บทที่ 4

การทดลองและผลการทดลอง

บทนี้จะกล่าวถึงรายละเอียดของการทดลองและผลการทดลองที่ได้จากการใช้เทคนิคการแตกครึ่งตามสารสนเทศโดยวัดประสิทธิภาพจากค่าจำนวนครั้งในการจำแนกข้อมูลเฉลี่ยและความถูกต้องในการจำแนกข้อมูลซึ่งจะทำการเปรียบเทียบกับเทคนิคอื่นๆได้แก่ การแตกครึ่งแบบสมดุล อาร์เอดีเอจีและแมกซ์วิน

เนื้อหาในบทนี้จะแบ่งออกเป็น 3 ส่วนด้วยกัน โดยเนื้อหาในส่วนแรกจะกล่าวถึงชุดข้อมูลที่ใช้ในการทดลอง ส่วนต่อมากล่าวถึงรายละเอียดของวิธีการทดลอง และเนื้อหาในส่วนสุดท้ายจะกล่าวถึงผลการทดลอง

4.1 ชุดข้อมูลที่ใช้ในการทดลอง

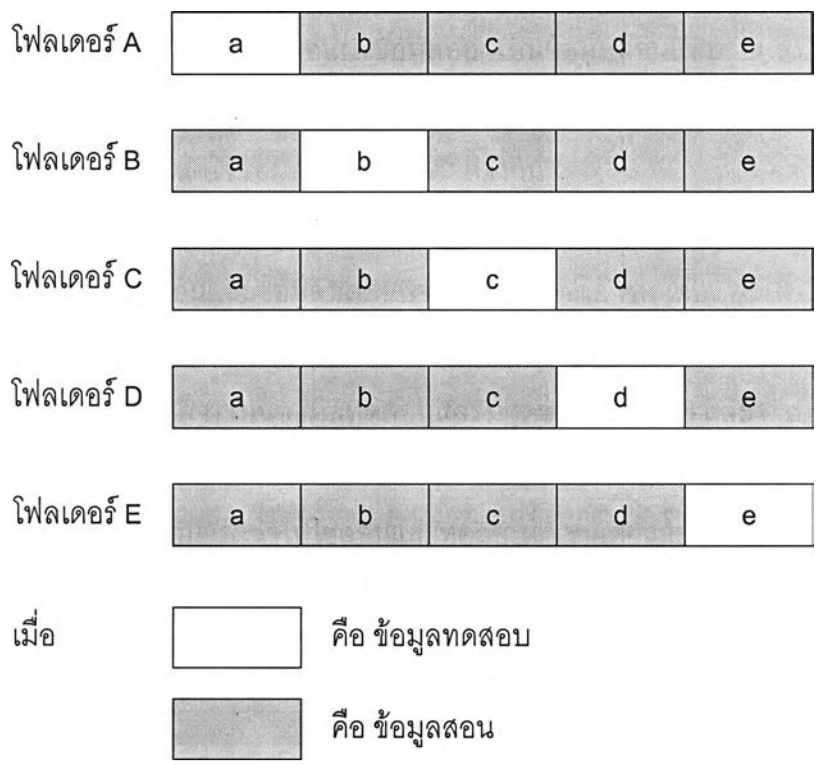
สำหรับการทดลองได้ใช้ข้อมูลทดสอบจาก UCI Machine Learning Repository [8] จำนวน 8 ชุด และ Thai printed character recognition dataset [9] จำนวน 2 ชุด รวมข้อมูลทดสอบจำนวน 10 ชุด ดังตารางที่ 1 โดยในตารางจะแสดงรายละเอียดของชุดข้อมูลที่นำมาทดสอบ ได้แก่ จำนวนข้อมูลสอน (Training Data) จำนวนข้อมูลทดสอบ (Test Data) จำนวนประเภทข้อมูล (Class) และจำนวนคุณสมบัติ (Feature) ของชุดข้อมูลต่าง ๆ

ตารางที่ 1: ข้อมูลที่ใช้ในการทดลองจาก UCI Machine Learning Repository และ Thai printed character recognition dataset

No	Dataset	#training set	#test set	#class	#feature
1.	Glass	214	5-fold	6	9
2.	Satimage	4,435	2,000	6	36
3.	Segment	2,310	5-fold	7	18
4.	Shuttle	43,500	14,500	7	9
5.	Vowel	528	462	11	10
6.	Soybean	290	340	15	35
7.	Letter	15,963	4,037	26	16
8.	Isolet	6,238	1,559	26	617
9.	Thaiprinted Character1	3,264	3,264	68	128
10.	Thaiprinted Character2	3,264	3,264	68	128

สำหรับชุดข้อมูล Satimage, Shuttle, Vowel, Soybean, Letter และ Isolet ลักษณะของข้อมูลที่นำมาทดสอบแต่ละชุดจะประกอบด้วย ข้อมูลสอน (Training data) และข้อมูลทดสอบ (Test data) สำหรับ Glass และ Segment มีข้อมูลเพียงชุดเดียวจึงใช้การตรวจสอบไขว้ 5 พับ (5-folds Cross Validation) เพื่อแบ่งข้อมูลออกเป็นข้อมูลสอนและข้อมูลทดสอบ ส่วน Thai printed character recognition dataset ซึ่งประกอบด้วยข้อมูลจำนวน 2 ชุดนั้นข้อมูลทั้งสองชุดใช้ข้อมูลสอนชุดเดียวกันแต่ใช้ข้อมูลทดสอบที่แตกต่างกัน

การตรวจสอบไขว้ 5 พับทำได้โดยแบ่งข้อมูลที่มีออกเป็น 5 ส่วน คือ a, b, c, d และ e เพื่อสร้างเป็นชุดข้อมูลสอนและข้อมูลทดสอบจำนวน 5 ชุด คือ โฟลเดอร์ A, B, C, D และ E โดยที่โฟลเดอร์ A จะมีชุดข้อมูล a เป็นข้อมูลทดสอบและมีชุดข้อมูล b, c, d และ e เป็นข้อมูลสอน ส่วนโฟลเดอร์ B จะมีชุดข้อมูล b เป็นข้อมูลทดสอบและมีชุดข้อมูล a, c, d และ e เป็นข้อมูลสอน เป็นแบบนี้ตามลำดับไปจนถึงโฟลเดอร์ E ก็จะได้ข้อมูลสอนและข้อมูลทดสอบจำนวน 5 ชุด ดังแสดงในรูปที่ 19 ซึ่งสามารถนำไปใช้สอนและทดสอบได้เช่นเดียวกับข้อมูลชุดอื่นๆ เพียงแต่กรณีนี้ผลลัพธ์จะหาได้จากค่าเฉลี่ยของคำตอบในทุกๆ โฟลเดอร์



รูปที่ 19 การตรวจสอบไขว้ 5 พับ

4.2 วิธีการทดลอง

การทดลองกับชุดข้อมูลแต่ละชุดซึ่งประกอบด้วย ข้อมูลสอนและข้อมูลทดสอบ จะทำการแบ่งชุดข้อมูลสอนออกเป็น ข้อมูลสอนจริงและข้อมูลทดสอบความถูกต้อง (Validation data) ด้วยการตรวจสอบไขว้ 5 พับเช่นเดียวกับการสร้างชุดทดลองของ Glass และ Segment เพื่อใช้ในการหาค่าพารามิเตอร์ที่เหมาะสม ซึ่งประกอบด้วยค่า P และ R

เมื่อ	P	คือค่าร้อยละในการตัดเส็ม โดยที่ $0 \le P \le 10$
	R	คือ ค่าขอบเขตความผิดพลาด โดยที่ $x_{min} \le R \le x_{mean+sd}$
	x_{min}	คือ ค่าต่ำสุดของขอบเขตความผิดพลาดของตัวจำแนกทั้งหมด
	$x_{mean+sd}$	คือ ผลรวมของค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของขอบเขตความผิดพลาดของตัวจำแนกทั้งหมด

การเลือกค่า P และ R จะพิจารณาจากการโหวตของทั้ง 5 โฟลเดอร์ โดยในเบื้องต้นจะเลือกพารามิเตอร์ที่ดีที่สุดของแต่ละโฟลเดอร์มาก่อน โดยพิจารณาเลือกพารามิเตอร์ที่ให้ค่าความถูกต้องสูงสุดเป็นอันดับแรก หากมีค่าความถูกต้องสูงสุดเท่ากันหลายตัวจึงเลือกตัวที่ให้จำนวนครั้งเฉลี่ยในการจำแนกต่ำสุด และถ้ามีคู่แข่งหลายตัวอีกจึงเลือกตัวที่ให้ค่า R ต่ำสุดและค่า P ต่ำสุดตามลำดับ เมื่อได้ตัวแทนของแต่ละโฟลเดอร์มาแล้วจึงนำมาโหวตเพื่อเลือกตัวที่ดีที่สุดซึ่งจะเลือกพารามิเตอร์ที่มีความถี่สูงสุดมาเป็นคำตอบ และหากมีหลายพารามิเตอร์ที่มีความถี่เท่ากันหลายตัวจึงใช้หลักในการเลือกเช่นเดียวกับการเลือกตัวแทนของแต่ละโฟลเดอร์

หลังจากได้ค่า P และ R แล้วจึงนำไปสร้างตัวจำแนกแบบหลายประเภทจากชุดข้อมูลสอนเดิมโดยในขั้นตอนการสร้างตัวจำแนกแบบสองประเภทจะใช้ฟังก์ชันเคอร์เนล 2 ฟังก์ชันคือ Polynomial Kernel และ RBF kernel เมื่อได้ตัวจำแนกแบบหลายประเภทแล้วจึงใช้ชุดข้อมูลทดสอบเพื่อหาค่าความถูกต้องและค่าจำนวนครั้งเฉลี่ยในการจำแนกข้อมูล

4.3 ผลการทดลอง

ในหัวข้อนี้จะแสดงผลที่ได้จากการทดลองในข้อมูลทดลอง 10 ชุดข้อมูล โดยชุดข้อมูลแต่ละชุดจะแสดงการเปรียบเทียบ 4 ส่วนได้แก่ ตารางเปรียบค่าความถูกต้องของการจำแนก รูปแสดงความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภท ตารางเปรียบเทียบค่าจำนวนครั้งของการจำแนกที่คาดหวังของวิธีต่างๆ และตารางเปรียบเทียบค่าจำนวนครั้งของการจำแนกที่ได้จากการทดลอง

ในส่วนของเทคนิคการแตกครึ่งตามสารสนเทศจำนวนครั้งในการจำแนกที่คาดหวังจะขึ้นอยู่กับค่าความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูลสอน ภายใต้สมมติฐานที่ว่าความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของข้อมูลสอนและข้อมูลทดสอบมีค่าใกล้เคียงกัน และค่าจำนวนครั้งในการจำแนกดังกล่าวสามารถคำนวณได้จากสูตร $\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i)$ เมื่อ $P(m_i)$ คือ ความน่าจะเป็นในการเกิดข้อมูลประเภท i สำหรับเทคนิคการแตกครึ่งแบบสมดุลสามารถคำนวณค่าจำนวนครั้งในการจำแนกที่คาดหวังได้จากสูตร $\log_2 k$ เมื่อ k คือจำนวนประเภทของข้อมูล เมื่อพิจารณาค่าจำนวนครั้งในการจำแนกที่คาดหวังของเทคนิคการแตกครึ่งตามสารสนเทศและการแตกครึ่งแบบสมดุลจะได้ว่า 2 เทคนิคนี้จะให้ค่าที่เท่ากันหากค่าความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของข้อมูลสอนมีค่าเท่ากัน

ในส่วนของวิธีอาร์เอดีเอจีและแมกซ์วินค่าจำนวนครั้งในการจำแนกที่คาดหวังและค่าจำนวนครั้งที่ได้จากการทดลองจะมีค่าเท่ากันโดยวิธีอาร์เอดีเอจีคำนวณได้จากสูตร k-1 และแมกซ์วินคำนวณได้จากสูตร $k(k\text{-}1)/2$

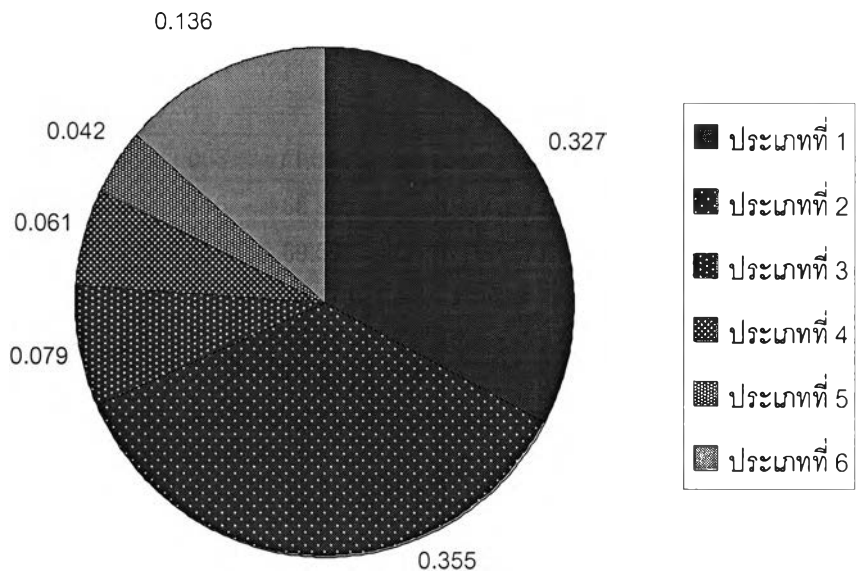
ผลการทดลองแสดง ดังตารางที่ 2-11 โดยกำหนดให้ IBD คือ Information-Based dichotomization / BD คือ Balanced Dichotomization / Expected Value คือ ค่าจำนวนครั้งในการจำแนกที่คาดหวัง / ตัวอักษร c และ d คือ พารามิเตอร์ของฟังก์ชันเคอร์เนลโดย c เป็นค่าดีกรีของ Polynomial Kernel และ d เป็นค่าดีกรีของ RBF Kernel สำหรับตัวอักษรเข้มในตารางแสดงค่าที่ดีที่สุดในแต่ละชุดข้อมูล และค่าที่สูงที่สุดในแต่ละ Kernel จะใส่เครื่องหมาย * กำกับไว้ สำหรับผลการทดลองโดยละเอียดและค่าพารามิเตอร์ P และ R ที่หาได้ในแต่ละโฟลเดอร์ จะได้แสดงไว้ในภาคผนวกท้ายเล่มต่อไป

4.3.1 ชุดข้อมูล Glass

ตารางที่ 2(ก): การเปรียบเทียบค่าความถูกต้องของการจำแนกชุดข้อมูล Glass

Polynomial kernel				
d	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
2	69.646	70.202	71.063	**71.078***
3	68.250	67.772	68.716	**68.989**
4	69.181	66.932	68.239	**69.598**
5	69.635	67.114	69.158	**70.173**
6	67.741	67.192	69.158	**69.553**
7	66.357	67.096	67.298	**67.902**
8	68.217	67.380	68.217	**68.463**

RBF kernel				
c	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
0.01	**72.436**	69.442	71.971	71.871
0.02	**71.506**	70.288	**71.506**	70.918
0.03	**71.506**	69.725	**71.506**	70.918
0.04	71.030	70.377	**71.041**	70.616
0.05	**71.517**	70.472	71.052	69.943
0.06	**73.378**	69.905	72.912	72.088
0.07	72.913	70.560	**73.377**	72.385
0.08	72.447	70.556	**73.377**	72.506
0.09	72.924	69.440	**74.319***	73.238
0.10	**72.447**	69.723	**72.447**	71.367



รูปที่ 20 ความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูล Glass

I22777684

ตารางที่ 2(ข): การเปรียบเทียบจำนวนครั้งในการจำแนกที่คาดหวังของชุดข้อมูล Glass

IBD $(\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i))$	BD $(\log_2 k)$	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
2.177	2.585	5	15

ตารางที่ 2(ค): การเปรียบเทียบจำนวนครั้งของการจำแนกชุดข้อมูล Glass

Polynomial kernel				
d	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
2	**4.207**	4.444	5	15
3	**4.101**	4.329	5	15
4	**4.027**	4.281	5	15
5	**3.837**	3.969	5	15
6	**4.069**	4.089	5	15
7	4.125	**4.074**	5	15
8	**3.993**	4.076	5	15

RBF kernel				
c	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
0.01	**3.837**	4.534	5	15
0.02	**4.319**	4.462	5	15
0.03	**4.144**	4.450	5	15
0.04	**4.216**	4.483	5	15
0.05	**4.435**	4.508	5	15
0.06	**4.430**	4.492	5	15
0.07	**4.196**	4.455	5	15
0.08	**4.294**	4.451	5	15
0.09	**4.443**	4.479	5	15
0.10	4.612	**4.457**	5	15

ชุดข้อมูล Glass ที่มีข้อมูลทดสอบเพียงชุดเดียวดังนั้นจึงต้องมีการตรวจสอบไขว้ 5 พับ เพื่อหาค่าความถูกต้องของการจำแนกเฉลี่ยและจำนวนครั้งเฉลี่ยในการจำแนกข้อมูล ผลลัพธ์ค่าความถูกต้องของการจำแนกชุดข้อมูล Glass ด้วย RBF kernel โดยภาพรวมให้ค่าที่สูงกว่า Polymomial kernel และเมื่อเปรียบเทียบค่าความถูกต้องของแต่ละวิธีพบว่าให้ค่าความถูกต้องใกล้เคียงกัน

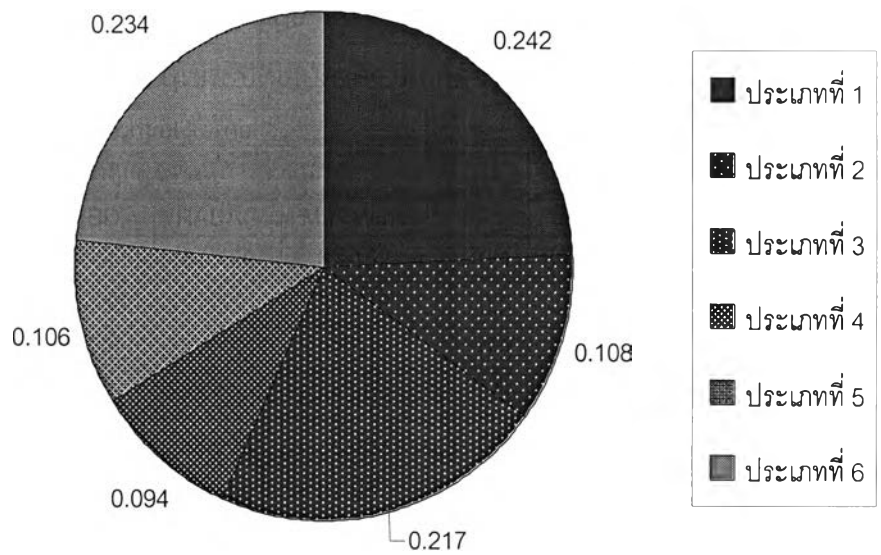
ในส่วนของค่าจำนวนครั้งในการจำแนกเฉลี่ยเมื่อพิจารณาความน่าจะเป็นของการเกิดข้อมูลประเภทต่างๆ พบว่า แต่ละประเภทมีความน่าจะเป็นในการเกิดที่แตกต่างกัน ซึ่งเป็นปัจจัยที่ชี้ให้เห็นว่าแนวโน้มของค่าจำนวนครั้งในการจำแนกที่คาดหวังของวิธีการแตกครึ่งตามสารสนเทศมีค่าต่ำสุดเมื่อเปรียบเทียบกับวิธีอื่นและเมื่อพิจารณาผลการทดลองก็พบว่าโดยส่วนใหญ่วิธีการแตกครึ่งตามสารสนเทศให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีอื่น

4.3.2 **ชุดข้อมูล** Satimage

ตารางที่ 3(ก): การเปรียบเทียบค่าความถูกต้องของการจำแนกชุดข้อมูล Satimage

Polynomial kernel				
d	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
2	87.350	86.330	**87.450**	87.152
3	**87.500**	87.480	87.350	87.248
4	**88.100**	87.550	87.750	87.545
5	**88.550**	87.710	88.500	88.198
6	88.850	87.840	**88.900***	88.453
7	**88.150**	87.710	**88.150**	87.957
8	**88.050**	87.640	87.850	87.757

RBF kernel				
c	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
0.1	**89.200**	89.040	89.150	89.123
0.2	90.150	89.710	**90.350**	90.340
0.3	90.900	90.450	**90.950**	90.834
0.4	91.550	90.880	**91.600**	91.542
0.5	91.950	91.320	91.950	**91.984***
1.0	91.800	91.350	91.800	**91.820**
1.5	**91.800**	91.150	91.500	91.496



รูปที่ 21 ความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูล Satimage

ตารางที่ 3(ข): การเปรียบเทียบจำนวนครั้งในการจำแนกที่คาดหวังของชุดข้อมูล Satimage

IBD $(\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i))$	BD $(\log_2 k)$	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
2.474	2.585	5	15

ตารางที่ 3(ค): การเปรียบเทียบจำนวนครั้งของการจำแนกชุดข้อมูล Satimage

Polynomial kernel				
d	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG (k -1)	Max Wins ($k\ (k$ -1)/2)
2	**4.594**	4.650	5	15
3	**4.603**	4.718	5	15
4	**4.554**	4.789	5	15
5	4.990	**4.712**	5	15
6	4.999	**4.847**	5	15
7	**4.722**	4.898	5	15
8	**4.682**	4.723	5	15

RBF kernel				
c	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG (k -1)	Max Wins ($k\ (k$ -1)/2)
0.1	4.911	**4.774**	5	15
0.2	**4.552**	4.663	5	15
0.3	4.970	**4.834**	5	15
0.4	4.992	**4.928**	5	15
0.5	4.734	**4.714**	5	15
1.0	4.696	**4.577**	5	15
1.5	**3.940**	4.239	5	15

ค่าความถูกต้องของการจำแนกชุดข้อมูล Satimage ด้วย RBF kernel โดยภาพรวมให้ค่าที่สูงกว่า Polymomial kernel และเมื่อเปรียบเทียบค่าความถูกต้องของแต่ละวิธีพบว่าให้ค่าความถูกต้องใกล้เคียงกัน

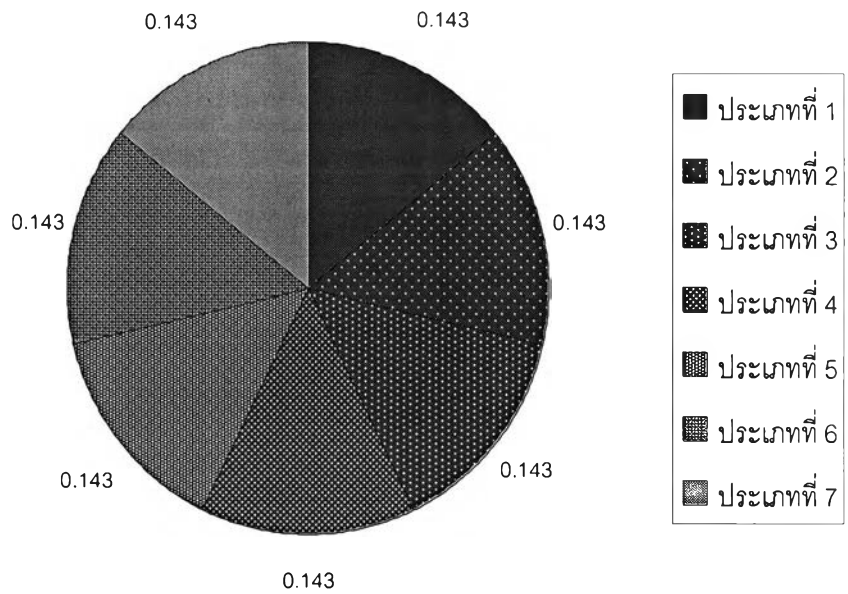
ในส่วนของค่าจำนวนครั้งในการจำแนกเฉลี่ยเมื่อพิจารณาความน่าจะเป็นของการเกิดข้อมูลประเภทต่างๆ พบว่า แต่ละประเภทมีความน่าจะเป็นในการเกิดที่แตกต่างกัน ซึ่งเป็นปัจจัยที่ชี้ให้เห็นว่าแนวโน้มของค่าจำนวนครั้งในการจำแนกที่คาดหวังของวิธีการแตกครึ่งตามสารสนเทศมีค่าต่ำสุดเมื่อเปรียบเทียบกับวิธีอื่นและเมื่อพิจารณาผลการทดลองก็พบว่าโดยส่วนใหญ่วิธีการแตกครึ่งตามสารสนเทศให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีอื่นซึ่งแม้ว่าในบางดีกรีวิธีแตกครึ่งแบบสมดุลจะให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าแต่เมื่อพิจารณาค่าความถูกต้องร่วมด้วยพบว่าทุกดีกรีวิธีแตกครึ่งตามสารสนเทศจะให้ค่าความถูกต้องสูงกว่าแบบสมดุลทุกกรณี

4.3.3 **ชุดข้อมูล** Segment

ตารางที่ 4(ก): การเปรียบเทียบค่าความถูกต้องของการจำแนกชุดข้อมูล Segment

Polynomial kernel				
d	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
2	**97.013**	96.883	96.970	96.743
3	**96.710**	96.571	96.623	96.428
4	**96.926**	96.623	96.840	96.658
5	97.013	96.615	**97.056**	96.942
6	97.100	96.719	**97.143**	97.080
7	97.229	96.771	**97.230**	97.083
8	**97.489 ***	96.788	**97.489 ***	97.379

RBF kernel				
c	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
0.5	97.056	96.823	97.013	**97.071**
0.6	97.013	96.823	97.056	**97.089**
0.7	97.273	96.857	97.273	**97.298 ***
0.8	97.186	96.857	**97.273**	97.241
0.9	97.143	96.736	**97.186**	97.171
1.0	96.926	96.719	97.056	**97.077**
1.5	96.840	96.554	**97.056**	97.046



รูปที่ 22 ความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูล Segment

ตารางที่ 4(ข): การเปรียบเทียบจำนวนครั้งในการจำแนกที่คาดหวังของชุดข้อมูล Segment

IBD $(\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i))$	BD $(\log_2 k)$	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
2.807	**2.807**	6	21

ตารางที่ 4(ค): การเปรียบเทียบจำนวนครั้งของการจำแนกชุดข้อมูล Segment

Polynomial kernel				
d	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
2	**4.676**	4.837	6	21
3	4.979	**4.835**	6	21
4	**4.865**	4.872	6	21
5	**4.857**	4.912	6	21
6	4.963	**4.823**	6	21
7	**4.751**	4.881	6	21
8	**4.666**	4.886	6	21

RBF kernel				
c	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
0.5	**4.452**	4.966	6	21
0.6	**4.468**	4.932	6	21
0.7	5.094	**4.944**	6	21
0.8	**4.945**	4.945	6	21
0.9	**4.650**	4.920	6	21
1.0	**4.859**	4.903	6	21
1.5	4.620	**4.525**	6	21

ชุดข้อมูล Segment ที่มีข้อมูลทดสอบเพียงชุดเดียวดังนั้นจึงต้องมีการตรวจสอบไขว้ 5 พับ เพื่อหาค่าความถูกต้องของการจำแนกเฉลี่ยและจำนวนครั้งเฉลี่ยในการจำแนกข้อมูล ผลลัพธ์ค่าความถูกต้องของการจำแนกชุดข้อมูล Segment ด้วย RBF kernel โดยภาพรวมให้ค่าที่ใกล้เคียงกับ Polymomial kernel และเมื่อเปรียบเทียบค่าความถูกต้องของแต่ละวิธีพบว่าให้ค่าความถูกต้องใกล้เคียงกันโดยกรณีดีสุดที่ให้ค่าดีที่สุดเมื่อใช้ Polymomial kernel ในวิธีการแตกครึ่งตามสารสนเทศและอาร์เอดีเอจี

ในส่วนของค่าจำนวนครั้งในการจำแนกเฉลี่ยเมื่อพิจารณาความน่าจะเป็นของการเกิดข้อมูลประเภทต่างๆ พบว่า แต่ละประเภทมีความน่าจะเป็นในการเกิดที่เท่ากัน ซึ่งเป็นปัจจัยที่ชี้ให้เห็นว่าแนวโน้มของค่าจำนวนครั้งในการจำแนกที่คาดหวังของวิธีการแตกครึ่งตามสารสนเทศมีค่าเท่ากับแบบสมดุลและต่ำสุดเมื่อเปรียบเทียบกับวิธีอื่นและเมื่อพิจารณาผลการทดลองก็พบว่าโดยส่วนใหญ่วิธีการแตกครึ่งตามสารสนเทศให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีอื่นซึ่งแม้ว่าในบางดีกรีวิธีแตกครึ่งแบบสมดุลจะให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าแต่เมื่อพิจารณาค่า

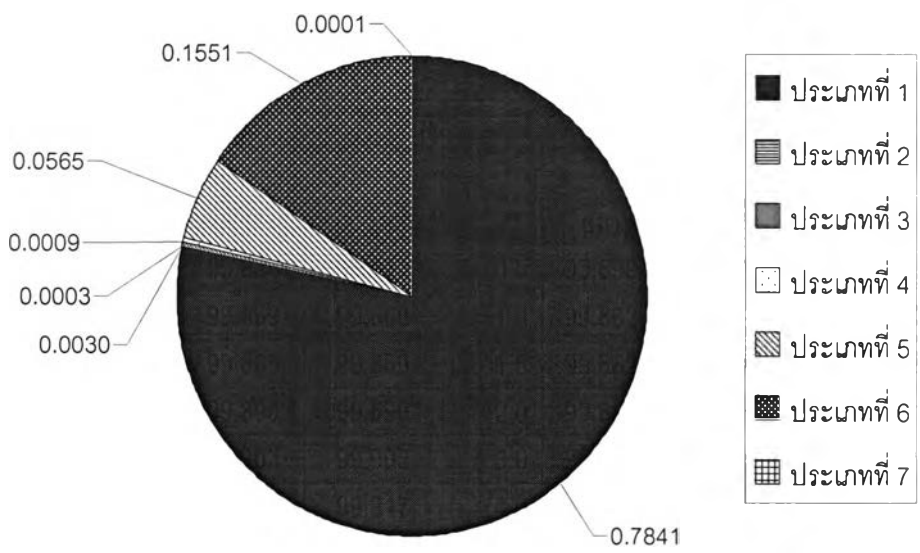
ความถูกต้องร่วมด้วยพบว่ากรณีดังกล่าวในทุกดีกรี วิธีแตกครึ่งตามสารสนเทศจะให้ค่าความถูกต้องสูงกว่าแบบสมดุลทุกกรณี

4.3.4 **ชุดข้อมูล** Shuttle

ตารางที่ 5(ก): การเปรียบเทียบค่าความถูกต้องของการจำแนกชุดข้อมูล Shuttle

Polynomial kernel				
d	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
2	**99.834**	99.823	**99.834**	**99.834**
3	**99.869**	99.847	**99.869**	**99.869**
4	99.855	99.868	**99.869**	**99.869**
5	**99.890**	**99.890**	**99.890**	**99.890**
6	**99.903**	99.892	**99.903**	**99.903**
7	**99.917**	99.916	**99.917**	**99.917**
8	**99.924***	**99.924***	**99.924***	**99.924***

RBF kernel				
c	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
0.5	**99.890**	99.841	99.834	99.834
1.0	**99.862**	99.848	**99.862**	**99.862**
1.5	99.862	99.869	99.876	**99.879**
2.0	**99.883**	99.877	**99.883**	**99.883**
3.0	99.890	99.886	**99.897**	**99.897**
4.0	**99.931***	99.897	99.890	99.897
5.0	**99.931***	99.890	99.890	99.892



รูปที่ 23 ความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูล Shuttle

ตารางที่ 5(ข): การเปรียบเทียบจำนวนครั้งในการจำแนกที่คาดหวังของชุดข้อมูล Shuttle

IBD ($\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i)$)	BD ($\log_2 k$)	RADAG (k -1)	Max Wins (k (k -1)/2)
0.965	2.807	6	21

ตารางที่ 5(ค): การเปรียบเทียบจำนวนครั้งของการจำแนกชุดข้อมูล Shuttle

Polynomial kernel				
d	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG (k -1)	Max Wins (k (k -1)/2)
2	**4.997**	5.277	6	21
3	**5.296**	5.301	6	21
4	**4.777**	5.263	6	21
5	5.396	**5.382**	6	21
6	5.396	**5.362**	6	21
7	**5.352**	5.368	6	21
8	**5.359**	5.378	6	21

RBF kernel				
c	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG (k -1)	Max Wins (k (k -1)/2)
0.5	5.872	**5.862**	6	21
1.0	**5.841**	5.863	6	21
1.5	**5.011**	5.825	6	21
2.0	**4.799**	5.807	6	21
3.0	**4.982**	5.758	6	21
4.0	5.783	**5.746**	6	21
5.0	5.830	**5.773**	6	21

ค่าความถูกต้องของการจำแนกชุดข้อมูล Shuttle ด้วย RBF kernel โดยภาพรวมให้ค่าที่ใกล้เคียงกับ Polymomial kernel และเมื่อเปรียบเทียบค่าความถูกต้องของแต่ละวิธีพบว่าให้ค่าความถูกต้องใกล้เคียงกันโดยกรณีดีสุดที่ให้ค่าดีที่สุดเมื่อใช้ RBF kernel ในวิธีการแตกครึ่งตามสารสนเทศ

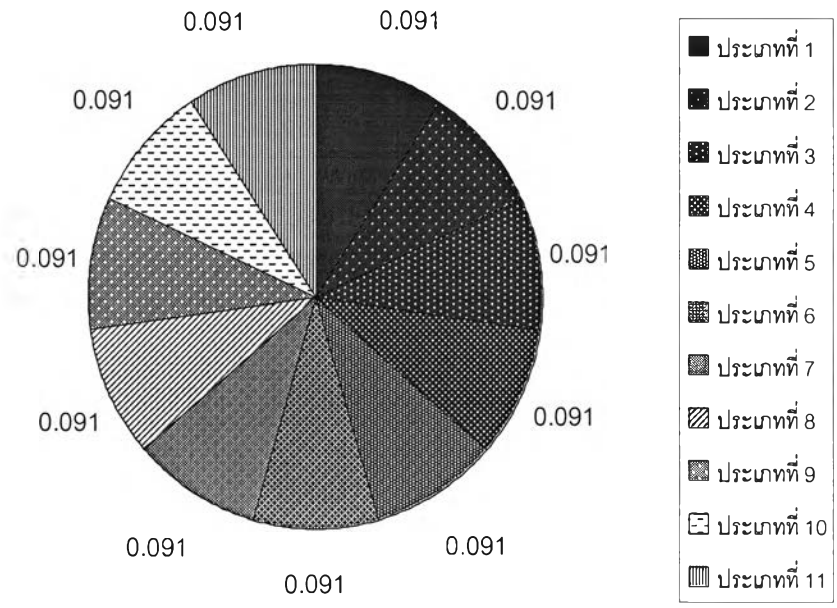
ในส่วนของค่าจำนวนครั้งในการจำแนกเฉลี่ยเมื่อพิจารณาความน่าจะเป็นของการเกิดข้อมูลประเภทต่างๆ พบว่า แต่ละประเภทมีความน่าจะเป็นในการเกิดที่แตกต่างกันมาก ซึ่งเป็นปัจจัยที่ชี้ให้เห็นว่าแนวโน้มของค่าจำนวนครั้งในการจำแนกที่คาดหวังของวิธีการแตกครึ่งตามสารสนเทศมีค่าต่ำสุดเมื่อเปรียบเทียบกับวิธีอื่น และเมื่อพิจารณาผลการทดลองก็พบว่าโดยส่วนใหญ่วิธีการแตกครึ่งตามสารสนเทศให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีอื่นซึ่งแม้ว่าในบางดีกรีวิธีแตกครึ่งแบบสมดุลจะให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าแต่เมื่อพิจารณาค่าความถูกต้องร่วมด้วยพบว่าในทุกดีกรี ไม่มีกรณีใดที่วิธีแตกครึ่งตามสารสนเทศจะให้ค่าความถูกต้องต่ำกว่าแบบสมดุล

4.3.5 **ชุดข้อมูล** Vowel

ตารางที่ 6(ก): การเปรียบเทียบค่าความถูกต้องของการจำแนกชุดข้อมูล Vowel

Polynomial kernel				
d	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
2	59.307	**65.022***	63.420	63.918
3	**64.935**	62.121	64.502	64.329
4	62.771	58.615	61.905	**62.773**
5	59.740	56.407	**60.606**	60.346
6	56.277	56.320	**58.874**	58.649
7	56.277	54.113	56.710	**56.725**
8	53.030	52.771	55.411	**55.647**

RBF kernel				
c	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
0.1	61.039	58.052	**62.771**	61.169
0.2	62.554	62.900	**67.100***	65.340
0.3	**66.450**	62.857	**66.450**	65.203
0.4	64.286	61.429	**65.368**	64.108
0.5	63.203	60.216	**64.286**	63.677
1.0	57.143	59.134	**61.255**	61.214
1.5	60.606	56.017	**61.472**	60.833



รูปที่ 24 ความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูล Vowel

ตารางที่ 6(ข): การเปรียบเทียบจำนวนครั้งในการจำแนกที่คาดหวังของชุดข้อมูล Vowel

IBD $(\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i))$	BD $(\log_2 k)$	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
3.459	**3.459**	10	55

ตารางที่ 6(ค): การเปรียบเทียบจำนวนครั้งของการจำแนกชุดข้อมูล Vowel

Polynomial kernel				
d	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
2	**5.180**	5.665	10	55
3	**5.385**	5.461	10	55
4	5.299	**5.289**	10	55
5	5.320	**5.170**	10	55
6	5.301	**5.069**	10	55
7	**4.833**	5.289	10	55
8	**5.370**	5.404	10	55

RBF kernel				
c	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
0.1	**5.262**	5.707	10	55
0.2	**5.097**	5.819	10	55
0.3	5.628	**5.612**	10	55
0.4	**5.561**	5.794	10	55
0.5	5.810	**5.342**	10	55
1.0	**5.048**	5.351	10	55
1.5	5.340	**5.227**	10	55

ค่าความถูกต้องของการจำแนกชุดข้อมูล Vowel ด้วย RBF kernel โดยภาพรวมให้ค่าที่สูงกว่า Polymomial kernel และเมื่อเปรียบเทียบค่าความถูกต้องของแต่ละวิธีพบว่าให้ค่าความถูกต้องใกล้เคียงกันโดยกรณีดีสุดที่ให้ค่าดีที่สุดเมื่อใช้ RBF kernel ในวิธีอาร์เอดีเอจี

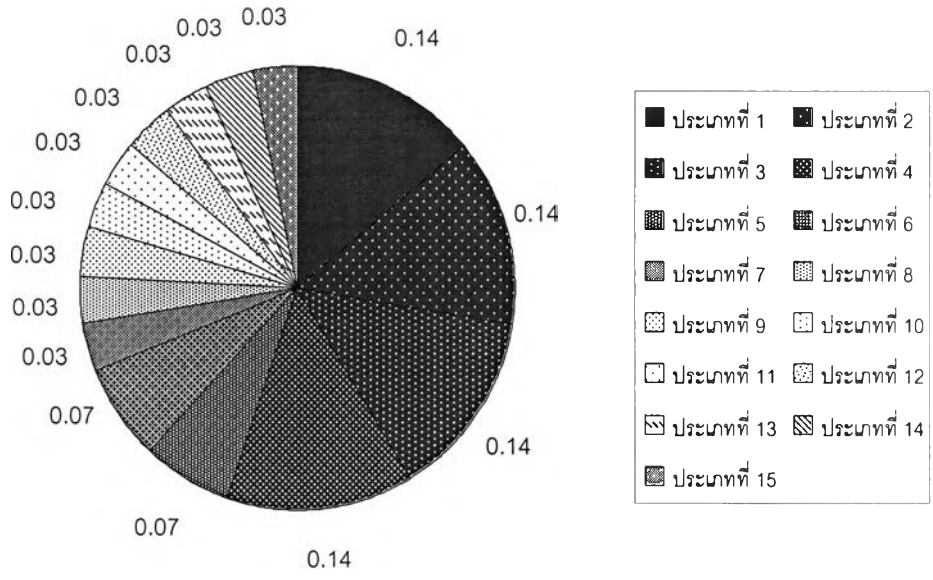
ในส่วนของค่าจำนวนครั้งในการจำแนกเฉลี่ยเมื่อพิจารณาความน่าจะเป็นของการเกิดข้อมูลประเภทต่างๆ พบว่า แต่ละประเภทมีความน่าจะเป็นในการเกิดที่เท่ากัน ซึ่งเป็นปัจจัยที่ชี้ให้เห็นว่าแนวโน้มของค่าจำนวนครั้งในการจำแนกที่คาดหวังของวิธีการแตกครึ่งตามสารสนเทศมีค่าเท่ากับแบบสมดุลและต่ำสุดเมื่อเปรียบเทียบกับวิธีอื่น และเมื่อพิจารณาผลการทดลองก็พบว่าโดยส่วนใหญ่วิธีการแตกครึ่งตามสารสนเทศให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีอื่นซึ่งแม้ว่าในบางดีกรีวิธีแตกครึ่งแบบสมดุลจะให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าแต่เมื่อพิจารณาค่าความถูกต้องร่วมด้วยพบว่าโดยส่วนใหญ่วิธีแตกครึ่งตามสารสนเทศจะให้ค่าความถูกต้องสูงกว่าแบบสมดุล

4.3.6 ชุดข้อมูล Soybean

ตารางที่ 7(ก): การเปรียบเทียบค่าความถูกต้องของการจำแนกชุดข้อมูล Soybean

Polynomial kernel				
d	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
2	90.000	89.235	**90. 588**	89.768
3	90.882	89.059	**91. 176***	90.471
4	87.059	89.529	**90. 588**	89.968
5	89.412	89.471	**90. 588**	90.353
6	87.941	88.529	**89. 706**	89.506
7	86.765	87.000	88.529	**88. 668**
8	83.824	87.294	87.647	**87. 785**

RBF kernel				
c	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
0.04	88.235	89.118	**90. 294**	89.738
0.05	87.647	88.588	**90. 882**	90.379
0.06	90.000	88.529	**90. 588**	90.068
0.07	**91.471***	89.000	90.882	90.362
0.08	88.235	87.706	**90. 588**	90.468
0.09	90.000	88.647	**90. 588**	90.468
0.10	88.529	88.118	**90. 294**	90.174
0.20	86.471	85.471	**86. 765**	86.682



รูปที่ 25 ความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูล Soybean

ตารางที่ 7(ข): การเปรียบเทียบจำนวนครั้งในการจำแนกที่คาดหวังของชุดข้อมูล Soybean

IBD ($\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i)$)	BD ($\log_2 k$)	RADAG (k -1)	Max Wins (k (k -1)/2)
3.617	3.907	14	105

ตารางที่ 7(ค): การเปรียบเทียบจำนวนครั้งของการจำแนกชุดข้อมูล Soybean

Polynomial kernel				
d	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG (k -1)	Max Wins (k (k -1)/2)
2	6.791	**6.656**	14	105
3	6.971	**6.381**	14	105
4	**6.647**	6.783	14	105
5	7.288	**7.263**	14	105
6	8.021	**7.425**	14	105
7	**5.932**	7.112	14	105
8	7.897	**7.785**	14	105

RBF kernel				
c	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG (k -1)	Max Wins (k (k -1)/2)
0.04	7.232	**6.591**	14	105
0.05	**5.015**	6.544	14	105
0.06	7.159	**6.764**	14	105
0.07	**5.400**	6.216	14	105
0.08	8.212	**6.963**	14	105
0.09	**6.147**	7.082	14	105
0.10	7.962	**7.389**	14	105
0.20	9.218	**8.914**	14	105

ค่าความถูกต้องของการจำแนกชุดข้อมูล Soybean ด้วย RBF kernel โดยภาพรวมให้ค่าที่สูงกว่า Polymomial kernel และเมื่อเปรียบเทียบค่าความถูกต้องของแต่ละวิธีพบว่าแนวโน้มส่วนใหญ่วิธีอาร์เอดีเอจีให้ค่าความถูกต้องสูงกว่าวิธีอื่น แต่กรณีดีสุดที่ให้ค่าดีที่สุดเมื่อใช้ RBF kernel ในวิธีแตกครึ่งตามสารสนเทศ

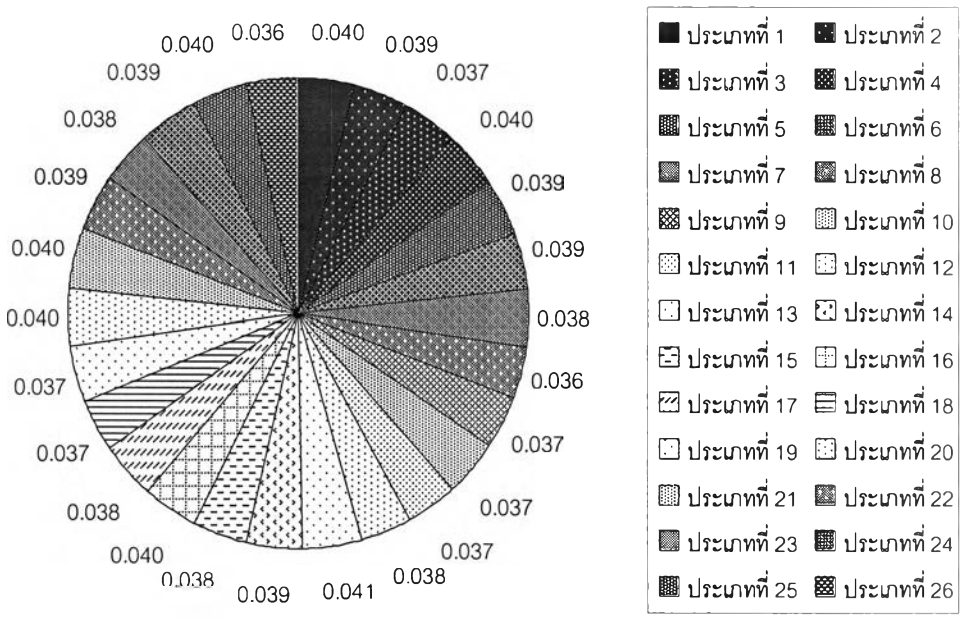
ในส่วนของค่าจำนวนครั้งในการจำแนกเฉลี่ยเมื่อพิจารณาความน่าจะเป็นของการเกิดข้อมูลประเภทต่างๆ พบว่า แต่ละประเภทมีความน่าจะเป็นในการเกิดที่แตกต่างกัน ซึ่งเป็นปัจจัยที่ชี้ให้เห็นว่าแนวโน้มของค่าจำนวนครั้งในการจำแนกที่คาดหวังของวิธีการแตกครึ่งตามสารสนเทศมีค่าเท่ากับต่ำสุดเมื่อเปรียบเทียบกับวิธีอื่น และเมื่อพิจารณาผลการทดลองก็พบว่าโดยส่วนใหญ่วิธีการแตกครึ่งตามสารสนเทศให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีอื่นซึ่งแม้ว่าในบางดีกรีวิธีแตกครึ่งแบบสมดุลจะให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าแต่เมื่อพิจารณาค่าความถูกต้องร่วมด้วยพบว่าโดยส่วนใหญ่วิธีแตกครึ่งตามสารสนเทศจะให้ค่าความถูกต้องสูงกว่าแบบสมดุล

4.3.7 ชุดข้อมูล Letter

ตารางที่ 8(ก): การเปรียบเทียบค่าความถูกต้องของการจำแนกชุดข้อมูล Letter

Polynomial kernel				
d	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
2	93.075	90.870	95.400	**95.522**
3	93.650	91.055	**96.125**	**96.125***
4	93.100	91.640	95.975	**96.120**
5	92.800	90.815	**95.725**	94.622
6	92.400	90.480	95.400	**95.551**
7	91.950	90.060	94.875	**94.975**
8	91.975	89.490	94.050	**94.072**

RBF kernel				
c	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
0.5	94.700	92.515	**97.050**	96.674
1.0	94.600	93.375	**97.600**	97.427
1.5	95.150	93.640	**97.650**	97.629
2.0	95.050	94.040	**97.825**	97.661
3.0	95.175	94.605	**97.950***	97.918
4.0	94.450	94.695	**97.825**	97.815
5.0	94.700	94.575	**97.700**	97.689



รูปที่ 26 ความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูล Letter

ตารางที่ 8(ข): การเปรียบเทียบจำนวนครั้งในการจำแนกที่คาดหวังของชุดข้อมูล Letter

IBD $(\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i))$	BD $(\log_2 k)$	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
4.700	**4.700**	25	325

ตารางที่ 8(ค): การเปรียบเทียบจำนวนครั้งของการจำแนกชุดข้อมูล Letter

Polynomial kernel				
d	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
2	**11.125**	12.136	25	325
3	**11.262**	12.427	25	325
4	**10.829**	12.179	25	325
5	**10.918**	12.265	25	325
6	**11.140**	12.599	25	325
7	**11.222**	12.464	25	325
8	**11.029**	12.469	25	325

RBF kernel				
c	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
0.5	**10.846**	12.056	25	325
1.0	**11.160**	12.473	25	325
1.5	**11.125**	13.051	25	325
2.0	**11.381**	14.086	25	325
3.0	**11.588**	14.785	25	325
4.0	**11.282**	15.216	25	325
5.0	**11.266**	15.096	25	325

ค่าความถูกต้องของการจำแนกชุดข้อมูล Letter ด้วย RBF kernel โดยภาพรวมให้ค่าที่สูงกว่า Polymomial kernel และเมื่อเปรียบเทียบค่าความถูกต้องของแต่ละวิธีพบว่าแนวโน้มส่วนใหญ่กรณีใช้ Polymomial kernel จะได้ว่าวิธีแมกซ์วินให้ค่าความถูกต้องสูงกว่าวิธีอื่น และเมื่อใช้ RBF kernel จะได้ว่าวิธีอาร์เอดีเอจีให้ค่าความถูกต้องสูงกว่าวิธีอื่น แต่กรณีดีสุดที่ให้ค่าดีที่สุดเมื่อใช้ RBF kernel ในวิธีอาร์เอดีเอจี

ในส่วนของค่าจำนวนครั้งในการจำแนกเฉลี่ยเมื่อพิจารณาความน่าจะเป็นของการเกิดข้อมูลประเภทต่างๆ พบว่า แต่ละประเภทมีความน่าจะเป็นในการเกิดที่เท่ากัน ซึ่งเป็นปัจจัยที่ชี้ให้เห็นว่าแนวโน้มของค่าจำนวนครั้งในการจำแนกที่คาดหวังของวิธีการแตกครึ่งตามสารสนเทศมีค่าเท่ากับแบบสมดุลและต่ำสุดเมื่อเปรียบเทียบกับวิธีอื่น และเมื่อพิจารณาผลการทดลองก็พบว่าวิธีการแตกครึ่งตามสารสนเทศให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีอื่นทุกกรณี และกรณีดีสุดด้านจำนวนครั้งของการจำแนกพบว่าสามารถลดจำนวนครั้งในการจำแนกลดลงจาก

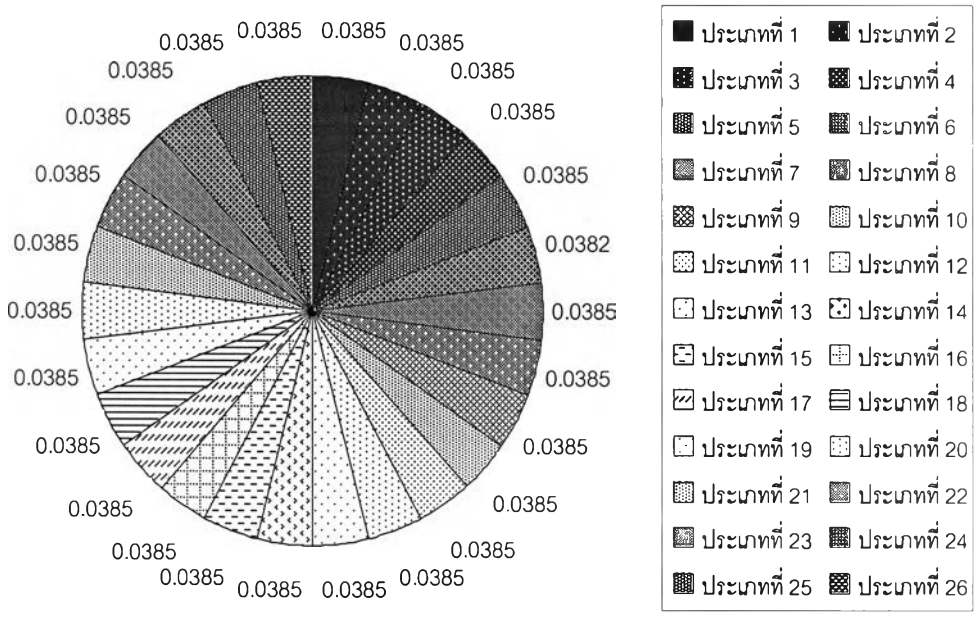
วิธีอาร์เอดีเอจีได้เกินครึ่งของเวลาที่อาร์เอดีเอจีใช้ และลดเวลาของการจำแนกด้วยวิธีแมกซ์วินลงได้มากกว่าสามสิบเท่า

4.3.8 ชุดข้อมูล Isolet

ตารางที่ 9(ก): การเปรียบเทียบค่าความถูกต้องของการจำแนกชุดข้อมูล Isolet

Polynomial kernel				
d	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
2	95.318	94.997	**96.600**	96.586
3	96.344	95.523	**97.049 ***	97.040
4	96.216	95.843	96.985	**97.024**
5	95.702	95.497	96.665	**96.695**
6	95.959	95.253	**96.729**	96.666
7	**96.280**	95.510	96.151	96.133
8	94.869	94.779	**95.510**	95.488

RBF kernel				
c	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
0.001	95.510	94.689	**96.665**	96.554
0.002	95.189	94.830	**96.793**	96.619
0.003	95.318	95.112	**97.115 ***	96.889
0.004	95.253	95.151	**96.857**	96.726
0.005	95.638	95.446	**96.857**	96.681
0.010	96.151	95.895	**96.985**	96.916
0.020	96.023	95.497	96.729	**96.731**
0.030	**96.087**	95.330	95.767	95.680



รูปที่ 27 ความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูล Isolet

ตารางที่ 9(ข): การเปรียบเทียบจำนวนครั้งในการจำแนกที่คาดหวังของชุดข้อมูล Isolet

IBD $(\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i))$	BD $(\log_2 k)$	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
4.700	**4.700**	25	325

ตารางที่ 9(ค): การเปรียบเทียบจำนวนครั้งของการจำแนกชุดข้อมูล Isolet

Polynomial kernel				
d	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
2	**7.813**	9.382	25	325
3	**7.917**	9.344	25	325
4	**7.719**	9.382	25	325
5	**7.962**	9.556	25	325
6	**7.544**	10.637	25	325
7	**8.528**	9.994	25	325
8	**8.010**	9.692	25	325

RBF kernel				
c	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\ (k$ -1)/2)
0.001	**8.033**	9.101	25	325
0.002	**7.893**	9.499	25	325
0.003	**8.107**	9.673	25	325
0.004	**8.067**	9.202	25	325
0.005	**8.067**	8.770	25	325
0.010	**8.013**	10.009	25	325
0.020	**8.061**	10.508	25	325
0.030	**8.366**	10.515	25	325

ค่าความถูกต้องของการจำแนกชุดข้อมูล Isolet ด้วย RBF kernel โดยภาพรวมให้ค่าที่สูงกว่า Polymomial kernel และเมื่อเปรียบเทียบค่าความถูกต้องของแต่ละวิธีพบว่าแนวโน้มโดยส่วนใหญ่วิธีอาร์เอดีเอจีให้ค่าความถูกต้องสูงกว่าวิธีอื่น

ในส่วนของค่าจำนวนครั้งในการจำแนกเฉลี่ยเมื่อพิจารณาความน่าจะเป็นของการเกิดข้อมูลประเภทต่างๆ พบว่า แต่ละประเภทมีความน่าจะเป็นในการเกิดที่เท่ากัน ซึ่งเป็นปัจจัยที่ชี้ให้เห็นว่าแนวโน้มของค่าจำนวนครั้งในการจำแนกที่คาดหวังของวิธีการแตกครึ่งตามสารสนเทศมีค่าเท่ากับแบบสมดุลและต่ำสุดเมื่อเปรียบเทียบกับวิธีอื่น และเมื่อพิจารณาผลการทดลองก็พบว่าวิธีการแตกครึ่งตามสารสนเทศให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีอื่นทุกกรณี และกรณีดีสุดด้านจำนวนครั้งของการจำแนกพบว่าสามารถลดจำนวนครั้งในการจำแนกลดลงจากวิธีอาร์เอดีเอจีลงได้มากกว่าสามเท่าของเวลาที่อาร์เอดีเอจีใช้ และลดเวลาของการจำแนกด้วยวิธีแมกซ์วินลงได้มากกว่าสี่สิบเท่า

4.3.9 ชุดข้อมูล Thaiprinted Character1

ตารางที่ 10(ก): การเปรียบเทียบค่าความถูกต้องของการจำแนกชุดข้อมูล Thaiprinted Character1

Polynomial kernel				
d	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
2	67.433	70.919	**93.444***	93.280
3	68.658	71.366	**82.077**	81.927
4	67.310	65.098	63.817	**63.858**
5	**60.417**	56.189	48.775	48.760

RBF kernel				
c	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
0.0007	80.453	71.299	96.538	**96.616**
0.0008	82.475	71.495	**96.599**	96.595
0.0009	79.075	72.249	96.599	**96.626**
0.0010	81.648	72.904	96.661	**96.688**
0.002	80.515	70.735	**97.028**	97.018
0.003	78.922	72.132	97.028	**97.080***
0.004	76.869	71.893	**96.783**	96.782

กรณีความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูล Thaiprinted Character 1 และ 2 ซึ่งใช้ข้อมูลสอนชุดเดียวกันข้อมูลทุกประเภทมีความน่าจะเป็นในการเกิดข้อมูลเท่ากัน คือ 0.015

ค่าความถูกต้องของการจำแนกชุดข้อมูล Thaiprinted Character 1 ด้วย RBF kernel โดยภาพรวมให้ค่าที่สูงกว่า Polymomial kernel และเมื่อเปรียบเทียบค่าความถูกต้องของแต่ละวิธี พบว่าแนวโน้มโดยส่วนใหญ่กรณีใช้ Polymomial kernel วิธีอาร์เอดีเอจีให้ค่าความถูกต้องสูงกว่าวิธีอื่นและกรณีใช้ RBF kernel วิธีแมกซ์วินให้ค่าความถูกต้องใกล้เคียงกับวิธีวิธีอาร์เอดีเอจี

ในส่วนของค่าจำนวนครั้งในการจำแนกเฉลี่ยเมื่อพิจารณาความน่าจะเป็นของการเกิดข้อมูลประเภทต่างๆ พบว่า แต่ละประเภทมีความน่าจะเป็นในการเกิดที่เท่ากัน ซึ่งเป็นปัจจัยที่ชี้ให้เห็นว่าแนวโน้มของค่าจำนวนครั้งในการจำแนกที่คาดหวังของวิธีการแตกครึ่งตามสารสนเทศมีค่าเท่ากับแบบสมดุลและต่ำสุดเมื่อเปรียบเทียบกับวิธีอื่น และเมื่อพิจารณาผลการทดลองแม้ว่าวิธีแตกครึ่งตามสารสนเทศจะให้ค่าความถูกต้องที่ต่ำกว่าวิธีแมกซ์วินและอาร์เอดีเอจีในหลายดีกรีแต่ก็พบว่าวิธีการแตกครึ่งตามสารสนเทศให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีอื่นทุกกรณี และกรณีดีสุดด้านจำนวนครั้งของการจำแนกพบว่าสามารถลดจำนวนครั้งในการจำแนกลดลงจากวิธีอาร์เอดีเอจีลงได้มากกว่าแปดเท่าของเวลาที่อาร์เอดีเอจีใช้ และลดเวลาของการจำแนกด้วยวิธีแมกซ์วินลงได้มากกว่าสองร้อยเจ็ดสิบเท่า

ตารางที่ 10(ข): การเปรียบเทียบจำนวนครั้งในการจำแนกที่คาดหวังของชุดข้อมูล Thaiprinted Character 1 และ 2

IBD $(\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i))$	BD $(\log_2 k)$	RADAG $(k$ -1)	Max Wins $(k\,(k$ -1)/2)
6.087	**6.087**	67	2278

ตารางที่ 10(ค): การเปรียบเทียบจำนวนครั้งของการจำแนกชุดข้อมูล Thaiprinted Character 1

Polynomial kernel				
d	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\,(k$ -1)/2)
2	**9.857**	22.698	67	2278
3	**9.675**	24.895	67	2278
4	**13.014**	35.480	67	2278
5	**12.230**	35.616	67	2278

RBF kernel				
c	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG $(k$ -1)	Max Wins $(k\,(k$ -1)/2)
0.0007	8.503	**8.443**	67	2278
0.0008	8.554	**8.471**	67	2278
0.0009	8.499	**8.459**	67	2278
0.0010	8.417	**8.295**	67	2278
0.002	**8.312**	8.480	67	2278
0.003	**8.385**	8.715	67	2278
0.004	**8.471**	8.896	67	2278

4.3.10 **ชุดข้อมูล** Thaiprinted Character 2

ตารางที่ 11(ก): การเปรียบเทียบค่าความถูกต้องของการจำแนกชุดข้อมูล Thaiprinted Character 2

Polynomial kernel				
d	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
2	87.377	83.009	**99.112***	98.989
3	89.645	84.449	**98.254**	98.123
4	90.564	85.110	**94.761**	94.709
5	**89.982**	81.765	88.634	88.584

RBF kernel				
c	ค่าความถูกต้องของการจำแนก			
	IBD	BD	RADAG	Max Wins
0.0007	93.076	81.967	**99.173**	99.058
0.0008	93.597	82.500	**99.265**	99.192
0.0009	92.188	82.482	**99.234**	99.192
0.0010	93.076	82.506	**99.265**	99.223
0.002	93.107	82.574	**99.357**	99.263
0.003	93.627	82.886	**99.357**	99.304
0.004	93.719	82.457	**99.418***	99.385

ตารางที่ 11(ข): การเปรียบเทียบจำนวนครั้งของการจำแนกชุดข้อมูล Thaiprinted Character2

Polynomial kernel				
d	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG (k -1)	Max Wins (k (k -1)/2)
2	**9.225**	21.450	67	2278
3	**8.971**	23.044	67	2278
4	**10.672**	29.915	67	2278
5	**9.841**	29.763	67	2278

RBF kernel				
c	จำนวนครั้งของการจำแนก			
	IBD	BD	RADAG (k -1)	Max Wins (k (k -1)/2)
0.0007	8.524	**8.516**	67	2278
0.0008	8.554	**8.538**	67	2278
0.0009	8.526	**8.495**	67	2278
0.0010	**8.426**	8.514	67	2278
0.002	**8.338**	8.690	67	2278
0.003	**8.418**	8.740	67	2278
0.004	**8.487**	8.760	67	2278

ค่าความถูกต้องของของการจำแนกชุดข้อมูล Thaiprinted Character 2 ด้วย RBF kernel โดยภาพรวมให้ค่าที่สูงกว่า Polymomial kernel และเมื่อเปรียบเทียบค่าความถูกต้องของแต่ละวิธี พบว่าแนวโน้มโดยส่วนใหญ่วิธีอาร์เอดีเอจีให้ค่าความถูกต้องสูงกว่าวิธีอื่น

การจำแนกชุดข้อมูล Thaiprinted Character 2 ซึ่งใช้ชุดข้อมูลสอนชุดเดียวกับ Thaiprinted Character 1 แต่ใช้ชุดทดสอบคนละชุดซึ่งแสดงให้ทราบว่าความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของข้อมูลสอนทั้งสองชุดเหมือนกันด้วยคือ ข้อมูลแต่ละประเภทมีความน่าจะเป็นในการเกิดที่เท่ากันซึ่งเป็นปัจจัยที่ชี้ให้เห็นว่าแนวโน้มของค่าจำนวนครั้งในการจำแนกที่คาดหวังของวิธีการแตกครึ่งตามสารสนเทศมีค่าเท่ากับแบบสมดุลและต่ำสุดเมื่อเปรียบเทียบกับวิธีอื่น และเมื่อพิจารณาผลการทดลองแม้ว่าวิธีแตกครึ่งตามสารสนเทศจะให้ค่าความถูกต้องที่ต่ำกว่าวิธีอาร์เอดีเอจีในหลายดีกรีแต่ก็พบว่าวิธีการแตกครึ่งตามสารสนเทศให้ค่าจำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีอื่น แม้จะมีบางกรณีที่แบบสมดุลใช้เวลาในการจำแนกน้อยกว่าแต่กรณีดังกล่าวทุกกรณีวิธีการแตกครึ่งตามสารสนเทศจะให้ค่าความถูกต้องที่สูงกว่า และกรณีดีสุดด้านจำนวนครั้งของการจำแนกพบว่าสามารถลดจำนวนครั้งในการจำแนกลดลงจากวิธีอาร์เอดีเอจีลงได้มากกว่าแปดเท่าของเวลาที่อาร์เอดีเอจีใช้ และลดเวลาของการจำแนกด้วยวิธีแมกซ์วินลงได้มากกว่าสองร้อยเจ็ดสิบเท่า

4.4 เวลาที่ใช้ในการสอน

ในด้านเวลาสอนวิธีแมกซ์วินใช้เวลาเฉพาะในส่วนของการสร้างระนาบ $k*(k-1)/2$ ตัว สำหรับปัญหา k ประเภท วิธีอาร์เอดีเอจีจะใช้เวลาสอนเพิ่มเติมจากแมกซ์วินในส่วนของอัลกอริทึมการจับคู่ที่ดีที่สุดเพื่อเลือกตัวจำแนกที่มีค่าขอบเขตความผิดพลาดต่ำสุด สำหรับวิธีการแตกครึ่งตามสารสนเทศจะใช้เวลาสอนเพิ่มขึ้นจากแมกซ์วินในส่วนของการหาพารามิเตอร์ที่เหมาะสมแล้วจึงนำไปสร้างต้นไม้สำหรับการจำแนกแบบหลายประเภทซึ่งคล้ายกับกรณีของวิธีการแตกครึ่งแบบสมดุล แต่ละวิธีที่กล่าวมาจะใช้เวลาสอนเรียงลำดับจากน้อยไปหามากโดยแม้ว่าวิธีการในลำดับหลังจะใช้เวลาสอนน้อยกว่าแต่พบว่าใช้เวลาในการจำแนกที่รวดเร็วกว่ามากซึ่งมีความเหมาะสมกรณีนำไปใช้ในการจำแนกข้อมูลในปัญหาจริง

เมื่อนำผลการทดลองจากทั้ง 10 ชุดข้อมูลมาสรุปผลการทดลองกรณีที่ดีที่สุดของแต่ละวิธีจะแสดงได้ดังตารางที่ 12-14

ตารางที่ 12: การเปรียบเทียบค่าความถูกต้องใน Polynomial kernel ของการจำแนกกรณีที่ให้ค่าความถูกต้องสูงสุดของแต่ละวิธีในข้อมูลทั้ง 10 ชุด

Dataset	d	IBD	d	BD	d	RADAG	d	Max Wins
Glass	2	69.646	2	70.202	2	71.063	2	**71.078**
Satimage	5	88.550	3	87.480	6	**88.900**	6	88.453
Segment	8	**97.489**	8	96.788	8	**97.489**	8	97.379
Shuttle	8	**99.924**	8	**99.924**	8	**99.924**	8	**99.924**
Vowel	3	64.935	2	**65.022**	3	64.502	4	62.773
Soybean	3	90.882	4	89.529	3	**91.176**	3	90.471
Letter	3	93.650	4	91.640	3	**96.125**	3	**96.125**
Isolet	7	96.280	4	95.843	3	**97.049**	3	97.040
Thaiprinted Character1	3	68.658	3	71.366	2	**93.444**	2	93.280
Thaiprinted Character2	4	90.564	4	85.110	2	99.112	2	**99.989**
เปรียบเทียบ IBD vs วิธีอื่น (ชนะ-แพ้-เสมอ)			6-3-1		1-7-2		4-5-1	

ตารางที่ 13: การเปรียบเทียบค่าความถูกต้องใน RBF kernel ของการจำแนกกรณีที่ให้ค่าความถูกต้องสูงสุดของแต่ละวิธีในข้อมูลทั้ง 10 ชุด

Dataset	c	IBD	c	BD	c	RADAG	c	Max Wins
Glass	0.006	73.378	0.08	70.556	0.09	**74.319**	2	71.078
Satimage	0.5	91.950	1.0	91.350	0.5	91.950	0.5	**91.984**
Segment	0.7	97.273	0.7	96.857	0.7	97.273	0.7	**97.298**
Shuttle	4.0	**99.931**	4.0	99.897	3.0	99.897	3.0	99.897
Vowel	0.3	66.450	0.2	62.900	0.2	**67.100**	0.2	65.340
Soybean	0.007	**91.471**	0.007	89.000	0.05	90.882	0.08	90.468
Letter	3.0	95.175	4.0	94.695	3.0	**97.950**	5.0	97.689
Isolet	0.003	96.087	0.010	95.895	0.003	**97.115**	0.002	96.731
Thaiprinted Character1	0.0008	82.475	0.001	72.904	0.002	97.028	0.003	**97.080**
Thaiprinted Character2	0.004	93.719	0.003	82.886	0.004	**99.418**	0.004	99.385
เปรียบเทียบ IBD vs วิธีอื่น (ชนะ-แพ้-เสมอ)			10-0-0		2-6-2		4-6-0	

ตารางที่ 14: การเปรียบเทียบค่าจำนวนครั้งในการจำแนกกรณีที่ให้ค่าความถูกต้องสูงสุดของแต่ละวิธีในข้อมูลทั้ง 10 ชุด

Dataset	Expected Value				Experiment			
					Polynomial Kernel		RBF Kernel	
	IBD $(\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i))$	BD $(\log_2 k)$	RADAG $(k\text{-}1)$	Max wins $(k^*(k\text{-}1)/2)$	IBD	BD	IBD	BD
Glass	2.177	2.585	5	15	**4.207**	4.444	**4.430**	4.451
Satimage	2.474	2.585	5	15	4.990	**4.718**	4.734	**4.577**
Segment	2.807	2.807	6	21	**4.666**	4.886	5.094	**4.944**
Shuttle	0.965	2.807	6	21	**5.359**	5.378	5.783	**5.746**
Vowel	3.459	3.459	10	55	**5.385**	5.665	**5.628**	5.819
Soybean	3.617	3.907	14	105	6.791	**6.783**	**5.400**	6.216
Letter	4.700	4.700	25	325	**11.262**	12.179	**11.588**	15.216
Isolet	4.700	4.700	25	325	**8.528**	9.382	**8.366**	10.009
Thaiprinted Character1	6.087	6.087	67	2278	**9.675**	24.835	8.554	**8.295**
Thaiprinted Character2	6.087	6.087	67	2278	**10.672**	29.915	**8.487**	8.740
เปรียบเทียบ IBD vs วิธีอื่น(ชนะ-แพ้-เสมอ)		4-0-6	10-0-0	10-0-0		8-2-0		6-4-0

การประมาณค่าความแตกต่างของค่าความถูกต้องระหว่างวิธีการแตกครึ่งตามสารสนเทศกับวิธีอื่น โดยข้อมูลที่นำมาวิเคราะห์ใช้ข้อมูลจากตารางที่ 13-14 ภายใต้สมมติฐานที่ว่าข้อมูลดังกล่าวมีการกระจายแบบปกติ (Normal distribution) และใช้การทดสอบแบบทางเดียว (One-Sided Test)

ตามปกติค่าความผิดพลาดจริงเมื่อใช้สมมติฐาน h (True error: $\text{error}_D(h)$) สามารถประมาณค่าได้จากค่าความผิดพลาดของตัวอย่างเมื่อใช้สมมติฐาน h (Sample error: $\text{error}_S(h)$) โดยในการพิจารณาว่าสมมติฐานใดให้ค่าความถูกต้องสูงกว่า หรือกล่าวอีกนัยคือให้ความผิดพลาดต่ำกว่า สามารถหาคำตอบได้จากสูตรต่อไปนี้ [12]

$$\hat{d} \equiv \text{error}_S(h_1) - \text{error}_S(h_2) \tag{23}$$

เมื่อค่าความแปรปรวนของ $\hat{d}$ หาได้จากสูตร

$$\sigma_{\hat{d}} \approx \frac{\text{error}_{S1}(h_1)(1-\text{error}_{S1}(h_1))}{n_1} + \frac{\text{error}_{S2}(h_2)(1-\text{error}_{S2}(h_2))}{n_1} \tag{24}$$

กรณี $\text{error}_D(h_1) > \text{error}_D(h_2)$ หรือกล่าวได้ว่าค่าความถูกต้องของ h_2 มากกว่าหรือเท่ากับ h_1 จะต้องสอดคล้องกับเงื่อนไขต่อไปนี้

$$\hat{d} < \mu_{\hat{d}} + z\sigma_{\hat{d}} \tag{25}$$

เมื่อค่า z สอดคล้องกับระดับความเชื่อมั่น(Confidence Level)ที่กำหนด และ $\mu_{\hat{d}}$ มีค่าเป็นศูนย์ ที่ระดับความเชื่อมั่นมากกว่า 95 เปอร์เซ็นต์ ถือว่ามีความแตกต่างอย่างมีนัยสำคัญ

ตารางที่ 15: การเปรียบเทียบค่าระดับความเชื่อมั่น (Confidence Level) ของค่าความต่างของค่าความถูกต้องระหว่างวิธีแตกครึ่งตามสารสนเทศกับวิธีอื่น

Dataset	Polynomial			RBF		
	BD	Max Wins	RADAG	BD	Max Wins	RADAG
Glass	53.59	58.71	58.71	**67.00**	**64.06**	55.96
Satimage	**59.10**	**50.80**	53.19	**55.96**	50.40	**50.00**
Segment	**61.79**	**51.99**	**50.00**	**56.75**	50.40	**50.00**
Shuttle	**50.00**	**50.00**	**50.00**	**53.19**	**53.19**	**53.19**
Vowel	50.40	**62.55**	**52.39**	**70.19**	**56.75**	53.98
Soybean	**62.55**	**53.98**	52.79	**72.24**	**59.87**	**55.96**
Letter	**70.54**	78.81	78.81	**55.96**	83.15	85.99
Isolet	**56.36**	61.79	61.79	**52.79**	59.48	65.54
Thaiprinted Character1	66.28	>99.9995*	>99.9995*	**94.95**	>99.95*	>99.95*
Thaiprinted Character2	**88.10**	99.94*	99.73*	**99.22***	98.68*	98.75*

ผลการวิเคราะห์ค่าความแตกต่างของค่าความถูกต้องระหว่างวิธีการแตกครึ่งตามสารสนเทศกับวิธีอื่น แสดงดังตารางที่ 15 โดยตัวหนาหมายถึงกรณีนั้นๆ การแตกครึ่งตามสารสนเทศให้ค่าความถูกต้องมากกว่าหรือเท่ากับวิธีที่เปรียบเทียบด้วยระดับความเชื่อมั่นที่ระบุในเซลล์นั้นๆ

จากตารางที่ 12, 13 และ 15 เมื่อเปรียบเทียบค่าความถูกต้องในการจำแนกข้อมูลของวิธีการแตกครึ่งตามสารสนเทศกับวิธีอื่นๆ พบว่าไม่มีความแตกต่างอย่างมีนัยสำคัญ ยกเว้นในชุดข้อมูล Thaiprinted Character 1 และ Thaiprinted Character 2 สังเกตได้ที่เครื่องหมาย * แต่เมื่อพิจารณาค่าจำนวนครั้งในการจำแนกเฉลี่ยประกอบพบว่าวิธีการแตกครึ่งตามสารสนเทศให้จำนวนครั้งในการจำแนกที่ต่ำกว่าวิธีแมกซ์วินและวิธีอาร์เอดีเอจีในทุกกรณีโดยเฉพาะอย่างยิ่งเมื่อพิจารณาชุดข้อมูลที่มีจำนวนประเภทเป็นจำนวนมาก ยกตัวอย่างเช่น ชุดข้อมูล Isolet ซึ่งมีจำนวนประเภทเป็น 26 หากใช้วิธีอาร์เอดีเอจีจะมีจำนวนครั้งในการจำแนกเฉลี่ย 25 ครั้ง และหากใช้วิธีแมกซ์วินจะมีจำนวนครั้งในการจำแนกเฉลี่ยถึง 325 ครั้ง ขณะที่ใช้วิธีการแตกครึ่งตามสารสนเทศด้วย RBF kernel จะมีจำนวนครั้งในการจำแนกเฉลี่ยเพียง 8.366 ครั้ง

พิจารณาค่าจำนวนครั้งในการจำแนกที่คาดหวังและค่าจำนวนครั้งในการจำแนกจากการทดลองพบว่าสองค่านี้จะให้ค่าที่เท่ากันทั้งในวิธีอาร์เอดีเอจีและวิธีแมกซ์วิน แต่สำหรับวิธีการแตกครึ่งตามสารสนเทศและแบบสมดุลค่าจำนวนครั้งในการจำแนกที่คาดหวังสามารถคำนวณได้จากสูตร $\sum_{i=1}^{k} -P(m_i)\log_2 P(m_i)$ และ $\log_2 k$ ตามลำดับ ซึ่งจากตารางที่ 14 จะเห็นว่าสองวิธีนี้จะให้ค่าที่เท่ากันในหลายชุดข้อมูลโดยจะเท่ากันเมื่อความน่าจะเป็นในการเกิดข้อมูลแต่ละประเภทของชุดข้อมูลนั้นๆ มีค่าเท่ากัน กรณีดังกล่าวแม้ว่าค่าจำนวนครั้งในการจำแนกที่คาดหวังของสองวิธีนี้จะให้ค่าที่เท่ากันแต่เมื่อพิจารณาผลการทดลองพบว่าโดยส่วนใหญ่วิธีการแตกครึ่งตามสารสนเทศจะให้จำนวนครั้งในการจำแนกเฉลี่ยต่ำกว่าทั้งนี้เนื่องจากในการสร้างต้นไม้ในแต่ละระดับที่ถัดจากโนดรากเป็นต้นไปมีโอกาสในการเกิดข้อมูลแต่ละประเภทที่อาจไม่เท่ากัน เมื่อใช้วิธีการแตกครึ่งตามสารสนเทศในการสร้างต้นไม้จึงมีความได้เปรียบในการที่จะได้มาซึ่งต้นไม้ที่สั้นที่สุดมากกว่าวิธีการแตกครึ่งแบบสมดุล กล่าวโดยรวมจึงสรุปได้ว่าการแตกครึ่งตามสารสนเทศสามารถลดจำนวนครั้งในการจำแนกลงจากวิธีอื่นๆ ได้เป็นอย่างดี