รายงานการวิจัย

เรื่อง

ระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสม สำหรับนักศึกษาที่สมัครเข้าศึกษาต่อในระดับปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต โดยใช้เทคนิคเหมืองข้อมูล

The Development of Model for Online Autonomous Decision Support Systems in Selecting a Suitable Program for Entrance to Suan Dusit University by Applying a Data Mining

ผู้ช่วยศาสตราจารย์ ดร.ศิริพร ฉิมพลี

มหาวิทยาลัยสวนดุสิต

2560

ลิขสิทธิ์ของมหาวิทยาลัยสวนดุสิต

รายงานการวิจัย

เรื่อง

ระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสม สำหรับนักศึกษาที่สมัครเข้าศึกษาต่อในระดับปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต โดยใช้เทคนิคเหมืองข้อมูล

The Development of Model for Online Autonomous Decision Support Systems in Selecting a Suitable Program for Entrance to Suan Dusit University by Applying a Data Mining

ผู้ช่วยศาสตราจารย์ ดร.ศิริพร ฉิมพลี

คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยสวนดุสิต

2560

ลิขสิทธิ์ของมหาวิทยาลัยสวนดุสิต

(งานวิจัยนี้ได้รับทุนอุดหนุนจากมหาวิทยาลัยสวนดุสิต ปีงบประมาณ 2558)

หัวข้อวิจัย	ระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสม สำหรับนักศึกษาที่สมัครเข้าศึกษาต่อในระดับปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต โดยใช้เทคนิคเหมืองข้อมูล
ผู้ดำเนินการวิจัย	ผู้ช่วยศาสตราจารย์ ดร.ศิริพร ฉิมพลี
ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.สุชาดา โทผล
	ผู้ช่วยศาสตราจารย์ ดร.วิชชา ฉิมพลี
หน่วยงาน	คณะวิทยาศาสตร์และเทคโนโลยี
	มหาวิทยาลัยสวนดุสิต
ปี พ.ศ.	2560

การวิจัยครั้งนี้มีวัตถุประสงค์ ดังนี้ (1) เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมการจำแนกประเภท และ (2) เพื่อพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์ในการเลือกเรียนหลักสูตรที่เหมาะสม สำหรับนักศึกษาที่สมัครเข้าศึกษาต่อในระดับปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต ประชากรและกลุ่มตัวอย่างเป็นนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยี เครื่องมือที่ใช้ในงานวิจัย ได้แก่ (1) เทคนิคการทำเหมืองข้อมูลแบบการจำแนกประเภท โดยเป็นการศึกษาเปรียบเทียบประสิทธิภาพระหว่างการจำแนกประเภท 3 วิธี ประกอบด้วย การจำแนกประเภทแบบเบย์ด้วยอัลกอริทึมนาอีฟเบเซียน การจำแนกประเภทโดยใช้ต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 และอัลกอริทึม RandomForest และการจำแนกประเภทโดยใช้กฎด้วยอัลกอริทึม JRip (2) โปรแกรมเวกา (3) แบบสอบถามเพื่อประเมินความพึงพอใจและความเชื่อมั่นของระบบโดยผู้เชี่ยวชาญและผู้ใช้งาน สถิติที่ใช้ในงานวิจัย ได้แก่ ค่าเฉลี่ยเลขคณิตและค่าส่วนเบี่ยงเบนมาตรฐาน ผลการวิจัย มีดังนี้ (1) ต้นแบบที่ได้จากอัลกอริทึม JRip ให้ค่าความแม่นยำ (Accuracy) ความเที่ยง (Precision) คาการเรียกคืน (Recall) ค่าความถ่วงดุล (F-Measure) สูงที่สุด (2) ผลการประเมินประสิทธิภาพของระบบโดยผู้เชี่ยวชาญและผู้ใช้งานระบบโดยภาพรวม อยู่ในระดับมาก โดยมีค่าเฉลี่ยเท่ากับ 3.84 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.466 และค่าเฉลี่ยเท่ากับ 3.825 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.594 ตามลำดับ

Research Title	The Development of Model for Online Autonomous Decision Support Systems in Selecting a Suitable Program for Entrance to Suan Dusit University by Applying a Data Mining
Researcher	Assistant Professor Dr.Siriporn Chimphlee
Research Consultants	Assistant Professor Dr.Suchada Thophon
	Assistant Professor Dr.Witcha Chimphlee
Organization	Faculty of Science and Technology
	Suan Dusit University
Year	2017

This research has the following objectives: (1) to compare the performance of algorithms for classification, and (2) to develop an autonomous decision support systems in selecting a suitable program for entrance to Faculty of Science and Technology, Suan Dusit University. The population and sample are the student of Faculty of Science and Technology. Tools used in research are: (1) data mining techniques for classification. This study compared the efficacy of three methods including: Bayes' classification with naive bayesian algorithm classification, decision tree with J48 and RandomForest, and classification with JRip algorithm (2) WEKA program (3) questionnaires to evaluate the satisfaction and confidence of the system by experts and users. Statistics used in this research are the arithmetic mean and standard deviation. The results were as follows: (1) the model derived from JRip algorithm provides the highest accuracy, precision, recall and F-measure (2) the development of effective performance evaluation by experts and users were at good level $\bar{x}$ = 3.84, $S.D.$=0.466, and $\bar{x}$ = 3.825, $S.D.$=0.594, respectively.

กิตติกรรมประกาศ

รายงานวิจัยฉบับนี้สำเร็จลุล่วงด้วยดีด้วยความอนุเคราะห์จากผู้ช่วยศาสตราจารย์ ดร.สุชาดา โทผล และผู้ช่วยศาสตราจารย์ ดร.วิชชา ฉิมพลี ที่ให้คำแนะนำในการทำการวิจัย

ขอขอบคุณ ผู้ช่วยศาสตราจารย์ ดร.ธนากร ปามุนา ที่ให้คำแนะนำและข้อเสนอแนะในการปรับปรุงแก้ไขเครื่องมือที่ใช้ในการวิจัย

ขอขอบคุณบุคลากร และสำนักส่งเสริมวิชาการและงานทะเบียน ที่ให้ความอนุเคราะห์ข้อมูลที่ใช้ในการวิจัย

ขอขอบคุณนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยีที่ให้ความร่วมมือ ช่วยเหลือ และอำนวยความสะดวกแก่ผู้วิจัยเป็นอย่างดีในการทดลองและเก็บข้อมูลจน สำเร็จลุล่วงไปด้วยดี

ขอขอบคุณบุคลากรและสถาบันวิจัยและพัฒนาที่ให้ความร่วมมือ ช่วยเหลือ และอำนวยความสะดวกแก่ผู้วิจัยเป็นอย่างดีในการทำงานวิจัย

สุดท้ายนี้ขอขอบคุณมหาวิทยาลัยสวนดุสิตที่ให้ทุนสนับสนุนการวิจัยในครั้งนี้

ศิริพร ฉิมพลี

2560

สารบัญ

		หน้า
บทคัดย่อภาษาไทย		ก
บทคัดย่อภาษาอังกฤษ		ข
กิตติกรรมประกาศ		ค
สารบัญ		ง
สารบัญตาราง		ฉ
สารบัญภาพ		ช
บทที่ 1	**บทนำ**	1
	ความเป็นมาและความสำคัญ	1
	วัตถุประสงค์ของการวิจัย	2
	ขอบเขตการวิจัย	2
	คำจำกัดความที่ใช้ในงานวิจัย	3
	ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2	**แนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง**	4
	บริบทคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต	4
	การทำเหมืองข้อมูล	5
	การจำแนกประเภทของข้อมูล	9
	การพยากรณ์ข้อมูล	11
	การจำแนกประเภทโดยใช้นาอีฟเบเซียน	12
	การจำแนกประเภทโดยใช้ต้นไม้ตัดสินใจ	15
	การจำแนกประเภทโดยใช้อัลกอริทึม JRip	16
	ระบบสนับสนุนการตัดสินใจ	17
	โปรแกรมเวกา	19
	การพัฒนาเว็บแอปพลิเคชัน	21
	งานวิจัยที่เกี่ยวข้อง	24
	กรอบแนวคิดในการวิจัย	26

หน้า

บทที่ 3 วิธีดำเนินการวิจัย 28

ประชากรและการสุ่มกลุ่มตัวอย่าง 28

เครื่องมือในการวิจัยและการตรวจสอบคุณภาพเครื่องมือ 29

การเก็บรวบรวมข้อมูล 31

การวิเคราะห์ข้อมูล 32

การสร้างตัวแบบเพื่อการพยากรณ์ 33

การทดสอบตัวแบบเพื่อการพยากรณ์ 39

การพัฒนาระบบสนับสนุนการตัดสินใจ 41

บทที่ 4 ผลการวิจัย 43

ผลการประเมินประสิทธิภาพของอัลกอริทึม 43

ผลการวิเคราะห์ความพึงพอใจของระบบ 47

บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ 49

สรุปผลการวิจัย 49

อภิปรายผล 50

ข้อเสนอแนะในการนำผลการวิจัยไปใช้ 51

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป 51

บรรณานุกรม

บรรณานุกรมภาษาไทย 52

บรรณานุกรมภาษาต่างประเทศ 54

ภาคผนวก 56

ภาคผนวก ก ผลการวิเคราะห์ความเห็นของผู้เชี่ยวชาญและผู้ใช้งาน 57

ประวัติผู้วิจัย 61

สารบัญตาราง

ตารางที่		หน้า
2.1	ฐานข้อมูลของลูกค้าร้านเครื่องใช้ไฟฟ้าแห่งหนึ่ง	13
2.2	คำสั่งสำหรับการเรียกใช้งาน WEKA CLI	24
3.1	ตัวแปรที่ใช้ในงานวิจัย	34
3.2	ตัวอย่างข้อมูลที่ผ่านการเตรียมข้อมูลแล้วโดยไม่มีการแปลงข้อมูล	36
3.3	ตัวอย่างข้อมูลที่ผ่านการเตรียมข้อมูลแล้วโดยมีการแปลงข้อมูล	36
4.1	ค่าความแม่นยำเมื่อกำหนดระดับผลการเรียนเฉลี่ยเมื่อสำเร็จการศึกษา	44
4.2	ค่าความแม่นยำเมื่อมีการแปลงระดับผลการเรียนคะแนนเฉลี่ยเป็นคลาส	44
4.3	ผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึม ด้านความเที่ยง (Precision) ค่าการเรียกคืน (Recall) และค่าความถ่วงดุล (F-Measure)	46
4.4	ผลประเมินความพึงพอใจของระบบโดยผู้เชี่ยวชาญ	47
4.5	ผลประเมินความพึงพอใจของระบบโดยผู้ใช้งาน	48
ก-1	ผลการวิเคราะห์ความเห็นของผู้เชี่ยวชาญ	58
ก-2	ผลประเมินความพึงพอใจของระบบโดยผู้ใช้งาน	59

สารบัญภาพ

ภาพที่		หน้า
2.1	กระบวนการค้นหาความรู้ในฐานข้อมูล	6
2.2	กระบวนการสร้างแบบจำลองจำแนกประเภท	9
2.3	กระบวนการพยากรณ์	11
2.4	การสร้างต้นไม้ตัดสินใจจากข้อมูลลูกค้าที่ซื้อคอมพิวเตอร์	16
2.5	กระบวนการตัดสินใจและแก้ไขปัญหา	18
2.6	หน้าจอหลักของโปรแกรม Weka Explore	20
2.7	เมนูหลักของ Explorer	21
2.8	กรอบแนวคิดของการวิจัย	27
3.1	กระบวนการสร้างต้นแบบการพยากรณ์สำหรับการเลือกหลักสูตร	34
3.2	ตัวอย่างรูปแบบของไฟล์นามสกุล ARFF ที่ใช้ในการประมวลผลด้วยโปรแกรมเวกา	37
4.1	ผลการเปรียบเทียบค่าความแม่นยำของอัลกอริทึม	45
4.2	ผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึม ด้านความเที่ยง (Precision) ค่าการเรียกคืน (Recall) และค่าความถ่วงดุล (F-Measure)	46
4.3	ผลการเปรียบเทียบค่ากำลังสองของความคลาดเคลื่อน (Root Mean Square Error: RMSE)	47

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญ

การศึกษาต่อในระดับอุดมศึกษามีความสำคัญต่อนักเรียนเป็นอย่างมาก จะเห็นได้จากจำนวนนักศึกษาใหม่ ปีการศึกษา 2557 ระดับปริญญาตรี ในสถาบันอุดมศึกษารัฐ ที่มีจำนวนทั้งสิ้น 491,023 คน (สารสนเทศอุดมศึกษา. 2558) การเลือกหลักสูตรในการศึกษาต่อในระดับปริญญาตรีในสถาบันอุดมศึกษาต่าง ๆ จึงมีความสำคัญต่อนักเรียน เพราะส่งผลโดยตรงกับผู้สมัครทางด้านคุณภาพหลังสำเร็จการศึกษา อย่างไรก็ตามมีนักศึกษาจำนวนมากที่ไม่สามารถสำเร็จการศึกษาตามหลักสูตรที่มหาวิทยาลัยได้กำหนดไว้ เนื่องจากนักศึกษาเลือกเรียนต่อในหลักสูตรที่ตนเองรู้จักไม่ดีพอ การเลือกหลักสูตรที่เรียนโดยขาดข้อมูลที่สำคัญเกี่ยวกับหลักสูตรนั้น ๆ และนักศึกษาไม่ทราบว่าหลักสูตรใดที่เหมาะสมกับลักษณะเฉพาะของตนเอง นักศึกษาบางคนเลือกเรียนต่อในหลักสูตรที่ตนเองชอบเป็นการส่วนตัว เลือกเรียนตามเพื่อน เลือกเรียนตามคำแนะนำของผู้ปกครอง โดยไม่ได้พิจารณาถึงความสามารถและคุณสมบัติของตนเอง เมื่อเข้าไปศึกษาแล้วนักศึกษาจำนวนไม่น้อยจึงพบว่าตนเองไม่เหมาะสมกับหลักสูตรดังกล่าว ไม่สามารถเรียนต่อไปได้ ผลที่ตามมาคือ นักศึกษาจำนวนมากต้องขอย้ายหลักสูตร ขอพักการเรียน ขอลาออกจากการเรียน และพ้นสภาพจากการเรียนเนื่องจากสอบไม่ผ่านตามเกณฑ์ที่กำหนดไว้ในแต่ละรายวิชาของแต่ละหลักสูตร และทำระดับคะแนนเฉลี่ยสะสมอยู่ในระดับต่ำ และสุดท้ายนักศึกษาไม่สำเร็จการศึกษาตามหลักสูตรที่มหาวิทยาลัยวางไว้ นอกจากสาเหตุดังกล่าวข้างต้นแล้ว นักศึกษาเองก็ไม่มั่นใจว่าจะวางแผนเตรียมตัวสำหรับความพร้อมในการตัดสินใจเลือกเรียนในสาขาวิชาใดถึงจะประสบผลสำเร็จ ดังนั้นถ้าสามารถที่จะคาดการณ์ได้ว่านักศึกษาที่เข้ามาศึกษานั้นมีความเหมาะสมกับหลักสูตรที่เลือกหรือไม่ ก็จะสามารถให้คำแนะนำในการเลือกหลักสูตรที่เหมาะสมให้กับนักศึกษาได้ เป็นการลดจำนวนของนักศึกษาที่ไม่สามารถสำเร็จการศึกษาได้อย่างมีประสิทธิภาพ

การพยากรณ์หรือเลือกหลักสูตรที่เหมาะสมใช้ข้อมูลของนักศึกษาที่สถาบันการศึกษาต่าง ๆ ได้เก็บรวบรวมข้อมูลไว้และมีจำนวนมาก โดยที่สถาบันการศึกษาส่วนใหญ่ไม่ได้นำข้อมูลเหล่านั้นมาวิเคราะห์ หรือนำไปใช้ให้เกิดประโยชน์มากนัก ดังนั้นหากสามารถนำข้อมูลเกี่ยวกับนักศึกษาที่มีอยู่นี้มาวิเคราะห์ เพื่อหาความรู้ที่เป็นประโยชน์ ก็จะเป็นวิธีที่สามารถช่วยนักศึกษาในการเลือกสาขาที่เหมาะสมกับตนเองมากที่สุดได้ การทำเหมืองข้อมูลเป็นวิธีที่ถูกนำมาใช้ในการค้นหารูปแบบความสัมพันธ์ องค์ความรู้ จากฐานข้อมูลขนาดใหญ่ หรือที่เรียกว่า Knowledge Discovery in Database: KDD โดยมุ่งเน้นข้อมูลในเชิงลึกตามวัตถุประสงค์ จึงสามารถนำมาใช้ในการเลือก

หลักสูตรที่เหมาะสมให้กับนักศึกษา (อัจฉราภรณ์ จุฑาผาด กานต์ เจริญจิตร และพยุง มีสัจ. 2558; สุธีรา วงศ์อนันทรัพย์ ต้องใจ แย้มผกา และอรวรรณ มูสิกะ. 2559; เยาวภา ภารสำเร็จ. 2556) โดยการนำข้อมูลของนักศึกษาที่เรียนในแต่ละหลักสูตรมาสร้างต้นแบบโดยใช้วิธีการจำแนกประเภท (Classification) พร้อมกับการค้นหาคุณสมบัติของข้อมูลที่มีผลต่อการสำเร็จการศึกษาในหลักสูตรที่เลือกเรียน ต้นแบบที่มีความน่าเชื่อถือที่สร้างขึ้นนี้สามารถนำมาใช้กับนักศึกษาที่เข้ามาใหม่ โดยการนำข้อมูลของนักศึกษาใหม่มาทดสอบกับต้นแบบนี้ ก็จะสามารถพยากรณ์ได้ว่านักศึกษาใหม่มีความเหมาะสมกับหลักสูตรใด หรือมีความเหมาะสมกับหลักสูตรที่ได้เลือกเรียนหรือไม่ มีโอกาสของการสำเร็จการศึกษาตามหลักสูตรที่ตนเองได้เลือกไว้หรือไม่ และหากมีนักศึกษาคนใดที่มีแนวโน้มหรือเลือกศึกษาในหลักสูตรที่ไม่เหมาะสมกับตนเองแล้ว หลักสูตรหรืออาจารย์ที่ปรึกษาจะได้ให้คำแนะนำหรือให้ความช่วยเหลือที่เหมาะสมให้แก่นักศึกษาต่อไป เป็นการป้องกันการเกิดปัญหาในอนาคต

งานวิจัยนี้เป็นการพัฒนาต้นแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสม สำหรับนักศึกษาที่สมัครเข้าศึกษาต่อในระดับปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต ข้อมูลที่นำมาใช้ ได้แก่ ข้อมูลส่วนตัวนักศึกษา และข้อมูลผลการเรียน แล้วนำมาวิเคราะห์เพื่อหาประสิทธิภาพของอัลกอริทึมการจำแนกประเภท อัลกอริทึมที่เหมาะสมที่สุดจะถูกนำมาสร้างเป็นต้นแบบสำหรับการพัฒนาระบบสนับสนุนการตัดสินใจสำหรับการเลือกหลักสูตรที่เหมะสม และสามารถนำมาใช้ประกอบการตัดสินใจของนักเรียนในการเลือกหลักสูตรที่เหมาะสมได้

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมการจำแนกประเภท
2. เพื่อพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์ในการเลือกเรียนหลักสูตรที่เหมาะสม สำหรับนักศึกษาที่สมัครเข้าศึกษาต่อในระดับปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต

ขอบเขตการวิจัย

การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสมสำหรับนักศึกษา มีขอบเขตการวิจัย ดังนี้

1. ใช้ข้อมูลของนักศึกษาที่สำเร็จการศึกษาไปแล้วย้อนหลัง 4 ปี ตั้งแต่ปี พ.ศ. 2553-2556
2. ใช้ข้อมูลของนักศึกษาที่กำลังศึกษาอยู่ในชั้นปีที่ 1 คณะวิทยาศาสตร์และเทคโนโลยี ในการทดสอบการทำนาย

คำจำกัดความที่ใช้ในงานวิจัย

การทำเหมืองข้อมูล (Data mining) คือ การค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลจำนวนมาก

ระบบสนับสนุนการตัดสินใจ เป็นระบบที่ช่วยในการตัดสินใจเกี่ยวกับการจัดการ การรวบรวมข้อมูล การวิเคราะห์ข้อมูล และการสร้างตัวแบบที่ซับซ้อน ภายใต้ซอฟต์แวร์เดียวกัน

ตัวแบบ หมายถึง แบบอย่างของสิ่งใดสิ่งหนึ่งเพื่อเป็นแนวทางในการสร้างหรือทำซ้ำ เป็นตัวอย่างเพื่อการเลียนแบบ เป็นแผนภูมิหรือรูปสามมิติซึ่งเป็นตัวแทนของสิ่งใดสิ่งหนึ่งหรือหลักการหรือแนวคิด เป็นชุดของปัจจัย หรือตัวแปรที่มีความสัมพันธ์ซึ่งกันและกัน

ประโยชน์ที่คาดว่าจะได้รับ

1. นักเรียนมีระบบที่จะช่วยในการตัดสินใจในการเข้าศึกษาต่อในระดับอุดมศึกษา
2. มหาวิทยาลัยสวนดุสิตมีระบบให้บริการด้านการแนะแนวการศึกษาต่อในระดับอุดมศึกษา
3. มหาวิทยาลัยสวนดุสิตมีระบบที่ช่วยในการวางแผนและบริหารงานด้านการรับสมัครนักศึกษา

บทที่ 2
แนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ได้ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง เพื่อนำมาใช้ในการพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสมสำหรับนักศึกษา ได้แก่ บริบทคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต การทำเหมืองข้อมูล การจำแนกประเภทของข้อมูล การพยากรณ์ข้อมูล การจำแนกประเภทโดยใช้นาอีฟเบเซียน การจำแนกประเภทโดยใช้ต้นไม้ตัดสินใจ การจำแนกประเภทโดยใช้อัลกอริทึม JRip ระบบสนับสนุนการตัดสินใจ โปรแกรมเวกา การพัฒนาเว็บแอปพลิเคชัน งานวิจัยที่เกี่ยวข้อง และกรอบแนวคิดในการวิจัย

บริบทคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต

ในปี พ.ศ. 2552 คณะวิทยาศาสตร์และเทคโนโลยี เปิดสอนนักศึกษาระดับปริญญาตรี หลักสูตรวิทยาศาสตรบัณฑิต จำนวน 7 หลักสูตร คือ วิทยาการคอมพิวเตอร์ เทคโนโลยีสารสนเทศ วิทยาศาสตร์และเทคโนโลยีการอาหาร อุตสาหกรรมอาหารและการบริการ การจัดการสิ่งแวดล้อม วิทยาการความปลอดภัย และวิทยาศาสตร์เครื่องสำอาง

ต่อมาในปี พ.ศ. 2554 คณะวิทยาศาสตร์และเทคโนโลยี เปิดรับนักศึกษาภาคปกติหลักสูตรวิทยาศาสตรบัณฑิต จำนวน 5 หลักสูตร คือ วิทยาการคอมพิวเตอร์ เทคโนโลยีสารสนเทศ การจัดการสิ่งแวดล้อมเมืองและอุตสาหกรรม วิทยาการความปลอดภัย และวิทยาศาสตร์เครื่องสำอาง

พุทธศักราช 2555 คณะวิทยาศาสตร์และเทคโนโลยีเปิดรับนักศึกษาภาคปกติ หลักสูตรวิทยาศาสตรบัณฑิต จำนวน 6 หลักสูตร คือ วิทยาการคอมพิวเตอร์ เทคโนโลยีสารสนเทศ การจัดการสิ่งแวดล้อมเมืองและอุตสาหกรรม วิทยาการความปลอดภัย วิทยาศาสตร์เครื่องสำอาง และเทคโนโลยีเคมี

พุทธศักราช 2557 คณะวิทยาศาสตร์และเทคโนโลยีเปิดรับนักศึกษาภาคปกติ หลักสูตรวิทยาศาสตรบัณฑิต จำนวน 6 หลักสูตร คือ วิทยาการคอมพิวเตอร์ เทคโนโลยีสารสนเทศ การจัดการสิ่งแวดล้อมเมืองและอุตสาหกรรม วิทยาการความปลอดภัย และวิทยาศาสตร์เครื่องสำอาง และเปิดรับนักศึกษาภาคปกติหลักสูตรศึกษาศาสตรบัณฑิต จำนวน 2 หลักสูตร คือ ศึกษาศาสตร์สาขาวิชาฟิสิกส์ และศึกษาศาสตร์ สาขาวิชาคณิตศาสตร์

พุทธศักราช 2558 มหาวิทยาลัยราชภัฏสวนดุสิต ได้เปลี่ยนสถานะเป็นมหาวิทยาลัยในกำกับของรัฐ ตามพระราชบัญญัติมหาวิทยาลัยสวนดุสิต พ.ศ. 2558 ซึ่งประกาศในราชกิจจานุเบกษา เมื่อวันที่ 17 กรกฎาคม พ.ศ. 2558 และมีผลบังคับใช้วันที่ 18 กรกฎาคม พ.ศ. 2558 โดยใช้ชื่อว่า มหาวิทยาลัยสวนดุสิต

การทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Miming) เป็นการสืบค้นความรู้ที่เป็นประโยชน์และน่าสนใจบนฐานข้อมูลขนาดใหญ่ (Knowledge Discovery in Database: KDD) ข้อมูลจะถูกวิเคราะห์และค้นหาความรู้ที่ซ่อนอยู่ เช่น รูปแบบความสัมพันธ์ การเปลี่ยนแปลงโครงสร้างที่เด่นชัด หรือลักษณะที่ผิดปกติของข้อมูล การทำเหมืองข้อมูลสามารถนำไปใช้ในการวิเคราะห์การทำงานได้หลายลักษณะ เช่น การวิเคราะห์ด้านการตลาด การบริหารธุรกิจ รวมถึงระบบที่ช่วยสนับสนุนการตัดสินใจ เป็นต้น

มีผู้ให้ความหมายเกี่ยวกับการทำเหมืองข้อมูลไว้ ดังนี้

Linoff & Berry (2011:7) กล่าวว่า การทำเหมืองข้อมูลคือการสำรวจและวิเคราะห์ข้อมูลที่มีขนาดใหญ่ เพื่อค้นหารูปแบบหรือกฎที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่นั้นๆ และนำความรู้ที่ได้ไปใช้ให้เกิดประโยชน์ในการพัฒนาองค์กร

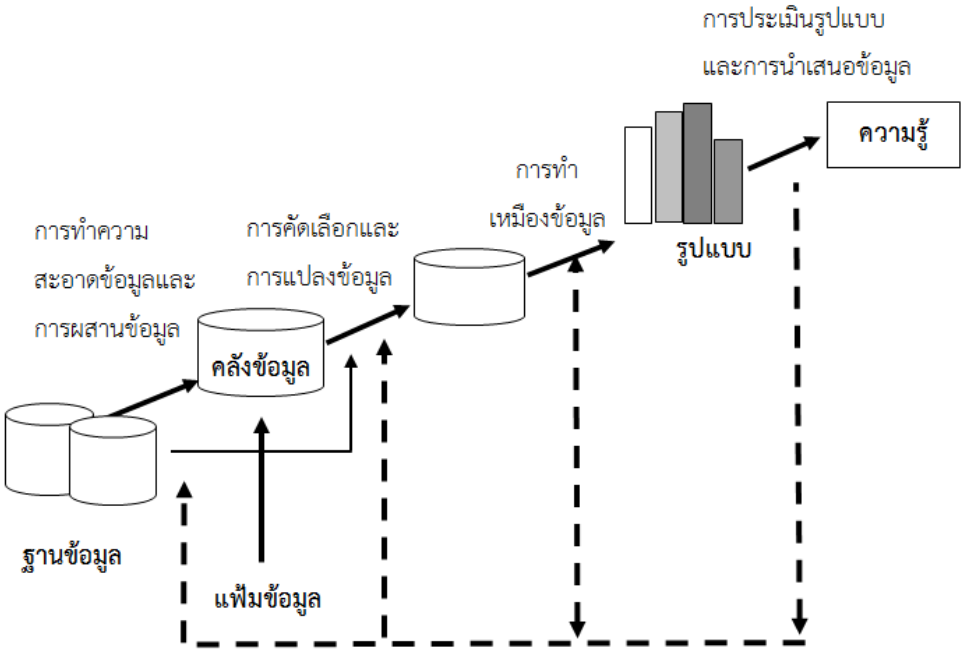
Han, Kamber & Pei (2011: 5) กล่าวว่า การทำเหมืองข้อมูล คือ การสกัดหรือขุดหาความรู้จากข้อมูลจำนวนมาก

Witten, Frank & Hall (2011: 5) กล่าวว่า การทำเหมืองข้อมูลเป็นการแก้ปัญหาโดยการวิเคราะห์ข้อมูลที่มีอยู่แล้วในฐานข้อมูล

จากความหมายที่มีผู้ให้ไว้ข้างต้น สามารถสรุปได้ว่า การทำเหมืองข้อมูล (Data mining) คือ การค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลจำนวนมาก

เป้าหมายของการทำเหมืองข้อมูล เป็นการสกัด หรือค้นหารูปแบบของข้อมูลที่ฝังลึก และซ่อนอยู่ภายในฐานข้อมูลขนาดใหญ่ โดยใช้เครื่องมือที่ทันสมัย สามารถแสดงผลแบบกราฟิก ผู้ใช้สามารถดูข้อมูลแบบเจาะลึก และสามารถสอบถามข้อมูลได้ โดยการเรียนรู้จากข้อมูลในอดีต หรือข้อมูลปัจจุบัน ผลลัพธ์ที่ได้จากการสังเคราะห์ของการทำเหมืองข้อมูลอาจจะเป็นความรู้ที่ไม่เคยรู้มาก่อน (Unknown) ความรู้ที่มีความถูกต้อง (Valid) และความรู้ที่นำไปใช้ตัดสินใจ (Actionable)

Han, Kamber & Pei (2011: 7) อธิบายกระบวนการค้นหาความรู้ในการทำเหมืองข้อมูลประกอบด้วย 7 ขั้นตอน ดังภาพที่ 2.1 ซึ่งเป็นการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ดังนี้



ภาพที่ 2.1 กระบวนการค้นหาความรู้ในฐานข้อมูล
ที่มา: Han, Kamber & Pei (2011: 7)

1. การทำความสะอาดข้อมูล (Data Cleaning) เป็นขั้นตอนในการขจัดข้อมูลที่ไม่ต้องการออกไป แล้วจึงนำไปประมวลผลเพราะข้อมูลที่เก็บรวมรวมมามีจำนวนมากอาจมีความผิดปกติ เช่น ข้อมูลขาดหายไป มีข้อมูลรบกวน ข้อมูลไม่มีความสอดคล้องกัน เป็นต้น การทำความสะอาดข้อมูลเป็นการเลือกข้อมูลให้ตรงประเด็น เพื่อคุณภาพของข้อมูลที่จะนำไปวิเคราะห์

2. การผสานข้อมูล (Data Integration) เป็นการรวมกันของข้อมูลที่มาจากแหล่งที่มาที่แตกต่างกันแล้วนำมาไว้ด้วยกัน อาจมีความซ้ำซ้อนของข้อมูล (Data Redundancies) จึงต้องมีการผสานข้อมูล เพื่อลดหรือหลีกเลี่ยงปัญหาความไม่สอดคล้องกันของข้อมูล (Data inconsistencies)

3. การคัดเลือกข้อมูล (Data Selection) เป็นการดึงเฉพาะข้อมูลที่ต้องการนำมาวิเคราะห์ การคัดเลือกข้อมูลเป็นการระบุถึงแหล่งข้อมูลที่จะนำมาทำเหมืองข้อมูล รวมถึงการนำข้อมูลที่ต้องการออกจากฐานข้อมูล เพื่อสร้างกลุ่มข้อมูลสำหรับพิจารณาในเบื้องต้น

4. การแปลงข้อมูล (Data Transformation) เป็นการเปลี่ยนรูปของข้อมูลให้เหมาะสมสำหรับการวิเคราะห์ เช่น การแปลงรูปแบบของข้อมูล การลดรูป และจัดข้อมูลให้อยู่ในรูปแบบเดียวกัน เป็นต้น เพื่อให้เป็นมาตรฐาน และเหมาะสมที่จะนำไปใช้กับการทำเหมืองข้อมูล

5. การทำเหมืองข้อมูล (Data Mining) เป็นการขุดค้นรูปแบบที่เป็นประโยชน์จากข้อมูลที่ให้มา โดยเลือกเทคนิคที่เหมาะสมกับลักษณะของข้อมูล และปัญหาที่ต้องการคำตอบ

6. การประเมินรูปแบบ (Pattern Evaluation) เป็นการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล เพื่อให้ได้รูปแบบที่น่าสนใจและเป็นตัวแทนของความรู้ที่ต้องการทำการค้นหา ในขั้นตอนนี้จะเป็นการวิเคราะห์ผลลัพธ์ที่ได้ แปลความหมาย และประเมินผลว่าผลลัพธ์นั้นเหมาะสมหรือตรงวัตถุประสงค์หรือไม่ จึงต้องอาศัยทักษะในการวิเคราะห์ข้อมูลทางธุรกิจเข้าช่วย

7. การนำเสนอความรู้ (Knowledge Presentation) เป็นขั้นตอนการนำเสนอความรู้ที่ได้ค้นพบต่อผู้ใช้ เพื่อนำความรู้ที่ค้นพบไปประยุกต์ใช้งานจริงต่อไป

เทคนิคของการทำเหมืองข้อมูลเหมาะสำหรับการทำงานที่มีลักษณะดังต่อไปนี้

1. การบรรยายลักษณะของข้อมูล (Description) เป็นการบรรยายลักษณะของประชากรผลิตภัณฑ์ หรือขบวนการเพื่อสร้างความเข้าใจให้มากขึ้น เป็นการป้องกันความผิดพลาดที่อาจเกิดจากการนำข้อมูลขนาดเล็กมาใช้ แล้วทำให้ข้อสรุปที่ได้ไม่น่าเชื่อถือ

2. การวิเคราะห์คุณสมบัติและการแยกแยะข้อมูล (Characterization and discrimination) การวิเคราะห์คุณสมบัติที่สำคัญเพื่อนำมาใช้ในการคัดเลือก เช่น การพิจารณารับพนักงานเข้าทำงานต้องพิจารณาคุณสมบัติจากข้อมูลในใบสมัคร และการสัมภาษณ์ ส่วนการแยกแยะข้อมูลจะทำให้ทราบถึงสมรรถนะของคนที่มาสมัคร เป็นต้น

3. การแบ่งประเภทข้อมูล (Classification) เป็นการหาชุดต้นแบบ หรือชุดของการทำงานที่อธิบายและแบ่งประเภทข้อมูล วัตถุประสงค์ของการแบ่งประเภทข้อมูล คือ ต้องการใช้เป็นต้นแบบทำนายประเภทของวัตถุ หรือข้อมูลที่ไม่มีการระบุประเภท หรือชนิดของข้อมูล ซึ่งต้นแบบสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (Training data) โดยอาจจะเป็นกลุ่มข้อมูลที่มีการระบุประเภทหรือกลุ่มเรียบร้อยแล้ว เครื่องมือที่ใช้ในการจัดแบ่งประเภทของข้อมูล ได้แก่ การสร้างต้นไม้ตัดสินใจ (Decision trees) โครงข่ายประสาทเทียม (Neural networks) ทฤษฎีของเบย์ (Bayes theorem) เป็นต้น ตัวอย่างของการจัดหมวดหมู่ ได้แก่ การจัดประเภทของผู้สมัครใช้บัตรเครดิตเป็นผู้ที่มีความเสี่ยงระดับต่ำ ผู้ที่มีความเสี่ยงระดับกลาง และผู้ที่มีความเสี่ยงระดับสูง เป็นต้น การแบ่งประเภทข้อมูลเป็นปัญหาพื้นฐานของการเรียนรู้แบบมีผู้สอน เป็นการทำนายประเภทของวัตถุจากคุณสมบัติต่าง ๆ ของวัตถุ ซึ่งการเรียนรู้แบบมีผู้สอนจะสร้างฟังก์ชันเชื่อมโยงระหว่างคุณสมบัติของวัตถุกับประเภทของวัตถุจากตัวอย่างสอน แล้วจึงใช้ฟังก์ชันทำนายประเภทของวัตถุที่ไม่เคยพบ

4. การทำนายล่วงหน้า (Prediction) มีลักษณะการทำงานคล้ายกับการจัดประเภทและการประเมินค่า แต่มีการใช้สถิติการบันทึกของการจัดประเภทในการทำนายอนาคตของพฤติกรรม หรือการประเมิน

ค่าที่จะเกิดขึ้นในอนาคต เช่น การทำนายการเปลี่ยนแปลงพฤติกรรมของตลาด หรือการทำนายจำนวนลูกค้าที่จะเพิ่มขึ้นในอีก 6 เดือนข้างหน้า เป็นต้น

5. การค้นหารูปแบบที่เกิดขึ้นบ่อย ความสัมพันธ์ และสหสัมพันธ์ (Mining Frequent Patterns, Associations, and Correlations) เป็นการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลตั้งแต่สองชุดขึ้นไป ความสัมพันธ์ที่ได้จะอยู่ในรูปของกฎความสัมพันธ์ กฎที่ได้จะเรียกว่าเป็นกฎที่สำคัญขึ้นกับค่าพารามิเตอร์ 2 ชนิด คือ ค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence) ค่าสนับสนุน คือ เปอร์เซ็นต์ของการดำเนินการที่กฎสามารถนำไปใช้ได้ สำหรับค่าความเชื่อมั่น (Confidence) คือ จำนวนของกรณีที่กฎสามารถทำนายได้ถูกต้อง โดยสัมพันธ์กับจำนวนของกรณีที่กฎสามารถนำไปใช้ได้ ในการหากฎความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน วิธีการนี้ถูกนำไปใช้ในการวิเคราะห์การซื้อสินค้า (Market Basket Analysis) คือ รายการทั้งหมดที่ลูกค้าซื้อต่อครั้ง การวิเคราะห์นี้สามารถนำมาใช้ประโยชน์ในการตัดสินใจ เช่น การเตรียมสินค้าคงเหลือ การวางแผนจัดชั้นวางสินค้า และการวางแผนเพื่อกระตุ้นการจำหน่าย เป็นต้น

6. การจัดกลุ่ม (Clustering) เป็นการนำข้อมูลที่มีคุณลักษณะเหมือนกัน หรือคล้ายกันจัดไว้ในกลุ่มเดียวกัน ขั้นตอนที่ใช้ในการจัดกลุ่มจะอาศัยความเหมือน หรือความใกล้ชิด โดยคำนวณจากการวัดระยะห่างเวกเตอร์ของข้อมูลเข้า ซึ่งมีการวัดระยะแบบต่าง ๆ เช่น การวัดระยะแบบยูคลีเดียน (Euclidean distance) การวัดระยะแบบแมนฮัตตัน (Manhattan distance) การวัดระยะแบบ เชบิเชฟ (Chebychev distance) เป็นต้น เพื่อช่วยในการลดขนาดข้อมูล การจัดกลุ่มนี้ไม่มีการกำหนดจำนวนกลุ่มไว้ล่วงหน้า และการจัดกลุ่มจะเกิดขึ้นขณะที่กำลังเรียนรู้ จึงไม่จำเป็นต้องอาศัยตัวอย่าง หรือข้อมูลสอน

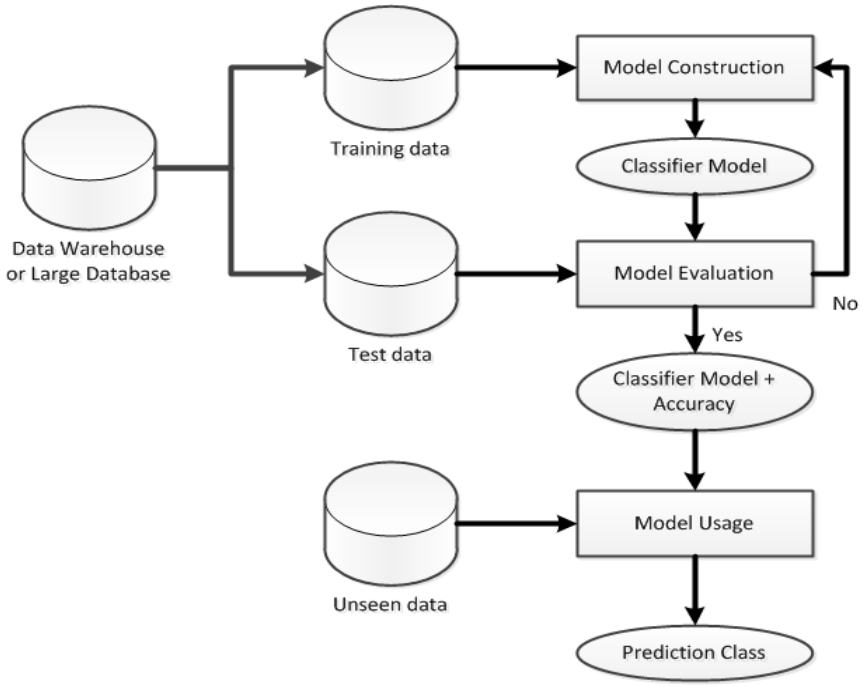
7. การวิเคราะห์ข้อมูลที่มีความผิดปกติ (Outlier analysis) ฐานข้อมูลอาจประกอบด้วยข้อมูลที่ไม่สอดคล้องกับลักษณะการทำงานหรือรูปแบบทั่วไปของข้อมูล ข้อมูลเหล่านั้นจะถูกเรียกว่าข้อมูลผิดปกติ

8. การวิเคราะห์วิวัฒนาการ (Evolution analysis) อธิบายและกำหนดรูปแบบหรือแนวโน้มของข้อมูลที่มีการเปลี่ยนแปลงพฤติกรรมตลอดเวลา รวมถึงการจำแนกลักษณะ การเลือกปฏิบัติ การวิเคราะห์ความสัมพันธ์และสหสัมพันธ์ การจำแนกประเภท และการจัดกลุ่มข้อมูล

การจำแนกประเภทของข้อมูล

การจำแนกประเภทเป็นเทคนิคที่ใช้ในการจำแนกคลาสของข้อมูลด้วยคุณลักษณะต่าง ๆ ที่ได้มีการกำหนดไว้แล้ว โดยอาศัยข้อมูลส่วนหนึ่งมาสอนระบบให้เรียนรู้และสามารถจำแนกข้อมูลนั้นออกมาเป็นกลุ่ม ๆ ตามที่กำหนดไว้ ผลที่ได้จากการเรียนรู้นี้ คือ ตัวจำแนกกลุ่มข้อมูล เมื่อมีข้อมูลใหม่เข้ามาในระบบ ตัวจำแนกประเภทจะสามารถทำนายประเภทหรือคลาสของข้อมูลที่ควรจะเป็นได้โดยอัตโนมัติ

การสร้างแบบจำลองจำแนกประเภทข้อมูล จะเกิดขึ้นมาจากการหาความสัมพันธ์ของข้อมูลในฐานข้อมูลขนาดใหญ่ โดยข้อมูลทั้งหมดจะมีการแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มข้อมูลเรียนรู้หรือข้อมูลที่ใช้สำหรับสอน (Training set) เป็นชุดข้อมูลที่มีบทบาทในการสร้างจำลองจำแนกประเภทข้อมูลขึ้นมา และกลุ่มข้อมูลทดสอบ (Test set) เป็นชุดข้อมูลที่ใช้ประเมินความถูกต้องของแบบจำลองที่สร้างขึ้น กระบวนการสร้างแบบจำลองจำแนกประเภท ประกอบด้วย 3 ขั้นตอน ดังภาพที่ 2.2 โดยมีรายละเอียด ดังนี้



ภาพที่ 2.2 กระบวนการสร้างแบบจำลองจำแนกประเภท
ที่มา: ชินพัฒน์ แก้วชินพร (2553)

1. การสร้างแบบจำลอง

การสร้างแบบจำลอง (Model construction) จะใช้ข้อมูลที่ได้เรียนรู้เสร็จเรียบร้อยแล้ว ซึ่งแต่ละตัวจะมีคุณลักษณะหนึ่งที่บอกค่าประเภทที่กำหนดไว้ล่วงหน้า หรือมีการระบุคลาสไว้เรียบร้อยแล้ว ข้อมูลชุดนี้ถูกเรียกว่า ชุดข้อมูลสอน (Training data) แบบจำลองที่ได้มาจากผลการใช้อัลกอริทึมใดอัลกอริทึมหนึ่งที่จัดอยู่ในการทำเหมืองข้อมูลจำแนกประเภท เช่น ต้นไม้ตัดสินใจ (Decision tree) แบบจำลองที่ได้จะมีลักษณะคล้ายต้นไม้จริงกลับหัวที่มีโหนดรากอยู่ด้านบนสุดและโหนดใบอยู่ล่างสุดของต้นไม้ แต่ละโหนดบนต้นไม้จะมีคุณลักษณะ (Attribute) เป็นตัวเลือกทดสอบ ซึ่งจะมีกิ่งซึ่งเป็นค่าที่เป็นไปได้ของคุณลักษณะ (Attribute value) ที่ถูกเลือกทดสอบและมีโหนดใบแสดงคลาสที่กำหนดไว้

2. การทดสอบความแม่นยำของแบบจำลอง

แบบจำลองที่ได้จากการเรียนรู้ต้องถูกตรวจสอบความถูกต้อง (Evaluation of model accuracy) โดยอาศัยข้อมูลที่ใช้สำหรับทดสอบเรียกว่า ชุดข้อมูลทดสอบ (Test data) ซึ่งข้อมูลในชุดนี้จะต้องเป็นคนละชุดกับที่ใช้ในการเรียนรู้ กลุ่มที่ถูกต้องของข้อมูลทดสอบจะถูกนำมาเปรียบเทียบกับกลุ่มที่หามาได้จากแบบจำลองจำแนกประเภท เพื่อทดสอบว่าแบบจำลองจำแนกประเภทนี้สามารถจัดกลุ่มประเภทข้อมูลได้อย่างถูกต้องมากน้อยเพียงใด

3. การนำแบบจำลองที่ได้ไปใช้

การนำแบบจำลองที่ได้ไปใช้ (Model usage) เป็นขั้นตอนการนำแบบจำลองที่สร้างขึ้นมาไปใช้กับข้อมูลที่ไม่เคยเห็นมาก่อน (Unseen data) เพื่อทำนายและกำหนดกลุ่มให้กับข้อมูลนั้น เป็นการจำแนกประเภทข้อมูลในอนาคต โดยจะต้องมีการหาค่าความแม่นยำ (Accuracy) ของแบบจำลองที่ได้ก่อนนำไปใช้ โดยเปรียบเทียบค่าฉลากประเภท (Class label) ที่ทราบล่วงหน้าของตัวอย่างในชุดทดสอบกับค่าผลลัพธ์การจำแนกประเภทที่ได้จากแบบจำลอง เรียกว่า อัตราความแม่นยำ (Accuracy rate) เป็นเปอร์เซ็นต์ของการจำแนกประเภทได้อย่างถูกต้อง ชุดข้อมูลทดสอบเป็นอิสระไม่ขึ้นต่อชุดข้อมูลสอน เพื่อป้องกันการเกิดปรากฏการที่แบบจำลองหรือตัวจำแนกประเภทที่ได้มีความพอดีเกินไปกับชุดข้อมูลสอน (Over fitting) กล่าวคือ จะให้ผลการจำแนกประเภทที่ถูกต้องสำหรับตัวอย่างในชุดข้อมูลสอน แต่ใช้ได้ไม่ดีกับกรณีตัวอย่างทั่วไป สาเหตุอาจมาจากเซตข้อมูลมีขนาดเล็กเกินไป หรือข้อมูลที่ใช้สอนมีความผิดปกติ เป็นต้น

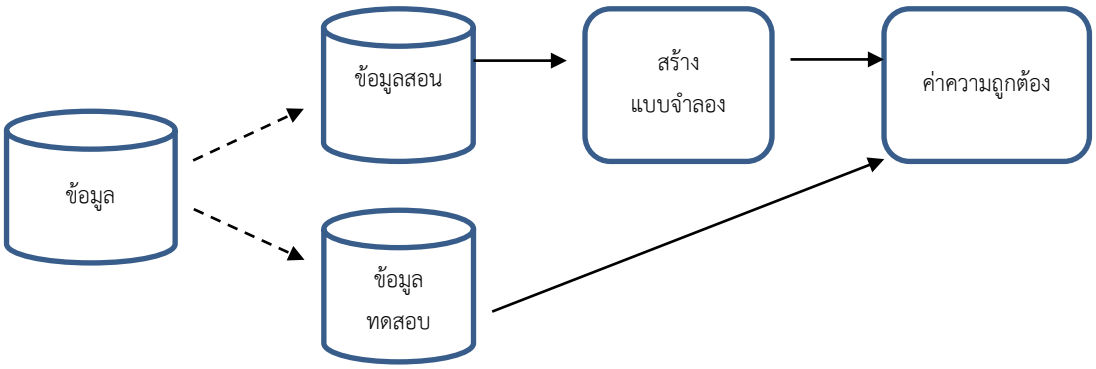
การพยากรณ์ข้อมูล

การพยากรณ์ข้อมูล (Data prediction) เป็นกระบวนการสร้างโมเดล เพื่อทำนายหาค่าที่ต้องการจากข้อมูลที่มีอยู่ โดยมีกระบวนการสร้างโมเดลคล้ายกับการจำแนกประเภทข้อมูลดังที่ได้กล่าวมาข้างต้นต่างกันตรงที่การพยากรณ์ข้อมูลไม่มีการจัดข้อมูลเข้ากลุ่มตามที่ได้กำหนด แต่การพยากรณ์ข้อมูลนี้เป็นการหาค่าที่ต้องการออกมาเป็นตัวเลข ตัวอย่างเช่น หายอดขายของเดือนถัดไปจากข้อมูลการขายทั้งหมดที่ผ่านมา หรือทำนายเกรดเฉลี่ยของนักเรียนในปีการศึกษาหน้าจากข้อมูลการลงทะเบียนของนักศึกษาทั้งหมด เป็นต้น การพยากรณ์ข้อมูลมีกระบวนการ ดังภาพที่ 2.3 มีรายละเอียดดังนี้ (นงเยาว์ ในอรุณ. 2557: 34; Han, Kamber & Pei. 2011: 364)

1. การเรียนรู้ เพื่อหาโมเดล เป็นขั้นตอนในการสร้างโมเดล โดยการเรียนรู้จากชุดข้อมูลสอนโมเดลที่ได้สามารถเป็นไปได้หลายรูปแบบ

2. การประมาณค่าความถูกต้องของโมเดล เป็นขั้นตอนการประมาณค่าความถูกต้องโดยใช้ชุดข้อมูลทดสอบ ค่าที่ได้นี้จะนำไปเปรียบเทียบค่าที่ได้จากโมเดล เพื่อประเมินค่าความถูกต้อง

3. การนำโมเดลไปใช้ เป็นการนำโมเดลมาใช้กับข้อมูลที่ไม่เคยเห็นมาก่อน โดยการกำหนดค่าผลลัพธ์ หรือคลาสให้กับข้อมูลที่ไม่เคยเห็นมาก่อนนี้



ภาพที่ 2.3 กระบวนการพยากรณ์

ที่มา: Han, Kamber & Pei (2011: 364)

การจำแนกประเภทโดยใช้นาอีฟเบเซียน

การจำแนกประเภทโดยใช้นาอีฟเบเซียน (Naïve Bayesian Classification) เป็นวิธีการเรียนรู้ที่ใช้หลักการของความน่าจะเป็น ซึ่งมีพื้นฐานมาจากทฤษฎีของเบย์ (Bayes theorem) เข้ามาช่วยในการเรียนรู้ เพื่อสร้างแบบจำลองที่อยู่ในรูปของความน่าจะเป็น จากค่าที่บันทึกได้จากการสังเกต เพื่อหาแบบจำลองที่ถูกต้องที่สุด (Linnoff & Berry. 2011) นาอีฟเบเซียนเป็นการคำนวณหาความน่าจะเป็นของแต่ละประเภทหรือกลุ่ม ซึ่งในที่นี้ เรียกว่า คลาส เมื่อกำหนดคุณลักษณะและค่าคุณลักษณะแต่ละตัวที่จะใช้ในการทำนายแล้วมาคำนวณหาค่าความน่าจะเป็นของทุกคลาสเพื่อนำมาเปรียบเทียบกัน ค่าความน่าจะเป็นที่สูงที่สุดของคลาสใด ๆ จะเป็นผลของการทำนายเพียงค่าเดียว ซึ่งกำหนดให้คุณลักษณะแต่ละตัวมีความเป็นอิสระต่อกัน (Class Conditional Independence) โดยมีการทำงานดังนี้ (Han, Kamber & Pei. 2011: 297, เพชรรัตน์ ปัญญาภาณุวัฒน์, พรชัย มงคลนาม, และวิทิดา จงศุภชัยสิทธิ์. 2552)

1. ให้ X เป็นเซตของเหตุการณ์หนึ่งๆ จะได้ $X = \{x_1, x_2,...,x_n\}$ เมื่อกำหนดให้ x คือ ค่าของ X ในลักษณะประจำที่ $A_1, A_2,..., A_n$ ตามลำดับ

2. สมมติให้มีคลาสที่เป็นไปได้จำนวน m คลาส คือ $C_1, C_2,...,C_n$ สำหรับเหตุการณ์ X โดยการทำนายจะเลือกค่าความน่าจะเป็นของคลาสที่สูงที่สุดที่ทำให้เกิดเหตุการณ์ X มาเป็นผลของการทำนาย $P(C_i|X) > P(C_j|X)$ สำหรับ $1 \le j \le m, j \ne i$ จากทฤษฎีของเบย์ ดังสมการที่ 2.1

$$P(C_i/X) = \frac{P(X/C_i)P(C_i)}{P(X)} \quad (2.1)$$

3. เมื่อ $P(X)$ เป็นค่าคงที่สำหรับทุกคลาส ดังนั้นการพิจารณา จึงเลือกจากค่า $P(X|C_i)P(C_i)$ ที่สูงที่สุด โดย $P(C_i)= s_i/s$ เมื่อ s_i คือ จำนวนข้อมูลที่เป็นคลาส C_i และ S คือจำนวนข้อมูลทั้งหมด

4. เนื่องจากนาอีฟเบเซียนมีเงื่อนไขว่า ลักษณะประจำแต่ละตัวเป็นอิสระต่อกัน ดังนั้นค่าของ $P(X)$ จะหาได้จากสมการที่ 2.2

$$P(X/C_i) = \prod_{k=1}^{n} P(x_k/C_i) \quad (2.2)$$

ค่าความน่าจะเป็น $P(x_1|C_i), P(x_2|C_i),..., P(x_n|C_i)$ สามารถคำนวณได้จากลักษณะประจำของข้อมูลที่นำมาทำการเรียนรู้โดย แบ่งได้ออกเป็น 2 กรณี คือ

ถ้า A_k เป็นข้อมูลเชิงคุณภาพ $P(X_k/C_i) = \frac{s_{ik}}{s_i}$ เมื่อ s_{ik} คือ จำนวนข้อมูลที่เป็นคลาส C_i ซึ่งมีค่าเท่ากับ x_k สำหรับลักษะประจำที่ A_k และ s_i คือจำนวนข้อมูลที่เป็นคลาส C_i

ถ้า A_kเป็นข้อมูลเชิงปริมาณ จะได้

$$P(X_k/C_i) = \mathrm{g}(x_k, \mu_{C_i}, \sigma_{C_i}) \tag{2.3}$$

$$P(x_k/C_i) = \frac{1}{\sqrt{2\pi\sigma_{C_i}}} e^{\frac{-(x_k-\mu_{C_i})^2}{2\sigma_{C_i}^2}} \tag{2.4}$$

เมื่อ $\mathrm{g}(x_k, \mu_{C_i}, \sigma_{C_i})$ คือ ฟังก์ชันเกาส์ (Gaussian function) สำหรับลักษณะประจำ A_k, μ_{C_i} และ σ_{C_i} คือ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของลักษณะประจำ A_k ที่เป็นคลาส C_i ตามลำดับ

5. ในการทำนายคลาสของเหตุการณ์ X จะทำการคำนวณ $P(X|C_i)P(C_i)$ สำหรับคลาสทุกคลาสและเหตุการณ์ X จะถูกจัดอยู่ในคลาสที่มีค่า $P(X|C_i)P(C_i)$ สูงสุด

ตารางที่ 2.1 ฐานข้อมูลของลูกค้าร้านเครื่องใช้ไฟฟ้าแห่งหนึ่ง

อายุ	รายได้	เป็นนักเรียน	สถานะทางการเงิน	ซื้อคอมพิวเตอร์
≤30	สูง	ไม่ใช่	ปานกลาง	ไม่ซื้อ
≤30	สูง	ไม่ใช่	ดี	ไม่ซื้อ
31-40	สูง	ไม่ใช่	ปานกลาง	ซื้อ
>40	ปานกลาง	ไม่ใช่	ปานกลาง	ซื้อ
>40	ต่ำ	ใช่	ปานกลาง	ซื้อ
>40	ต่ำ	ใช่	ดี	ไม่ซื้อ
31-40	ต่ำ	ใช่	ดี	ซื้อ
≤30	ปานกลาง	ไม่ใช่	ปานกลาง	ไม่ซื้อ
≤30	ต่ำ	ใช่	ปานกลาง	ซื้อ
>40	ปานกลาง	ใช่	ปานกลาง	ซื้อ
≤30	ปานกลาง	ใช่	ดี	ซื้อ
31-40	ปานกลาง	ไม่ใช่	ดี	ซื้อ

อายุ	รายได้	เป็นนักเรียน	สถานะทางการเงิน	ซื้อคอมพิวเตอร์
31-40	สูง	ใช่	ปานกลาง	ซื้อ
>40	ปานกลาง	ไม่	ดี	ไม่ซื้อ

จากข้อมูลในตารางที่ 2.1 ข้อมูลประกอบด้วยคุณลักษณะจำนวน 5 ตัว ได้แก่ อายุ (Age) รายได้ (Income) นักเรียน (Student) สถานะทางการเงิน (Credit rating) และกลุ่มของข้อมูล (Class) คือ กลุ่มที่ซื้อคอมพิวเตอร์ (Class = ซื้อ) และกลุ่มที่ไม่ซื้อคอมพิวเตอร์ (Class = ไม่ซื้อ) เซตของข้อมูลในตาราง 2.1 ประกอบด้วยข้อมูลจำนวน 14 ตัวอย่าง แบ่งข้อมูลออกเป็น 2 กลุ่ม คือ ข้อมูลที่ตัดสินใจซื้อคอมพิวเตอร์ (*Class* = ซื้อ) จำนวน 9 ตัวอย่าง และตัดสินใจไม่ซื้อคอมพิวเตอร์ (*Class* = ไม่ซื้อ) จำนวน 5 ตัวอย่าง นำมาสร้างแบบจำลองจำแนกประเภทข้อมูล ได้ดังนี้ กำหนดให้

C_1 แทน class buys_computer = yes (หมายถึง กลุ่ม C_1 แทนกลุ่มที่ซื้อคอมพิวเตอร์)

C_2 แทน class buys_computer = no (หมายถึง กลุ่ม C_2 แทนกลุ่มที่ไม่ซื้อคอมพิวเตอร์)

เมื่อมีตัวอย่างที่เข้ามาใหม่ (new instance) จะจัดให้อยู่ในกลุ่มใด

X = (age = “≤ 30”, income = “medium”, student = “yes”, credit_rating = “fair”)

ต้องการหาค่า $P(X|C_i)P(C_i)$ ที่สูงสุด สำหรับ i = 1, 2. $P(C_i)$ คือ ค่าความน่าจะเป็นที่ทราบแล้วที่ได้จากข้อมูลในตาราง

P(buys_computer = “yes” = 9/14 = 0.643

P(buys_computer = “no” = 5/14 = 0.357

ต้องการหาค่า $P(X|C_i)$ สำหรับ i = 1, 2 สามารถคำนวณหาค่าความน่าจะเป็นแบบมีเงื่อนไข (condition probability) ได้ดังนี้

P(age = “≤30”|buys_computer = “yes”) = 2/9 = 0.222

P(age = “<30”|buys_computer = “no”) = 3/5 = 0.6

P(income = “medium”|buys_computer = “yes”) = 4/9 = 0.444

P(income = “medium”|buys_computer = “no”) = 2/5 = 0.4

P(student = “yes”|buys_computer = “yes”) = 6/9 = 0.667

P(student = “yes”|buys_computer = “no”) = 1/5 = 0.2

P(credit_rating = “fair”|buys_computer = “yes”) = 6/9 = 0.667

P(credit_rating = “fair”|buys_computer = “no”) = 2/5 = 0.4

จากค่าความน่าจะเป็นที่คำนวณได้ สามารถนำมาหาค่า

P(X/buys_computer = “yes”) = 0.222 x 0.444 x 0.667 = 0.044

P(X/buys_computer = “no”) = 0.6 x 0.4 x 0.2 x 0.4 = 0.019

P(X/buys_computer = “yes”) P(buys_computer = “yes”)

= 0.044 x 0.643 = 0.028

P(X/buys_computer = “no”) P(buys_computer = “no”)

= 0.019 x 0.357 = 0.007

ดังนั้น ใช้นาอีฟเบเซียนจำแนกกลุ่มสำหรับ *X* ได้ว่าอยู่ใน คลาส buys_computer = “yes” ดังนั้น ถ้ามีอายุน้อยกว่าหรือเท่ากับ 30 ปี มีรายได้ปานกลาง เป็นนักเรียน มีสถานะทางการเงินอยู่ในระดับปานกลาง จะซื้อคอมพิวเตอร์

การจำแนกประเภทโดยใช้ต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจ (Decision tree) เป็นขั้นตอนวิธีการเรียนรู้ที่มีการอนุมานความรู้ (Inference engine) ที่ได้ในรูปกฎ “ถ้า....แล้ว...” ซึ่งเข้าใจได้ง่าย และมีความเหมาะสมสำหรับการแก้ปัญหาที่มีข้อมูล 2 กลุ่ม และเป็นข้อมูลที่ไม่มีความต่อเนื่อง เช่น ในงานวิจัยของดิษฐพล มั่นธรรม (2553) ได้ประยุกต์ใช้ต้นไม้ตัดสินใจในการวินิจฉัยโรคระบบทางเดินหายใจ สอดคล้องกับนิเวศ จิระวิชิตชัย (2553) ที่นำต้นไม้ตัดสินใจไปใช้สร้างโมเดลการวิเคราะห์โรคอัตโนมัติ นอกจากนี้ต้นไม้ตัดสินใจยังสามารถนำไปช่วยสนับสนุนการตัดสินใจในการเลือกสาขาการเรียนของนักศึกษา ระดับปริญญาตรีในงานวิจัยของไพฑูรย์ จันทร์เรือง (2550) เป็นต้น ในการแก้ปัญหาทำโดยเลือกคุณสมบัติของข้อมูล (Attributes) ที่มีคุณภาพดี สำหรับสร้างเป็นโหนดราก (Root node) และเลือกคุณสมบัติของข้อมูลที่มีคุณภาพรองลงมาสำหรับสร้างโหนดลูกของโหนดรากในชั้นต่อๆ ไปและไปสิ้นสุดที่โหนดใบ (Leaf node)

การสร้างต้นไม้ตัดสินใจ มี 2 ขั้นตอนหลัก คือ การสร้างแบบจำลอง (Model construction) และการนำแบบจำลองที่ได้ไปใช้ (Model usage)

จากตารางที่ 2.1 ข้อมูลลูกค้าที่ซื้อคอมพิวเตอร์ สามารถนำมาสร้างเป็นโมเดลได้จำนวน 5 โมเดล ดังภาพที่ 2 ซึ่งสามารถนำมาแปลงเป็นกฎจากต้นไม้ตัดสินใจ ได้ทั้งหมด 5 กฎ คือ

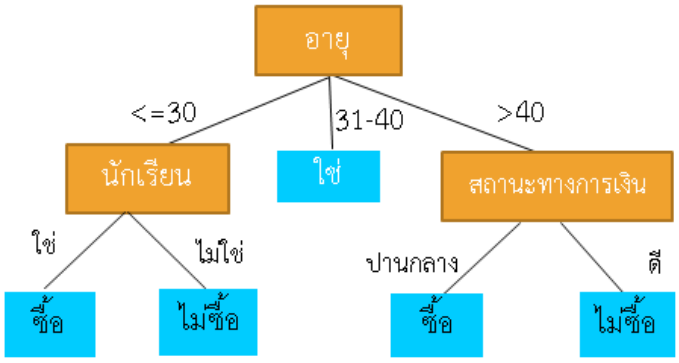
กฎข้อที่ 1: ถ้า (อายุ ≤ 30) และ เป็นนักเรียนแล้ว จะซื้อคอมพิวเตอร์

กฎข้อที่ 2: ถ้า (อายุ ≤ 30) และไม่ได้เป็นนักเรียนแล้ว จะไม่ซื้อคอมพิวเตอร์

กฎข้อที่ 3: ถ้า (อายุ 31-40) แล้ว จะซื้อคอมพิวเตอร์

กฎข้อที่ 4: ถ้า (อายุ > 40) และสถานะทางการเงินปานกลาง แล้วจะซื้อคอมพิวเตอร์

กฎข้อที่ 5: ถ้า (อายุ > 40) และสถานะทางการเงินดี แล้วจะไม่ซื้อคอมพิวเตอร์



ภาพที่ 2.4 การสร้างต้นไม้ตัดสินใจจากข้อมูลลูกค้าที่ซื้อคอมพิวเตอร์

จากภาพที่ 2.4 เมื่อมีลูกค้าใหม่ที่มีลักษณะเหมือนกับลูกค้าที่ได้จากการสร้างแบบจำลอง คือ มีอายุอยู่ระหว่าง 31-40 ปี และมีรายได้สูง ก็จะทำนายว่าลูกค้าคนใหม่ที่เข้ามาจะซื้อคอมพิวเตอร์ด้วย

การจำแนกประเภทโดยใช้อัลกอริทึม JRip

อัลกอริทึม JRip เป็นเทคนิคการจำแนกประเภทข้อมูล (ปราลี มณีรัตน์. 2555; Xu. 2014) ประกอบด้วย 3 ขั้นตอน คือ

ขั้นตอนที่ 1 การสร้างกฎเพื่อใช้ในการจำแนกประเภทข้อมูล เป็นการทำงานแบบวนรอบจนกว่าข้อมูลที่อยู่ในคลาสถูกนำมาสร้างเป็นกฎครบทุกตัว

ขั้นตอนที่ 2 เป็นขั้นตอนการปรับแต่งกฎที่ได้จากขั้นตอนที่ 1 ในขั้นตอนนี้มีการสร้างกฎเพิ่มอีก 2 แบบ คือ กฎแบบแทนที่ (Replacement) และกฎแบบแก้ไข (Revision)

- กฎแบบแทนที่เป็นการสร้างกฎที่มีการตัดออก โดยพิจารณาจากผลรวมของความถูกต้องของกฎที่ได้ทั้งหมด

- กฎแบบแก้ไข มีการเพิ่มแอททริบิวท์เข้าไปในกฎที่สร้างได้จากขั้นตอนที่ 1 แล้วมีการตัดออกโดยพิจารณาจากผลรวมของความถูกต้องของกฎที่ได้ทั้งหมด

ขั้นตอนที่ 3 การเลือกกฎที่ดีที่สุด โดยตัดทิ้งบางกฎออกไปโดยค่าความยาวของคำอธิบายที่น้อยที่สุด (Minimum description length) จะเป็นเกณฑ์ในการหยุดกระบวนการค้นหากฎ (พรพิรุณ โอ่งอินทร์ และวรรณา ศิริแสงตระกูล. 2558)

ระบบสนับสนุนการตัดสินใจ

ระบบสนับสนุนการตัดสินใจ (Decision Support Systems) ถูกพัฒนาขึ้นเพื่อการจัดการที่ช่วยในการตัดสินใจ โดยระบบรวบรวมข้อมูล วิเคราะห์ข้อมูล และสร้างแบบจำลองในการตัดสินใจ (ณัฏภัทรศญา ทับทิมเทศ. 2550) โดยมีส่วนประสานการทำงานระหว่างบุคลากรกับเทคโนโลยีทางด้านซอฟแวร์ที่มีการโต้ตอบกันเพื่อแก้ปัญหาแบบกึ่งโครงสร้างและไม่มีโครงสร้าง ระบบสนับสนุนการตัดสินใจจะไม่ทำการตัดสินใจแทนผู้ตัดสินใจ แต่จะนำเสนอข้อมูลสำคัญที่ช่วยในการตัดสินใจ ผู้ตัดสินใจต้องเป็นผู้ทำการตัดสินใจโดยใช้ข้อมูลสำคัญที่ได้รับจากระบบ ร่วมกับเหตุผล ประสบการณ์และความคิดของตนเป็นหลัก โดยอาศัยกระบวนการตัดสินใจและแก้ไขปัญหา ลักษณะของปัญหาที่ผู้ตัดสินใจต้องเผชิญมี 3 ลักษณะ ดังนี้

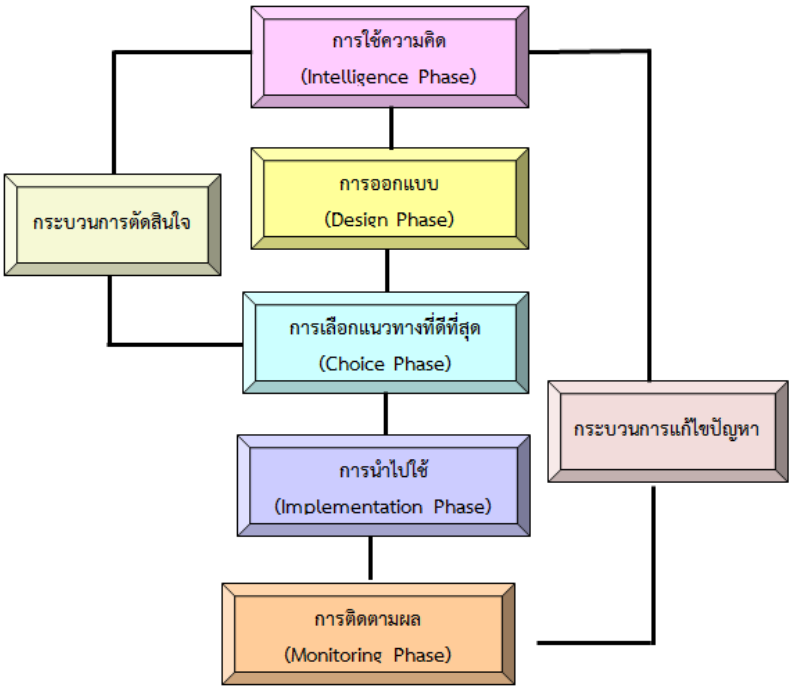
1. ปัญหาแบบมีโครงสร้าง (Structured problem) เป็นปัญหาที่มีวิธีการแก้ไขปัญหาได้อย่างชัดเจนแน่นอนหรือสามารถจำลองปัญหาได้ด้วยแบบจำลองทางคณิตศาสตร์ และแทนค่าในสูตรจนสามารถคำนวณหาคำตอบได้อย่างชัดเจน หรือกล่าวอีกนัยหนึ่งคือ ปัญหาที่ผู้ตัดสินใจมีข้อมูลและสารสนเทศประกอบการตัดสินใจอย่างครบถ้วนและสามารถนำไปใช้ในการแก้ปัญหาได้โดยการเขียนโปรแกรม เช่น ปัญหาการกำหนดระดับสินค้าคงคลัง ปริมาณการสั่งซื้อที่ประหยัดที่สุด และจุดสั่งซื้อ ผู้ตัดสินใจก็สามารถคำนวณหาผลลัพธ์ได้อย่างแน่นอน

2. ปัญหาแบบไม่มีโครงสร้าง (Unstructured problem) เป็นปัญหาที่ไม่สามารถหาวิธีการแก้ไขได้อย่างชัดเจนและแน่นอน ไม่สามารถจำลองได้ด้วยสูตรทางคณิตศาสตร์ หรือกล่าวอีกนัยหนึ่งคือ ปัญหาที่ผู้ตัดสินใจมีข้อมูลและสารสนเทศไม่เพียงพอต่อการแก้ไขปัญหา จึงต้องอาศัยประสบการณ์ของผู้ตัดสินใจแก้ไขปัญหา เช่น ปัญหาการเลือกลงทุนกับหุ้นที่อยู่ในตลาดหลักทรัพย์ เนื่องจากผู้ตัดสินใจไม่สามารถทราบได้แน่นอนว่าหุ้นที่ตัดสินใจลงทุนไปนั้นจะให้ผลตอบแทนสูงสุดหรือไม่เมื่อถึงสิ้นปี

3. ปัญหาแบบกึ่งมีโครงสร้าง (Semi-structure problem) เป็นปัญหาที่มีลักษณะเฉพาะส่วนมากจะไม่เกิดซ้ำและไม่มีกระบวนการดำเนินการมาตรฐาน หรือเป็นปัญหาที่มีวิธีการแก้ไขเพียงบางส่วนเท่านั้น ส่วนที่เหลือจะต้องอาศัยประสบการณ์หรือความชำนาญในการตัดสินใจแก้ไขปัญหา ส่วนเทคโนโลยีสารสนเทศให้ได้แค่การสนับสนุนเท่านั้น เช่น จากปัญหาการกำหนดระดับสินค้าคงคลัง ผู้ตัดสินใจไม่สามารถทราบปริมาณความต้องการสินค้าที่แน่นอนได้ จึงไม่สามารถคำนวณหาผลลัพธ์ได้อย่างถูกต้อง แต่อาศัยประสบการณ์ในการคาดการณ์ปริมาณความต้องการที่จะเกิดขึ้นในอนาคต

กระบวนการตัดสินใจและแก้ไขปัญหา คือ การกำหนดขั้นตอนในการตัดสินใจแก้ไขปัญหาที่เกิดขึ้นภายในองค์กรอย่างมีหลักเกณฑ์ ด้วยการกำหนดขั้นตอนตั้งแต่ขั้นแรกจนถึงขั้นตอนสุดท้ายเพื่อให้ได้ผลลัพธ์ที่ต้องการ ดังภาพที่ 2.5 กระบวนการตัดสินใจตามแนวคิดของ Herbert Simon

แบ่งแยกระยะของการตัดสินใจออกเป็น 3 ขั้นตอน ได้แก่ การใช้ความคิด (Intelligence Phase) การออกแบบ (Design Phase) และการเลือกทางเลือกที่ดีที่สุด (Choice Phase) ต่อมา George Huber ได้นำมารวมเข้ากับกระบวนการแก้ไขปัญหา จึงทำให้การตัดสินใจและกระบวนการแก้ไขปัญหารวมแล้วมีจำนวนทั้งหมด 5 ขั้นตอน ได้แก่ 1) การใช้ความคิด (Intelligence Phase) 2) การออกแบบ (Design Phase) 3) การเลือกทางเลือกที่ดีที่สุด (Choice Phase) 4) การนำไปใช้ (Implementation Phase) และ 5) การติดตามผล (Monitoring Phase) โดยมีรายละเอียดดังนี้



ภาพที่ 2.5 กระบวนการตัดสินใจและแก้ไขปัญหา
ที่มา: กิตติ ภักดีวัฒนะกุล (2550)

ขั้นตอนที่ 1 การใช้ความคิด (Intelligence Phase) เป็นการค้นหาสาเหตุของปัญหา โดยการศึกษาถึงต้นเหตุของปัญหา ประเมินผลที่จะเกิดขึ้นหากไม่ทำการแก้ไขปัญหา วิเคราะห์ปัจจัยแวดล้อมของปัญหา เพื่อสร้างแบบจำลองที่ใช้อธิบายลักษณะและสาเหตุของปัญหา โดยอาจใช้การจำแนกปัญหาออกเป็นส่วนย่อยและคิดวิธีการแก้ไขปัญหา ซึ่งผลที่ได้รับจากขั้นตอนนี้ เรียกว่า “Decision Statement” หรือ “การระบุปัญหา” เช่น ในการตัดสินใจเลือกรายวิชาที่จะลงทะเบียนเรียนในภาคเรียนหน้า ต้องทำการสร้าง Decision Statement ซึ่งคือ รายวิชาที่ควรลงทะเบียน เป็นต้น สิ่งสำคัญ คือ ต้องทำการจำแนกหาสาเหตุที่แท้จริงของปัญหาก่อนทำการแก้ไข ไม่ควรแก้ไขที่ปลายเหตุ เช่น เมื่อได้รับการตำหนิจากลูกค้าเรื่องการส่งของล่าช้า แล้วทำการแก้ไขโดยจัดให้มีโทรศัพท์สายด่วน

เพื่อให้ลูกค้าแจ้งปัญหาการส่งสินค้า ซึ่งจักเป็นการเพิ่มงานให้กับแผนกจัดส่งสินค้า (เนื่องจากต้องจัดพนักงานรับโทรศัพท์สายด่วนจากลูกค้า) โดยที่ไม่ได้เพิ่มความเร็วในการส่งสินค้า ซึ่งเป็นการแก้ปัญหาที่ไม่ตรงจุด

ขั้นตอนที่ 2 การออกแบบ (Design Phase) เป็นการสร้างและวิเคราะห์ทางเลือกในการตัดสินใจ โดยทางเลือกที่สร้างขึ้นมาจะต้องมีความเป็นไปได้ในการแก้ไขปัญหาให้ได้ผลประโยชน์สูงสุด และมีการกำหนดวัตถุประสงค์ของการตัดสินใจ เพื่อให้ผู้ตัดสินใจสามารถสร้างทางเลือกได้หลาย ๆ ทาง ประกอบการตัดสินใจ ในขั้นตอนนี้อาจมีการสร้างแบบจำลอง (Model) แผนภาพการตัดสินใจแบบต้นไม้ (Decision Tree) หรือตารางการตัดสินใจ (Decision Table)

ขั้นตอนที่ 3 การเลือกทางเลือกที่ดีที่สุด (Choice Phase) เป็นขั้นตอนการค้นหาและประเมินทางเลือกต่าง ๆ ที่ได้จากขั้นตอนการออกแบบและคัดเลือกให้เหลือทางเลือกเดียว ผลลัพธ์ที่ได้ คือ ทางเลือกเพื่อการนำไปใช้จริงในการแก้ปัญหา และทำการประเมินทางเลือกต่าง ๆ ตามส่วนที่พิจารณาที่ช่วยให้ได้เลือกทางเลือกได้อย่างถูกต้อง

ขั้นตอนที่ 4 การนำไปใช้ (Implementation Phase) ขั้นตอนการนำทางเลือกในการแก้ไขปัญหาที่ได้จากข้อที่ 3 ไปลงมือปฏิบัติเพื่อแก้ไขปัญหาจริง ซึ่งอาจจะประสบความสำเร็จ หรือล้มเหลวก็ได้ หากล้มเหลวก็อาจย้อนกลับไปสู่ขั้นตอนใดขั้นตอนหนึ่ง เพื่อทบทวนกระบวนการใหม่

ขั้นตอนที่ 5 การติดตามผล (Monitoring Phase)

เป็นขั้นตอนสุดท้ายของกระบวนการตัดสินใจและแก้ไขปัญหา ผู้ตัดสินใจจะมีการประเมินผลหลังจากนำแนวทางที่ได้เลือกแล้วไปใช้มนการแก้ไขปัญหา หากผลลัพธ์ที่ได้ไม่เป็นที่น่าพอใจ ต้องพิจารณาถึงสาเหตุว่า เกิดขึ้นจากขั้นตอนใด เพื่อนำไปปรับปรุงการตัดสินใจแก้ไขปัญหาใหม่อีกครั้งหนึ่ง

โปรแกรมเวกา

เวกา หรือ WEKA ย่อมาจากคำว่า Waikato Environment for Knowledge Analysis เป็นซอฟต์แวร์สำเร็จรูปที่ได้รับการพัฒนามาตั้งแต่ปี ค.ศ.1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ อยู่ภายใต้การควบคุมของ GPL License ซอฟต์แวร์ WEKA ได้ถูกพัฒนาด้วยภาษาจาวาทั้งหมด เหมาะกับงานทางด้านการเรียนรู้ด้วยเครื่อง (Machine learning) และการทำเหมืองข้อมูล (Data mining) ซอฟต์แวร์ประกอบไปด้วยโมดูลย่อยสำหรับใช้ในการจัดการข้อมูล และเป็นซอฟต์แวร์ที่สามารถใช้ Graphic User Interface (GUI) ได้ สามารถทำงานได้หลายระบบปฏิบัติการ การเรียกใช้งานโปรแกรม WEKA สามารถทำได้ 2 วิธี คือ

1. การเรียกใช้งานผ่าน Start > Programs > Weka 3.8.0 > Weka 3.8 หลังจากนั้นจะปรากฏหน้าต่างของ Weka GUI Chooser ขึ้นมา

2. การเรียกใช้งานผ่าน Command line โดยการเข้าไปที่ไดเร็กทอรีที่เก็บ Weka อยู่หลังจากนั้นใช้คำสั่ง java –jar weka.jar แล้วจะปรากฏหน้าต่างของ Weka GUI Chooser ขึ้นมา ดังภาพที่ 2.6

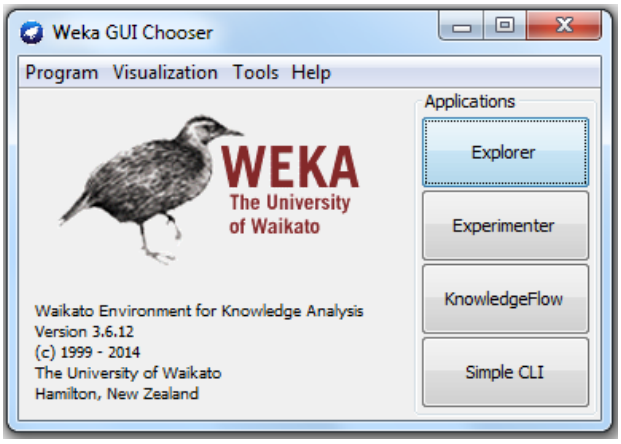
โปรแกรมหลักของซอฟต์แวร์ WEKA แสดงดังภาพที่ 2.6 ประกอบด้วย

1. Explorer เป็นโปรแกรมที่ออกแบบในลักษณะ GUI ที่ผู้ใช้สามารถเรียกใช้ฟังก์ชันการทำงานต่าง ๆ ของ WEKA ผ่านทางหน้าจอ GUI จึงเหมาะสำหรับผู้ที่เริ่มต้นใช้งาน WEKA เพราะผู้ใช้จะสามารถเรียกฟังก์ชันการทำงานต่าง ๆ โดยการคลิกและเปลี่ยนแปลงพารามิเตอร์ในหน้าฟอร์มเท่านั้น

2. Experimenter เป็นโปรแกรมที่ให้ผู้ใช้สามารถออกแบบการทดลองและการทดสอบผล จึงสามารถเปลี่ยนแปลงเทคนิคในการวิเคราะห์ข้อมูลและค่าพารามิเตอร์ต่าง ๆ ที่เกี่ยวข้องได้หลากหลายรูปแบบจนได้ผลลัพธ์ที่น่าพอใจ ซึ่งการใช้โปรแกรม Explorer นั้นไม่สะดวกเพราะต้องทำการเปลี่ยนแปลงค่าต่าง ๆ เอง

3. KnowledgeFlow เป็นโปรแกรมออกแบบผังการไหลของความรู้ที่ยอมให้ผู้ใช้สามารถนำเทคนิคต่าง ๆ ของ WEKA มาเรียงต่อกัน เพื่อช่วยให้การวิเคราะห์ข้อมูลทำงานได้ตามที่ต้องการ

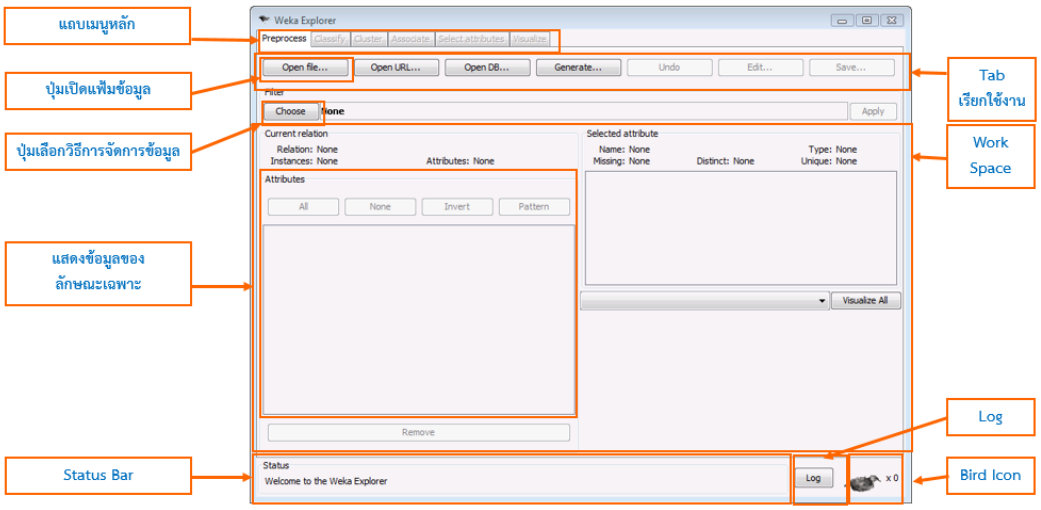
4. Simple CLI (Command Line Interface) เป็นโปรแกรมรับคำสั่งการทำงานผ่านทาง command line ซึ่งการเรียกใช้ฟังก์ชันผ่านทาง command line นี้จะช่วยให้ผู้ใช้เข้าใจการเรียกฟังก์ชันต่าง ๆ เบื้องหลังหน้าจอ GUI ของ WEKA ได้ และสามารถนำไปประยุกต์ใช้ในการเขียนโปรแกรมเพื่อเรียก WEKA ในการใช้งานได้อีกด้วย



ภาพที่ 2.6 หน้าจอหลักของโปรแกรม Weka Explore

สำหรับเมนูหลักของ Explorer ประกอบด้วย 6 เมนูย่อย ดังภาพที่ 2.7 โดยมีรายละเอียด ดังนี้

1. Preprocess การเตรียมข้อมูล
2. Classify รวมโมดูลการทำเหมืองข้อมูลแบบจัดจำแนกประเภท
3. Cluster รวมโมดูลการทำเหมืองข้อมูลแบบการเกาะกลุ่ม
4. Associate รวมโมดูลการทำเหมืองข้อมูลแบบกฎเชื่อมโยง
5. Select attributes รวมโมดูลสาหรับการวิเคราะห์ความเกี่ยวพันของลักษณะประจำ
6. Visualize นำเสนอข้อมูลด้วยภาพนามธรรมสองมิติ



ภาพที่ 2.7 เมนูหลักของ Explorer

ข้อมูลที่เหมาะใช้งานกับโปรแกรม WEKA ได้แก่ แฟ้มข้อมูลที่อยู่ในรูปแบบของ arff และ csv แฟ้มข้อมูลที่อยู่ในเครือข่ายโดยผู้ใช้เรียกใช้ผ่าน URL และข้อมูลที่อยู่ในฐานข้อมูลที่เชื่อมโยงผ่าน JDBC

การพัฒนาเว็บแอปพลิเคชัน

เว็บแอปพลิเคชัน (Web application) คือ โปรแกรมประยุกต์ที่เข้าถึงได้ด้วยโปรแกรม ที่ช่วยค้นหาหรืออ่านเอกสารบนอินเทอร์เน็ต (Internet browser) ซึ่งเหมาะสำหรับงานที่ต้องการข้อมูลตามเวลาจริง (Real time) ข้อดีของเว็บแอปพลิเคชัน คือ ข้อมูลต่าง ๆ ที่อยู่ในระบบที่มีการไหลเวียนในแบบออนไลน์ (Online) จึงสามารถโต้ตอบกับผู้ใช้บริการแบบตามเวลาจริง สามารถใช้งานได้ง่ายและสามารถใช้งานผ่านการเชื่อมต่ออินเทอร์เน็ต ผู้ใช้บริการจึงสามารถใช้งานโปรแกรมได้ทุกที่ ภาษาที่ใช้ในการพัฒนาเว็บแอปพลิเคชัน เช่น HTML, ASP/ASP.Net, PHP, และ JAVA script เป็นต้น

นอกจากนี้ยังมีเครื่องอื่นที่ช่วยในการพัฒนาระบบ เช่น Macromedia Dreamweaver, Microsoft Visual Studio.net เป็นต้น สำหรับงานวิจัยนี้ได้ศึกษาในหัวข้อต่อไปนี้ เพื่อใช้ในการพัฒนาระบบสนับสนุนการตัดสินใจ

1. ภาษา PHP

คำว่า PHP ย่อมาจาก Personal Home Page เป็นภาษาคำสั่ง (Script Language) คำสั่งต่าง ๆ จะเก็บอยู่ในไฟล์ที่เรียกว่า สคริปต์ (Script) เมื่อต้องการใช้งานจะใช้ตัวแปลชุดคำสั่งที่ทำงานโดยการสั่งงานจากเว็บเพจแล้วไปประมวลผลที่เว็บเซิร์ฟเวอร์ (Web Server) จึงถูกจัดให้อยู่ในกลุ่ม Server Side Script เพราะการทำงานอยู่ในฝั่งเซิร์ฟเวอร์ (Server) แล้วส่งผลลัพธ์มายัง Browser ของตัว Client เมื่อผู้ใช้ส่งความต้องการผ่านเว็บ Browser ทาง HTTP อาจจะเป็นในรูปแบบของการกรอกแบบฟอร์ม หรือการใส่ข้อมูล โดยจะแสดงเป็นเอกสาร PHP เมื่อเอกสาร PHP เข้ามาถึง เว็บเซิร์ฟเวอร์ก็จะถูกส่งไปให้ PHP เพื่อทำการแปลคำสั่งแล้ว Execute คำสั่งนั้น จากนั้น PHP จะสร้างผลลัพธ์ในรูปแบบเอกสาร HTML ทำให้ไม่เห็นคำสั่งที่ผู้ใช้เขียน แล้วจะส่งกลับไปให้เว็บเซิร์ฟเวอร์เพื่อส่งต่อไปให้กับ Browser แสดงผลทางฝั่งผู้ใช้งานต่อไป (กิตติ ภัคดีวัฒนะกุล. 2548) นอกจากนี้ยังเป็น Script ที่ฝังอยู่บน HTML ตัวอย่างของภาษาสคริปต์ เช่น JavaScript, Perl, ASP (Active Server Page) เป็นต้น

ความสามารถในการประมวลผลของภาษา PHP ได้แก่ การสร้างเนื้อหาอัตโนมัติ จัดการคำสั่ง การอ่านข้อมูลจากผู้ใช้และประมวลผล การอ่านข้อมูลจากฐานข้อมูล ความสามารถจัดการกับ Cookies ซึ่งทำงานเช่นเดียวกับโปรแกรมในลักษณะ CGI คุณสมบัติอื่น เช่น การประมวลผลตามบรรทัดคำสั่ง (Command line scripting) ทำให้ผู้เขียนโปรแกรมสร้าง PHP script ทำงานผ่าน PHP Parser โดยไม่ต้องผ่านเซิร์ฟเวอร์หรือเบราว์เซอร์

PHP สามารถทำงานได้หลากหลายระบบปฏิบัติการและเว็บเซิร์ฟเวอร์ รวมทั้งสามารถทำงานร่วมกับฐานข้อมูลได้หลายชนิด เช่น Oracle, dBase, PostgreSQL, IBM, DB2, MySQL, Informix, และODBC (Open Database Connection) เป็นต้น

2. โปรแกรมจัดการฐานข้อมูล

โปรแกรมจัดการฐานข้อมูล MySQL คือ โปรแกรมจัดการฐานข้อมูลที่พัฒนาโดยบริษัท MySQLAB ทำหน้าที่จัดเก็บข้อมูลอย่างเป็นระบบรองรับคำสั่ง SQL ซึ่งเป็นเครื่องมือสำหรับเก็บข้อมูลที่ต้องใช้ร่วมกับเครื่องมือหรือโปรแกรมอื่นอย่างบูรณาการ เพื่อให้ได้ระบบงานที่รองรับความต้องการของผู้ใช้ เช่น ทำงานร่วมกับเว็บเซิร์ฟเวอร์เพื่อให้บริการแก่ภาษาสคริปต์ที่ทำงานฝั่งเครื่องเซิร์ฟเวอร์ (Server-side script) เช่น ภาษา PHP ภาษา APS.net และภาษา JSP เป็นต้น นอกจากนี้ยังสามารถทำงานร่วมกับโปรแกรมประยุกต์ (Application program) เช่น ภาษา Visual.net ภาษา C# เป็นต้น ได้อีกด้วย โปรแกรมถูกออกแบบให้สามารถทำงานได้บนระบบปฏิบัติการที่หลากหลาย และเป็น

ระบบฐานข้อมูลโอเพนซอร์ส (Open Source) ที่ถูกนำไปใช้งานมากที่สุด การทำงานของโปรแกรมจัดการฐานข้อมูล MySQL มีลักษณะดังต่อไปนี้

1) ฐานข้อมูลมีลักษณะเป็นโครงสร้างของการเก็บรวบรวมข้อมูล การที่จะเพิ่มเติม เข้าถึงหรือประมวลผลข้อมูลที่เก็บในฐานข้อมูลจำเป็นจะต้องอาศัยระบบจัดการฐานข้อมูล ซึ่งจะทำหน้าที่เป็นตัวกลางในการจัดการกับข้อมูลในฐานข้อมูลทั้งสำหรับการใช้งานเฉพาะและรองรับการทำงานของแอปพลิเคชันอื่น ๆ ที่ต้องการใช้งานข้อมูลในฐานข้อมูล เพื่อให้ได้รับความสะดวกในการจัดการกับข้อมูลจำนวนมาก MySQL ทำหน้าที่เป็นทั้งตัวฐานข้อมูลและระบบจัดการฐานข้อมูล

2) MySQL เป็นระบบจัดการฐานข้อมูลเชิงสัมพันธ์ที่เก็บข้อมูลทั้งหมดในรูปแบบของตารางแทนการเก็บข้อมูลทั้งหมดลงในไฟล์เพียงไฟล์เดียว ทำให้ทำงานได้รวดเร็วและมีความยืดหยุ่น นอกจากนั้นแต่ละตารางที่เก็บข้อมูลสามารถเชื่อมโยงเข้าหากันทำให้สามารถรวมหรือจัด กลุ่มข้อมูลได้ตามต้องการ โดยอาศัยภาษา SQL ที่เป็นส่วนหนึ่งของโปรแกรม MySQL ซึ่งเป็นภาษามาตรฐานในการเข้าถึงฐานข้อมูล

งานวิจัยนี้มีการพัฒนาเว็บแอปพลิเคชันด้วย WEKA และภาษา PHP โดยมีขั้นตอนการทำงาน ดังนี้

1. สร้างโมเดลด้วยโปรแกรมประยุกต์ WEKA Explorer
2. บันทึกโมเดลที่เหมาะสมที่สุดสาหรับการพยากรณ์ (เป็นนามสกุล .model)
3. สร้างไฟล์ที่มีนามสกุล .arff สำหรับข้อมูลที่ใช้ในการพยากรณ์
4. พัฒนาเว็บแอปพลิเคชันด้วยภาษา PHP โดยมีการรับข้อมูลผ่านทางหน้าเว็บ แล้วระบบจะทำการบันทึกข้อมูลใหม่ลงในไฟล์ที่มีนามสกุล .arff
5. เรียกใช้คำสั่ง WEKA Command Line Interface (WEKA CLI) เพื่อทำการเชื่อมไฟล์ .arff และไฟล์ .model แล้วทำการการพยากรณ์ข้อมูลใหม่

การทำงานร่วมกันของ WEKA และภาษา PHP โดยใช้คำสั่งสำหรับการเรียกใช้งาน WEKA CLI แสดงในตาราง 2.2

ตาราง 2.2 คำสั่งสำหรับการเรียกใช้งาน WEKA CLI

คำสั่ง	การทำงาน (Function)
-t<training file>	สำหรับระบุตำแหน่งของไฟล์ที่จะใช้เป็น Training file
-T<testing file>	สำหรับระบุตำแหน่งของไฟล์ที่จะใช้เป็น Testing file
-c<class index>	สำหรับระบุว่าจะใช้ attribute ใดเป็น class คำตอบ
-x<number of folds>	สำหรับระบุว่าจะใช้การทดสอบแบบ cross-validation ที่ Fold
-d<output file>	สำหรับระบุตำแหน่งของไฟล์ที่ใช้เก็บไฟล์โมเดลที่สร้างได้
-l<input file>	สำหรับระบุตำแหน่งของไฟล์ที่ใช้โหลดโมเดลเข้ามาใช้
-p<attribute range>	สำหรับการแสดง class ทำนายที่ได้

งานวิจัยที่เกี่ยวข้อง

ชุติมา อุตมะมุณีย์ และประสงค์ ประณีตพลกรัง (2553) ได้ทำงานวิจัยเพื่อสร้างตัวแบบสำหรับหาปัจจัยที่มีผลต่อการตัดสินใจในการเลือกสาขาวิชาเรียนของนักศึกษา และพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจอย่างอัตโนมัติแบบออนไลน์สำหรับแนะแนวทางการเลือกสาขาวิชาเรียนของนักศึกษาในระดับอุดมศึกษา ใช้ข้อมูลด้านการเรียนของกลุ่มตัวอย่างจากนักศึกษาในระดับอุดมศึกษาทั้งภาครัฐและเอกชนจำนวน 9 มหาวิทยาลัย โดยใช้วิธีของข่ายงานเบย์ พบว่า ตัวแปรสำคัญที่มีผลต่อการตัดสินใจเลือกสาขาวิชาเรียนของนักศึกษา ได้แก่ เกรดเฉลี่ยวิชาคณิตศาสตร์ก่อนศึกษา เกรดเฉลี่ยวิชาคณิตศาสตร์ขณะกำลังศึกษา เกรดเฉลี่ยวิชาเขียนโปรแกรม ความรู้ในการพัฒนาระบบคอมพิวเตอร์และซอฟต์แวร์ ความรู้ด้านวิชาเรียน 1 ความรู้ด้านวิชาเรียน 2 และความถนัด โดยมีความแม่นยำในการทำนายสูงถึง ร้อยละ 91.35 โดยให้ผู้ใช้กรอกข้อมูลเข้าระบบและระบบทำนาย % แต่ละสาขาว่ามีโอกาสจบกี่เปอร์เซ็นต์

ธนาคม จุ้ยศิริ (2557) ได้ทำการพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกสาขาวิชาและวิชาเอกของผู้ที่สมัครเข้าศึกษาต่อในระดับปริญญาตรีของมหาวิทยาลัยสุโขทัยธรรมาธิราชด้วยเหมืองข้อมูล โดยใช้ข้อมูลการสำเร็จการศึกษา 3 ปีการศึกษาย้อนหลัง โดยนำข้อมูลพื้นฐาน ได้แก่ วุฒิการศึกษา สาขาวิชา/วิชาเอก ปีการศึกษาที่สำเร็จการศึกษา สถาบันการศึกษา อายุ เพศ มาช่วยในการสนับสนุนการตัดสินใจในการเลือกสาขาวิชาและวิชาเอก เพื่อเป็นข้อมูลพื้นฐานสำหรับผู้สมัคร

พรรณิภา บุตรเอก และสุรเดช บุญลือ (2557: 40-49) ศึกษาโอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน ข้อมูลที่ใช้ทดสอบเป็นข้อมูลของนักศึกษาระดับปริญญาตรี

หลักสูตร 4 ปี สาขาวิชาวิทยาการคอมพิวเตอร์ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยนอร์ทกรุงเทพ ระหว่างปีการศึกษา 2547-2551 จำนวนทั้งสิ้น 138 ระเบียน 18 คุณลักษณะ ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (BP-ANN) มาสร้างตัวแบบพยากรณ์แล้วนำมาเปรียบเทียบ เพื่อคัดเลือกตัวแบบพยากรณ์ที่มีประสิทธิภาพสูงสุด ผลการวิจัยพบว่าตัวแบบการพยากรณ์ที่ใช้เทคนิค SVM-PK ให้ความแม่นยำในการพยากรณ์ คิดเป็นร้อยละ 89.13 ค่าความเที่ยง คิดเป็น ร้อยละ 87.8 ค่าการเรียกคืน คิดเป็นร้อยละ 89.1 และค่าความถูกต้องโดยรวม คิดเป็นร้อยละ 86.0

ธีรพงษ์ สังข์ศรี (2014) ได้วิเคราะห์พฤติกรรมสำหรับการเลือกสมัครสาขาวิชาเรียนโดยใช้เทคนิคการจัดกลุ่มแบบ k-means clustering และเทคนิคกฎของความสัมพันธ์ และศึกษาเปรียบเทียบตัวแบบพยากรณ์จำนวนนักศึกษาใหม่โดยใช้เทคนิคต้นไม้ตัดสินใจ และเทคนิคโครงข่ายประสาทเทียม ผลการวิจัย พบว่า ตัวแบบที่ได้จากเทคนิคต้นไม้ตัดสินใจให้ค่าความถูกต้อง เท่ากับ 93.76% ซึ่งสูงกว่าเทคนิคโครงข่ายประสาทเทียม ที่ให้ค่าค่าความถูกต้อง เท่ากับ 93.60%

เสกสรรค์ วิลัยลักษณ์ วิภา เจริญภัณฑารักษ์ และดวงดาว วิชาดากุล (2558) ได้พัฒนาคลังข้อมูล และสร้างตัวแบบพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน โดยใช้ข้อมูลนักเรียนระดับมัธยมศึกษา ระหว่างปีการศึกษา 2548 – 2556 โดยใช้โครงสร้างแบบสโนว์เฟลกสกีมา (Snowflake schema) จากนั้นใช้ข้อมูลนักเรียนระดับมัธยมศึกษาปีที่ 4 ระหว่าง ปีการศึกษา 2553 – 2556 จำนวน 525 ระเบียน ประกอบด้วย 16 คุณลักษณะ มาสร้างตัวแบบพยากรณ์ผลการเรียนโดย ใช้ชุดข้อมูล 2 แบบ คือ ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และข้อมูลแบบจัดกลุ่ม (Cluster Data) คุณลักษณะของข้อมูลถูกคัดเลือกโดยใช้วิธี Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) แล้วจึงใช้เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree) มาสร้างตัวแบบพยากรณ์ และใช้วิธีการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation ผลการวิจัยพบว่าชุดข้อมูลแบบไม่จัดกลุ่มที่คัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) และใช้เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ค่าความถูกต้องสูงที่สุด คิดเป็นร้อยละ 94.48 โดยคุณลักษณะที่มีความสำคัญ ได้แก่ แผนการเรียน ผลการเรียนเฉลี่ยวิทยาศาสตร์ ผลการเรียนเฉลี่ยสังคมศึกษา ผลการเรียนเฉลี่ยภาษาอังกฤษ และผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3

อนันต์ ปินะเต (2016) ใช้เทคนิคต้นไม้ตัดสินใจ อัลกอริทึม C4.5 เพื่อสร้างแบบจำลองและและกฎการตัดสินใจด้วยต้นไม้ตัดสินใจ อัลกอริทึม C4.5 และพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกสมัครเรียน ในสาขาวิชาในระดับปริญญาตรี มหาวิทยาลัยมหาสารคาม ผลการวิจัย พบว่าสามารถสร้างเป็นกฎการตัดสินใจได้ทั้งหมด 333 กฎ จาก 51 สาขาวิชา

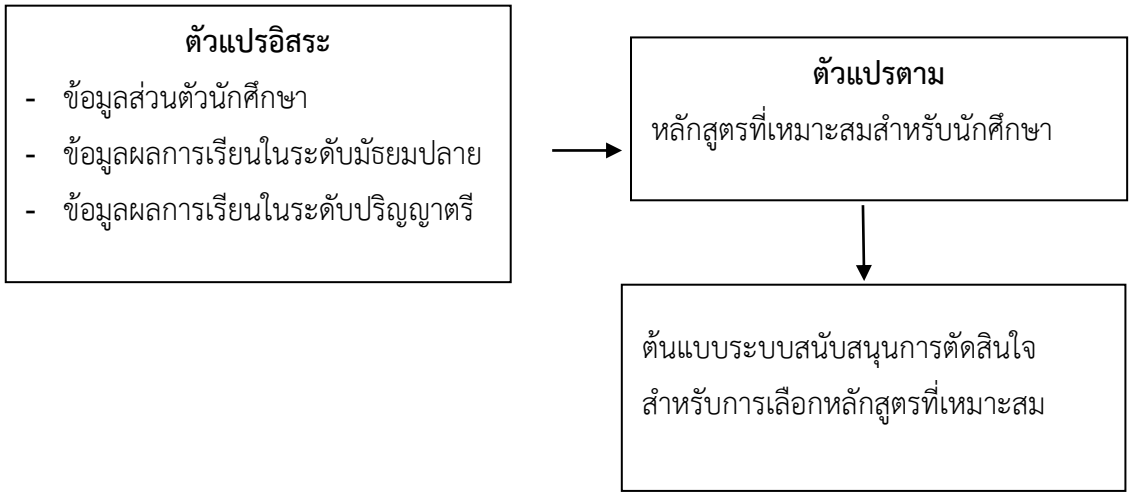
Delavari, Beikzadeh, & Phon-Amnuaisuk (2005) ทำวิจัยโดยใช้เหมืองข้อมูลในการนำเสนอโมเดลการวิเคราะห์ชื่อ Data Mining Education University (DM_EDU) โดยผู้วิจัยสร้างแบบจำลองในการทำนายว่าผู้เรียนจะเรียนสำเร็จหรือไม่ โดยใช้ต้นไม้ในการจำแนกประเภทข้อมูล มีจำนวนข้อมูลในการทดลองทั้งสิ้น 841 คน พบว่าตัวแปรที่มีผลต่อการเรียนสำเร็จหรือไม ได้แก เชื้อชาติ การได้รับทุน เชื้อชาติของผู้สอน สาขาที่เรียน จำนวนครั้งที่ลงทะเบียนเรียนวิชา Computer Programming II และเกรดวิชาภาษาอังกฤษ

Hsu and Laia (2003) ได้ทำวิจัยเพื่อทำนายว่าผู้เรียนจะเรียนผ่านหรือไม่ผ่านในรายวิชา Computer Programming II โดยใช้ข้อมูลคะแนนสอบกลางภาคร่วมกับข้อมูลส่วนตัว โดยการประยุกต์ใช้เทคนิค Association Base GA (AGA) เพื่อแนะแนวการเลือกเรียนต่อในระดับปริญญาตรี หลังจากนั้นจะนำกฎที่ได้ไปสร้างเป็นต้นไม้ ร่วมกับการประยุกต์ใช้จีเนติกอัลกอริทึม (Genetic Algorithms) เพื่อหาผลการทำนายที่ถูกต้องที่สุด

Dole and Rajurkar (2014) ได้ทำวิจัยเปรียบเทียบประสิทธิภาพในการพยากรณ์ผลสัมฤทธิ์ในการเรียนของนักศึกษาโดยใช้นาอีฟเบเซียน และต้นไม้ตัดสินใจ พบว่า นาอีฟเบเซียนให้ค่าความถูกต้องสูงกว่า

กรอบแนวคิดในการวิจัย

ในการวิจัยนี้ต้องการสร้างตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสมสำหรับนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยี ได้กำหนดกรอบแนวคิดของการวิจัย ดังภาพที่ 2.8 ซึ่งแสดงตัวแปรที่ใช้ในการวิจัย ได้แก่ ข้อมูลส่วนตัวนักศึกษา ข้อมูลผลการเรียนในระดับมัธยมปลาย ข้อมูลผลการเรียนในระดับปริญญาตรี เพื่อสร้างแบบจำลองในการทำนายผลสำเร็จการศึกษาของนักศึกษา



ภาพที่ 2.8 กรอบแนวคิดของการวิจัย

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยนี้เป็นงานวิจัยและพัฒนา โดยมีวัตถุประสงค์เพื่อพัฒนาระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสมในการเข้าศึกษาต่อมหาวิทยาลัยสวนดุสิต โดยใช้เทคนิคเหมืองข้อมูล มีการสร้างต้นแบบที่ใช้ในการพยากรณ์จากเทคนิคเหมืองข้อมูล แล้วนำต้นแบบที่ได้มาพัฒนาเป็นระบบสนับสนุนการตัดสินใจ โดยผู้วิจัยได้ศึกษาทฤษฎี หลักการ งานวิจัยที่เกี่ยวข้อง แล้วนำความรู้ที่ได้มาใช้ในการกำหนดแผนการวิจัยและขั้นตอนการดำเนินงานตลอดจนประชากรและกลุ่มตัวอย่างสำหรับการวิจัยโดยมีรายละเอียด ดังนี้

ประชากรและการสุ่มกลุ่มตัวอย่าง

ประชากรที่นำมาใช้ในการวิจัยนี้เป็นนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยีที่สำเร็จการศึกษาแล้ว โดยแบ่งออกเป็น 5 หลักสูตร ได้แก่ หลักสูตรวิทยาการคอมพิวเตอร์ หลักสูตรเทคโนยีสารสนเทศ หลักสูตรอาชีวอนามัยและความปลอดภัย หลักสูตรสิ่งแวดล้อมเมืองและอุตสาหกรรม และหลักสูตรเครื่องสำอาง ปีการศึกษา 2551-2553 จำนวน 1,119 คน ข้อมูลที่ใช้ในงานวิจัยนี้ประกอบด้วย

(1) ข้อมูลส่วนตัวของนักศึกษา ได้แก่ รหัสประจำตัวนักศึกษา เพศ อายุ ที่อยู่ และโรงเรียนเดิม

(2) ข้อมูลผลการเรียน ได้แก่ ผลการเรียนเฉลี่ยกลุ่มวิชาภาษาไทย ผลการเรียนเฉลี่ยกลุ่มวิชาคณิตศาสตร์ ผลการเรียนเฉลี่ยกลุ่มวิชาวิทยาศาสตร์ ผลการเรียนเฉลี่ยกลุ่มวิชาสังคม ผลการเรียนเฉลี่ยกลุ่มวิชาศิลปะ ผลการเรียนเฉลี่ยกลุ่มวิชาภาษาต่างประเทศ ผลการเรียนเฉลี่ยกลุ่มวิชาการงานอาชีพและเทคโนโลยี ผลการเรียนเฉลี่ยรวมระดับมัธยมปลาย ผลการเรียนเฉลี่ยรวมระดับปริญญาตรี และหลักสูตรที่สำเร็จการศึกษา สำหรับข้อมูลที่นำมาใช้ในการทดสอบประสิทธิภาพของระบบเป็นนักศึกษาปัจจุบันที่กำลังศึกษาอยู่ใน 5 หลักสูตรดังกล่าวข้างต้น จำนวน 1,168 คน (ข้อมูล ณ วันที่ 31 กรกฎาคม 2559) ในงานวิจัยนี้ได้ใช้วิธีการสุ่มตัวอย่างโดยใช้วิธีสุ่มอย่างแบบง่าย เพื่อหากลุ่มตัวอย่างของนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต โดยขนาดของกลุ่มตัวอย่างคำนวณโดยใช้สูตรของ Taro Yamane ที่ระดับความเชื่อมั่น 95% โดยมีสูตรดังต่อไปนี้

$$n = \frac{N}{1+Ne^2} \tag{3.1}$$

โดยที่ n = จำนวนกลุ่มตัวอย่าง

N = จำนวนประชากรทั้งหมด

e = ความคลาดเคลื่อนของการสุ่มตัวอย่าง

จำนวนกลุ่มตัวอย่างที่ได้จากการคำนวณ มีค่าเท่ากับ 298

เครื่องมือในการวิจัยและการตรวจสอบคุณภาพเครื่องมือ

เครื่องมือที่ใช้ในการวิจัยนี้ประกอบด้วย

1. เทคนิคการทำเหมืองข้อมูลแบบการจำแนกประเภท โดยเป็นการศึกษาเปรียบเทียบประสิทธิภาพระหว่างการจำแนกประเภท 3 วิธี ประกอบด้วย การจำแนกประเภทแบบเบย์ด้วยอัลกอริทึมนาอีฟเบเซียน การจำแนกประเภทโดยใช้ต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 และอัลกอริทึม RandomForest และการจำแนกประเภทโดยใช้กฎด้วยอัลกอริทึม JRip

2. โปรแกรมเวกา (WEKA 3.8.1) โปรแกรมสำเร็จรูปที่พัฒนาด้วยภาษาจาวา ประกอบด้วยโปรแกรมหลัก 4 โปรแกรม ได้แก่ Explorer Experimenter KnowledgeFlow และ Simple CLI สำหรับเมนูหลักของ Explorer ประกอบด้วย 6 เมนูย่อย คือ Preprocess Classify Cluste Associate Select attributes และ Visualize สำหรับงานวิจัยนี้จะใช้เมนูย่อย Classify ซึ่งเป็นโมดูลของการทำเหมืองข้อมูลแบบจำแนกประเภท

3. แบบสอบถามเพื่อประเมินความพึงพอใจและความเชื่อมั่นของระบบโดยผู้เชี่ยวชาญและผู้ใช้งาน การประเมินประสิทธิภาพพิจารณาจาก 5 ด้าน คือ

3.1 การประเมินด้านการทำงานของระบบตรงตามความต้องการของผู้ใช้ (Functional requirement test) เป็นการประเมินว่าระบบที่พัฒนามีความถูกต้อง และมีประสิทธิภาพตรงตามความต้องการของผู้ใช้

3.2 การประเมินระบบด้านการใช้งานระบบ (Functional test) เป็นการประเมินว่าระบบที่พัฒนามีความถูกต้องและมีประสิทธิภาพสามารถทำงานได้ตามหน้าที่ที่กำหนดไว้ในระบบ

3.3 การประเมินด้านการออกแบบและความง่ายในการใช้งานระบบ (Usability test) เป็นการประเมินว่าระบบที่ได้พัฒนามีความสามารถในการใช้งาน และมีความสะดวกในการใช้งาน

3.4 การประเมินด้านประสิทธิภาพของระบบ (Performance test) เป็นการประเมินว่ามีประสิทธิภาพตรงตามที่ต้องการ

3.5 การประเมินด้านความปลอดภัยของระบบ (Security test) เป็นการประเมินว่าระบบที่ได้พัฒนานั้นมีความปลอดภัยของข้อมูล

การตรวจสอบคุณภาพเครื่องมือเป็นการประเมินว่าเครื่องมือที่ใช้มีประสิทธิภาพในการนำไปใช้หรือไม่ ในที่นี้จะกล่าวถึงการตรวจสอบแบบสอบถามที่ใช้ประเมินความพึงพอใจในการใช้งาน โดยใช้ 2 วิธี คือ

วิธีที่ 1 การหาค่าความเที่ยงตรงเชิงเนื้อหา (Content validity) เป็นการนำแบบสอบถามที่จัดทำขึ้นเพื่อใช้ประเมินความพึงพอใจในการใช้งานให้ผู้เชี่ยวชาญตรวจสอบว่าข้อคำถามแต่ละข้อมีความสอดคล้องกับวัตถุประสงค์ที่ได้กำหนดไว้หรือไม่ (Item-Objective Congruence Index: IOC) โดยมีเกณฑ์ในการให้คะแนนความสอดคล้องของแบบสอบถามแต่ละข้อกับวัตถุประสงค์ ไว้ดังนี้

คะแนน +1 ถ้าเห็นว่าสามารถวัดได้ตรงตามวัตถุประสงค์หรือมีความสอดคล้อง

คะแนน 0 ถ้าไม่แน่ใจว่าสามารถวัดได้ตรงตามวัตถุประสงค์หรือไม่มีความสอดคล้อง

คะแนน -1 ถ้าเห็นว่าไม่สามารถวัดตรงตามวัตถุประสงค์หรือไม่มีความสอดคล้อง

หลังจากนั้นนำแบบทดสอบมาวิเคราะห์หาค่า IOC เฉลี่ยในแต่ละข้อ สามารถคำนวณได้จาก

$$IOC = \frac{\sum R}{N} \tag{3.2}$$

เมื่อ IOC คือ ค่าดัชนีความสอดคล้อง มีค่าอยู่ระหว่าง -1 ถึง 1

$\sum R$ คือ ผลรวมของการพิจารณาของผู้เชี่ยวชาญ

N คือ จำนวนผู้เชี่ยวชาญ

เกณฑ์ที่ใช้ในการประเมินความสอดคล้องของแบบสอบถามที่ใช้ประเมินความพึงพอใจในการใช้งาน เป็นดังนี้

ค่าเฉลี่ย	0.00-0.49	ความสอดคล้องของแบบสอบถามอยู่ในระดับต่ำ
ค่าเฉลี่ย	0.50-0.69	ความสอดคล้องของแบบสอบถามอยู่ในระดับยอมรับ
ค่าเฉลี่ย	0.70-0.79	ความสอดคล้องของแบบสอบถามอยู่ในระดับดี
ค่าเฉลี่ย	0.80-1.00	ความสอดคล้องของแบบสอบถามอยู่ในระดับดีมาก

วิธีที่ 2 การหาค่าความเชื่อมั่นของแบบทดสอบ ด้วยวิธีวิเคราะห์ค่าสัมประสิทธิ์แอลฟ่า (α-Coefficient) ของครอนบาค (Cronbach) โดยค่าที่ได้ต้องไม่น้อยกว่า 0.70

$$\propto = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum s_i^2}{s_t^2}\right) \tag{3.3}$$

α คือ ค่าความเชื่อมั่นของเครื่องมือ

K คือ จำนวนคำถาม

s_i^2 คือ ความแปรปรวนของคะแนนแต่ละข้อ

s_t^2 คือ ความแปรปรวนของคะแนนทั้งฉบับ

4. ซอฟต์แวร์ (Software) อื่น ๆ ได้แก่
   - ระบบปฏิบัติการ Windows 10
   - ภาษา PHP
   - เว็บเซิร์ฟเวอร์ใช้โปรแกรม Apache
   - ระบบฐานข้อมูล MySQL
   - Sublime Text 2
   - Adobe Dreamweaver CS5
   - Xampp 5.6.12
5. ฮาร์ดแวร์ (Hardware) ได้แก่
   - หน่วยความจำ RAM อย่างน้อย 4.00 GB
   - หน่วยความจำสำรอง CPU Intel Core i3 1.90 GHz
   - ระบบเครือข่ายอินเทอร์เน็ต
   - หน่วยความจำสำรอง Hard disk อย่างน้อย 100 GB

การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลของนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยีที่ถูกเก็บรวบรวมไว้ในฐานข้อมูลของสำนักส่งเสริมวิชาการและงานทะเบียน มหาวิทยาลัยสวนดุสิต ประกอบด้วย ข้อมูลส่วนตัวของนักศึกษา ได้แก่ รหัสประจำตัวนักศึกษา เพศ อายุ ที่อยู่ โรงเรียนเดิม และข้อมูลผลการเรียนในระดับมัธยมศึกษาตอนปลายที่ได้จากระเบียนแสดงผลการเรียนหลักสูตรการศึกษาขั้นพื้นฐาน ช่วงชั้นที่ 4 มัธยมศึกษาปีที่ 4-6 (ปพ.1) และรายงานผลการศึกษา ระดับปริญญาตรี ที่นักศึกษาจะได้รับเมื่อสำเร็จการศึกษา

การวิเคราะห์ข้อมูล

แบบสอบถามที่สร้างขึ้นเป็นแบบมาตราส่วนประมาณค่า (Rating Scale) ลักษณะข้อความประกอบด้วยข้อความที่เป็นการให้ความสำคัญในแต่ละเรื่อง แต่ละคำตอบมีคำตอบให้เลือก 5 ระดับ ตามมาตรการวัดของลิเคอร์ท(Likert Scale) โดยแบ่งระดับคะแนนเป็น 5 ระดับ กำหนดช่วงของการวัดได้ดังนี้ (ศิริชัย พงษ์วิชัย. 2551; ธานินทร์ ศิลป์จารุ. 2555) คือ พอใจมากที่สุด พอใจมาก พอใจปานกลาง พอใจน้อย และพอใจน้อยที่สุด โดยมีเกณฑ์ในการให้ค่าคะแนนดังนี้

พอใจมากที่สุด	มีค่าระดับคะแนนเท่ากับ 5
พอใจมาก	มีค่าระดับคะแนนเท่ากับ 4
พอใจปานกลาง	มีค่าระดับคะแนนเท่ากับ 3
พอใจน้อย	มีค่าระดับคะแนนเท่ากับ 2
พอใจน้อยที่สุด	มีค่าระดับคะแนนเท่ากับ 1

การวิเคราะห์ความพึงพอใจของผู้ใช้คำนวณได้จากข้อมูลที่ได้จากแบบสอบถามที่เก็บรวบรวมมา แล้วนำมาคำนวณเพื่อหาค่าเฉลี่ยเลขคณิตและค่าส่วนเบี่ยงเบนมาตรฐาน ดังนี้

ค่าเฉลี่ยเลขคณิต $\bar{x} = \frac{\sum x}{n}$ (3.4)

โดยที่ $\bar{x}$ คือ ค่าเฉลี่ยเลขคณิต

$\sum x$ คือ ผลรวมของข้อมูลทั้งหมด

n คือ จำนวนข้อมูล

ค่าส่วนเบี่ยงเบนมาตรฐาน $S.D. = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$ (3.5)

$S.D.$ คือ ส่วนเบี่ยงเบนมาตรฐาน

x คือ ค่าข้อมูล

$\bar{x}$ คือ ค่าเฉลี่ยเลขคณิต

n คือ จำนวนข้อมูล

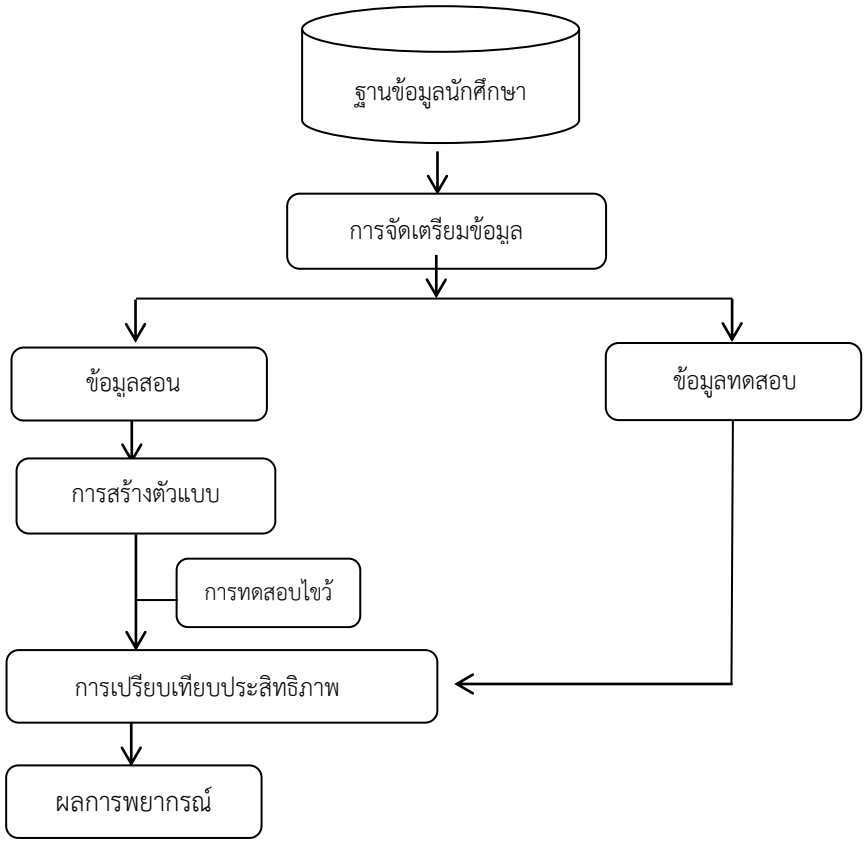
การแปลค่าความหมายของค่าความพึงพอใจใช้เกณฑ์การแปลความหมายข้อมูลของ Likert Scale โดยแบ่งระดับคะแนนเป็น 5 ระดับ กำหนดช่วงของการวัดได้ดังนี้ (ศิริชัย พงษ์วิชัย, 2551)

คะแนนเฉลี่ยตั้งแต่ 4.21-5.00 หมายถึง เห็นด้วยมากที่สุด
คะแนนเฉลี่ยตั้งแต่ 3.41-4.20 หมายถึง เห็นด้วยมาก
คะแนนเฉลี่ยตั้งแต่ 2.61-3.40 หมายถึง เห็นด้วยปานกลาง
คะแนนเฉลี่ยตั้งแต่ 1.81-2.60 หมายถึง เห็นด้วยน้อย
คะแนนเฉลี่ยตั้งแต่ 1.00-1.80 หมายถึง เห็นด้วยน้อยที่สุด

การสร้างตัวแบบเพื่อการพยากรณ์

ข้อมูลที่ผ่านการคัดเลือกและจัดเตรียมให้มีความเหมาะสมจะถูกนำไปสร้างตัวแบบของการพยากรณ์ โดยมีขั้นตอนการสร้างตัวแบบการพยากรณ์สำหรับการเลือกหลักสูตรที่เหมาะสมในการเข้าศึกษาต่อมหาวิทยาลัยสวนดุสิตโดยใช้เทคนิคเหมืองข้อมูล ดังภาพที่ 3.1 การสร้างตัวแบบเพื่อการพยากรณ์ มีรายละเอียดดังนี้

1. ข้อมูลถูกจัดเก็บในฐานข้อมูลสำหรับนำมาใช้ในงานวิจัย ประกอบด้วย (1) ข้อมูลส่วนตัวของนักศึกษา ได้แก่ รหัสประจำตัวนักศึกษา เพศ อายุ ที่อยู่ และโรงเรียนเดิม (2) ข้อมูลผลการเรียน ได้แก่ ผลการเรียนเฉลี่ยกลุ่มวิชาภาษาไทย ผลการเรียนเฉลี่ยกลุ่มวิชาคณิตศาสตร์ ผลการเรียนเฉลี่ยกลุ่มวิชาวิทยาศาสตร์ ผลการเรียนเฉลี่ยกลุ่มวิชาสังคม ผลการเรียนเฉลี่ยกลุ่มวิชาศิลปะ ผลการเรียนเฉลี่ยกลุ่มวิชาภาษาต่างประเทศ ผลการเรียนเฉลี่ยกลุ่มวิชาการงานอาชีพและเทคโนโลยี ผลการเรียนเฉลี่ยรวมระดับมัธยมปลาย ผลการเรียนเฉลี่ยรวมระดับปริญญาตรี และหลักสูตรที่สำเร็จการศึกษา สำหรับข้อมูลที่นำมาใช้ในการทดสอบประสิทธิภาพของระบบเป็นนักศึกษาปัจจุบันที่กำลังศึกษาอยู่ใน 5 หลักสูตรดังกล่าวข้างต้น



ภาพที่ 3.1 กระบวนการสร้างต้นแบบการพยากรณ์สำหรับการเลือกหลักสูตร

ตารางที่ 3.1 ตัวแปรที่ใช้ในงานวิจัย

ที่	แอททริบิวท์	รายละเอียด
1	ID	รหัสประจำตัวนักศึกษา
2	Sex	เพศ
3	Age	อายุ
4	Address	ที่อยู่
5	School name	ชื่อโรงเรียนเดิม
6	Thai	เกรดเฉลี่ยกลุ่มวิชาภาษาไทย
7	Math	เกรดเฉลี่ยกลุ่มวิชาคณิตศาสตร์

ที่	แอททริบิวท์	รายละเอียด
8	Science	เกรดเฉลี่ยกลุ่มวิชาวิทยาศาสตร์
9	Social	เกรดเฉลี่ยกลุ่มวิชาสังคม
10	Art	คะแนนเฉลี่ยกลุ่มวิชาศิลปะ
11	Foreign	เกรดเฉลี่ยกลุ่มวิชาภาษาต่างประเทศ
12	Occupation	เกรดเฉลี่ยกลุ่มวิชาการงานอาชีพและเทคโนโลยี
13	High school Grade Point Average	เกรดเฉลี่ยตลอดหลักสูตรในระดับมัธยมปลาย
14	University Grade Point Average	เกรดเฉลี่ยเมื่อสำเร็จการศึกษาในระดับปริญญาตรี
15	Program	หลักสูตรที่สำเร็จการศึกษาในระดับปริญญาตรี ประกอบด้วย 5 หลักสูตรได้แก่ - วิทยาการคอมพิวเตอร์ - อาชีวอนามัยและความปลอดภัย - สิ่งแวดล้อมเมือง - เทคโนโลยีสารสนเทศ - เครื่องสำอาง

2. การจัดเตรียมข้อมูล ข้อมูลที่ได้จากการเก็บรวบรวมข้อมูลจะถูกนำมาจัดเตรียมข้อมูลตามเทคนิคการเตรียมข้อมูล ดังนี้

2.1 การลดขนาดของข้อมูล ในข้อมูลที่เก็บรวบรวมมามีแอททริบิวท์ที่ไม่เกี่ยวข้อง จึงต้องมีการตัดบางแอททริบิวท์ออกไป ได้แก่ รหัสประจำตัวนักศึกษา เพศ อายุ ที่อยู่ เป็นต้น

2.2 การทำความสะอาดข้อมูลเนื่องจากข้อมูลจริงไม่สมบูรณ์ มีค่าที่ขาดหายไป (Missing value) จึงต้องนำมาปรับข้อมูลโดยพิจารณาว่า ถ้าข้อมูลจริงที่ไม่สมบูรณ์นั้นเป็นข้อมูลเชิงตัวเลขจะใช้ค่าเฉลี่ยของคุณลักษณะนั้นเติมค่าข้อมูลที่ขาดหายไปแทน และถ้าค่าข้อมูลนั้นไม่ใช่ข้อมูลเชิงตัวเลขจะใช้ค่าฐานนิยม (Mode) ในโปรแกรมเวกาสามารถทำได้โดยเลือกที่ ReplacingMissingValue

2.3 การแปลงข้อมูลเป็นการจัดการข้อมูลให้อยู่ในรูปที่เหมาะสมข้อมูลจริงที่นำมาใช้ในการทำเหมืองข้อมูลมีทั้งที่เป็นข้อมูลแบบประเภท และข้อมูลเชิงตัวเลข จึงต้องมีการนำข้อมูลเชิงตัวเลขมาแปลงเพื่อให้เหมาะกับอัลกอริทึมที่ถูกนำมาใช้ในการวิจัยนี้โดยการแทนค่าของข้อมูล

ข้อมูลผลการเรียนในระดับมัธยมศึกษาตอนปลาย

ข้อมูลผลการเรียนจะถูกนำมาแทนค่าโดยเปลี่ยนจากข้อมูลเชิงตัวลขเป็นข้อมูลประเภท ดังนี้

- ผลการเรียน 0.00-2.50 แปลงเป็นผลการเรียนระดับต่ำ เขียนแทนด้วย Low (L)
- ผลการเรียน 2.51-3.24 แปลงเป็นผลการเรียนระดับปานกลาง เขียนแทนด้วย Medium (M)
- ผลการเรียน 3.25-4.00 แปลงเป็นผลการเรียนระดับสูง เขียนแทนด้วย High (H)

ข้อมูลที่ผ่านการจัดเตรียมแล้วจะมีลักษณะดังตารางที่ 3.2 และ 3.3

ตารางที่ 3.2 ตัวอย่างข้อมูลที่ผ่านการเตรียมข้อมูลแล้วโดยไม่มีการแปลงข้อมูล

Thai	Math	Sci	Social	Art	Foreign	Occ	Sch	Course
2.75	2.56	2.45	3.53	3.75	3.75	3.72	3.05	Computer science
2.81	2.71	2.77	3.55	3.55	3.25	3.71	3.14	Computer science
2.75	2.65	2.55	3.23	3.25	3.05	3.12	3.25	Computer science

ตารางที่ 3.3 ตัวอย่างข้อมูลที่ผ่านการเตรียมข้อมูลแล้วโดยมีการแปลงข้อมูล

Thai	Math	Sci	Social	Art	Foreign	Occ	Sch	Course
M	M	L	H	H	H	H	M	Computer science
M	M	M	H	H	H	H	M	Computer science
M	M	L	H	H	H	H	M	Computer science

2.4 การจัดเตรียมข้อมูลให้อยู่ในรูปแบบของไฟล์นามสกุล ARFF เป็นการจัดการข้อมูลเพื่อนำไปประมวลผลด้วยโปรแกรมเวกา (Waikato Environment for Knowledge Analysis: WEKA) เพื่อสร้างตัวแบบระบบสนับสนุนการตัดสินใจต่อไป โดยในแฟ้มข้อมูลจะประกอบด้วยส่วนต่าง ๆ ได้แก่ ชื่อตารางข้อมูลเชิงสัมพันธ์ (Relation name) ชื่อลักษณะเฉพาะและชนิด (Attribute name) และข้อมูล (Data) ดังภาพที่ 3.2

จากภาพที่ 3.2 เป็นตัวอย่างรายชื่อและค่าของลักษณะประจำหรือแอททริบิวท์ทั้งหมด รวมทั้งขอบเขตของข้อมูลที่เลือกมาใช้สำหรับการสร้างต้นแบบและทดสอบต้นแบบระบบ

สนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสมในการเข้าศึกษาต่อมหาวิทยาลัยสวนดุสิต โดยใช้เทคนิคเหมืองข้อมูลโดยลักษณะประจำสุดท้าย คือ คะแนนเฉลี่ยเมื่อสำเร็จการศึกษาในระดับปริญญาตรี (Uni) จะเป็นคลาสผลลัพธ์ ส่วนข้อมูลจะอยู่ถัดไปหลังจากบรรทัด @data ซึ่งค่าลักษณะประจำ แต่ละค่าในชุดข้อมูลจะคั่นด้วยเครื่องหมายจุลภาค (,) เช่น 2.75, 2.56, 2.45, 3.53, 3.75, 3.75, 3.72, 3.05, 2.91, Computer science เป็นต้น

```
@relation student-data
@attribute thai numeric
@attribute math numeric
@attribute sci numeric
@attribute social numeric
@attribute art  numeric
@attribute foreign  numeric
@attribute occ numeric
@attribute sch numeric
@attribute Uni numeric
@attribute Course {Computer science, Health, Environment, Information technology, Chemical, Cosmetic}
@data
2.75, 2.56, 2.45, 3.53, 3.75, 3.75, 3.72, 3.05, 2.91, Computer science
2.81, 2.71, 2.77, 3.55, 3.55, 3.25, 3.71, 3.14, 3.21, Computer science
2.75, 2.65, 2.55, 3.23, 3.25, 3.05, 3.12, 3.25, 3.12, Computer science
```

ภาพที่ 3.2 ตัวอย่างรูปแบบของไฟล์นามสกุล ARFF ที่ใช้ในการประมวลผลด้วยโปรแกรมเวกา

3. การแบ่งข้อมูลเพื่อการทดสอบ

การพัฒนาตัวแบบมีวิธีการสร้างที่แตกต่างกัน 3 วิธี ซึ่งมาจากวิธีการแบ่งชุดข้อมูลเพื่อการเรียนรู้ ดังนี้

3.1 วิธีสร้างตัวแบบด้วยชุดข้อมูลสอน (Training set) วิธีนี้ข้อมูลที่ใช้ในการสร้างต้นแบบ และข้อมูลที่ใช้ในการทดสอบต้นแบบเป็นข้อมูลชุดเดียวกัน วิธีการนี้เริ่มต้นจากการสร้างต้นแบบด้วยชุดข้อมูลสอน หลังจากนั้นนำต้นแบบที่ได้มาทำนายข้อมูลที่ชุดข้อมูลสอนชุดเดิม การวัดประสิทธิภาพด้วยวิธีนี้จะให้ผลการวัดประสิทธิภาพที่มีค่าสูงมาก เนื่องจากเป็นข้อมูลชุดเดิมที่ระบบได้ทำการเรียนรู้มาแล้ว วิธีการนี้เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพเพื่อดูแนวโน้มของต้นแบบ

ที่ได้สร้างขึ้น ถ้าได้ผลการวัดที่น้อย แสดงว่าต้นแบบไม่เหมาะสมกับข้อมูล จึงไม่ควรจะนำไปทดสอบด้วยวิธีการแบ่งข้อมูลแบบอื่น

3.2 วิธีการทดสอบไขว้ (Cross validation) เป็นการแบ่งกลุ่มข้อมูลออกเป็นชุดข้อมูลสอนและชุดข้อมูลทดสอบ วิธีนี้เหมาะกับชุดข้อมูลที่มีจำนวนไม่มาก โดยใช้วิธีการสุ่มตัวอย่าง เริ่มจากจำนวนข้อมูลทั้งหมดมี N จำนวน ข้อมูลถูกแบ่งออกเป็นส่วน ๆ จำนวน k ส่วน (Fold) แต่ละส่วนจะมีจำนวนเท่ากับ N หารด้วย k และทำการคำนวณค่าความแม่นยำจากการทำนาย k รอบ โดยแต่ละรอบจะมีการสร้างต้นแบบในการจำแนกประเภทหนึ่งตัว จากข้อมูลเรียนรู้ k-1 ชุด และข้อมูลอีก 1 ชุดจะเป็นข้อมูลทดสอบ ซึ่งยังไม่ได้นำมาเรียนรู้ โปรแกรมเวกาจะกำหนดค่าอัตโนมัติเป็น 10 ถ้าไม่มีการกำหนดค่าใหม่เอง เช่น ถ้ามีข้อมูลจำนวน 100 ตัวอย่าง แบ่งชุดข้อมูลออกเป็น 5 ส่วน แต่ละชุดจะมีขนาดของข้อมูลเท่ากับ 20 ตัวอย่าง และการคำนวณค่าความแม่นยำจากการทำนายมีทั้งหมด 5 รอบ โดยที่

รอบที่ 1 ใช้ข้อมูลชุดที่ 1 เป็นชุดข้อมูลทดสอบ ชุดที่ 2-5 เป็นชุดข้อมูลสอน
รอบที่ 2 ใช้ข้อมูลชุดที่ 2 เป็นชุดข้อมูลทดสอบ ชุดที่ 1, 3-5 เป็นชุดข้อมูลสอน
รอบที่ 3 ใช้ข้อมูลชุดที่ 3 เป็นชุดข้อมูลทดสอบ ชุดที่ 1-2, 4-5 เป็นชุดข้อมูลสอน
รอบที่ 4 ใช้ข้อมูลชุดที่ 4 เป็นชุดข้อมูลทดสอบ ชุดที่ 1-3, 5 เป็นชุดข้อมูลสอน
รอบที่ 5 ใช้ข้อมูลชุดที่ 5 เป็นชุดข้อมูลทดสอบ ชุดที่ 1- 4 เป็นชุดข้อมูลสอน

3.3 วิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ (Percentage split) เป็นการแบ่งชุดข้อมูลแบบสุ่มออกเป็นสองชุด ได้แก่ ชุดข้อมูลสอนและชุดทดสอบตามค่าร้อยละที่กำหนด เช่น ถ้ากำหนดค่าร้อยละเท่ากับ 60% หมายความว่าให้แบ่งชุดข้อมูลสอนเป็น 60% และชุดทดสอบเป็น 40% โปรแกรมเวกาจะกำหนดค่าอัตโนมัติ เป็นร้อยละ 66 ถ้าไม่มีการกำหนดค่าใหม่เอง (Witten, & Frank. 2011)

ในงานวิจัยนี้จะใช้วิธีการทดสอบไขว้ 10 ส่วน (10-folds Cross validation) เป็นการแบ่งข้อมูลออกเป็น 10 ส่วน ในแต่ละรอบจะนำข้อมูลจำนวน 9 ส่วน เป็นข้อมูลสอน และอีก 1 ส่วนเป็นข้อมูลทดสอบ การทดสอบจะทำทั้งหมด 10 ครั้ง เพื่อให้ข้อมูลทุกส่วนได้เป็นข้อมูลทดสอบ แล้วนำผลลัพธ์ที่ได้มาหาค่าเฉลี่ยเพื่อเปรียบเทียบประสิทธิภาพ อัลกอริทึมที่ให้ผลลัพธ์ที่ดีที่สุดจะนำไปพัฒนาเป็นต้นแบบการพยากรณ์ สำหรับการเลือกหลักสูตรที่เหมาะสมในการเข้าศึกษาต่อ

4. การสร้างตัวแบบพยากรณ์ ข้อมูลสอนจะนำไปสร้างต้นแบบการพยากรณ์โดยใช้ตัวจำแนกประเภท 3 วิธี คือ การจำแนกประเภทแบบเบย์ด้วยอัลกอริทึมนาอีฟเบเซียน การจำแนกประเภทโดยใช้ต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 และ อัลกอริทึม RandomForest และการจำแนกประเภทโดยใช้กฎด้วยอัลกอริทึม JRip และมีการแบ่งข้อมูลเพื่อการเรียนรู้โดยใช้วิธีการทดสอบไขว้

5. การเปรียบเทียบประสิทธิภาพของตัวแบบในการจำแนกประเภท ได้แก่ ค่าความแม่นยำ (Accuracy) ความเที่ยง (Precision) คาการเรียกคืน (Recall) ค่าความถ่วงดุล (F-Measure) และ

ค่ากำลังสองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) ต้นแบบที่ได้มาจากการแบ่งข้อมูลเพื่อการเรียนรู้ด้วยวิธีการทดสอบไขว้ ต้นแบบที่ดีที่สุดจะนำไปสร้างระบบการพยากรณ์

6. สร้างระบบการพยากรณ์สำหรับการเลือกหลักสูตรที่เหมะสม

การทดสอบตัวแบบเพื่อการพยากรณ์

ตัวแบบในการจำแนกประเภทจะถูกนำไปประเมินประสิทธิภาพ เพื่อประเมินความถูกต้องและความแม่นยำของแบบจำลอง สามารถประเมินได้ 4 วิธี คือ ค่าความแม่นยำ (Accuracy) ความเที่ยง (Precision) คาการเรียกคืน (Recall) ค่าความถ่วงดุล (F-Measure) และค่ากำลังสองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) วิธีการวัดประสิทธิภาพของตัวแบบโดยใช้ข้อมูลทดสอบ มีรายละเอียด ดังนี้ (ชิดชนก ศรีชัยวงศ์, ไพศาล ตระกูลสุข และสุรเดช บุญลือ. 2557: 1-14; พรรณิภา บุตรเอก และสุรเดช บุญลือ. 2557: 40-49; อนันต์ ปินะเต. 2016; Stehl, Ghosh, & Mooney. 2000)

1. ค่าความแม่นยำ (Accuracy) คือ ค่าที่ใช้วัดประสิทธิภาพของการจำแนกประเภทของตัวแบบโดยใช้ชุดข้อมูลทดสอบ แล้วนำมาเปรียบเทียบกับค่าฉลากประเภท (Class label) ที่ทราบล่วงหน้าของตัวอย่างในชุดข้อมูลทดสอบกับค่าผลลัพธ์การจำแนกประเภทที่ได้จากแบบจำลองเพื่อหาผลลัพธ์จากกระบวนการเรียนรู้ที่สามารถทำนายกลุ่มของตัวอย่างใหม่ได้อย่างถูกต้อง ดังสมการที่ 3.6

$$\text{ค่าความแม่นยำ} = \frac{\text{จำนวนข้อมูลที่ทำนายถูกต้อง}}{\text{จำนวนข้อมูลทั้งหมด}} \tag{3.6}$$

2. ค่าความเที่ยง (Precision) คือ อัตราส่วนของการทำนายข้อมูลในคลาสได้อย่างถูกต้องจากจำนวนข้อมูลทั้งหมดในคลาสนั้น ดังนั้น ค่าความเที่ยง คือ ระดับความเที่ยงตรงของการทำซ้ำ ๆ กันว่าในการทำงานแต่ละครั้งมีค่าใกล้เคียงกันมากน้อยเพียงใด สามารถคำนวณได้ ดังสมการที่ 3.7

$$\text{ค่าความเที่ยง} = \frac{\text{จำนวนข้อมูลที่ทำนายได้ถูกต้องในคลาส}}{\text{จำนวนข้อมูลทั้งหมดที่ทำนายในคลาสนั้น}} \tag{3.7}$$

3. ค่าการเรียกคืน (Recall) คือ อัตราส่วนของการค้นพบคลาสที่ใกล้เคียงกับคำขอและมีการค้นคืนให้กับผู้ใช้กับเอกสารที่ตรงกับคำขอทั้งหมด สามารถคำนวณได้ ดังสมการที่ 3.8

$$\text{ค่าการเรียกคืน} = \frac{\text{จำนวนข้อมูลที่ทำนายได้ถูกต้องในคลาส}}{\text{จำนวนข้อมูลที่ถูกต้องทั้งหมด}} \quad (3.8)$$

4. ค่าความถ่วงดุล (F-Measure) คือ ค่าที่แสดงความสัมพันธ์ระหว่างค่าความเที่ยง และค่าการเรียกคืน เพื่อหาความถูกต้องโดยรวม ผลลัพธ์ของค่าความถูกต้องโดยรวมจะมีค่าระหว่าง 0-1 ซึ่งมีความหมายดังนี้ ถ้าค่าที่ได้เข้าใกล้ 1 หมายถึง การให้ผลในการจำแนกประเภทในกลุ่มนั้น มีประสิทธิภาพสูง และมีความถูกต้องในการพยากรณ์สูง และหากค่าที่ได้เข้าใกล้ 0 หมายถึง การให้ผลในการจำแนกประเภทในกลุ่มนั้นมีประสิทธิภาพต่ำ และมีความถูกต้องในการพยากรณ์ต่ำ สามารถคำนวณได้ดังสมการที่ 3.9

$$\text{ค่าความถูกต้องโดยรวม } (F) = \frac{2 \times (\text{ค่าความแม่นยำ} \times \text{ค่าการเรียกคืน})}{(\text{ค่าความแม่นยำ} + \text{ค่าการเรียกคืน})} \quad (3.9)$$

5. ค่ากำลังสองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) คือ ค่าที่ใช้วัดความแตกต่างระหว่างค่าจริงและค่าที่ได้จากการประมาณ ถ้าค่า RMSE มีค่าน้อยหรือมีค่าเข้าใกล้ศูนย์ แสดงว่าต้นแบบนั้นใกล้เคียงกับความเป็นจริง สามารถคำนวณได้ ดังสมการที่ 3.10

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i' - Y_i)^2} \quad (3.10)$$

โดยที่ $\boldsymbol{Y_i'}$ แทน ค่าของข้อมูลที่เกิดขึ้นจริง

Y_i แทน ค่าที่ได้จากการพยากรณ์

n แทน จำนวนนำเข้าทั้งหมด n

เมื่อดำเนินการทดสอบตัวแบบเพื่อการพยากรณ์แล้วจะทราบค่าความแม่นยำ ความเที่ยง คาการเรียกคืน ค่าความถ่วงดุล และค่ากำลังสองของความคลาดเคลื่อน อัลกอริทึมที่ให้ผลลัพธ์ที่ดีที่สุดจะนำไปพัฒนาเป็นต้นแบบการพยากรณ์ สำหรับการเลือกหลักสูตรที่เหมาะสมในการเข้าศึกษาต่อ

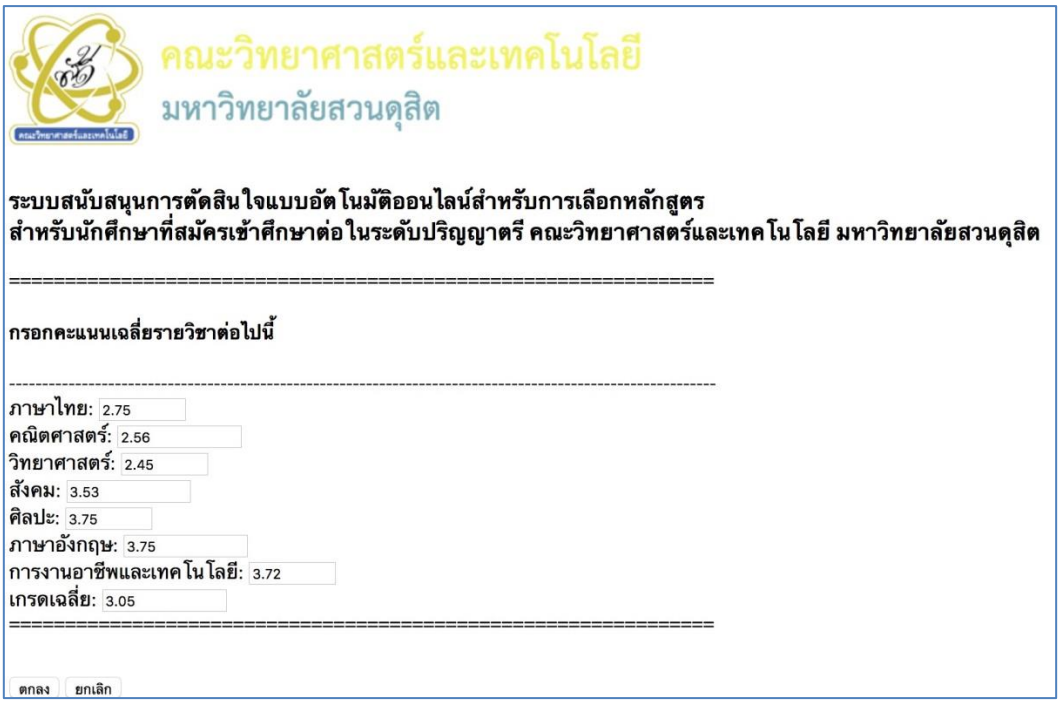
การพัฒนาระบบสนับสนุนการตัดสินใจ

ระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสมในการเข้าศึกษาต่อมหาวิทยาลัยสวนดุสิต เพื่อแนะแนวให้กับนักศึกษาแต่ละคนได้ทราบข้อมูลในด้านการเรียนของตนเอง จะนำต้นแบบที่มีค่าประสิทธิภาพมากที่สุดมาพัฒนาระบบเป็นเว็บแอปพลิเคชันโดยใช้ภาษา PHP ร่วมกับ WEKA CLI โดยมีรายละเอียดที่เกี่ยวข้องในการพัฒนาระบบ ดังนี้

1. การออกแบบและพัฒนาระบบ

จากการวิเคราะห์ระบบมีบุคคลที่เกี่ยวข้องประกอบด้วย ผู้ดูแลระบบและผู้ใช้ระบบ ผู้ดูแลระบบเป็นผู้ที่ทำหน้าที่ในการดูแลและจัดการระบบ ส่วนผู้ใช้ระบบเป็นผู้ที่ใช้งานระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์ สำหรับการเลือกหลักสูตรที่เหมาะสมในการเข้าศึกษาต่อมหาวิทยาลัยสวนดุสิตที่นำเข้าข้อมูลเพื่อให้ระบบพยากรณ์และเลือกหลักสูตรที่เหมาะสมในการเข้าศึกษาต่อมหาวิทยาลัยสวนดุสิต

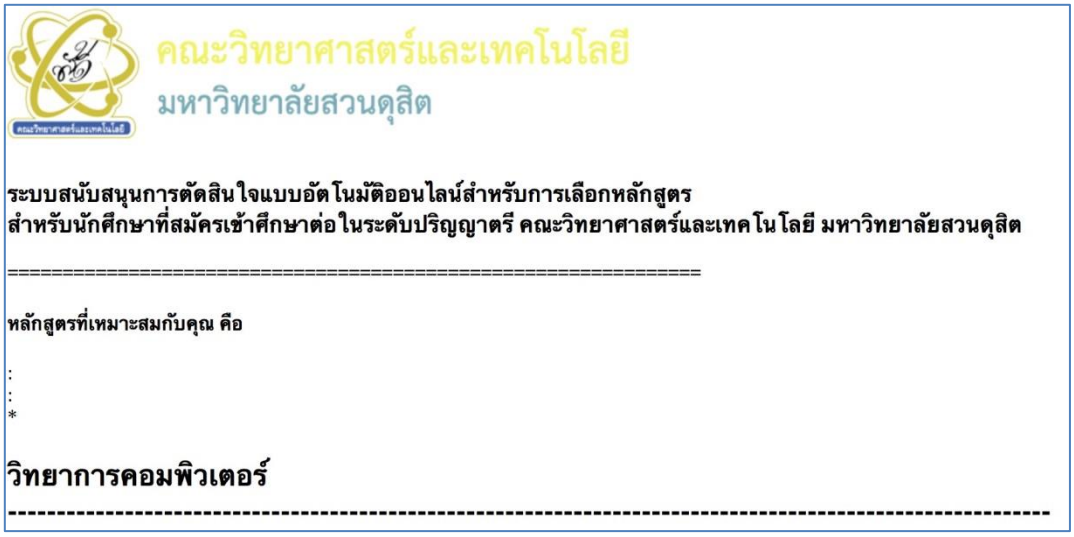
1.1 หน้าจอนำเข้าข้อมูล หน้าจอของระบบถูกออกแบบเพื่อรับข้อมูลเข้าสำหรับการพยากรณ์ผู้ใช้งานรายใหม่ ประกอบด้วยข้อมูลสำคัญที่ต้องใช้ในการพยากรณ์ที่ผู้ใช้งานกรอกข้อมูลของตนเองเข้าไป จำนวน 8 ข้อมูล ข้อมูลดังกล่าวประกอบด้วย คะแนนเฉลี่ยกลุ่มวิชาภาษาไทย (Thai average) คะแนนเฉลี่ยกลุ่มวิชาคณิตศาสตร์ (Math average) คะแนนเฉลี่ยกลุ่มวิชาวิทยาศาสตร์ (Science average) คะแนนเฉลี่ยกลุ่มวิชาสังคม (Social average) คะแนนเฉลี่ยกลุ่มวิชาศิลปะ (Art average) คะแนนเฉลี่ยกลุ่มวิชาภาษาต่างประเทศ (Foreign average) คะแนนเฉลี่ยกลุ่มวิชาการงานอาชีพและเทคโนโลยี (Occupation and Technology average) และคะแนนเฉลี่ยตลอดหลักสูตร (High school average)



ภาพที่ 3.3 หน้าจอนำเข้าข้อมูล

1.2 หน้าจอผลการทำนาย

เมื่อบันทึกข้อมูลเรียบร้อยแล้ว ระบบจะทำการคำนวณแสดงผลการเลือกหลักสูตรทางหน้าจอ ดังภาพที่ 3.4



ภาพที่ 3.4 หน้าจอแสดงผลการทำนาย

บทที่ 4
ผลการวิจัย

งานวิจัยนี้ได้ศึกษาเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ที่ได้จากการจำแนกประเภท ซึ่งจัดเป็นประเภทหนึ่งของการทำเหมืองข้อมูล พร้อมกับการค้นหาคุณสมบัติของข้อมูลที่มีผลต่อการพัฒนาตัวแบบการพยากรณ์ของระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์ในการเลือกเรียนหลักสูตรที่เหมาะสม สำหรับนักศึกษาที่สมัครเข้าศึกษาต่อในระดับปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต ที่มีความถูกต้องและเหมาะสมที่สุด

ต้นแบบการพยากรณ์โดยใช้ตัวจำแนกประเภท 3 วิธี ประกอบด้วย การจำแนกประเภทโดยใช้ต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 และอัลกอริทึม RandomForest การจำแนกประเภทโดยใช้กฎด้วยอัลกอริทึม JRip การจำแนกประเภทแบบเบย์ด้วยอัลกอริทึมนาอีฟเบเซียน และมีการแบ่งข้อมูลเพื่อการเรียนรู้โดยใช้วิธีการทดสอบไขว้แบบ 10 ส่วน ตัวแบบจะถูกประเมินเพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมด้วยค่าความแม่นยำ (Accuracy) ความเที่ยง (Precision) คาการเรียกคืน (Recall) ค่าความถ่วงดุล (F-Measure) และค่ากำลังสองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) ต้นแบบที่ได้มาจากการแบ่งข้อมูลเพื่อการเรียนรู้ด้วยวิธีการทดสอบไขว้ต้นแบบที่ดีที่สุดจะนำไปสร้างระบบการพยากรณ์ ผลการวิจัยสามารถแบ่งออกเป็น 2 ส่วน คือ ผลการประเมินประสิทธิภาพของอัลกอริทึม และผลการวิเคราะห์ความพึงพอใจของระบบ โดยมีรายละเอียด ดังนี้

ผลการประเมินประสิทธิภาพของอัลกอริทึม

1. ผลการประเมินเปรียบค่าความแม่นยำ เมื่อกำหนดระดับผลการเรียนเฉลี่ยเมื่อสำเร็จการศึกษามากกว่า 2.5 และระดับผลการเรียนเฉลี่ยเมื่อสำเร็จการศึกษามากกว่าหรือเท่ากับ 2.00 ผลการประเมินเป็นดังนี้

ตารางที่ 4.1 ค่าความแม่นยำเมื่อกำหนดระดับผลการเรียนเฉลี่ยเมื่อสำเร็จการศึกษา

Algorithm	ค่าความแม่นยำ GPA>2.5	ค่าความแม่นยำ GPA > 2.00
J48	68.2243	63.941
Random Forest	67.2897	63.0027
Jrip	68.785	63.6729
NaiveBayes	45.6075	46.5147

จากตารางที่ 4.1 เมื่อมีการจัดการข้อมูลโดยกำหนดระดับผลการเรียนเฉลี่ยเมื่อสำเร็จการศึกษาเป็น 2 กลุ่ม คือ ระดับผลการเรียนเฉลี่ยเมื่อสำเร็จการศึกษา มากกว่า 2.5 และระดับผลการเรียนเฉลี่ยเมื่อสำเร็จการศึกษา มากกว่าหรือเท่ากับ 2.00 พบว่า ต้นแบบการพยากรณ์ที่ได้จากการจำแนกประเภทด้วยอัลกอริทึม J48 อัลกอริทึม RandomForest และ อัลกอริทึม JRip เมื่อกำหนดระดับผลการเรียนเฉลี่ยเมื่อสำเร็จการศึกษา มากกว่า 2.5 ให้ค่าความแม่นยำสูงกว่า แต่ผลที่ได้จากการจำแนกประเภทด้วยอัลกอริทึม Naïve Bayes ให้ค่าความแม่นยำต่ำกว่าเล็กน้อย

2. ผลการประเมินเปรียบค่าความแม่นยำ เมื่อมีการแปลงระดับผลการเรียนคะแนนเฉลี่ยเป็นคลาส จำนวน 3 คลาส คือ

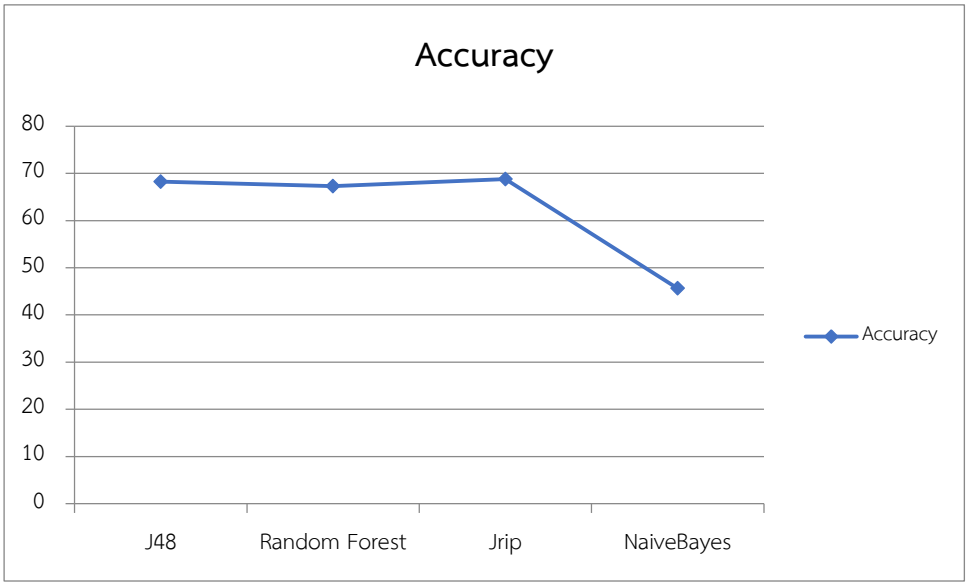
- ผลการเรียน 0.00-2.50 แปลงเป็นผลการเรียนเฉลี่ยระดับต่ำ
- ผลการเรียน 2.51-3.24 แปลงเป็นผลการเรียนเฉลี่ยระดับปานกลาง
- ผลการเรียน 3.25-4.00 แปลงเป็นผลการเรียนเฉลี่ยระดับสูง

ตารางที่ 4.2 ค่าความแม่นยำเมื่อมีการแปลงระดับผลการเรียนคะแนนเฉลี่ยเป็นคลาส

Algorithm	ค่าความแม่นยำ ผลการเรียนคะแนนเฉลี่ย	ค่าความแม่นยำ ผลการเรียนเฉลี่ยเป็นคลาส
J48	68.2243	64.2991
Random Forest	67.2897	59.2523
Jrip	68.785	64.2991
NaiveBayes	45.6075	58.6916

จากตารางที่ 4.2 ผลการประเมินเปรียบค่าความแม่นยำของอัลกอริทึมเมื่อมีการแปลงระดับผลการเรียนคะแนนเฉลี่ยเป็นคลาส จำนวน 3 คลาส คือ ผลการเรียนเฉลี่ยระดับต่ำ ผลการเรียนเฉลี่ยระดับปานกลาง และผลการเรียนเฉลี่ยระดับสูง พบว่า ต้นแบบการพยากรณ์ที่ได้จากการจำแนกประเภทด้วยอัลกอริทึม J48 อัลกอริทึม RandomForest และ อัลกอริทึม JRip เมื่อมีการแปลงระดับผลการเรียนคะแนนเฉลี่ยเป็นคลาส ให้ค่าความแม่นยำสูงกว่า แต่ผลที่ได้จากการจำแนกประเภทด้วยอัลกอริทึม Naïve Bayes ให้ค่าความแม่นยำสูงกว่าเมื่อมีการแปลงระดับผลการเรียนคะแนนเฉลี่ยเป็นคลาส

ภาพที่ 4.1 แสดงผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึมด้วยค่าความแม่นยำ (Accuracy) ของต้นแบบการพยากรณ์โดยใช้อัลกอริทึม J48 อัลกอริทึม RandomForest อัลกอริทึม JRip และอัลกอริทึมนาอีฟเบเซียน และมีการแบ่งข้อมูลเพื่อการเรียนรู้โดยใช้วิธีการทดสอบไขว้แบบ 10 ส่วน พบว่า อัลกอริทึม JRip ให้ค่าสูงที่สุด รองลงมาคือ อัลกอริทึม J48 อัลกอริทึม RandomForest และอัลกอริทึมนาอีฟเบเซียน ตามลำดับ

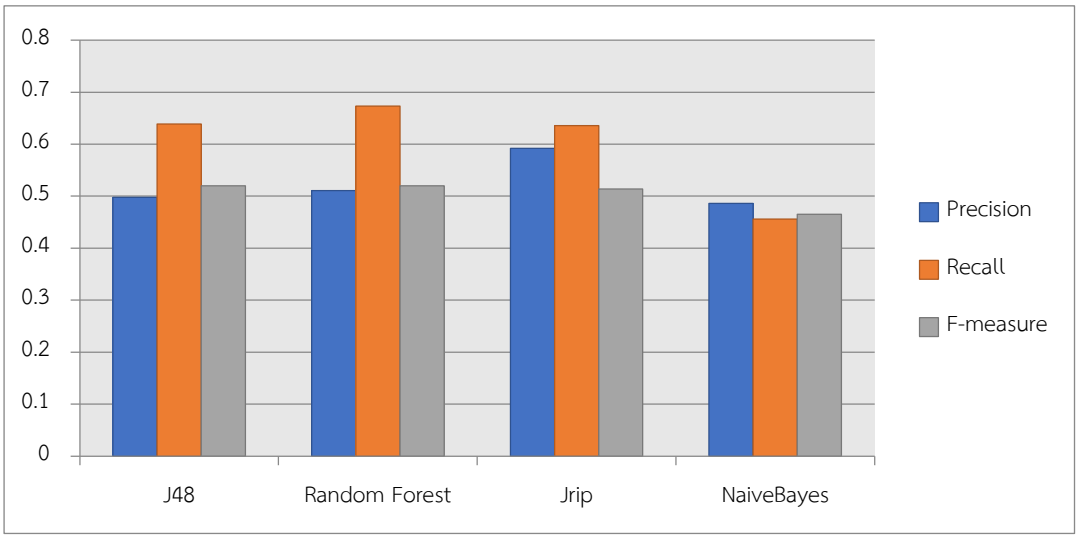


ภาพที่ 4.1 ผลการเปรียบเทียบค่าความแม่นยำของอัลกอริทึม

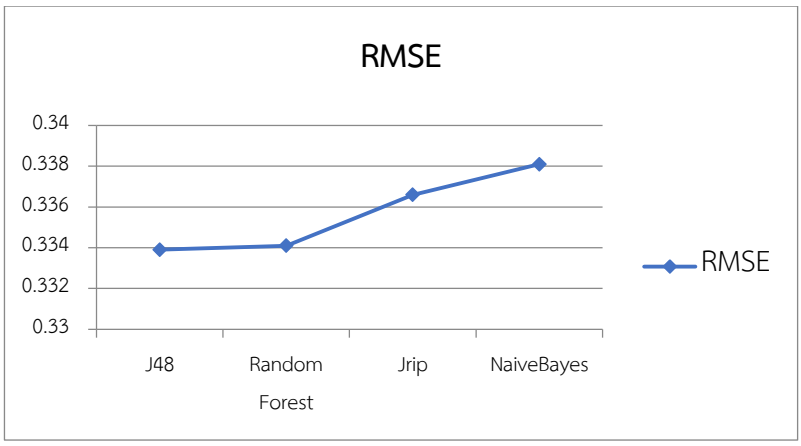
ตารางที่ 4.3 ผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึม ด้านความเที่ยง (Precision) คาการเรียกคืน (Recall) และค่าความถ่วงดุล (F-Measure)

Algorithm	**Precision**	**Recall**	**F-measure**
J48	0.498	0.639	0.52
Random Forest	0.511	0.673	0.521
Jrip	0.561	0.688	0.593
NaiveBayes	0.486	0.456	0.465

จากตารางที่ 4.3 และภาพที่ 4.2 พบว่า ต้นแบบที่ได้จากอัลกอริทึม JRip ให้ค่าความเที่ยง (Precision) คาการเรียกคืน (Recall) และค่าความถ่วงดุล (F-Measure) สูงที่สุด โดยให้ค่าความเที่ยง (Precision) เท่ากับ 0.561 ค่าการเรียกคืน (Recall) เท่ากับ 0.688 และค่าความถ่วงดุล เท่ากับ 0.593 นอกจากนี้ยังมีการเปรียบเทียบค่ารากที่สองของความคลาดเคลื่อน (RMSE) พบว่า ตัวแบบพยาการณ์ที่ได้จากอัลกอริทึม J48 ให้ค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดจากการวัด เท่ากับ 0.3339 ดังภาพที่ 4.3



ภาพที่ 4.2 ผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึม ด้านความเที่ยง (Precision) คาการเรียกคืน (Recall) และค่าความถ่วงดุล (F-Measure)



ภาพที่ 4.3 ผลการเปรียบเทียบค่ากำลังสองของความคลาดเคลื่อน (Root Mean Square Error: RMSE)

ผลการวิเคราะห์ความพึงพอใจของระบบ

เมื่อได้พัฒนาระบบสนับสนุนการตัดสินใจในการเลือกหลักสูตรที่เหมาะสมแล้ว จะต้องมีการประเมินประสิทธิภาพของระบบโดยผู้เชี่ยวชาญและผู้ใช้งาน การวิเคราะห์ความพึงพอใจพิจารณาจาก 5 ด้าน คือ ด้านการทำงานตรงตามความต้องการ (Functional requirement test) ด้านการใช้งานระบบ (Function test) ด้านการใช้งานระบบ (Usability) ด้านประสิทธิภาพของระบบ (Performance test) และด้านความปลอดภัยของระบบ (Security test) ผลประเมินประสิทธิภาพระบบโดยผู้เชี่ยวชาญจำนวน 5 คน และผู้ใช้งาน จำนวน 30 คน ทดลองใช้งานระบบแล้ว มีการเก็บข้อมูลด้านความพึงพอใจในการใช้งาน ผลการประเมินประสิทธิภาพระบบเป็นดังตารางที่ 4.4-4.5

ตารางที่ 4.4 ผลประเมินความพึงพอใจของระบบโดยผู้เชี่ยวชาญ

Algorithm	**ค่าเฉลี่ย**	**ค่าเบี่ยงเบนมาตรฐาน**	**ความหมาย**
1. ด้านการทำงานตรงตามความต้องการ	4.00	0.471	พอใจมาก
2. ด้านการใช้งานระบบ	3.86	0.516	พอใจมาก
3. ด้านการใช้งานระบบ	3.84	0.473	พอใจมาก
4. ด้านประสิทธิภาพของระบบ	3.80	0.422	พอใจมาก
5. ด้านความปลอดภัยของระบบ	3.73	0.457	พอใจมาก
ภาพรวมเฉลี่ย	3.84	0.466	พอใจมาก

จากตารางที่ 4.4 พบว่า โดยภาพรวมผู้เชี่ยวชาญมีความพึงพอใจในการใช้งานระบบอยู่ในระดับมาก ค่าเฉลี่ยเท่ากับ 3.84 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.466 สำหรับหัวข้อประเมินที่ผู้เชี่ยวชาญมีความพึงพอใจสูงที่สุด คือ ด้านการใช้งาน มีค่าเฉลี่ยความพึงพอใจเท่ากับ 3.86 และค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.516 หัวข้อประเมินที่ผู้เชี่ยวชาญมีความพึงพอใจน้อยที่สุด คือ ด้านความปลอดภัยของระบบ มีค่าเฉลี่ยความพึงพอใจเท่ากับ 3.73 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.466

ตารางที่ 4.5 ผลประเมินความพึงพอใจของระบบโดยผู้ใช้งาน

Algorithm	**ค่าเฉลี่ย**	**ค่าเบี่ยงเบนมาตรฐาน**	**ความหมาย**
1. ด้านการทำงานตรงตามความต้องการ	3.933	0.406	พอใจมาก
2. ด้านการใช้งานระบบ	3.822	0.646	พอใจมาก
3. ด้านการใช้งานระบบ	3.746	0.667	พอใจมาก
4. ด้านประสิทธิภาพของระบบ	3.926	0.632	พอใจมาก
5. ด้านความปลอดภัยของระบบ	3.822	0.680	พอใจมาก
ภาพรวมเฉลี่ย	3.825	0.594	พอใจมาก

ผลประเมินประสิทธิภาพระบบโดยผู้ใช้งาน จากการให้ผู้ใช้จำนวน 30 คน ทดลองใช้งานระบบแล้ว จึงเก็บข้อมูลด้านความพึงพอใจในการใช้งาน ผลการประเมินประสิทธิภาพระบบเป็นดังตารางที่ 4.5 พบว่า โดยภาพรวมผู้ใช้งานมีความพึงพอใจในการใช้งานระบบอยู่ในระดับมาก ค่าเฉลี่ยเท่ากับ 3.825 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.594 สำหรับหัวข้อประเมินที่ผู้เชี่ยวชาญมีความพึงพอใจสูงที่สุด คือ ด้านการทำงานตรงตามความต้องการ มีค่าเฉลี่ยเท่ากับ 3.933 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.406 หัวข้อประเมินที่ผู้เชี่ยวชาญมีความพึงพอใจน้อยที่สุด คือ ด้านการใช้งานระบบ มีค่าเฉลี่ยเท่ากับ 3.746 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.667

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

งานวิจัยนี้เป็นการพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสม สำหรับนักศึกษาที่สมัครเข้าศึกษาต่อในระดับปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต โดยใช้เทคนิคเหมืองข้อมูล แล้วนำผลที่ได้มาสรุป อภิปรายผลและข้อเสนอแนะ มีรายละเอียด ดังนี้

สรุปผลการวิจัย

การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสม สำหรับนักศึกษาที่สมัครเข้าศึกษาต่อในระดับปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต ใช้ข้อมูลของนักศึกษาที่สำเร็จการศึกษาไปแล้วย้อนหลัง 4 ปี ตั้งแต่ปี พ.ศ. 2553-2556 และข้อมูลของนักศึกษาที่กำลังศึกษาอยู่ในชั้นปีที่ 1 คณะวิทยาศาสตร์และเทคโนโลยี ประกอบด้วยข้อมูลส่วนตัวนักศึกษา และข้อมูลผลการเรียนในระดับมัธยมศึกษาตอนปลาย ข้อมูลที่ได้นำมาสร้างต้นแบบการพยากรณ์โดยใช้ตัวจำแนกประเภท 3 วิธี ประกอบด้วย การจำแนกประเภทโดยใช้ต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 และอัลกอริทึม RandomForest การจำแนกประเภทโดยใช้กฎด้วยอัลกอริทึม JRip การจำแนกประเภทแบบเบย์ด้วยอัลกอริทึมนาอีฟเบเซียน และมีการแบ่งข้อมูลเพื่อการเรียนรู้โดยใช้วิธีการทดสอบไขว้แบบ 10 ส่วน ตัวแบบจะถูกประเมินเพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมด้วยค่าความแม่นยำ (Accuracy) ความเที่ยง (Precision) คาการเรียกคืน (Recall) ค่าความถ่วงดุล (F-Measure) และค่ากำลังสองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) ต้นแบบที่ได้มาจากการแบ่งข้อมูลเพื่อการเรียนรู้ด้วยวิธีการทดสอบไขว้ ต้นแบบที่ดีที่สุดจะนำไปพัฒนาเป็นระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์ในการเลือกเรียนหลักสูตรที่เหมาะสม สำหรับนักศึกษาที่สมัครเข้าศึกษาต่อในระดับปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต

ผลการวิจัยพบว่าต้นแบบที่ได้จากอัลกอริทึม JRip ให้ค่าความแม่นยำ (Accuracy) ความเที่ยง (Precision) คาการเรียกคืน (Recall) ค่าความถ่วงดุล (F-Measure) สูงที่สุด แต่เมื่อเปรียบเทียบค่ารากที่สองของความคลาดเคลื่อน (RMSE) พบว่า ตัวแบบพยาการณ์ที่ได้จากอัลกอริทึม J48 ให้ค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดจากการวัด เท่ากับ 0.3339 ซึ่งมีความใกล้เคียงกับค่ารากที่สองของความคลาดเคลื่อนที่ได้จากอัลกอริทึม JRip ที่ทีค่าเท่ากับ 0.3366

เมื่อได้พัฒนาระบบสนับสนุนการตัดสินใจในการเลือกหลักสูตรที่เหมาะสมแล้ว มีการประเมินประสิทธิภาพของระบบโดยผู้เชี่ยวชาญและผู้ใช้งาน การวิเคราะห์ความพึงพอใจพิจารณาจาก 5 ด้าน คือ ด้านการทำงานตรงตามความต้องการ (Functional requirement test) ด้านการใช้งานระบบ (Functional test) ด้านการใช้งานระบบ(Usability test) ด้านประสิทธิภาพของระบบ (Performance test) และด้านความปลอดภัยของระบบ (Security test) ผลประเมินประสิทธิภาพระบบโดยผู้เชี่ยวชาญ และผู้ใช้งาน พบว่า ผู้เชี่ยวชาญมีความพึงพอใจในการใช้งานระบบโดยภาพรวมอยู่ในระดับมาก ค่าเฉลี่ยเท่ากับ 3.84 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.466 สำหรับหัวข้อประเมินที่ผู้เชี่ยวชาญมีความพึงพอใจสูงที่สุด คือ ความสามารถในการนำไปใช้ได้จริง/เป็นแนวทางในการหลักสูตรที่เหมาะสม มีค่าเฉลี่ยเท่ากับ 4.2 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.447 และการใช้งานง่าย สะดวก ไม่ซับซ้อน มีค่าเฉลี่ยเท่ากับ 4.2 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.447 หัวข้อประเมินที่ผู้เชี่ยวชาญมีความพึงพอใจน้อยที่สุด คือ มีคำอธิบายในการใช้งานระบบมีค่าเฉลี่ยเท่ากับ 3.6 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.547 สำหรับผลประเมินประสิทธิภาพระบบโดยผู้ใช้งาน พบว่า โดยภาพรวมผู้ใช้งานมีความพึงพอใจในการใช้งานระบบอยู่ในระดับมาก ค่าเฉลี่ยเท่ากับ 3.825 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.594 สำหรับหัวข้อประเมินที่ผู้เชี่ยวชาญมีความพึงพอใจสูงที่สุด คือ ความสามารถในการแสดงข้อมูลมีค่าเฉลี่ยเท่ากับ 4.00 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.447 หัวข้อประเมินที่ผู้เชี่ยวชาญมีความพึงพอใจน้อยที่สุด คือ มีขนาด สี และตัวอักษรมีความเหมาะสม ชัดเจน อ่านง่าย มีค่าเฉลี่ยเท่ากับ 3.5 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.731

อภิปรายผล

จากผลการวิจัย พบว่า ตัวแปรอิสระที่มีผลต่อการสร้างตัวแบบ ได้แก่ ผลการเรียนเฉลี่ยกลุ่มวิชาภาษาไทย ผลการเรียนเฉลี่ยกลุ่มวิชาคณิตศาสตร์ ผลการเรียนเฉลี่ยกลุ่มวิชาวิทยาศาสตร์ ผลการเรียนเฉลี่ยกลุ่มวิชาสังคม ผลการเรียนเฉลี่ยกลุ่มวิชาศิลปะ ผลการเรียนเฉลี่ยกลุ่มวิชาภาษาต่างประเทศ ผลการเรียนเฉลี่ยกลุ่มวิชาการงานอาชีพและเทคโนโลยี ผลการเรียนเฉลี่ยรวมระดับมัธยมปลาย ตัวแบบที่ได้จากอัลกอริทึม JRip เมื่อนำมาวัดประสิทธิภาพของอัลกอริทึมด้วยค่าความแม่นยำ (Accuracy) ความเที่ยง (Precision) คาการเรียกคืน (Recall) ค่าความถ่วงดุล (F-Measure) และค่ากำลังสองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) ให้ผลดีที่สุด เพราะอัลกอริทึม JRip เป็นการเรียนรู้กฎที่มีการตัดแต่งกิ่งที่เพิ่มขึ้น ซ้ำ ๆ จึงลดข้อผิดพลาดที่เกิดขึ้นจากการสร้างกฎ อย่างไรก็ตามจากผลการวิจัยประสิทธิภาพของอัลกอริทึมยังอยู่ในระดับที่ไม่สูงมากนัก เนื่องจากนักศึกษาในแต่ละหลักสูตรมีจำนวนไม่เท่ากันนักศึกษาที่มีจำนวนนักศึกษามากที่สุด คือ หลักสูตรเทคโนโลยีสารสนเทศ คิดเป็น 50% ของจำนวนนักศึกษาทั้งหมด ในขณะที่จำนวนของ

นักศึกษาหลักสูตรวิทยาการความปลอดภัย มีจำนวนคิดเป็น 10 % จากข้อมูลดังกล่าวจึงทำให้การจำแนกประเภทมีความคลาดเคลื่อน ทั้งนี้เพราะการจำแนกประเภทจะคิดจากจำนวนของข้อมูลที่มีมากที่สุด จะจัดให้อยู่ในกลุ่มนั้นก่อน หรือก็คือถ้าหลักสูตรใดมีจำนวนนักศึกษามาก ต้นแบบที่ได้จากการจำแนกประเภทด้วยอัลกอริทึมต่าง ๆ ก็จะพยากรณ์ว่านักศึกษาจะเหมาะสมกับหลักสูตรนั้น จึงทำให้ค่าความแม่นยำในการพยากรณ์ไม่สูง

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

ระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกหลักสูตรที่เหมาะสมในการเข้าศึกษาต่อมหาวิทยาลัยสวนดุสิตโดยใช้เทคนิคเหมืองข้อมูลที่พัฒนาขึ้น สามารถนำไปเป็นแนวทางในการเลือกหลักสูตรที่เหมาะสมและสามารถนำไปใช้ได้จริง โดยผู้ใช้ควรเป็นนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยี เนื่องจากบริบททางด้านการศึกษาในระดับมัธยมศึกษาจะมีลักษณะคล้ายคลึงกันและนักศึกษามาจากแผนการเรียนวิทยาศาสตร์-คณิตศาสตร์ จึงทำให้การนำไปใช้เพื่อช่วยแนะแนวสามารถนำไปใช้ได้อย่างมีประสิทธิภาพ

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

จากผลการวิจัยถ้าต้องการเพิ่มประสิทธิภาพของอัลกอริทึมให้สูงขึ้น ต้องพิจารณาจากการเก็บข้อมูลของนักศึกษาเพิ่มขึ้นทุกหลักสูตร เพื่อนำมาใช้ในการสร้างตัวแบบพยากรณ์ ร่วมกับการทำการคัดเลือกคุณลักษณะของข้อมูลที่จะนำมาใช้ในการสร้างต้นแบบ โดยใช้อัลกอริทึมต่าง ๆ

บรรณานุกรม

บรรณานุกรมภาษาไทย

กิตติ ภักดีวัฒนะกุล. (2548). *การวิเคราะห์และออกแบบระบบ*. กรุงเทพ: ซีเอ็ดยูเคชั่น.

กิตติ ภักดีวัฒนะกุล. (2550). *คัมภีร์ระบบสนับสนุนการตัดสินใจและระบบผู้เชี่ยวชาญ*. กรุงเทพฯ: เคทีพี คอมพ์ แอนด์ คอนซัลท์.

กิตติ ภักดีวัฒนะกุล. (2552). *PHP*. กรุงเทพฯ: เคทีพี คอมพ์ แอนด์ คอนซัลท์.

ชิดชนก ศรีชัยวงค์ ไพศาล ตระกูลสุข และสุรเดช บุญลือ. (2557). ระบบสนับสนุนการตัดสินใจเพื่อวินิจฉัยโรคใบลำไยด้วยเทคนิคต้นไม้ตัดสินใจ. *Veridian E-Journal Science and Technology Silpakorn University*, 1(6), 1-14.

ชินพัฒน์ แก้วชินพร.(2553). *การจำแนกประเภทข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจและการจัดกลุ่ม*. ปริญญานิพนธ์มหาบัณฑิต, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. กรุงเทพฯ.

ชุติมา อุตมะมุณีย์ และประสงค์ ปราณีตพลกรัง. (2553). การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา. *Journal of Information Science and Technology*, 1(2), 39-48.

ณัฎภัทรศญา ทับทิมเทศ. 2550. *ระบบสนับสนุนการตัดสินใจ*. www.no-poor.com/dssandos/Chapter5-dss.htm.

ดิษฐพล มั่นธรรม. (2010). *การประยุกต์ขั้นตอนวิธีต้นไม้ตัดสินใจกับการวินิจฉัยโรคระบบการหายใจ: กรณีศึกษาที่โรงพยาบาลพระนครศรีอยุธยา*. ปริญญานิพนธ์มหาบัณฑิต. มหาวิทยาลัยเกษตรศาสตร์. กรุงเทพฯ.

ธนาคม จุ้ยศิริ. (2557). การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกสาขาวิชาและวิชาเอกของผู้ที่สมัครเข้าศึกษาต่อในระดับปริญญาตรีของมหาวิทยาลัยสุโขทัยธรรมาธิราชด้วยเหมืองข้อมูล. *การประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 36*, 11-13 ธันวาคม 2557 ณ เฟลิกซ์ ริเวอร์แคว รีสอร์ท จังหวัดกาญจนบุรี. กรุงเทพ ฯ: มหาวิทยาลัยธุรกิจบัณฑิต.

ธานินทร์ ศิลป์จารุ. (2555). *การวิจัยและวิเคราะห์ข้อมูลทางสถิติด้วย SPSS*. พิมพ์ครั้งที่ 11. กรุงเทพฯ: บิสซิเนสอาร์แอนด์ดี.

ธีรพงษ์ สังข์ศรี. (2557). การวิเคราะห์พฤติกรรมสำหรับการเลือกสมัครสาขาวิชาเรียนและการเปรียบเทียบตัวแบบพยากรณ์จำนวนนักศึกษาใหม่โดยใช้เทคนิคการทำเหมืองข้อมูล. *การประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 10*, 8-9 พฤษภาคม 2557 ณ โรงแรมอังสนา ลากูนา จังหวัดภูเก็ต. กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

นงเยาว์ ในอรุณ. (2557). *การประยุกต์ใช้เทคนิคเหมืองข้อมูลในการพยากรณ์โรคเบาหวาน: กรณีศึกษาโรงพยาบาลศูนย์สวรรค์ประชารักษ์*. ปริญญานิพนธ์ดุษฎีบัณฑิต. มหาวิทยาลัยนเรศวร. กรุงเทพฯ.

นิเวศ จิระวิชิตชัย. (2553). *การค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ*. กรุงเทพ ฯ: มหาวิทยาลัยราชภัฏสวนสุนันทา.

ปราลี มณีรัตน์. (2555). การใช้เทคนิคเหมืองข้อมูลในการหาปัจจัยที่มีผลต่อความพึงพอใจของนักศึกษาในระบบบริหารความสัมพันธ์นักศึกษา. *วารสารวิจัยรามคำแหง (วิทยาศาสตร์และเทคโนโลยี)*, 15, 56-70.

พรรณิภา บุตรเอก และสุรเดช บุญลือ. (2557). การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน. Veridian E-Journal Science and Technology Silpakorn University, 1(6), 40-49.

พรพิรุณ โอ่งอินทร์ และวรรณา ศิริแสงตระกูล. (2558). การวิเคราะห์ปัจจัยที่มีความสัมพันธ์ต่อการเกิดอาชญากรรมด้วยวิธีการทำเหมืองข้อมูล. *การประชุมวิชาการเสนอผลงานวิจัยระดับบัณฑิตศึกษาแห่งชาติ ครั้งที่ 34*, 27 มีนาคม 2558. มหาวิทยาลัยขอนแก่น.

เพชรรัตน์ ปัญญาภาณุวัฒน์, พรชัย มงคลนาม, และวิทิดา จงศุภชัยสิทธิ์. (2552). การจัดสรรทรัพยากรในการพัฒนาเว็บแอบพลิเคชันโดยใช้แผนภาพการตัดสินใจ. *The 5th National Conference on Computing and Information Technology (NCCIT 2009)*. กรุงเทพ.

ไพฑูรย์ จันทร์เรือง. (2550). *ระบบการสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคต้นไม้ตัดสินใจ.* ปริญญานิพนธ์มหาบัณฑิต. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. กรุงเทพฯ.

เยาวภา ภารสำเร็จ. (2556). การเปรียบเทียบอัลกอริทึมเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อระดับผลการเรียนของนักศึกษา. *วารสารวิทยาศาสตร์และเทคโนโลยี*. มหาวิทยาลัยมหาสารคาม, ฉบับพิเศษ, 281-289.

ศิริชัย พงษ์วิชัย. (2551). *การวิเคราะห์ข้อมูลทางสถิติด้วยคอมพิวเตอร์*. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

สารสนเทศอุดมศึกษา. (2558). สืบค้นเมื่อ 10 กันยายน 2558, จาก www.info.mua.go.th/information)

สุธีรา วงศ์อนันทรัพย์ ต้องใจ แย้มผกา และอรวรรณ มูสิกะ. (2559). การใช้เทคนิคเหมืองข้อมูลในการประเมินความรู้ และหาความถนัด เพื่อพัฒนาศักยภาพของนักศึกษา. *วารสารสังคมศาสตร์*. 5(1), 12-19.

เสกสรรค์ วิลัยลักษณ์ วิภา เจริญภัณฑารักษ์ และดวงดาว วิชาดากุล. (2558). การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา. *Veridian E-journal Science and Technology Silpakorn University*. สาขาวิทยาศาสตร์และเทคโนโลยี, 2 (2), 1-17.

อนันต์ ปินะเต. (2559). การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกสมัครในสาขาวิชาโดยใช้เทคนิคต้นไม้ตัดสินใจ. *วารสารวิทยาศาสตร์และเทคโนโลยี*. มหาวิทยาลัยมหาสารคาม, 35 (4), 413-421.

อัจฉราภรณ์ จุฑาผาด กานต์ เจริญจิตร และพยุง มีสัจ. (2558). การใช้เทคนิคการทำเหมืองข้อมูลเพื่อค้นหาสาขาวิชาที่เหมาะสมให้กับนักศึกษา คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏเพชรบุรี. *Joint Conference on ACTIS & NCOBA 2015*, 30-31 มกราคม 2558 นครพนม.

บรรณานุกรมภาษาต่างประเทศ

Delavari, N., Beikzadeh, M.R. & Phon-Amnuaisuk, S. (2005). Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System. *In 6th International Conference on Information Technology Based Higher Education and Training*, Santo Domingo, Dominant Republic.

Dole, L., & Rajurkar, J. (2014). A Decision Support System for Predicting Student Performance. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(12), 7232-7237.

Hsu, P.L., Lai R., Chiu, C.C., Hsu C.I. (2003). The hybrid of association rule algorithms and genetic algorithms for tree induction: an example of predicting the student course performance. *Expert Systems with Applications*, 25(1): 51–62.

Han, J., Kamber, M. & Pei, J. (2011). *Data Mining: Concepts and Techniques*. (3th ed.). Morgan Kaufmann Publisher.

Linoff, G. S., & Berry M. J. A. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. New York: Wiley Publishing.

WEKA. (2013). Weka 3: *Data Mining Software in Java*. Cited 2014 September 10. Available from: http://www.cs.waikato.ac.nz/ml/weka/.

Witten, I.H. & Frank, E. (2011). *Data Mining Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann Publishers.

Xu, X. (2014). *Class JRip*. Cited 2014 September 10. Available from: http://weka.sourceforge.net/doc.stable/weka/classifiers/rules/JRip.html.

ภาคผนวก

ภาคผนวก ก
ผลการวิเคราะห์ความเห็นของผู้เชี่ยวชาญและผู้ใช้งาน

ตาราง ก-1 ผลการวิเคราะห์ความเห็นของผู้เชี่ยวชาญ

หัวข้อประเมิน	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ความหมาย
1. ด้านการทำงานตรงตามความต้องการ			
1.1 ความสามารถในการแสดงข้อมูล	3.80	0.447	พอใจมาก
1.2 ความสามารถในการนำไปใช้ได้จริง/เป็นแนวทางในการหลักสูตรที่เหมาะสม	4.20	0.447	พอใจมาก
2. ด้านการใช้งานระบบ			
2.1 ความถูกต้องในการทำงานของระบบ	3.80	0.447	พอใจมาก
2.2 ความถูกต้องในการแสดงข้อมูล	4.00	0.447	พอใจมาก
2.3 ความถูกต้องของผลลัพธ์ที่แสดง	3.80	0.447	พอใจมาก
3. ด้านการใช้งานระบบ			
3.1 การออกแบบหน้าจอและจัดวางองค์ประกอบต่าง ๆ ของระบบได้เหมาะสม	3.80	0.447	พอใจมาก
3.2 ภาษาเข้าใจง่าย สะดวกต่อการใช้งาน	3.80	0.447	พอใจมาก
3.3 ขนาด สี และตัวอักษรมีความเหมาะสม ชัดเจน อ่านง่าย	3.60	0.447	พอใจมาก
3.4 องค์ประกอบ สี พื้นหลัง ภาพประกอบมีความเหมาะสม	3.80	0.447	พอใจมาก
3.5 การใช้งานง่าย สะดวก ไม่ซับซ้อน	4.20	0.447	พอใจมาก
4. ด้านประสิทธิภาพของระบบ			
4.1 มีความแม่นยำและความเสถียรภาพในการเข้าใช้งาน	3.80	0.447	พอใจมาก
4.2 ความเร็วการประมวลผล	3.80	0.447	พอใจมาก
5. ด้านความปลอดภัยของระบบ			
5.1 มีการควบคุมความปลอดภัยในการเข้าถึงข้อมูลในระบบ	3.80	0.447	พอใจมาก
5.2 มีการกำหนดสิทธิการเข้าใช้ระบบของผู้ใช้	3.80	0.447	พอใจมาก

หัวข้อประเมิน	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ความหมาย
5.3 มีคำอธิบายในการใช้งานระบบ	3.60	0.547	พอใจมาก
ภาพรวมเฉลี่ย	3.825	0.594	พอใจมาก

ตาราง ก-2 ผลประเมินความพึงพอใจของระบบโดยผู้ใช้งาน

หัวข้อประเมิน	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ความหมาย
1. ด้านการทำงานตรงตามความต้องการ			
1.1 ความสามารถในการแสดงข้อมูล	4.000	0.447	พอใจมาก
1.2 ความสามารถในการนำไปใช้ได้จริง /เป็นแนวทางในการหลักสูตรที่เหมาะสม	3.867	0.571	พอใจมาก
2. ด้านการใช้งานระบบ			
2.1 ความถูกต้องในการทำงานของระบบ	3.833	0.699	พอใจมาก
2.2 ความถูกต้องในการแสดงข้อมูล	3.767	0.568	พอใจมาก
2.3 ความถูกต้องของผลลัพธ์ที่แสดง	3.867	0.681	พอใจมาก
3. ด้านการใช้งานระบบ			
3.1 การออกแบบหน้าจอและจัดวางองค์ประกอบต่าง ๆ ของระบบได้เหมาะสม	3.900	0.607	พอใจมาก
3.2 ภาษาเข้าใจง่าย สะดวกต่อการใช้งาน	3.700	0.750	พอใจมาก
3.3 ขนาด สี และตัวอักษรมีความเหมาะสม ชัดเจน อ่านง่าย	3.500	0.731	พอใจมาก
3.4 องค์ประกอบ สี พื้นหลัง ภาพประกอบมีความเหมาะสม	3.767	0.626	พอใจมาก
3.5 การใช้งานง่าย สะดวก ไม่ซับซ้อน	3.867	0.571	พอใจมาก
4. ด้านประสิทธิภาพของระบบ			
4.1 มีความแม่นยำและความเสถียรภาพในการเข้าใช้งาน	3.833	0.699	พอใจมาก
4.2 ความเร็วการประมวลผล	3.767	0.568	พอใจมาก

หัวข้อประเมิน	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ความหมาย
5. ด้านความปลอดภัยของระบบ			
5.1 มีการควบคุมความปลอดภัยในการเข้าถึงข้อมูลในระบบ	3.867	0.681	พอใจมาก
5.2 มีการกำหนดสิทธิการเข้าใช้ระบบของผู้ใช้	3.900	0.607	พอใจมาก
5.3 มีคำอธิบายในการใช้งานระบบ	3.700	0.749	พอใจมาก
ภาพรวมเฉลี่ย	3.825	0.594	พอใจมาก

ประวัติผู้วิจัย

ชื่อ	นางศิริพร ฉิมพลี
วันเดือนปีเกิด	10 กันยายน 2516
วุฒิการศึกษา	ปีการศึกษา 2538: พยาบาลศาสตรบัณฑิต (พยาบาลศาสตร์) วิทยาลัยพยาบาลเกื้อการุณย์ ปีการศึกษา 2544: วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์) สถาบันบัณฑิตพัฒนบริหารศาสตร์ ปีการศึกษา 2555: Doctor of Philosophy (Computer Science) University Technology of Malaysia
ตำแหน่ง	ผู้ช่วยศาสตราจารย์
ทุนวิจัย	ปีงบประมาณ 2558: ทุนสนับสนุนการวิจัย มหาวิทยาลัยสวนดุสิต

ผลงานวิชาการ

Thanakorn Pamutha, Chom Kimpan, Siriporn Chimplee and Parinya Sanguansat. (2012). Improving Web Page Prediction Using Default Rule Selection. *International Journal of Advanced Computer Science and Applications* (IJACSA). 3(11): 159 - 164.

Thanakorn Pamutha, Siriporn Chimphlee, Chom Kimpan and Parinya Sanguansat. (2013). Effective Sequence Representation in Predicting Web Page Accessibility. *Advances in Information Sciences and Service Sciences* (AISS). 5(12): 168 - 179.

Thanakorn Pamutha, Siriporn Chimphlee, Chom Kimpan and Parinya Sanguansat. (2014). Web Page Access Prediction on Server Side. *Journal of Convergence Information Technology* (JCIT). 9(5): 81 - 91.

ประสบการณ์การทำงาน	2544-ปัจจุบัน: อาจารย์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต 2538-2544: พยาบาลวิชาชีพ โรงพยาบาลกลาง กรุงเทพมหานคร