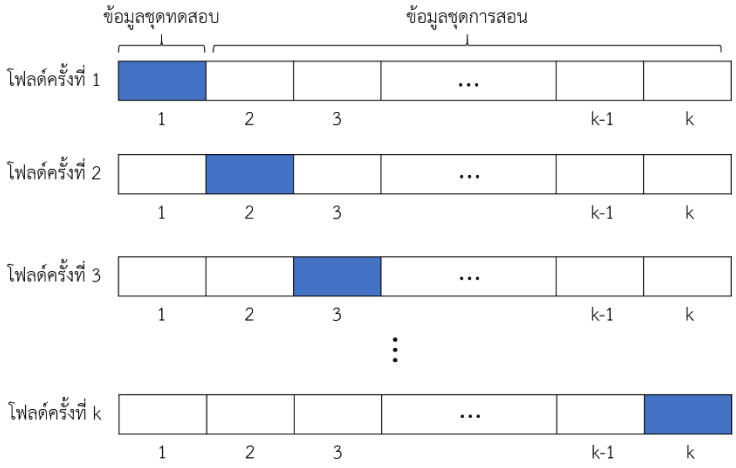
การตรวจสอบแบบไขว้มีด้วยกันหลากหลายวิธี เช่น การตรวจสอบไขว้แบบแบ่งออกเป็นส่วน (k-fold cross-validation) การตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัว (Leave-one-out cross-validation) การตรวจสอบไขว้แบบ Hold-out และการตรวจสอบไขว้แบบ Time series เป็นต้น เนื่องจากการตรวจสอบไขว้แบบแบ่งออกเป็นส่วน และการตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัวเป็นวิธีที่นิยมใช้สำหรับวัดประสิทธิภาพแบบจำลองของการจำแนกประเภทข้อมูลและเป็นวิธีที่มีค่าเบี่ยงเบนน้อย [16] ดังนั้นในการศึกษานี้จึงนำ 2 วิธีที่ได้กล่าวไปข้างต้นมาประยุกต์ใช้ในงานวิจัยนี้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.1.6.1. การตรวจสอบไขว้แบบแบ่งออกเป็นส่วน

การตรวจสอบไขว้แบบแบ่งออกเป็นส่วน เป็นการแบ่งชุดข้อมูลที่ไม่ซ้ำกันออกเป็นจำนวน k ชุดในปริมาณข้อมูลที่เท่า ๆ กัน และมีการสุ่มแบบไม่คืนที่ของชุดข้อมูล k-1 ชุด สำหรับกำหนดให้เป็นชุดข้อมูลการสอน และอีก 1 ชุดที่เหลือจะถูกนำไปเป็นชุดข้อมูลการทดสอบ เมื่อสามารถกำหนดชุดข้อมูลได้แล้วจึงนำข้อมูล 2 ชุดนี้เข้าสู่กระบวนการสร้างและวัดประสิทธิภาพแบบจำลอง หลังจากนั้นจะทำการสุ่มและแบ่งข้อมูลซ้ำในการสร้างและวัดประสิทธิภาพแบบจำลองวนจนกระทั่งชุดข้อมูลทดสอบถูก

นำไปใช้ทดสอบครบจำนวน k ครั้ง ตัวอย่างการตรวจสอบไขว้แบบแบ่งออกเป็นส่วน ดังรูปภาพที่ 4



รูปภาพที่ 4 ตัวอย่างภาพการแบ่งชุดข้อมูลการสอนและทดสอบของการตรวจสอบไขว้แบบแบ่งออกเป็นส่วน

หลังจากสร้างและวัดประสิทธิภาพแบบจำลองครบ k แบบแล้ว ให้นำค่าผิดพลาดที่ได้จากการทดสอบประสิทธิภาพแบบจำลองทั้งหมด k แบบนี้มาหาค่าเฉลี่ย ดังสมการที่ 10 และจะถือว่าค่าเฉลี่ยนี้เป็นผลการทดสอบประสิทธิภาพแบบจำลองในที่สุด

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

$$Error = \frac{1}{k}\sum_{i=1}^{k} Error_i \quad (10)$$

เมื่อ k คือ จำนวนก้อนชุดข้อมูลที่ต้องการแบ่ง หรือโฟลด์

$Error_i$ คือ ค่าผิดพลาดที่เกิดจากค่าที่แท้จริงเทียบกับค่าประมาณครั้งที่ i โฟลด์

2.1.6.2. การตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัว

การตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัว มีกระบวนการคล้ายกับการตรวจสอบไขว้แบบแบ่งออกเป็นส่วน เนื่องจากเป็นกรณีพิเศษของการตรวจสอบไขว้แบบแบ่งออกเป็นส่วน โดยจากจะแบ่งข้อมูลออกเป็น k ชุดให้นำจำนวนข้อมูลทั้งหมดมากำหนดเป็นข้อมูลชุดทดสอบ กล่าวคือหาก

ข้อมูลมีจำนวน n ข้อมูล n-1 ข้อมูลจะถูกกำหนดเป็นข้อมูลการสอนและอีก 1 ข้อมูลที่เหลือจะถูกกำหนดเป็นข้อมูลทดสอบ และจะทำการกำหนดข้อมูลทดสอบซ้ำจนกระทั่งครบจำนวน n ข้อมูล หลังจากนั้นคำนวณค่าเฉลี่ยของค่าผิดพลาดที่ได้จากแบบจำลองทั้งหมด n แบบและกำหนดค่าที่ได้เป็นผลลัพธ์สุดท้าย ซึ่งการกำหนดข้อมูลทดสอบในลักษณะนี้ส่งผลให้การตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัวเหมาะสำหรับข้อมูลที่มีขนาดเล็ก

2.2. งานวิจัยที่เกี่ยวข้อง

Shu Hui Cheng และคณะ [5] ศึกษาเกี่ยวกับอิทธิพลที่ส่งผลต่อการคุณภาพการนอนหลับที่แย่ของนักเรียนในมหาวิทยาลัยไตหวันโดยใช้ปัจจัยทางด้านประวัติทางการแพทย์ส่วนตัว (Personal medical history) ความเคยชินของไลฟ์สไตล์ (Lifestyle habits) การวัดการทำงานสนับสนุน (Measurement of support functions - MSF) แบบสอบถามคุณภาพการนอนหลับของพิตต์สเบิร์ก (Pittsburgh sleep quality index - PSQI) แบบสอบถามการแก้ไขมาตราส่วนการเสพติดอินเทอร์เน็ตของจีน (Chinese internet addiction scale-revision – CIAS-R) มาตราส่วนย่อยความไม่เสถียรทางอารมณ์จากแบบทดสอบบุคลิกภาพ (Neuroticism subscale of the Maudsley personality inventory - MPI) และแบบสอบถามสุขภาพของจีน 12 ข้อ (12-item Chinese health questionnaire – CHQ-12) ซึ่งใช้การทดสอบที (T-test) ในการทดสอบตัวแปรระหว่าง 2 กลุ่มโดยแบ่งกลุ่มตามคุณภาพการนอนหลับกล่าวคือ กลุ่มนอนหลับที่ดีมีค่า PSQI น้อยกว่า 6 และกลุ่มนอนหลับที่ไม่ดีมีค่า PSQI ตั้งแต่ 6 ขึ้นไป เพื่อทดสอบหาตัวแปรที่มีอิทธิพลต่อการนอนหลับที่ไม่ดีพบว่านักเรียนเพศหญิง มีค่าดัชนีมวลกาย (Body Mass Index - BMI) ต่ำ นักเรียนระดับปริญญาตรี ไม่รับประทานอาหารเช้า ดื่มชา และมีระดับคะแนน CIAS-R ที่สูง CHQ-12 ที่สูง และ MSF ที่ต่ำมีนัยสำคัญต่อคุณภาพการนอนหลับที่แย่ หลังจากนั้นนำตัวแปรที่ได้ข้างต้นเข้าวิเคราะห์ด้วยการถดถอยโลจิสติกพหุคูณ (Multivariate logistic regression) พบว่าตัวแปรเพศหญิง นักเรียนระดับปริญญาตรี ไม่รับประทานอาหารเช้า ดื่มชา ค่า CIAS-R ที่สูง ค่า CHQ-12 ที่สูง ค่า MSF ที่ต่ำ และการเสพติดอินเทอร์เน็ตมีนัยสำคัญต่อการทำนายการนอนหลับที่ไม่ดี

Aarti Sathyanarayana และคณะ [6] ศึกษาเกี่ยวกับการพยากรณ์คุณภาพการนอนหลับดีหรือไม่ดีด้วยประสิทธิภาพการนอนหลับ โดยกำหนดป้ายกำกับ (Label) ด้วยค่าประสิทธิภาพการนอนหลับถ้ามีค่ามากกว่าเท่ากับ 85 เปอร์เซ็นถือว่ามีคุณภาพการนอนหลับที่ดี หากต่ำกว่านี้จะถือว่าเป็นคุณภาพการนอนหลับที่ไม่ดี มีการเก็บข้อมูลเกี่ยวกับกิจกรรมทางกาย (Physical activity) และรูปแบบการนอนหลับ (Sleep patterns) จากอุปกรณ์สวมใส่ (Wearable device) ในแต่ละบุคคล นอกจากนี้มีการกำหนดให้สวมใส่อุปกรณ์ตลอดวันและใช้ระยะเวลาเก็บข้อมูล 1 สัปดาห์ โดยการพยากรณ์จะใช้เทคนิคการเรียนรู้ของโปรแกรมด้วยตัวเอง (Machine learning) ทั้งหมด 6 ตัวเข้ามาเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ ซึ่งประกอบด้วยการเรียนรู้แบบไม่เชิงลึก (Non-deep learning) คือ การถดถอยโลจิสติส และการเรียนรู้แบบเชิงลึก (Deep learning) คือ เพอร์เซปตรอนหลายชั้น (Multilayer perceptrons) โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional neural network) โครงข่ายประสาทเทียมแบบวนกลับ (Recurrent neural network) หน่วยความจำระยะสั้นยาว (Long short-term memory) และหน่วยความจำระยะสั้นยาวแบทช์ไทม์ (Time-batched long short-term memory) มีการแบ่งสัดส่วนข้อมูลสำหรับการเรียนรู้, ทดสอบและวัดประสิทธิภาพด้วยอัตราส่วน 0.7:0.15:0.15 โดยการตรวจสอบประสิทธิภาพแบบจำลองวัดด้วยเมตริก (Metrics) ได้แก่ ค่าความแม่นยำ (Precision) ค่าระลึก (Recall), ค่าความถูกต้อง (Accuracy) ค่าคะแนนเอฟวัน (F1-score) และพื้นที่ใต้โค้งอาร์โอซี (Area under the roc curve - AUC) จากการทดลองพบว่าแบบจำลองจากวิธีการถดถอยโลจิสติกมีค่าเมตริกและพื้นที่ใต้โค้งอาร์โอซีต่ำกว่าแบบจำลองจากวิธีการเรียนรู้เชิงลึกซึ่งมีค่าความแม่นยำ ค่าระลึก ค่าความถูกต้อง ค่าคะแนนเอฟวัน และพื้นที่ใต้โค้งอาร์โอซีเท่ากับ 0.7083 0.9714 0.7321 0.8193 และ 0.6463 ตามลำดับ ยกเว้นวิธีโครงข่ายประสาทเทียมแบบวนกลับที่มีค่าเมตริกทั้งหมดน้อยกว่าการถดถอยโลจิสติก โดยมีค่าความแม่นยำ, ค่าระลึก, ค่าความถูกต้อง, ค่าคะแนนเอฟวัน เท่ากับ 0.6667 0.9143 0.6607 และ 0.7711 ตามลำดับ สาเหตุที่วิธีจากการเรียนรู้ไม่เชิงลึกมีประสิทธิภาพน้อยกว่าวิธีจากการเรียนรู้เชิงลึก เนื่องมาจากข้อมูลดิบที่เก็บมาเป็นข้อมูลแบบแอคเซเลอโรมิเตอร์ (Accelerometer data) ทำให้วิธีการถดถอยโลจิสติกไม่สามารถทนทานต่อการเรียนรู้ของข้อมูลลักษณะนี้ได้และมีขีดจำกัดในการเรียนรู้งานที่มีคุณลักษณะ (Features) ที่เฉพาะเจาะจง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 3
แนวคิดและวิธีดำเนินงาน

ในบทนี้จะกล่าวถึงแนวคิดและวิธีการดำเนินงาน ดังนี้ สภาพแวดล้อมและเครื่องมือ แนวคิดและการออกแบบการทดลอง การจัดการข้อมูลดิบ และการวัดประสิทธิภาพแบบจำลองการจำแนกพฤติกรรมการนอนหลับ

3.1. สภาพแวดล้อมและเครื่องมือที่ใช้

สภาพแวดล้อม

- ใช้ระบบปฏิบัติการวินโดว์ 11 แบบ 64 บิต
- หน่วยการประมวลผลแบบ Intel core i7
- หน่วยความจำเครื่อง 8 กิกะไบต์

เครื่องมือ

- กูเกิลโคแล็บ (Google colab)
- โปรแกรมไซคิตเลิร์น (Scikit-learn)
- ภาษาไพทอน (Python)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.2. แนวคิดและการออกแบบการทดลอง

เนื่องจากการศึกษานี้ผู้วิจัยสนใจศึกษาพฤติกรรมการนอนหลับรายวัน โดยใช้ข้อมูลตัวแปรที่สำคัญและส่งผลโดยตรงต่อการนอนหลับอย่างที่ได้กล่าวและอธิบายไว้ในบทก่อนหน้า ซึ่งข้อมูลประกอบด้วยระยะเวลาที่ใช้อุปกรณ์อิเล็กทรอนิกส์ก่อนนอน เวลาตื่นนอน เวลาที่เข้านอน จำนวนครั้งที่ตื่นระหว่างคืน ระยะเวลาทั้งหมดที่นอนหลับ การนอนหลับแบบแฝง ประสิทธิภาพการนอนหลับ การตื่นในช่วงเวลานอนหลับ และอื่น ๆ โดยผู้วิจัยใช้การถดถอยโลจิสติกและการป่าแบบสุ่มสำหรับจำแนกพฤติกรรมการนอนหลับ ผู้วิจัยมีความเชื่อว่าพฤติกรรมการนอนหลับของแต่ละบุคคลมีความแตกต่างกันขึ้นอยู่กับลักษณะการดำรงชีวิตของแต่ละคน อีกทั้งพฤติกรรมการนอนหลับรายวันของแต่ละคนนี้อาจมีการเปลี่ยนแปลงและไม่แน่นอนที่อาจส่งผลต่อการจำแนก

ดังนั้นผู้วิจัยจึงได้ออกแบบการทดลองโดยใช้ข้อมูล 3 รูปแบบ ได้แก่

1. ข้อมูลทั้งหมด
2. ข้อมูลรายบุคคล
3. ข้อมูลรายกลุ่มที่ได้จากการจัดกลุ่มด้วยอัลกอริทึมเคมีน

สำหรับสร้างแบบจำลองการจำแนกเพื่อใช้ในการเปรียบเทียบและวัดประสิทธิภาพของแบบจำลอง และเนื่องจากรูปแบบการนอนหลับที่เปลี่ยนแปลงไปตามรายวันนี้ทำให้ผู้วิจัยตัดสินใจกำหนดการติดป้ายกำกับสำหรับการจำแนกโดยใช้ค่าประสิทธิภาพการนอนหลับจากคุณภาพการนอนหลับของแต่ละวันมาเปรียบเทียบกัน ซึ่งจะเปรียบเทียบระหว่างค่าของวันถัดไปกับค่าของวันก่อนหน้าตามรายบุคคล ดังรูปภาพที่ 5

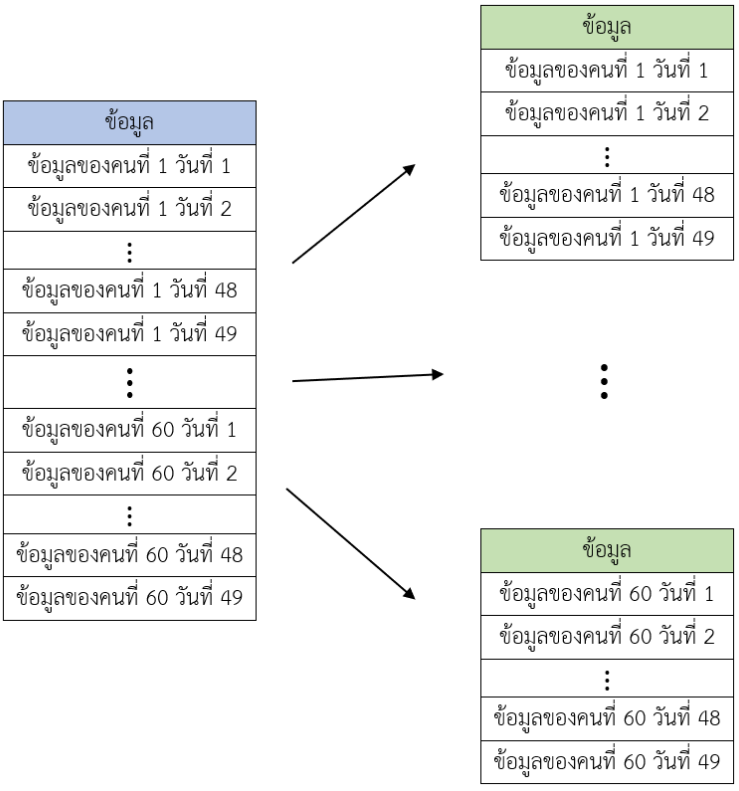
ค่าประสิทธิภาพการนอนหลับ	เปรียบเทียบค่า	ติดป้ายกำกับสำหรับการจำแนก
ค่าวันที่ 1 ของคนที่ 1	-	-
ค่าวันที่ 2 ของคนที่ 1	ค่าเพิ่มขึ้น	1
ค่าวันที่ 3 ของคนที่ 1	ค่าคงที่	0
⋮	⋮	⋮
ค่าวันที่ 48 ของคนที่ 1	ค่าคงที่	0
ค่าวันที่ 49 ของคนที่ 1	ค่าแย่ลง	0
⋮	⋮	⋮
ค่าวันที่ 1 ของคนที่ 31	-	-
ค่าวันที่ 2 ของคนที่ 31	ค่าเพิ่มขึ้น	1
ค่าวันที่ 3 ของคนที่ 31	ค่าคงที่	0
⋮	⋮	⋮
ค่าวันที่ 48 ของคนที่ 31	ค่าเพิ่มขึ้น	1
ค่าวันที่ 49 ของคนที่ 31	ค่าเพิ่มขึ้น	1
⋮	⋮	⋮
ค่าวันที่ 1 ของคนที่ 60	-	-
ค่าวันที่ 2 ของคนที่ 60	ค่าแย่ลง	0
ค่าวันที่ 3 ของคนที่ 60	ค่าเพิ่มขึ้น	1
⋮	⋮	⋮
ค่าวันที่ 48 ของคนที่ 60	ค่าเพิ่มขึ้น	1
ค่าวันที่ 49 ของคนที่ 60	ค่าแย่ลง	0

รูปภาพที่ 5 รูปภาพแสดงการเปรียบเทียบค่าประสิทธิภาพการนอนหลับระหว่างของวันถัดไปกับวันก่อนหน้าของรายบุคคลและการติดป้ายกำกับการจำแนก

โดยผู้วิจัยสนใจเฉพาะค่าประสิทธิภาพการนอนหลับที่เพิ่มขึ้น กล่าวคือ หากค่าประสิทธิภาพการนอนหลับมีค่าที่ดีขึ้นจากวันก่อนหน้าย่อมหมายความว่าพฤติกรรมการ

นอนหลับของวันถัดไปนั้นมีการนอนหลับที่ดีขึ้นจากเดิม ดังนั้นผู้วิจัยจึงกำหนดให้เหตุการณ์สนใจที่จะเกิดขึ้น คือ การที่ค่าประสิทธิภาพการนอนหลับเพิ่มขึ้นและติดป้ายกำกับการจำแนกให้เท่ากับ 1 ส่วนค่าที่นอกเหนือจากนี้จะถือว่าเป็นเหตุการณ์สนใจที่จะไม่เกิดขึ้นซึ่งจะติดป้ายกำกับการจำแนกเท่ากับ 0

นอกจากนี้ข้อมูลที่ใช้ในการศึกษานี้ประกอบด้วยพารามิเตอร์พฤติกรรมการนอนหลับทั้งหมด 18 พารามิเตอร์ จากการเก็บข้อมูลทั้งหมด 60 คน โดยมีระยะเวลาในการเก็บข้อมูล 7 สัปดาห์ ดังนั้นข้อมูลการนอนหลับแบบรายบุคคลจึงมีข้อมูลคนละ 49 ข้อมูล รวมเป็นจำนวนข้อมูลทั้งหมด 2,940 ข้อมูล และเนื่องจากจำนวนพารามิเตอร์ที่เยอะผู้วิจัยจึงตัดสินใจใช้เทคนิคการวิเคราะห์องค์ประกอบหลักเข้ามาช่วยในการสร้างตัวแปรใหม่และเลือกเฉพาะตัวแปรที่มีความสำคัญต่อข้อมูลเพื่อลดปริมาณตัวแปรในการวิเคราะห์ให้น้อยลง และช่วยเพิ่มประสิทธิภาพของแบบจำลอง ซึ่งการใช้เทคนิคการวิเคราะห์องค์ประกอบหลักจะถูกใช้กับข้อมูลทั้งหมดหลังจากนั้นจะนำผลลัพธ์ที่ได้นี้ไปแยกข้อมูลให้ออกมาให้เป็นข้อมูลเฉพาะรายบุคคลโดยจะได้ข้อมูลออกมาทั้งหมด 60 ชุด เพื่อที่จะนำข้อมูลทั้งหมดนี้ไปใช้สำหรับการจำแนกประเภทรายบุคคลต่อไป ตัวอย่างการแบ่งชุดข้อมูลรายบุคคลดังรูปภาพที่ 6



รูปภาพที่ 6 ตัวอย่างรูปภาพการแบ่งข้อมูลรายบุคคลจากข้อมูลทั้งหมด 60 คน

ส่วนข้อมูลสำหรับการจำแนกประเภทรายกลุ่มจะใช้ข้อมูลที่ได้จากเทคนิคการวิเคราะห์องค์ประกอบหลักมาเข้ากระบวนการจัดกลุ่มโดยใช้อัลกอริทึมเคมีนและวิธี Elbow สำหรับคำนวณค่า Within cluster sum square error เพื่อเลือกค่า k ที่เหมาะสมสำหรับการกำหนดจำนวนกลุ่ม และเมื่อได้ผลลัพธ์การจัดกลุ่มข้อมูลแล้วขั้นตอนถัดไป คือ การคำนวณเปอร์เซ็นต์สัดส่วนของกลุ่มในรายบุคคลเพื่อกำหนดว่าบุคคลนั้นมีความน่าจะเป็นที่จะถูกจัดให้อยู่ในกลุ่มใดมากที่สุด เนื่องจากข้อมูลทั้งหมดประกอบด้วยข้อมูลรายบุคคลจำนวน 60 คน หลังจากได้ผลลัพธ์การติดป้ายกำกับจากอัลกอริทึมเคมีนของข้อมูลทั้งหมดส่งผลให้ข้อมูลแต่ละคนอาจประกอบด้วยหมายเลขกลุ่มต่าง ๆ หรือกล่าวได้ว่าข้อมูลหนึ่งคนอาจถูกจัดให้อยู่ได้มากกว่า 1 กลุ่ม ดังนั้นจึงต้องคำนวณสัดส่วนความน่าจะเป็นของกลุ่มในข้อมูลรายบุคคลเพื่อค้นหาความน่าจะเป็นที่บุคคลนั้นจะถูกจัดให้อยู่ในกลุ่มใดสูงสุด ตัวอย่างการคำนวณสัดส่วนความน่าจะเป็นของกลุ่มในข้อมูลรายบุคคลเป็นไปดังรูปภาพที่ 7

ข้อมูลรายบุคคล	กลุ่ม
วันที่ 1 ของคนที่ 1	A
วันที่ 2 ของคนที่ 1	A
วันที่ 3 ของคนที่ 1	A
วันที่ 4 ของคนที่ 1	B
วันที่ 5 ของคนที่ 1	C
วันที่ 6 ของคนที่ 1	B
⋮	⋮
วันที่ 47 ของคนที่ 1	C
วันที่ 48 ของคนที่ 1	B
วันที่ 49 ของคนที่ 1	A

↓

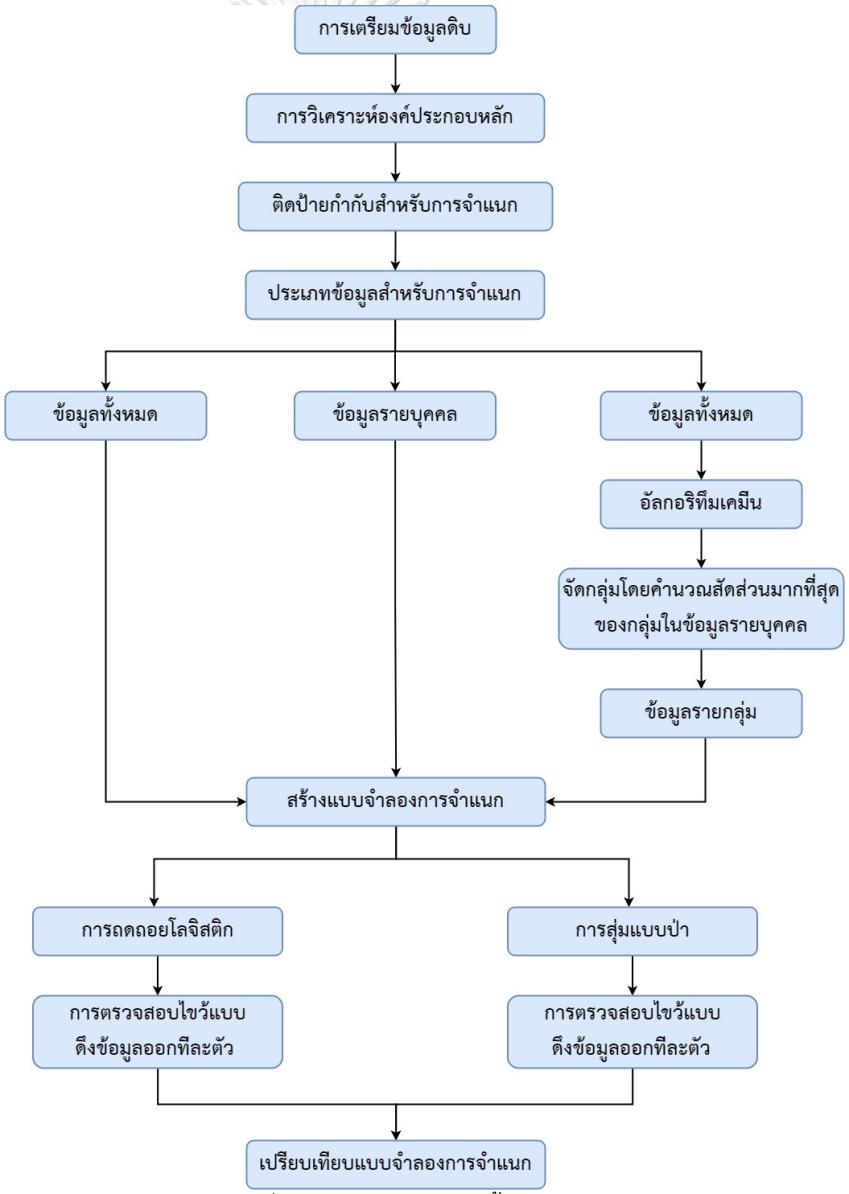
คนที่ 1 : กลุ่ม	เปอร์เซ็นต์สัดส่วนของจำนวนกลุ่ม
A	70%
B	20%
C	10%

↓

ดังนั้น คนที่ 1 มีความน่าจะเป็นที่จะอยู่กลุ่ม A มากที่สุด

รูปภาพที่ 7 ตัวอย่างรูปภาพการคำนวณสัดส่วนความน่าจะเป็นของกลุ่มในข้อมูลรายบุคคล

และเนื่องจากผู้วิจัยได้ใช้การเปรียบเทียบค่าประสิทธิภาพการนอนหลับของวันก่อนหน้ากับวันถัดไปส่งผลให้จำนวนข้อมูลเดิมจาก 2,940 ข้อมูลลดลงเหลือ 2,880 ข้อมูล เนื่องจากเมื่อติดป้ายกำกับการจำแนกการนอนหลับส่งผลให้ข้อมูลในวันที่ 1 ของทุกคนไม่มีการติดป้ายกำกับจึงทำให้ข้อมูลที่ใช้สำหรับสร้างแบบจำลองและทดสอบประสิทธิภาพลดลง 60 ข้อมูลจึงเหลือข้อมูลทั้งหมด 2,880 ข้อมูล ดังนั้นเพื่อเลือกใช้วิธีการตรวจสอบประสิทธิแบบจำลองที่เหมาะสมผู้วิจัยจึงตัดสินใจเลือกการตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัว สาเหตุเพราะการตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัวเหมาะสำหรับข้อมูลที่มีขนาดเล็กมากกว่า โดยสรุปแล้วขั้นตอนการทดลองทั้งหมดแสดงให้เห็นดังรูปภาพที่ 8



รูปภาพที่ 8 รูปภาพไดอะแกรมขั้นตอนการทดลอง

วัถตุประสงค์หลักของการออกแบบการทดลองนี้ ผู้วิจัยเชื่อว่าการใช้ข้อมูลรายกลุ่มที่ประกอบด้วยข้อมูลของคนที่มีรูปแบบหรือพฤตกรรมการนอนหลับที่คล้ายคลึงกันจะให้ผลลัพธ์การจำแนกที่ดีกว่าการใช้ข้อมูลทั้งหมดหรือข้อมูลรายบุคคล เนื่องจากสาเหตุที่ได้กล่าวไว้ข้างต้นนั่นคือ ข้อมูลรายบุคคลอาจมีรูปแบบการนอนหลับที่ผกผันหรือเปลี่ยนแปลงไปอย่างชัดเจนตามแต่ละวัน ซึ่งข้อมูลในลักษณะนี้อาจส่งผลอย่างมีนัยสำคัญต่อการสร้างแบบจำลองการจำแนกได้ อีกทั้งข้อมูลรายกลุ่มที่ได้จากอัลกอริทึมเคมีนและการคำนวณสัดส่วนมากที่สุดของกลุ่มตามรายบุคคลย่อมให้ผลลัพธ์แบบจำลองที่ดีกว่าแบบจำลองที่สร้างด้วยข้อมูลทั้งหมด

3.3. การจัดการข้อมูลดิบ

การศึกษานี้ได้รับข้อมูลดิบเกี่ยวกับพฤติกรรมการนอนหลับ จาก the Service Solution Business (SSB) ที่ SCG CBM ตัวอย่างเช่น ระยะเวลานอนหลับทั้งหมด การนอนหลับแบบแฝง ประสิทธิภาพการนอนหลับ เวลาตื่นนอน เวลาที่เข้านอน จำนวนครั้งที่ตื่นระหว่างคืน เป็นต้น ซึ่งเป็นข้อมูลพฤติกรรมการนอนหลับที่ส่งผลโดยตรงต่อการนอนหลับ รวมเป็นพารามิเตอร์ทั้งหมด 21 พารามิเตอร์ จากการเก็บข้อมูล 60 คน เป็นระยะเวลา 7 สัปดาห์

ซึ่งขั้นตอนการจัดการข้อมูลดิบมีดังนี้

3.1.1. จัดรูปแบบของข้อมูลที่มีการเก็บข้อมูลในลักษณะตามลำดับแถวให้เป็นลักษณะตามคอลัมน์ เพื่อให้สะดวกต่อการจัดการข้อมูลในลำดับถัดไป

3.1.2. ทำการตัดพารามิเตอร์ที่ไม่จำเป็นหรือมีค่าที่เหมือนกันออกไป เช่น วันที่ทำการบันทึกการนอน บันทึกกิจกรรมประจำวัน เป็นต้น ซึ่งจะเหลือจำนวนพารามิเตอร์ทั้งหมด 18 ตัว

3.1.3. ทำการแปลงข้อมูล (Data transformation) ของพารามิเตอร์ที่มีลักษณะเป็นหน่วยเวลาหรือพารามิเตอร์ที่เป็นตัวแปรนามบัญญัติ (Nominal Variable) ให้เป็นตัวแปรแบบช่วง (Interval Variable) เพื่อใช้สำหรับกระบวนการสร้างแบบจำลอง

3.1.4. ทำการเติมค่าที่หายไป (Missing value) ของแต่ละพารามิเตอร์ โดยคำนึกถึงข้อมูลของแต่ละบุคคล กล่าวคือ การเติมค่าที่หายไปจะทำการเติมตามข้อมูลรายบุคคลทั้งหมด 60 คน เพื่อให้ได้ค่าที่หายไปที่เหมาะสมตามรายบุคคล สาเหตุมาจากข้อมูลพฤติกรรมการนอนหลับของแต่ละคนไม่เหมือนกันซึ่งจะส่งผลให้หากใช้การเติมค่าที่

หายไปในข้อมูลทั้งหมดจะทำให้ค่าที่ได้นั่นอาจมีความไม่เหมาะสมที่จะนำไปวิเคราะห์ต่อ ดังนั้นการเติมค่าที่หายไปตามรายบุคคลย่อมส่งผลที่ดีกว่าการเติมจากข้อมูลทั้งหมด โดยสำหรับพารามิเตอร์ที่เป็นตัวแปรนามบัญญัติจะเติมค่าที่หายไปด้วยค่ามัธยฐาน และพารามิเตอร์ที่เป็นตัวแปรแบบช่วงจะเติมค่าที่หายไปด้วยค่าเฉลี่ย

3.4. การวัดประสิทธิภาพแบบจำลองการจำแนกพฤติกรรมการนอนหลับ

การวัดประสิทธิภาพแบบจำลองการจำแนกพฤติกรรมการนอนหลับ ด้วยการวัดค่าความถูกต้องจากวิธีการตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัว โดยการตรวจสอบนี้จะใช้การสุ่มข้อมูล 1 แถว กำหนดให้เป็นข้อมูลทดสอบ ส่วนข้อมูลของแถวที่เหลือทั้งหมดจะถูกกำหนดให้เป็นข้อมูลการสอน และจะทำการสร้างแบบจำลองและวัดประสิทธิภาพแบบจำลอง ต่อมาจะทำการสุ่มข้อมูล 1 แถวสำหรับเป็นข้อมูลทดสอบและจะสุ่มวนจนกระทั่งข้อมูลทุกแถวถูกสุ่มให้เป็นข้อมูลทดสอบครบทั้งหมด และจะนำค่าผิดพลาดจากการวัดประสิทธิภาพแบบจำลองมาคำนวณเฉลี่ยและกำหนดให้เป็นค่าความถูกต้องของแบบจำลอง ยกตัวอย่างการกำหนดข้อมูลทดสอบและการสอนของวิธีการตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัวของการจำแนกพฤติกรรมการนอนหลับตามรายบุคคล ดังนี้

รอบที่ 1 : กำหนดข้อมูลแถวที่ 1 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 2 : กำหนดข้อมูลแถวที่ 2 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 3 : กำหนดข้อมูลแถวที่ 3 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รอบที่ 4 : กำหนดข้อมูลแถวที่ 4 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 5 : กำหนดข้อมูลแถวที่ 5 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 6 : กำหนดข้อมูลแถวที่ 6 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 7 : กำหนดข้อมูลแถวที่ 7 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 8 : กำหนดข้อมูลแถวที่ 8 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รอบที่ 9 : กำหนดข้อมูลแถวที่ 9 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 10 : กำหนดข้อมูลแถวที่ 10 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 11 : กำหนดข้อมูลแถวที่ 11 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 12 : กำหนดข้อมูลแถวที่ 12 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 13 : กำหนดข้อมูลแถวที่ 13 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

รอบที่ 14 : กำหนดข้อมูลแถวที่ 14 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, ... , 47 และ 48 เป็นข้อมูลการสอน

จนกระทั่งถึงข้อมูลแถวสุดท้าย คือ

รอบที่ 48 : กำหนดข้อมูลแถวที่ 48 เป็นข้อมูลทดสอบ

กำหนดข้อมูลแถวที่ 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... , 46 และ 47 เป็นข้อมูลการสอน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4
การทดลองและผลการทดลอง

เป้าหมายของงานวิจัยนี้ คือ การใช้เทคนิคการเรียนรู้ของเครื่องประกอบด้วยการวิเคราะห์องค์ประกอบหลัก อัลกอริทึมเคมีน การถดถอยโลจิสติก และการป่าแบบสุ่มสำหรับการจำแนกพฤติกรรมการนอนหลับ เพื่อให้สามารถทำนายหรือพยากรณ์พฤติกรรมนอนหลับที่ส่งผลให้ผู้คนมีการนอนหลับที่ดีขึ้น และใช้การตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัวในการวัดและเปรียบเทียบประสิทธิภาพของแบบจำลอง ซึ่งในการทดลองนี้จะใช้โปรแกรมไซคิตเลิร์นและเขียนโปรแกรมด้วยภาษาไพทอน

โดยสามารถแบ่งการทดลองออกเป็น 3 ส่วนใหญ่ ได้แก่

1.การลดมิติข้อมูลด้วยเทคนิคการวิเคราะห์องค์ประกอบหลัก

2.การจัดรูปแบบข้อมูล

ซึ่งประกอบด้วยข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่มสำหรับสร้างแบบจำลอง โดยข้อมูลรายกลุ่มจะใช้อัลกอริทึมเคมีนในการจัดกลุ่ม

3.การสร้างและทดสอบประสิทธิภาพแบบจำลองการจำแนกพฤติกรรมการนอนหลับ

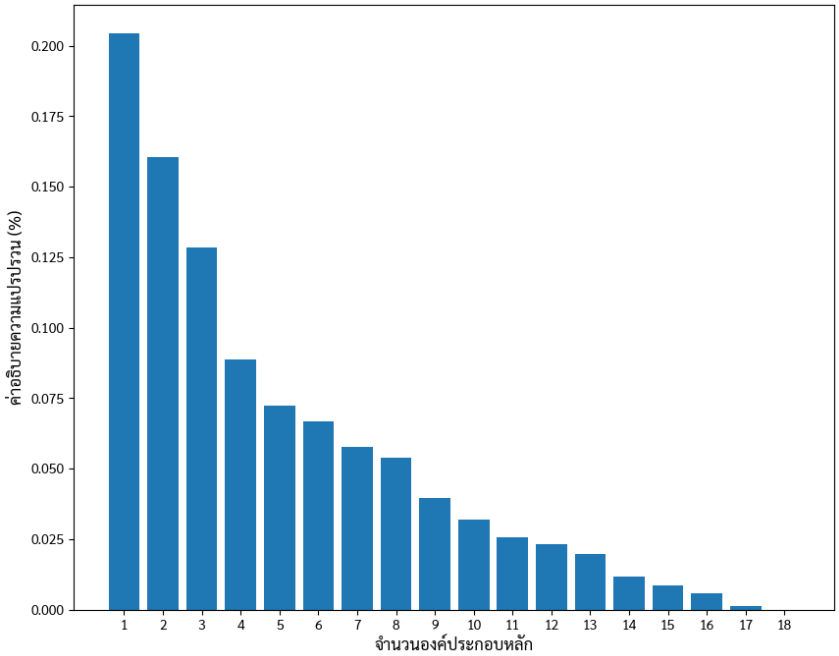
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.1. การทดลอง

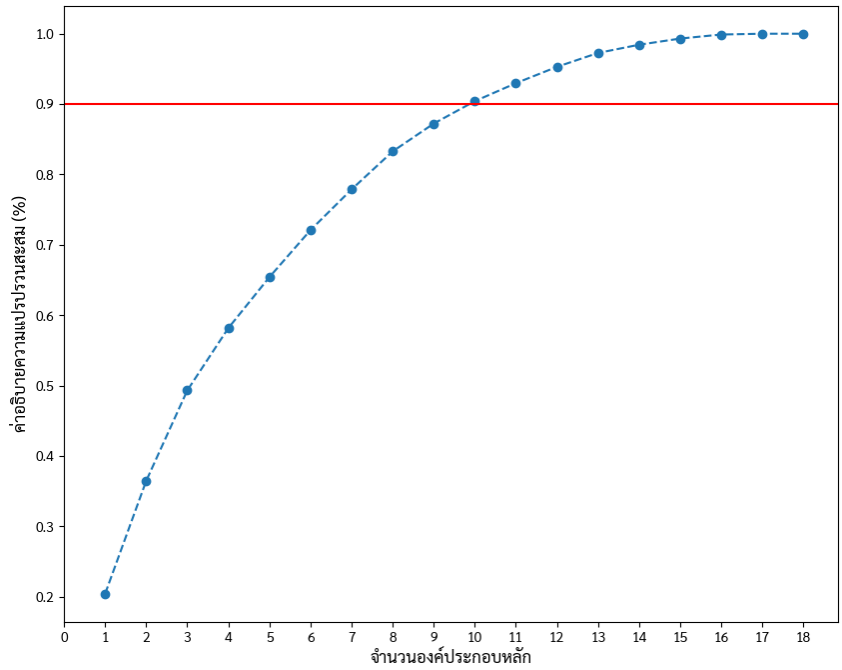
4.1.1.การลดมิติข้อมูลด้วยเทคนิคการวิเคราะห์องค์ประกอบหลัก

นำข้อมูลทั้งหมดที่ผ่านการเตรียมข้อมูลเข้ากระบวนการวิเคราะห์องค์ประกอบหลักและวิเคราะห์หาค่าเฉพาะเจาะจง หรือ Eigenvalue แล้ว ขั้นตอนที่สำคัญต่อมา คือ การเลือกจำนวนองค์ประกอบหลัก หรือ principal components สำหรับกำหนดจำนวนตัวแปรสำคัญที่จะใช้ในการวิเคราะห์ข้อมูลต่อไป โดยจะสามารถเลือกได้จากการวาดกราฟเทียบระหว่างค่าอธิบายความแปรปรวนสะสม (Cumulative explained variance) ที่ซึ่งจะเป็นค่าที่แสดงให้เห็นถึงความแปรปรวนที่สามารถอธิบายได้ด้วยองค์ประกอบหลักแต่ละตัว โดยยิ่งค่าความแปรปรวนสะสมสูง

แสดงถึงความสำคัญขององค์ประกอบหลัก เทียบกับจำนวนองค์ประกอบหลัก ซึ่งผลลัพธ์ที่ได้สามารถวาดกราฟออกมาดังรูปภาพที่ 9 และ 10



รูปภาพที่ 9 รูปภาพแสดงกราฟเปรียบเทียบระหว่างค่าอธิบายความแปรปรวนกับจำนวนองค์ประกอบหลัก



รูปภาพที่ 10 รูปภาพแสดงกราฟเปรียบเทียบระหว่างค่าอธิบายความแปรปรวนสะสมกับจำนวนองค์ประกอบหลัก

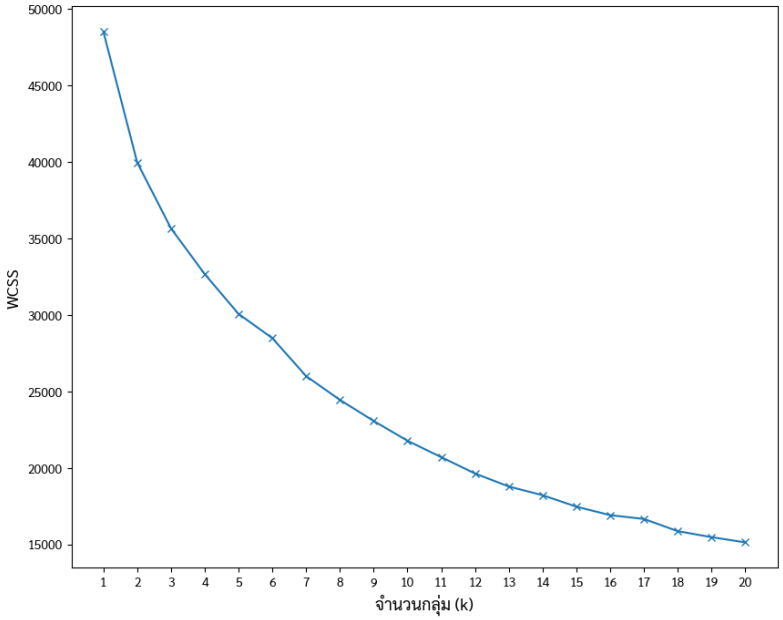
จากรูปภาพที่ 9 แสดงให้เห็นถึงค่าความแปรปรวนที่มีหน่วยเป็นเปอร์เซ็นต์ของแต่ละจำนวนองค์ประกอบหลักที่ถูกใช้ และเพื่อให้ง่ายต่อการวิเคราะห์และเลือกจำนวนองค์ประกอบหลักที่เหมาะสม จึงนำค่าความแปรปรวนที่ได้มาคิดเป็นค่าสะสมดังรูปภาพที่ 10 จากรูปภาพที่ 10 แสดงให้เห็นว่าจำนวนองค์ประกอบหลักเท่ากับ 10 มีระดับนัยสำคัญของค่าความแปรปรวนสะสมที่ 90% ซึ่งเป็นระดับที่ผู้วิจัยตัดสินใจเลือกใช้ ดังนั้นจึงได้ว่าจากผลการทดลองการวิเคราะห์องค์ประกอบหลักสำหรับการลดความซับซ้อนและลดมิติของข้อมูล จำนวนองค์ประกอบหลักเท่ากับ 10 เป็นจำนวนที่เหมาะสมในการกำหนดและนำไปสร้างตัวแปรใหม่ที่จะถูกใช้เป็นข้อมูลสำหรับการทดลองขั้นต่อไป

4.1.2.การจัดรูปแบบข้อมูล

การจัดรูปแบบข้อมูลให้อยู่ในลักษณะ 3 ประเภท คือ ข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่ม เนื่องจากข้อมูลทั้งหมดและข้อมูลรายบุคคลสามารถนำผลจากการทดลองที่ 4.1.1. มาใช้ในการสร้างแบบจำลองการจำแนกได้โดยตรง เพียงแต่ข้อมูลรายบุคคลจะต้องแบ่งข้อมูลทั้งหมดออกเป็น 60 ชุดตามจำนวนคนทั้งหมด 60 คนก่อน ส่วนข้อมูลรายกลุ่มจะต้องนำผลการทดลองที่ได้เข้าสู่กระบวนการจัดกลุ่มด้วยอัลกอริทึมเคมีนต่อไป

สำหรับการเลือกค่า k ที่เหมาะสมในการกำหนดจำนวนกลุ่มสำหรับอัลกอริทึมเคมีน ผู้วิจัยใช้การวาดกราฟเทียบระหว่างค่า WCSS กับจำนวนกลุ่มได้ผลลัพธ์ดังรูปภาพที่ 11

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปภาพที่ 11 รูปภาพแสดงกราฟเทียบระหว่างค่า WCSS กับจำนวนกลุ่ม (k)

จากรูปภาพที่ 11 สามารถอธิบายได้ว่าการเปรียบเทียบค่า WCSS กับจำนวนกลุ่ม (k) ที่ k=7 มีค่า WCSS ที่ลดลงอย่างมีนัยสำคัญและมีการลดลงในลักษณะของการหักศอกอย่างเห็นได้ชัด ดังนั้นผู้วิจัยจึงตัดสินใจเลือก k=7 สำหรับกำหนดจำนวนกลุ่ม เมื่อสามารถเลือกจำนวนกลุ่มที่เหมาะสมและนำข้อมูลเข้ากระบวนการอัลกอริทึมเคมีนแล้วขั้นตอนต่อไป คือ การจัดกลุ่มคนที่มีความน่าจะเป็นสูงสุดที่จะถูกจัดให้อยู่กลุ่มนั้นโดยอาศัยการคำนวณสัดส่วนของกลุ่มสูงสุดตามข้อมูลแต่ละคน ซึ่งได้ผลลัพธ์ดังตารางที่ 1 และตารางที่ 2

ตารางที่ 1 ตารางแสดงสัดส่วนของกลุ่มตามการติดป้ายกำกับจากอัลกอริทึมเคมีนตามรายบุคคล

คนที่	กลุ่มที่	จำนวนที่ถูกติดป้ายกำกับ	สัดส่วนของกลุ่ม (%)
1	4	36	73.5%
1	1	9	18.4%
1	7	3	6.1%
1	3	1	2.0%
2	1	35	71.4%
2	2	13	26.5%
2	7	1	2.0%
3	1	33	67.3%
3	2	10	20.4%

คนที่	กลุ่มที่	จำนวนที่ถูกติดป้ายกำกับ	สัดส่วนของกลุ่ม (%)
3	4	4	8.2%
3	7	1	2.0%
3	3	1	2.0%
4	1	39	79.6%
4	2	5	10.2%
4	4	3	6.1%
4	7	1	2.0%
4	3	1	2.0%
5	4	42	85.7%
5	3	3	6.1%
5	7	2	4.1%
5	2	1	2.0%
5	5	1	2.0%
6	1	22	44.9%
6	4	14	28.6%
6	2	7	14.3%
6	7	3	6.1%
6	3	2	4.1%
6	6	1	2.0%
7	4	47	95.9%
7	7	1	2.0%
7	1	1	2.0%
8	1	46	93.9%
8	4	2	4.1%
8	7	1	2.0%
9	1	47	95.9%
9	2	2	4.1%
10	4	37	75.5%
10	7	4	8.2%
10	2	4	8.2%
10	1	4	8.2%
11	4	42	85.7%
11	1	5	10.2%
11	3	1	2.0%
11	2	1	2.0%
12	2	20	40.8%
12	5	14	28.6%
12	7	5	10.2%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คนที่	กลุ่มที่	จำนวนที่ถูกติดป้ายกำกับ	สัดส่วนของกลุ่ม (%)
12	4	3	6.1%
12	3	3	6.1%
12	1	3	6.1%
12	6	1	2.0%
13	4	26	53.1%
13	1	20	40.8%
13	2	1	2.0%
13	7	1	2.0%
13	3	1	2.0%
14	1	23	46.9%
14	4	15	30.6%
14	2	6	12.2%
14	3	2	4.1%
14	7	2	4.1%
14	6	1	2.0%
15	1	25	51.0%
15	4	20	40.8%
15	7	3	6.1%
15	2	1	2.0%
16	4	37	75.5%
16	1	5	10.2%
16	2	5	10.2%
16	7	2	4.1%
17	4	20	40.8%
17	1	11	22.4%
17	2	7	14.3%
17	7	8	16.3%
17	3	3	6.1%
18	4	46	93.9%
18	7	2	4.1%
18	3	1	2.0%
19	2	44	89.8%
19	4	2	4.1%
19	3	2	4.1%
19	7	1	2.0%
20	4	48	98.0%
20	6	1	2.0%
21	1	40	81.6%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คนที่	กลุ่มที่	จำนวนที่ถูกติดป้ายกำกับ	สัดส่วนของกลุ่ม (%)
21	4	5	10.2%
21	3	4	8.2%
22	2	25	51.0%
22	5	15	30.6%
22	3	7	14.3%
22	1	2	4.1%
23	1	40	81.6%
23	2	6	12.2%
23	4	3	6.1%
24	2	23	46.9%
24	1	17	34.7%
24	3	5	10.2%
24	5	4	8.2%
25	4	25	51.0%
25	3	22	44.9%
25	1	2	4.1%
26	4	41	83.7%
26	1	3	6.1%
26	3	2	4.1%
26	5	2	4.1%
26	2	1	2.0%
27	4	32	65.3%
27	3	10	20.4%
27	1	7	14.3%
28	4	28	57.1%
28	3	11	22.4%
28	1	10	20.4%
29	4	32	65.3%
29	2	10	20.4%
29	5	3	6.1%
29	1	3	6.1%
29	3	1	2.0%
30	2	35	71.4%
30	1	13	26.5%
30	4	1	2.0%
31	4	42	85.7%
31	1	7	14.3%
32	4	15	30.6%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คนที่	กลุ่มที่	จำนวนที่ถูกติดป้ายกำกับ	สัดส่วนของกลุ่ม (%)
32	1	10	20.4%
32	2	10	20.4%
32	3	8	16.3%
32	5	6	12.2%
33	1	20	40.8%
33	2	20	40.8%
33	3	7	14.3%
33	4	1	2.0%
33	5	1	2.0%
34	4	47	95.9%
34	7	1	2.0%
34	1	1	2.0%
35	1	24	49.0%
35	4	20	40.8%
35	3	5	10.2%
36	2	46	93.9%
36	1	2	4.1%
36	5	1	2.0%
37	1	32	65.3%
37	2	9	18.4%
37	3	6	12.2%
37	4	2	4.1%
38	1	33	67.3%
38	2	7	14.3%
38	4	5	10.2%
38	3	4	8.2%
39	1	39	79.6%
39	4	7	14.3%
39	2	2	4.1%
39	3	1	2.0%
40	2	32	65.3%
40	3	15	30.6%
40	5	2	4.1%
41	2	45	91.8%
41	3	3	6.1%
41	1	1	2.0%
42	2	37	75.5%
42	1	11	22.4%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คนที่	กลุ่มที่	จำนวนที่ถูกติดป้ายกำกับ	สัดส่วนของกลุ่ม (%)
42	5	1	2.0%
43	4	33	67.3%
43	2	6	12.2%
43	1	6	12.2%
43	3	4	8.2%
44	2	44	89.8%
44	1	4	8.2%
44	7	1	2.0%
45	1	37	75.5%
45	2	6	12.2%
45	4	5	10.2%
45	3	1	2.0%
46	2	48	98.0%
46	7	1	2.0%
47	4	35	71.4%
47	1	11	22.4%
47	3	2	4.1%
47	5	1	2.0%
48	4	35	71.4%
48	3	12	24.5%
48	1	2	4.1%
49	2	18	36.7%
49	1	16	32.7%
49	4	8	16.3%
49	3	7	14.3%
50	2	28	57.1%
50	1	16	32.7%
50	3	5	10.2%
51	1	47	95.9%
51	5	2	4.1%
52	2	22	44.9%
52	1	16	32.7%
52	3	9	18.4%
52	4	2	4.1%
53	2	27	55.1%
53	1	17	34.7%
53	4	3	6.1%
53	3	2	4.1%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คนที่	กลุ่มที่	จำนวนที่ถูกติดป้ายกำกับ	สัดส่วนของกลุ่ม (%)
54	1	39	79.6%
54	4	6	12.2%
54	2	3	6.1%
54	3	1	2.0%
55	1	36	73.5%
55	4	12	24.5%
55	2	1	2.0%
56	1	48	98.0%
56	3	1	2.0%
57	1	48	98.0%
57	2	1	2.0%
58	2	20	40.8%
58	1	16	32.7%
58	5	6	12.2%
58	4	4	8.2%
58	3	3	6.1%
59	1	37	75.5%
59	3	7	14.3%
59	4	5	10.2%
60	1	29	59.2%
60	3	16	32.7%
60	2	3	6.1%
60	4	1	2.0%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 2 ตารางแสดงจำนวนคนที่ถูกจัดกลุ่มตามสัดส่วนของกลุ่มสูงสุดตามรายบุคคล

กลุ่มที่	จำนวนสมาชิกในกลุ่ม	เปอร์เซ็นต์เทียบของจำนวนสมาชิกในกลุ่ม กับสมาชิกทั้งหมด
1	23	38.3%
2	16	26.7%
3	0	-
4	21	35%
5	0	-
6	0	-
7	0	-

จากตารางที่ 1 และ 2 สามารถสรุปได้ว่าการจัดกลุ่มด้วยอัลกอริทึมเคมีนและการคำนวณสัดส่วนของกลุ่มสูงสุดของกลุ่มสามารถแบ่งกลุ่มออกมาได้ทั้งหมด 7 กลุ่ม ซึ่งแต่ละกลุ่มประกอบด้วยจำนวนสมาชิก 23 16 0 21 0 0 และ 0 ตามลำดับ เนื่องจากกลุ่มที่ 3 5 6 และ 7 ไม่มีจำนวนสมาชิกกลุ่ม ดังนั้นจึงได้ว่าจากจำนวนกลุ่มทั้งหมด 7 กลุ่มจะถูกลดลงเหลือเพียง 3 กลุ่มสำหรับนำไปสร้างแบบจำลองการจำแนกรายกลุ่มต่อไป อีกทั้งเปอร์เซ็นต์ของจำนวนสมาชิกในกลุ่มเทียบกับจำนวนสมาชิกทั้งหมดของกลุ่มที่ 1 2 และ 4 มีค่าเท่ากับ 38.3% 26.7% และ 35% ตามลำดับ

4.1.3.การสร้างและทดสอบประสิทธิภาพแบบจำลองการจำแนกพฤติกรรมการนอนหลับ

เมื่อเตรียมข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่มแล้วจึงนำข้อมูลเหล่านี้ไปสร้างแบบจำลองการจำแนกด้วยการถดถอยโลจิสติกและการป่าแบบสุ่มและทดสอบประสิทธิภาพแบบจำลองด้วยตรวจสอบไขว้แบบดึงข้อมูลออกทีละตัว ได้ผลลัพธ์ดังตารางที่ 3 4 5 และสามารถสรุปออกได้ดังตารางที่ 6 และ 7

ตารางที่ 3 ตารางแสดงค่าความถูกต้องจากแบบจำลองการจำแนกด้วยข้อมูลทั้งหมด

ค่าความถูกต้อง (%)	
การถดถอยโลจิสติก	การป่าแบบสุ่ม
60.3%	61.1%

ตารางที่ 4 ตารางแสดงค่าความถูกต้องจากแบบจำลองการจำแนกด้วยข้อมูลรายบุคคล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คนที่	ค่าความถูกต้อง (%)	
	การถดถอยโลจิสติก	การป่าแบบสุ่ม
1	66.0%	64.0%
2	64.0%	58.0%
3	50.0%	32.0%
4	54.0%	56.0%
5	66.0%	60.0%
6	58.0%	54.0%
7	64.0%	64.0%
8	66.0%	56.0%
9	56.0%	66.0%
10	62.0%	64.0%
11	64.0%	62.0%

คนที่	ค่าความถูกต้อง (%)	
	การถดถอยโลจิสติก	การป่าแบบสุ่ม
12	62.0%	68.0%
13	62.0%	60.0%
14	60.0%	68.0%
15	62.0%	62.0%
16	70.0%	76.0%
17	68.0%	62.0%
18	70.0%	64.0%
19	60.0%	56.0%
20	70.0%	66.0%
21	62.5%	72.9%
22	70.8%	56.3%
23	64.6%	60.4%
24	66.7%	64.6%
25	64.6%	64.6%
26	75.0%	64.6%
27	62.5%	70.8%
28	66.7%	64.6%
29	64.6%	68.8%
30	52.1%	56.3%
31	66.7%	52.1%
32	64.6%	72.9%
33	72.9%	70.8%
34	50.0%	60.4%
35	56.2%	60.4%
36	58.3%	54.2%
37	64.6%	56.3%
38	62.5%	54.2%
39	66.7%	81.3%
40	52.1%	68.8%
41	35.4%	64.6%
42	70.8%	72.9%
43	60.4%	56.3%
44	56.2%	56.3%
45	58.3%	56.3%
46	64.6%	52.1%
47	70.8%	68.8%
48	77.1%	60.4%
49	62.5%	58.3%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คนที่	ค่าความถูกต้อง (%)	
	การถดถอยโลจิสติก	การป่าแบบสุ่ม
50	62.5%	64.6%
51	70.8%	83.3%
52	62.5%	54.2%
53	66.7%	43.8%
54	68.8%	60.4%
55	66.7%	62.5%
56	66.7%	70.8%
57	60.4%	60.4%
58	68.8%	64.6%
59	56.2%	56.3%
60	52.1%	60.4%

ตารางที่ 5 ตารางแสดงค่าความถูกต้องจากแบบจำลองการจำแนกด้วยข้อมูลรายกลุ่ม

กลุ่มที่	ค่าความถูกต้อง (%)	
	การถดถอยโลจิสติก	การป่าแบบสุ่ม
1	58.5%	61.1%
2	64.1%	65.5%
3	62%	67.1%

ตารางที่ 6 ตารางแสดงการเปรียบเทียบค่าความถูกต้องระหว่างข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่มด้วยการถดถอยโลจิสติก

ข้อมูล	ค่าความถูกต้อง (%)			
	ค่าน้อยสุด	ค่ามากสุด	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน
ข้อมูลทั้งหมด	-	-	60.3%	-
ข้อมูลรายบุคคล	35.4%	77.1%	62.9%	7
ข้อมูลรายกลุ่ม	58.5%	64.1%	61.5%	2.9

ตารางที่ 7 ตารางแสดงการเปรียบเทียบค่าความถูกต้องระหว่างข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่มด้วยการป่าแบบสุ่ม

ข้อมูล	ค่าความถูกต้อง (%)			
	ค่าน้อยสุด	ค่ามากสุด	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน
ข้อมูลทั้งหมด	-	-	61.1%	-
ข้อมูลรายบุคคล	32%	83.3%	62%	8.2
ข้อมูลรายกลุ่ม	61.1%	67.1%	64.6%	3.1

จากตารางที่ 6 สามารถอธิบายได้ว่าการสร้างแบบจำลองการจำแนกด้วยการถดถอยโลจิสติกจากข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่มมีค่าเฉลี่ยของค่าความถูกต้องของแบบจำลอง คือ 60.3% 62.9% และ 61.5% ตามลำดับ มีค่าน้อยสุดจากข้อมูลรายบุคคลและข้อมูลรายกลุ่ม คือ 35.4% และ 58.5% ค่ามากสุด คือ 77.1% และ 64.1% และส่วนเบี่ยงเบนมาตรฐาน คือ 7 และ 2.9 ตามลำดับ

และจากตารางที่ 7 สามารถอธิบายการสร้างแบบจำลองการจำแนกด้วยการป่าแบบสุ่มจากข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่มได้ว่ามีค่าเฉลี่ยของค่าความถูกต้อง คือ 61.1% 62% และ 64.6% ตามลำดับ มีค่าน้อยสุดจากข้อมูลรายบุคคลและข้อมูลรายกลุ่ม คือ 32% และ 61.1% ค่ามากสุด คือ 83.3% และ 67.1% และส่วนเบี่ยงเบนมาตรฐาน คือ 8.2 และ 3.1 ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 5
สรุปผลการทดลอง

ในบทนี้จะกล่าวถึงบทสรุปของผลการทดลองจากบทที่ 4 ของการจำแนกพฤติกรรมนอนหลับที่ส่งผลให้มีการนอนหลับที่ดีขึ้นด้วยข้อมูล 3 ชุด ได้แก่ ข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่มด้วยการถดถอยโลจิสติกและการสุ่มแแบบป่า

5.1.สรุปผลการทดลอง

งานวิจัยนี้แสดงให้เห็นถึงวิธีการใช้เทคนิคการเรียนรู้ของเครื่องและเทคนิคอื่น ๆ สำหรับการลดมิติข้อมูลที่มีจำนวนพารามิเตอร์เยอะรวมถึงใช้ในการจำแนกพฤติกรรมการนอนหลับที่ดีขึ้น โดยได้นำเสนอการใช้ข้อมูลที่มีรูปแบบแตกต่างกัน 3 รูปแบบ คือ ข้อมูลทั้งหมด ข้อมูลรายบุคคล และข้อมูลรายกลุ่มในการสร้าง ทดสอบ และเปรียบเทียบประสิทธิภาพของแบบจำลอง

จากผลการทดลองที่ 4.1.3. สามารถสรุปได้ว่าค่าความถูกต้องจากแบบจำลองการจำแนกด้วยข้อมูลรายบุคคลให้ค่าความถูกต้องน้อยที่สุดจากการสร้างแบบจำลองด้วยวิธีการถดถอยโลจิสติกและการป่าแบบสุ่ม คือ 35.4% และ 32% ซึ่งเมื่อนำไปเปรียบเทียบกับแบบจำลองการจำแนกของข้อมูลรายกลุ่มจะพบว่าข้อมูลที่ถูกจัดกลุ่มด้วยอัลกอริทึมเคมีนสามารถส่งผลให้ค่าความถูกต้องของแบบจำลองมีค่าเพิ่มขึ้นจากแบบจำลองด้วยข้อมูลรายบุคคลของทั้ง 2 วิธีโดยสูงกว่าสูงสุดได้ถึงประมาณ 48% อีกทั้งยังให้ค่าความถูกต้องเฉลี่ยและค่าความถูกต้องสูงสุดมากกว่าแบบจำลองการจำแนกด้วยข้อมูลทั้งหมด

และจากผลการทดลองแสดงให้เห็นว่าการจำแนกพฤติกรรมการนอนหลับด้วยข้อมูลรายกลุ่มให้ผลลัพธ์ที่ดีกว่าการใช้ข้อมูลทั้งหมดและข้อมูลรายบุคคล อีกทั้งการจำแนกพฤติกรรมนอนหลับของข้อมูลรายกลุ่มด้วยวิธีการป่าแบบสุ่มให้ผลลัพธ์ที่ดีกว่าวิธีการถดถอยโลจิสติกอยู่ระหว่าง 61.1% ถึง 67.1% หรือมีค่าสูงกว่าระหว่าง 2.1% ถึง 7.6%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ดังนั้นจากงานวิจัยนี้สามารถสรุปได้ว่าการจัดรูปแบบพฤติกรรมนอนหลับของคนที่มีลักษณะการนอนหลับที่คล้ายคลึงกันมาอยู่กลุ่มเดียวกันย่อมส่งผลต่อการจำแนกการนอนหลับที่ดีกว่าการใช้ข้อมูลทั้งหมดรวมถึงการใช้ข้อมูลรายบุคคลมาวิเคราะห์

5.2. งานวิจัยในอนาคต

การนอนหลับเป็นสิ่งที่สำคัญมากต่อการดำรงชีวิต การมีการนอนหลับที่ดีย่อมส่งผลให้มีชีวิตที่ดีขึ้น ดังนั้นการต่อยอดงานวิจัยนี้จึงมีความน่าสนใจ สำหรับการศึกษาต่อยอดควรใช้ข้อมูลเพิ่มเติมที่มีความหลากหลายเข้ามาศึกษาร่วมด้วยมากขึ้น เช่น อัตราการเต้นของหัวใจขณะนอนหลับ ช่วงวงจรการนอนหลับ เป็นต้น รวมถึงการใช้การเรียนรู้ของเครื่องเชิงลึกเข้ามาวิเคราะห์พฤติกรรมนอนหลับที่ส่งผลให้นอนหลับได้ดีขึ้นหรือแย่ลง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บรรณานุกรม

1. Vyazovskiy, V.V., *Sleep, recovery, and metaregulation: explaining the benefits of sleep.* Nature and science of sleep, 2015: p. 171-184.
2. Grandner, M.A., *Sleep, health, and society.* Sleep medicine clinics, 2017. **12**(1): p. 1-22.
3. Hirshkowitz, M., et al., *National Sleep Foundation's updated sleep duration recommendations.* Sleep health, 2015. **1**(4): p. 233-243.
4. Luyster, F.S., et al., *Sleep: a health imperative.* Sleep, 2012. **35**(6): p. 727-734.
5. Cheng, S.H., et al., *A study on the sleep quality of incoming university students.* Psychiatry research, 2012. **197**(3): p. 270-274.
6. Sathyanarayana, A., et al., *Sleep quality prediction from wearable data using deep learning.* JMIR mHealth and uHealth, 2016. **4**(4): p. e6562.
7. Nelson, K.L., J.E. Davis, and C.F. Corbett. *Sleep quality: An evolutionary concept analysis*. in *Nursing forum*. 2022. Wiley Online Library.
8. Hasan, B.M.S. and A.M. Abdulazeez, *A review of principal component analysis algorithm for dimensionality reduction.* Journal of Soft Computing and Data Mining, 2021. **2**(1): p. 20-30.
9. Shlens, J., *A tutorial on principal component analysis.* arXiv preprint arXiv:1404.1100, 2014.
10. Frades, I. and R. Matthiesen, *Overview on techniques in cluster analysis.* Bioinformatics methods in clinical research, 2010: p. 81-107.
11. Kuraria, A., N. Jharbade, and M. Soni, *Centroid Selection Process Using WCSS and Elbow Method for K-Mean Clustering Algorithm in Data Mining.* International Journal of Scientific Research in Science, Engineering and Technology, 2018: p. 190-195.
12. Park, H.-A., *An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain.* Journal of Korean Academy of Nursing, 2013. **43**(2): p. 154-164.
13. Breiman, L., *Random forests.* Machine learning, 2001. **45**: p. 5-32.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

14. Cutler, A., D.R. Cutler, and J.R. Stevens, *Random forests.* Ensemble machine learning: Methods and applications, 2012: p. 157-175.
15. Berrar, D., *Cross-Validation*. 2019.
16. Wong, T.-T., *Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation.* Pattern Recognition, 2015. **48**(9): p. 2839-2846.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียน

ชื่อ-สกุล ปวรรัตน์ ขุมเงิน

วุฒิการศึกษา ปริญญาตรี วิทยาศาสตรบัณฑิต มหาวิทยาลัยธรรมศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University