

การเปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูบบุหรี่ของวัยรุ่น
ด้วยเทคนิคเหมืองข้อมูล

**The Comparing of Classification Efficiency for Predicting Adolescent Smoking
Behavior by Data Mining Techniques**



ศุภนิดา เปี่ยมมาลัย
นภัสกร นำพรชวาลรุ่ง
ลลิตา เกิดสันเทียะ

ภาคนิพนธ์เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาตรีวิทยาศาสตร์บัณฑิต
รายวิชาโครงการเทคโนโลยีสารสนเทศ
ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยสยาม
พ.ศ 2566

หัวข้อภาคนิพนธ์ : การเปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูบบุหรี่ของวัยรุ่นด้วยเทคนิคเหมืองข้อมูล
 : The Comparing of Classification Efficiency for Predicting Adolescent Smoking Behavior by Data Mining Techniques

หน่วยกิตของภาคนิพนธ์ : 3 หน่วยกิต

คณะผู้จัดทำ : นางสาวศุภนิตา เปี่ยมาลัย 6105100001
 : นางสาวนภัสกร นำพรชวาลรุ่ง 6105100005
 : นางสาวลลิตา เกิดสันเทียะ 6105100008

อาจารย์ที่ปรึกษา : อาจารย์ณรงค์ฤทธิ์ สุคนธ์สิงห์

ระดับการศึกษา : วิทยาศาสตรบัณฑิต

ภาควิชา : เทคโนโลยีสารสนเทศ

ปีการศึกษา : 2566

อนุมัติให้ภาคนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาตรีวิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

คณะกรรมการสอบภาคนิพนธ์

..... ประธานกรรมการ
 (ผศ.ดร.ศักดิ์ชาย ตั้งวรรณวิทย์)

..... กรรมการสอบ
 (อาจารย์อรรณพ กางกั้น)

..... อาจารย์ที่ปรึกษา
 (อาจารย์ณรงค์ฤทธิ์ สุคนธ์สิงห์)

หัวข้อโครงการ	: การเปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูญบุหรีของวัยรุ่นด้วยเทคนิคเหมืองข้อมูล		
หน่วยกิต	: 3 หน่วยกิต		
คณะผู้จัดทำ	: นางสาวศุภนิตา	เปี่ยมล้าย	6105100001
	: นางสาวนภัสกร	นำพรชวาลรุ่ง	6105100005
	: นางสาวลลิตา	เกิดสันเทียะ	6105100008
อาจารย์ที่ปรึกษา	: อาจารย์ณรงค์ฤทธิ์ สุคนธสิงห์		
ระดับการศึกษา	: วิทยาศาสตร์บัณฑิต		
ภาควิชา	: เทคโนโลยีสารสนเทศ		
คณะ	: เทคโนโลยีสารสนเทศ		
ปีการศึกษา	: 2566		

บทคัดย่อ

งานวิจัยนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูญบุหรีของวัยรุ่นโดยใช้เทคนิคเหมืองข้อมูล 3 วิธี คือ J48, Naïve Bayes และ โครงข่ายประสาทเทียม ในการสร้างแบบจำลอง ซึ่งใช้ข้อมูลการตอบแบบสอบถามของนักศึกษา มหาวิทยาลัยสยาม จำนวน 543 คน และนำมาเปรียบเทียบประสิทธิภาพของโมเดลการจำแนกหาตัวแบบที่เหมาะสมเพื่อใช้ทำนายการสูญบุหรีของวัยรุ่น โดยแบ่งข้อมูลออกเป็น 3 ชุด วิเคราะห์ข้อมูลบนพื้นฐานของวิธี 10-fold cross validation เพื่อใช้ในการทดสอบประสิทธิภาพของโมเดล โดยใช้โปรแกรม Weka นำข้อมูลเข้าโมเดล ในการสร้างแบบจำลอง ชุดข้อมูลที่ 1 จำนวน 22 Attribute พบว่า ประสิทธิภาพทำนายวิธี J48 มีค่าความถูกต้องมากที่สุดคือ 81.95% รองลงมาวิธี Naïve Bayes มีค่าความถูกต้องเท่ากับ 78.82% ลำดับสุดท้ายวิธี โครงข่ายประสาทเทียมมีค่าความถูกต้องเท่ากับ 78.45% , ชุดข้อมูลที่ 2 จำนวน 32 Attribute พบว่า ประสิทธิภาพทำนายวิธี J48 มีค่าความถูกต้อง (Accuracy) มากที่สุดคือ 80.85% รองลงมาวิธี โครงข่ายประสาทเทียมมีค่าความถูกต้องเท่ากับ 80.11% ลำดับสุดท้ายวิธี Naïve Bayes มีค่าความถูกต้องเท่ากับ 77.90% และชุดข้อมูลที่ 3 จำนวน 47 Attribute พบว่า ประสิทธิภาพทำนายวิธี โครงข่ายประสาทเทียม มีค่าความถูกต้องมากที่สุดคือ 81.40% รองลงมาวิธี J48 มีค่าความถูกต้องเท่ากับ 80.48

% ลำดับสุดท้ายวิธี Naïve Bayes มีค่าความถูกต้องเท่ากับ 77.72% จากการเปรียบเทียบค่าความถูกต้องของการใช้ข้อมูลชุดที่ 1, ข้อมูลชุดที่ 2 และ ข้อมูลชุดที่ 3 ทำนายผลการสูบบุหรี่ของวัยรุ่น พบว่า โมเดลทำนายของชุดข้อมูลทั้ง 3 ชุด มีค่าความถูกต้อง (Accuracy) เพิ่มขึ้นและลดลงแตกต่างกัน พิจารณารายวิธีพบว่า วิธี Naïve Bayes ลดลง 0.92% และ 0.18% , J48 ลดลง 1.1% และ 0.37% , โครงข่ายประสาทเทียม เพิ่มขึ้น 1.66% และ 1.29% สรุปได้ว่า โมเดลการทำนายผลการสูบบุหรี่ของวัยรุ่นด้วยวิธี J48 โดยใช้ข้อมูลชุดที่ 1 มีค่าความถูกต้องมากที่สุดคือ 81.95%

คำสำคัญ : การจำแนกข้อมูล, ปัจจัยการสูบบุหรี่, เหมืองข้อมูล



Project Title : The Comparing of Classification Efficiency for Predicting Adolescent Smoking Behavior by Data Mining Techniques

Project Credits : 3 Credits

By : Ms. Supanida Pearmalai 6105100001
 : Ms. Napatsakorn Numpornchawanrung 6105100005
 : Ms. Lalita Kerdsantia 6105100008

Advisor : Mr. Narongrit Sukonthasing

Degree : Bachelor of Science

Major : Information Technology

Faculty : Information Technology

Academic Year : 2023

Abstract

Objective of our research is comparing the classification performance of predicting adolescent smoking behavior using data mining techniques by J48, Naïve Bayes, and Artificial Neural Networks to produces research models. Datasets of 543 Siam University students who answers questionnaire was analysed for the classification efficiency for predicting models comparision. The data was divided into three sets, each sets were analyzed with 10-fold cross-validation method to examine the model's efficiency. All data was manage with Weka program. For the first dataset, with 22 attributes, the J48 method showed the highest accuracy at 81.95%, followed by Naïve Bayes, with 78.82%, and the Artificial Neural Network, with 78.45%. The second dataset included 32 attributes, J48 had the highest accuracy at 80.85%, followed by the Artificial Neural Network with 80.11%, and Naïve Bayes with 77.90%. The third dataset containing 47 attributes, the Artificial Neural Network had the highest accuracy at 81.40%, followed by J48 with 80.48%, and Naïve Bayes with 77.72%. Comparing the accuracy values for predicting adolescent smoking behavior across the three datasets, it was observed that the models' performance varied. Naïve Bayes decreased by 0.92% and 0.18%, J48 decreased by 1.1% and 0.37%, while the Artificial Neural Network increased by 1.66% and

1.29%. In summary, the J48 model using the first dataset achieved the highest accuracy of 81.95% in predicting adolescent smoking behavior.

Keywords: classification, data mining, smoking factor



Approved by

.....

กิตติกรรมประกาศ

งานวิจัยเรื่องการเปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูญพันธุ์ของวัยรุ่น ด้วยเทคนิคเหมืองข้อมูลนี้สามารถดำเนินได้สำเร็จเรียบร้อย เนื่องจากได้รับความกรุณาให้คำแนะนำ และการช่วยเหลือจากคณะอาจารย์ในภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยสยาม ให้ผู้วิจัยเกิดพลังใจสามารถดำเนินการวิจัยได้ลุล่วงตามวัตถุประสงค์

ขอขอบพระคุณ อาจารย์ณรงค์ฤทธิ์ สุคนธสิงห์ ที่ให้ความรู้ คำแนะนำ และความช่วยเหลือ ในการจัดทำงานวิจัยฉบับนี้จนเสร็จสมบูรณ์ไปด้วยดี

ขอขอบคุณนักศึกษา มหาวิทยาลัยสยาม ที่ให้ความร่วมมือในการตอบแบบสอบถามเพื่อเก็บ ข้อมูลในการทำงานวิจัยในครั้งนี้



คณะผู้จัดทำ

สารบัญ

หน้า

บทคัดย่อ.....	๗
กิตติกรรมประกาศ.....	๗
สารบัญ	๗
สารบัญรูปภาพ	๘
สารบัญตาราง	๘
บทที่ 1 บทนำ	
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์.....	2
1.3 ขอบเขตของการวิจัย.....	2
1.4 ขั้นตอนการดำเนินงาน	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ	4
บทที่ 2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	
2.1.ทฤษฎีการทำเหมืองข้อมูล.....	5
2.2 แนวความคิดที่เกี่ยวข้องกับการสูบบุหรี่	19
2.3 พฤติกรรมการสูบบุหรี่	21
2.4 กฎหมายที่เกี่ยวข้องกับบุหรี่	22
2.5 งานวิจัยที่เกี่ยวข้อง.....	23

สารบัญ(ต่อ)

หน้า

บทที่ 3 วิธีการดำเนินงานวิจัย

3.1 การศึกษาข้อมูลที่เกี่ยวข้อง	33
3.2 รวบรวม คัดเลือกและจัดเตรียมข้อมูล	34
3.3 เครื่องมือที่ใช้ในการเก็บข้อมูล	37
3.4 ทดสอบอัลกอริทึม	53

บทที่ 4 ผลของการวิจัย

4.1 ผลการวิเคราะห์ข้อมูล	63
4.2 ผลการเปรียบเทียบประสิทธิภาพของข้อมูล	76
4.3 สรุปกฎของอัลกอริทึม J48	81

บทที่ 5 สรุปผลและข้อเสนอแนะ

5.1 สรุปผลการวิจัย	87
5.2 ข้อเสนอแนะ	88
บรรณานุกรม	89
ภาคผนวก	
ภาคผนวก ก แบบสอบถามงานวิจัย	94
ประวัติคณะผู้จัดทำ	100

สารบัญรูปภาพ

ภาพที่	หน้า
2.1 กระบวนการทำเหมืองข้อมูลของ CRISP-DM.....	8
2.2 การตัดสินใจเลือกว่าจะอยู่บ้านหรือออกไปข้างนอก	12
2.3 ส่วนประกอบของสมการ	13
2.4 การทำนายระหว่างบุคคล	14
2.5 ส่วนประกอบของสมการ	15
2.6 ส่วนประกอบของโครงข่ายประสาทเทียม.....	17
2.7 CONFUSION MATRIX.....	18
3.1 หน้าจอแสดงแท็บเมนู	34
3.2 ข้อมูลที่ยังไม่ผ่านการ DATA CLEANING.....	35
3.3 ข้อมูลการสุ่มข้อมูลที่ผ่านการ DATA CLEANING แล้วแปลงเป็นไฟล์ .CSV.....	36
3.4 หน้าจอแสดงการนำข้อมูลเข้า WEKA.....	51
3.5 หน้าจอแสดงการประเภทข้อมูล	52
3.6 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 1 โดยใช้อัลกอริทึม NAÏVE BAYES.....	54
3.7 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 2 โดยใช้อัลกอริทึม NAÏVE BAYES.....	55
3.8 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 3 โดยใช้อัลกอริทึม NAÏVE BAYES.....	56
3.9 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 1 โดยใช้อัลกอริทึม J48.....	57
3.10 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 2 โดยใช้อัลกอริทึม J48.....	58
3.11 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 3 โดยใช้อัลกอริทึม J48.....	59
3.12 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 1 โดยใช้อัลกอริทึม โครงข่ายประสาทเทียม	60
3.13 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 2 โดยใช้อัลกอริทึม โครงข่ายประสาทเทียม	61
3.14 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 3 โดยใช้อัลกอริทึม โครงข่ายประสาทเทียม	62

สารบัญรูปภาพ (ต่อ)

ภาพที่	หน้า
4.1 การเปรียบเทียบประสิทธิภาพวิธี NAÏVE BAYES, J48 และ โครงข่ายประสาทเทียม ชุดข้อมูลที่ 1 จำนวน 22 ATTRIBUTE	76
4.2 การเปรียบเทียบประสิทธิภาพวิธี NAÏVE BAYES, J48 และ โครงข่ายประสาทเทียม ชุดข้อมูลที่ 2 จำนวน 32 ATTRIBUTE	78
4.3 การเปรียบเทียบประสิทธิภาพวิธี NAÏVE BAYES, J48 และ โครงข่ายประสาทเทียม ชุดข้อมูลที่ 3 จำนวน 47 ATTRIBUTE	79
4.4 การเปรียบเทียบค่าความถูกต้อง (ACCURACY) วิธี NAÏVE BAYES, J48 และ โครงข่าย ประสาทเทียม ชุดข้อมูลที่ 1, ชุดข้อมูลที่ 2 และ ชุดข้อมูลที่ 3	80
4.5 แสดงแผนภาพการจำแนกข้อมูล VISUALIZER TREE	86

สารบัญตาราง

ตารางที่	หน้า
2.1 ตารางเปรียบเทียบงานวิจัย.....	29
3.1 ข้อมูลทั่วไป ตัวแปรสำหรับการประมวลผลข้อมูลด้วยเทคนิคเหมืองข้อมูล	38
3.2 โทษและพิษภัยของบุนหรี.....	45
3.3 เจตคติและค่านิยมต่อการสูบบุหรี.....	47
4.1 จำนวนและร้อยละของคุณลักษณะของกลุ่มตัวอย่าง	64
4.2 การวิเคราะห์ความคิดเห็นเกี่ยวกับพิษภัยและโทษของบุนหรี	72
4.3 เจตคติต่อการสูบบุหรี	74
4.4 การเปรียบเทียบประสิทธิภาพความถูกต้อง (ACCURACY) ค่าความแม่นยำ (PRECISION) เฉลี่ย และ ค่าความระลึก (Recall) เฉลี่ย ชุดข้อมูลที่ 1 จำนวน 22 Attribute.....	76
4.5 การเปรียบเทียบประสิทธิภาพความถูกต้อง (ACCURACY) ค่าความแม่นยำ (PRECISION) เฉลี่ย และ ค่าความระลึก (RECALL) เฉลี่ย ชุดข้อมูลที่ 2 จำนวน 32 ATTRIBUTE.....	77
4.6 การเปรียบเทียบประสิทธิภาพความถูกต้อง (ACCURACY) ค่าความแม่นยำ (PRECISION) เฉลี่ย และ ค่าความระลึก (RECALL) เฉลี่ย ชุดข้อมูลที่ 3 จำนวน 47 ATTRIBUTE.....	79
4.7 ค่าความถูกต้อง (ACCURACY) การวิเคราะห์ข้อมูลชุดที่ 1, ข้อมูลชุดที่ 2 และ ข้อมูลชุดที่ 3....	80

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูลของการวิเคราะห์ ซึ่งช่วยพัฒนารูปแบบที่สามารถเผยแพร่เชื่อมต่อกายในข้อมูลนับล้าน หรือ พันล้านรายการและการค้นหาข้อมูลที่มีประโยชน์จากแหล่งข้อมูลที่มีเป็นจำนวนมากมายมหาศาล เพื่อดึงข้อมูลที่มีประโยชน์มาทำการวิเคราะห์ค้นหาหรือความสัมพันธ์ที่เกิดขึ้นในฐานข้อมูล และ จัดทำเป็นสารสนเทศเพื่อใช้ในการวางแผนบริหารจัดการธุรกิจ โดยการแยกข้อมูลที่มีประโยชน์ ออกมาใช้งาน ปัจจุบันคนที่สูบบุหรี่ในช่วงวัยรุ่นมีจำนวนมากมีแนวโน้มในการสูบบุหรี่เพิ่มขึ้น โดย เกิดจากหลายปัจจัยไม่ว่าจะเป็นความอยากรู้อยากลอง, ตามอย่างคนในครอบครัว และการเข้าถึงคม เป็นต้น “วัยรุ่นที่สูบบุหรี่ร้อยละ 38.4 เริ่มสูบบุหรี่เพราะ ความอยากรองตามอย่างเพื่อน วัยรุ่นที่สูบบุหรี่ร้อยละ 35.9 เริ่มสูบบุหรี่เพราะเพื่อนชวน” (สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ, 2561) และมีความเสี่ยงที่จะเป็นโรคอื่น ๆ เช่น “โรคมะเร็ง, โรคระบบทางเดินหายใจ, โรคระบบเส้นเลือด, หัวใจอักเสบ, โรคคลื่นหัวใจขาดเลือด, ความดันเลือดสูง, โรคหัวใจที่เกิดจากโรคปอด และโรค หลอดเลือดแดงใหญ่โป่งพอง เป็นต้น” (สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ, 2561) นอกจากการสูบบุหรี่จะมีโทษต่อผู้สูบแล้วควันบุหรี่ยังเป็นโทษต่อผู้ที่อยู่ใกล้ชิด หรือผู้ที่อยู่รอบข้าง อีกด้วย เพราะในควันบุหรี่จะมีสารเคมีอันตราย 2 ชนิด ได้แก่ แคมเฟียมและตะกั่ว นอกจากนี้สารเคมี อีกหลายชนิดถูกสร้างขึ้นระหว่างกระบวนการผลิต และเมื่อจุดบุหรี่การเผาไหม้ทำให้สารเคมีชนิด ใหม่เกิดขึ้นอีกภายในควันบุหรีมีสารเคมีหลายชนิด เช่น นิโคติน, ทาร์, คาร์บอนมอนอกไซด์, ไฮโดรเจนไซยาไนด์, ไนโตรเจนไดออกไซด์, และแอมโมเนีย เป็นต้น ซึ่งเมื่อสูดดมควันบุหรีเป็นประจำอาจเสี่ยงต่อโรคต่าง ๆ ควันบุหรีที่ออกมาทางปลายมวนที่ติดไฟไม่ผ่านการกรองส่วนนี้เองมี สารพิษค่อนข้างมาก ควันบุหรีเกิดได้จากสองแหล่งด้วยกัน คือ ควันบุหรีที่ผู้สูบพ่นออกมา และควันที่ ลอยจากตอนปลายมวนบุหรี ควันบุหรีเหล่านี้อันตราย ถึงแม้ไม่ได้เป็นผู้สูบโดยตรงก็ตาม แต่ก็จะทำให้ เกิดความระคายเคืองต่อ ตา จมูกคอ ส่งผลให้เกิดการคลื่นไส้ ปวดศีรษะ รวมถึงผู้ที่มิโรคประจำตัว เช่น หอบ ภูมิแพ้ อาจทำให้กำเริบได้

ปัญหาการสูบบุหรี่ในสถานศึกษา ทุกวันนี้การสูบบุหรี่ยังแพร่หลายในกลุ่มวัยรุ่น ถึงแม้ทราบกันดีถึงผลเสียของบุหรี่ รัฐบาลเองมีมาตรการป้องกัน และรณรงค์ เช่น การออกกฎหมายไม่ให้โชว์บุหรี่หน้าร้าน ช้างซองบุหรี่มีค่าเตือน และรูปที่น่ากลัวเกี่ยวกับโรคที่เกิดจากบุหรี่ซึ่งช่วยได้ในระดับหนึ่ง แต่ก็ยังมีคนที่ไม่กลัวและมีจำนวนวัยรุ่นที่สูบบุหรี่เพิ่มขึ้นเรื่อย ๆ ไม่ว่าจะเป็บบุหรี่แบบซอง หรือแบบบุหรี่ไฟฟ้าและการแก้ปัญหาซึ่งอาจจะไม่ได้ทำได้ภายในครั้งเดียว เพราะว่าเป็นสิ่งที่เกิดขึ้นมานานแล้วอาจจะต้องค่อย ๆ ปรับเปลี่ยนทัศนคติและมุมมองของผู้ที่สูบบุหรี่ให้รู้จักรักตัวเอง และเกิดความสำนึกที่ดีอาจจะติดป้ายรณรงค์ห้ามสูบบุหรี่ในที่สาธารณะและจัดกิจกรรม หรือชมรม ที่แสดงให้เห็นถึงโทษและพิษภัยอันร้ายแรงของบุหรี่

คณะผู้วิจัยได้ทำการรวบรวมข้อมูล และนำข้อมูลดังกล่าวมาทำการเปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูบบุหรี่ของวัยรุ่นด้วยเทคนิคเหมืองข้อมูล (Data Mining) เพื่อหาสาเหตุที่วัยรุ่นจะมีพฤติกรรมการสูบบุหรี่ ด้วยวิธีการทำแบบสอบถามเพื่อสำรวจพฤติกรรมความเสี่ยงของวัยรุ่นในจังหวัดกรุงเทพมหานครฯ ว่ามีแนวโน้มมากน้อยแค่ไหนที่วัยรุ่นในกรุงเทพมหานครฯ จะสูบบุหรี่หรือไม่สูบบุหรี่ โดยใช้เทคนิคเหมืองข้อมูลในการพยากรณ์ แต่ละเทคนิคที่ใช้จะนำมาเปรียบเทียบหาว่าเทคนิคใดมีประสิทธิภาพ น่าเชื่อถือได้มากที่สุด เพื่อนำมาประมวลผลว่าคนไหนมีแนวโน้มที่จะสูบบุหรี่ หรือไม่สูบบุหรี่

1.2 วัตถุประสงค์

1. เพื่อเปรียบเทียบประสิทธิภาพในการสร้างแบบจำลองในการทำนายข้อมูล โดยใช้เทคนิคเหมืองข้อมูล 3 วิธีคือ J48, Naïve Bayes และ โครงข่ายประสาทเทียม

1.3 ขอบเขตของการวิจัย

1. ขอบเขตด้านประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการวิจัยเพื่อทำการศึกษาค้นคว้าสำหรับการวิจัยครั้งนี้คือนักศึกษามหาวิทยาลัยสยาม พบว่ามีจำนวนมากกว่า 12,000 คน กลุ่มตัวอย่างที่มาทำการศึกษาวิจัย โดยทำการเก็บข้อมูลจากการตอบแบบสอบถามเกี่ยวกับพฤติกรรมการสูบบุหรี่ของวัยรุ่น จำนวน 543 คน

2. ขอบเขตการศึกษา

ผู้วิจัยดำเนินการศึกษาใช้วิธีการศึกษาค้นคว้า โดยใช้เทคนิคของเหมืองข้อมูลอัลกอริทึมต้นไม้ตัดสินใจ (J48) วิธีการเรียนรู้แบบอย่างง่าย (Naïve Bayes) และ โครงข่ายประสาทเทียม (Multilayer Perceptron)

3. ขอบเขตด้านระยะเวลา

มีระยะเวลาในการศึกษาวิจัยระหว่างเดือนมกราคม 2565 ถึงเดือน เมษายน 2565 เป็นระยะเวลา 3 เดือน

1.4 ขั้นตอนการดำเนินงาน

ขั้นตอนการทำเหมืองข้อมูล (Process of Data Mining)

ขั้นตอนในการวิเคราะห์ข้อมูลได้ใช้หลักการของกระบวนการหาความรู้แบบ Cross Industry Standard Process for Data Mining (CRISP-DM) ซึ่งเป็นแนวทางในการดำเนินงาน CRISP-DM ประกอบไปด้วย 6 ขั้นตอน

1. ทำความเข้าใจปัญหา (Business Understanding)

ศึกษาปัญหาที่พบในด้านการสูบบุหรี่ของเด็กวัยรุ่น เป็นปัญหาซึ่งองค์กรกำลังประสบและมีแนวโน้มว่าจะยิ่งทวีความรุนแรงในอนาคต งานวิจัยนี้จึงต้องการหาปัจจัยที่เป็นสาเหตุให้เด็กวัยรุ่นสูบบุหรี่ เพื่อนำมาวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล การจำแนกประเภทข้อมูล (Classification) และนำข้อมูลผลลัพธ์ที่ได้มาสรุปผล

2. ทำความเข้าใจข้อมูล (Data Understanding)

2.1 การเก็บรวบรวมข้อมูลขั้นต้น (Collect Initial Data) ข้อมูลที่นำมาใช้วิเคราะห์หาพฤติกรรมการสูบบุหรี่ของเด็กวัยรุ่นเป็นข้อมูลชั้นปฐมภูมิ คือ ข้อมูลจากการสำรวจการทำแบบสอบถาม ที่ได้ทำการเก็บข้อมูลจากเด็กวัยรุ่น ยกตัวอย่าง เช่น นักเรียน นักศึกษา แล้วนำข้อมูลพฤติกรรมการสูบบุหรี่ของเด็กวัยรุ่นที่ได้มาวิเคราะห์ในเบื้องต้น

2.2 การอธิบายข้อมูล (Data Describe) จากขั้นตอนการเก็บรวบรวมข้อมูลได้เพิ่มข้อมูลที่เกี่ยวข้อง ข้อมูลที่ใช้ในแบบจำลองจะใช้ข้อมูลพฤติกรรมการสูบบุหรี่ของเด็กวัยรุ่น

3. การเตรียมข้อมูล (Data Preparation)

3.1 การคัดเลือกข้อมูล (Data Select) ในขั้นตอนการเตรียมข้อมูลที่จะนำไปใช้ในการวิเคราะห์ เป็นการนำข้อมูลจากแบบสอบถามในการเก็บข้อมูล เพื่อทำการคัดเลือกข้อมูลที่ตรงตามเป้าหมายของงานวิจัย เช่น เพศ, อายุ, ระดับการศึกษา, พฤติกรรมการสูบบุหรี่, รายได้ เป็นต้น หรือเลือกตัดข้อมูลที่ไม่น่าจะเป็นต่อการวิเคราะห์ เพื่อลดขนาด ของข้อมูล

3.2 ทำความสะอาดข้อมูล (Data Clean) หลังจากการคัดกรองข้อมูลเป็นที่เรียบร้อยแล้ว ผู้วิจัยทำความสะอาดข้อมูลที่มีความซ้ำซ้อนหรือ ผิดปกติของข้อมูล เพื่อสร้างคุณภาพให้แก่ข้อมูลที่นำมาใช้อย่างแม่นยำ

3.3 การเปลี่ยนรูปแบบของข้อมูล (Data Format) เพื่อให้ข้อมูลเหมาะสมกับโปรแกรมที่นำไปใช้ และง่ายต่อการวิเคราะห์ จากข้อมูลพฤติกรรมการสูบบุหรี่ของเด็กวัยรุ่นที่มีลักษณะการเก็บต่างกัน นำมาแปลงเป็นข้อมูลที่เหมือนกัน

4. การสร้างตัวแบบพยากรณ์ (Modeling)

สำหรับงานวิจัยนี้จะกล่าวถึงเทคนิคการจำแนกประเภทข้อมูล (Classification) เพื่อจะนำไปวิเคราะห์พฤติกรรมการสูบบุหรี่ของเด็กวัยรุ่น ซึ่งใช้อัลกอริทึม J48, Naive Bayes และ โครงข่ายประสาทเทียม ทำการวิเคราะห์หาผลลัพธ์ว่าวัยรุ่นคนไหนจะสูบบุหรี่ หรือไม่สูบบุหรี่

5. การประเมินประสิทธิภาพ (Evaluation)

การประเมินประสิทธิภาพของโมเดลที่ได้นำมาเปรียบเทียบกับว่าโมเดลไหนให้ประสิทธิภาพและความแม่นยำสูงมากกว่ากัน และนำโมเดลนั้นไปใช้งาน

6. การนำไปใช้งาน (Deployment)

เป็นการนำโมเดลไปสร้างเป็นโปรแกรม เพื่อนำไปวิเคราะห์ว่าเด็กคนไหนมีแนวโน้มว่าจะสูบบุหรี่หรือไม่สูบบุหรี่ ซึ่งทางองค์กร หรือโรงเรียนสามารถนำโมเดลไปใช้เพื่อกำหนดแผนกลยุทธ์ เพื่อไม่ให้เด็กที่มีแนวโน้มจะสูบบุหรี่ ไม่สูบบุหรี่ได้และสามารถนำไปพัฒนาต่อได้

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ผลการประเมินประสิทธิภาพตัวแบบการจำแนกความเป็นไปได้ในการสูบบุหรี่
2. ทราบถึงปัจจัยที่มีความสัมพันธ์กับพฤติกรรมการสูบบุหรี่
3. สามารถนำโมเดลไปใช้ในการพัฒนาต่อยอดให้เป็นแอปพลิเคชันได้

บทที่ 2

แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การศึกษาวิจัยครั้งนี้ผู้วิจัยได้ทำการศึกษาเกี่ยวกับพฤติกรรมการสูบบุหรี่ เพื่อใช้เปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูบบุหรี่ของวัยรุ่นด้วยเทคนิคเหมืองข้อมูล แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง เพื่อเป็นแนวทางในการศึกษา โดยแยกเป็นประเด็นดังนี้

- 2.1 ทฤษฎีการทำเหมืองข้อมูล (Data Mining)
- 2.2 แนวความคิดที่เกี่ยวข้องกับการสูบบุหรี่
- 2.3 พฤติกรรมการสูบบุหรี่
- 2.4 กฎหมายที่เกี่ยวข้องกับการสูบบุหรี่
- 2.5 งานวิจัยที่เกี่ยวข้อง

2.1.ทฤษฎีการทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) หมายถึง กระบวนการสืบค้นความรู้ที่เป็นประโยชน์บนฐานข้อมูลขนาดใหญ่ (Knowledge Discovery from very large Databases : KDD) เป็นเทคนิคที่ใช้จัดการกับข้อมูลขนาดใหญ่ โดยนำข้อมูลที่สำคัญจากชุดข้อมูลที่มีอยู่และใช้เครื่องมือทางคอมพิวเตอร์ วิเคราะห์และสกัดข้อมูลที่มีความหมายออกมา เพื่อใช้ในการวิเคราะห์ หรือทำนายสิ่งต่าง ๆ ที่จะเกิดขึ้น ซึ่งการค้นหาคำรู้ ความจริงที่อยู่ในข้อมูล (Knowledge Discovery) เป็นกระบวนการขุดค้นข้อมูลที่น่าสนใจในกองข้อมูลที่มีอยู่

Data Mining มีความแตกต่างกันใน 2 มุม คือ มุมมองทางวิชาการและมุมมองเชิงธุรกิจ มุมมองเชิงวิชาการนั้น นักวิจัยจะอ้างถึงกระบวนการ “Knowledge discovery in database (KDD)” เพื่อบ่งชี้ว่า Data Mining เป็นขั้นตอนหนึ่งของกระบวนการที่เกี่ยวข้องกับการค้นหารูปแบบและความสัมพันธ์ของข้อมูลนั้น และในมุมมองเชิงธุรกิจ จะใช้คำว่า “Data Mining” แทนความหมายขั้นตอนทั้งหมด

2.1.1 วิวัฒนาการของการทำเหมืองข้อมูล

1. ปี 1960 Data Collection คือ การนำข้อมูลมาจัดเก็บในอุปกรณ์การเก็บข้อมูลที่เหมาะสม น่าเชื่อถือ และป้องกันการสูญหายได้เป็นอย่างดี

2. ปี 1980 Data Access คือ การเข้าถึงข้อมูลที่จัดเก็บและสร้างความสัมพันธ์ในข้อมูลเพื่อให้มีประโยชน์ในการนำข้อมูลไปวิเคราะห์และตัดสินใจอย่างมีคุณภาพ และการเข้าถึงข้อมูลมีความสำคัญในการนำมาใช้ในการตัดสินใจและการวางแผน

3. ปี 1990 Data Warehouse & Decision Support คือ การรวบรวมข้อมูลจากแหล่งต่าง ๆ และจัดเก็บไว้ในฐานข้อมูลขนาดใหญ่เพื่อครอบคลุมทุกแง่มุมขององค์กร และช่วยสนับสนุนการตัดสินใจในระดับองค์กร

4. ปี 2000 Data Mining เป็นการนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผล โดยการสร้างแบบจำลองและความสัมพันธ์ ทางสถิติ เพื่อค้นหาความรู้และข้อมูลที่มีค่าจากข้อมูลที่มีอยู่ (ศิริพร ครุบา, 2553)

2.1.2 กระบวนการทำเหมืองข้อมูล

การทำเหมืองข้อมูลในปัจจุบันใช้แนวคิดที่มาจากแบบจำลองการทำเหมืองข้อมูลของคริสตีดีเอ็ม (CRISP-DM Model) เพื่อสกัดความรู้หรือสร้างกฎความสัมพันธ์ระหว่างองค์ประกอบที่เกี่ยวข้องหรือมีส่วนเกี่ยวข้องกัน และนำความรู้ที่ได้ไปใช้ประโยชน์ ประกอบด้วยขั้นตอนต่าง ๆ ดังนี้

2.1.2.1 การเข้าใจธุรกิจ (Business Understanding)

เป็นขั้นตอนแรกในกระบวนการ CRISP-DM ซึ่งเป็นการเน้นไปที่ความเข้าใจถึงปัญหาและการแปลงปัญหาให้อยู่ในรูปของโจทย์ที่เหมาะสมสำหรับการวิเคราะห์ข้อมูล

2.1.2.2 การค้นคว้าข้อมูล (Data Understanding)

เริ่มจากการเก็บรวบรวมข้อมูลที่เป็นไปได้สำหรับโครงการ หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้รับเพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกใช้เฉพาะข้อมูลบางส่วนในการวิเคราะห์

2.1.2.3 การเตรียมข้อมูล (Data Preparation)

เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้รับมาจากการเก็บรวบรวม (raw data) เพื่อให้กลายเป็นข้อมูลที่เหมาะสมและพร้อมนำไปใช้ในการวิเคราะห์ในขั้นตอนถัดไป ขั้นตอนนี้อาจรวมถึงการทำความสะอาดข้อมูล (data cleaning) เพื่อแก้ไขข้อมูลที่ไม่ถูกต้องหรือข้อมูลที่ขาดหายไป และปรับรูปแบบข้อมูลให้อยู่ในรูปที่เหมาะสมสำหรับการวิเคราะห์ เช่น การแปลงข้อมูลให้อยู่ในช่วง (scale)

เดียวกัน ซึ่งขั้นตอนนี้ใช้เวลามากที่สุดในการบวนการ CRISP-DM เนื่องจากมักต้องใช้ในการปรับปรุงและการแปลงข้อมูลเพื่อให้ข้อมูลเป็นไปตามความต้องการของการวิเคราะห์ข้อมูล สามารถแบ่งออกได้เป็น 3 ขั้นตอน คือ

- การคัดเลือกข้อมูล (Data Selection)
- การกลั่นกรองข้อมูล (Data Cleaning)
- การแปลงรูปข้อมูล (Data Transformation)

2.1.2.4 การสร้างโมเดล (Modeling)

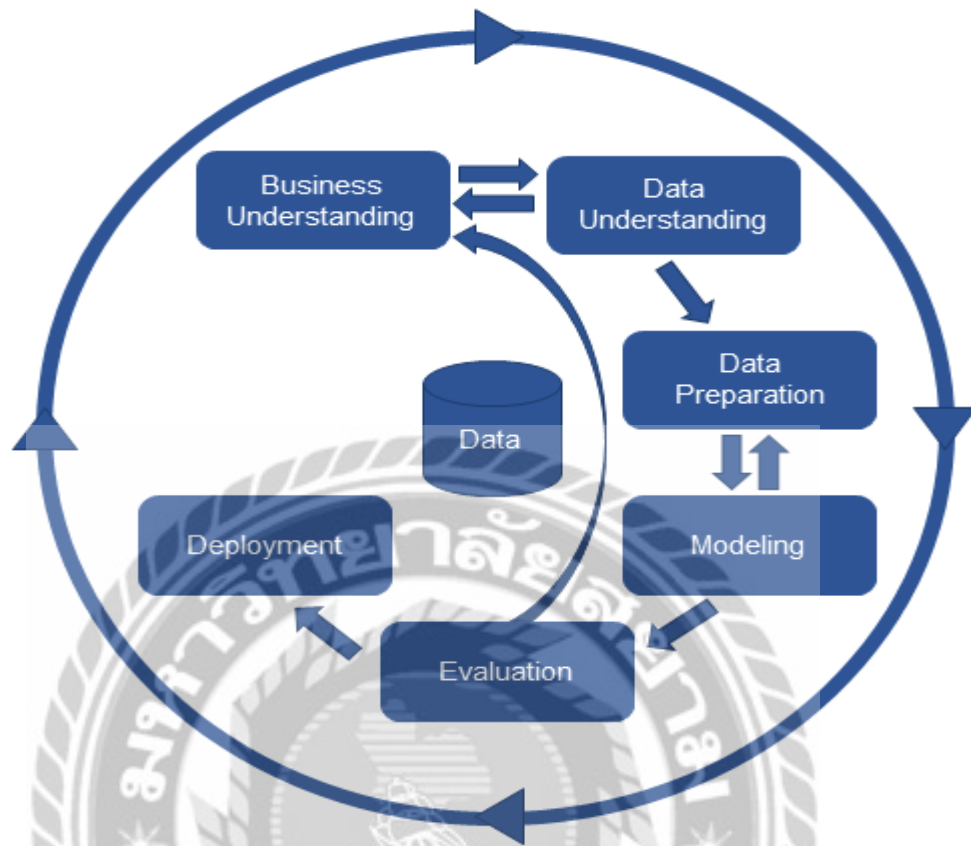
เป็นขั้นตอนที่ใช้เทคนิคทาง Data Mining ในการวิเคราะห์ข้อมูล โดยรวมการจำแนกประเภทข้อมูลหรือการแบ่งกลุ่มข้อมูล เพื่อสร้างโมเดลทางสถิติหรือ โมเดลทางเครื่องมือ (machine learning models) เพื่อการพยากรณ์ การจำแนก หรือการแสดงความสัมพันธ์ของข้อมูลในลักษณะที่เหมาะสมกับความต้องการของโครงการในขั้นตอนนี้ อาจจะต้องมีการย้อนกลับไปขั้นตอนนี้ "Data Preparation" เพื่อปรับปรุงหรือแปลงข้อมูลให้เหมาะสมกับเทคนิคการวิเคราะห์ข้อมูลที่ถูกเลือกใช้ เพื่อให้สามารถสร้างโมเดลที่มีประสิทธิภาพและความแม่นยำสูงสุด

2.1.2.5 การประเมินผล (Evaluation)

เป็นขั้นตอนที่ตรวจสอบประสิทธิภาพของโมเดล เพื่อให้ตรงกับวัตถุประสงค์ที่ต้องการให้โมเดลทำงานตามในขั้นตอนนี้จะวัดว่าโมเดลที่สร้างมีความแม่นยำและน่าเชื่อถือมากน้อยเพียงใด โดยใช้ตัวชี้วัดทางสถิติและเทคนิคทาง Data mining เช่น Precision, Recall, Accuracy, F-Measure และอื่น ๆ

2.1.2.6 การใช้งาน (Deployment)

ในกระบวนการ CRISP-DM ไม่ได้หยุดเพียงแค่การทำวิเคราะห์ข้อมูลด้วยเทคนิคทาง Data Mining เป็นขั้นตอนที่น่าผลลัพธ์และองค์ความรู้ที่ได้จากการวิเคราะห์ไปใช้งานจริงในองค์กรหรือบริษัท การนำองค์ความรู้ไปใช้งานสามารถทำได้ในหลายรูปแบบ เช่น การสร้างรายงานหรือแผนการดำเนินงานเพื่อให้ผู้บริหาร ตัดสินใจเข้าใจข้อมูลและตัดสินใจการสร้างแอปพลิเคชันหรือเครื่องมือที่ใช้ในการเลือกประเภทของผู้บริโภคและลูกค้าที่เหมาะสมกับสินค้าหรือบริการ หรือการนำความรู้ไปใช้ในการตรวจจับปัญหาหรือการทำนายเหตุการณ์ในอนาคต (บริษัท ทีวีบี อนาคตโลกดิจิทัล จำกัด, 2563)



ภาพที่ 2.1 กระบวนการทำเหมืองข้อมูลของ CRISP-DM

ที่มา: https://www.researchgate.net/figure/CRISP-DM-Model-Taylor-2017_fig1_326235288

การทำเหมืองข้อมูล เป็นขบวนการที่มีขั้นตอนต่าง ๆ ที่ถูกดำเนินขึ้นเพื่อให้ได้ผลลัพธ์ที่เป็นความรู้ที่เกี่ยวกับข้อมูลที่ต้องการและสามารถนำไปใช้ในการตัดสินใจและดำเนินงานได้โดยไม่เกิดความผิดพลาดหรือสร้างความเสียหายจากการนำไปใช้งาน ขบวนการทำเหมืองข้อมูลประกอบด้วยขั้นตอนต่าง ๆ ที่คล้ายกับการพัฒนาซอฟต์แวร์หรือการดำเนินโครงการทางเทคโนโลยีสารสนเทศ โดยมีการกำหนดขั้นตอนการทำเหมืองข้อมูลให้เป็นมาตรฐานเพื่อให้สามารถนำไปปฏิบัติ ทำให้การทำเหมืองข้อมูลมีประสิทธิภาพและได้ผลลัพธ์ที่มีความถูกต้อง มาตรฐานนี้เรียกว่า Cross-Industry Standard Process for Data Mining (CRISP-DM)

2.1.3 ประเภทการทำเหมืองข้อมูล

ประเภทของการทำเหมืองข้อมูล สามารถแบ่งออกได้เป็น 2 ประเภทใหญ่ๆ ดังนี้

2.1.3.1 การทำเหมืองข้อมูลแบบทำนาย (Predictive Mining) หรือการเรียนรู้แบบมีผู้สอน (Supervised learning) เป็นกระบวนการที่ใช้ข้อมูลฝึกหัด (training data) เพื่อสร้างตัวแบบ (model) ที่สามารถทำนายหรือจำแนกประเภทตัวอย่างในอนาคต โดยอัลกอริทึมประเภทนี้มุ่งเน้นการแบ่งแยกข้อมูลตามค่าคุณสมบัติของข้อมูลว่าการจำแนกประเภท (Classification) ถูกใช้เมื่อต้องการทำนายประเภทหรือกลุ่มของข้อมูล การถดถอย (Regression) ใช้เมื่อต้องการทำนายค่าต่อเนื่อง และการพยากรณ์ (Forecasting) ใช้ในการทำนายค่าในอนาคต ตัวแบบที่สร้างขึ้นจะใช้สำหรับการทำนายหรือการจำแนกในอนาคตในขณะที่มีการอนุมานจากข้อมูลในอดีตที่ใช้ในการสร้างตัวแบบเอง อัลกอริทึมของแบบจำลองการทำนายนี้สามารถปรับให้เหมาะสมกับการทำนายประเภท (Classification), การถดถอย (Regression), หรือการพยากรณ์ (Forecasting) ตามลักษณะของข้อมูลที่ใช้ในการฝึกตัวแบบ

2.1.3.2 การทำเหมืองข้อมูลในการบรรยาย (Descriptive Modeling) หรือเรียกว่าการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) มุ่งไปที่การค้นหาความสัมพันธ์และโครงสร้างในข้อมูล โดยไม่มีจุดมุ่งหมายในการทำนายผลลัพธ์เฉพาะ ซึ่งเน้นการวิเคราะห์และทำความเข้าใจข้อมูล การทำแบบจำลองนี้จะนำข้อมูลที่มีอยู่มาสืบค้นหาความสัมพันธ์, การจัดกลุ่มข้อมูล, หรือการค้นหาโครงสร้างที่ซ่อนอยู่ในข้อมูลที่ไม่มีการสอน (unsupervised) ด้วยคุณสมบัติที่มีอยู่ในข้อมูลภายใน ดังนี้

2.1.3.2.1 การจำแนกประเภทข้อมูล (Classification) เป็นกระบวนการสร้างตัวแบบ (model) จากข้อมูลที่มีการจำแนกประเภทแล้ว เพื่อใช้ในการจำแนกข้อมูลที่ไม่ทราบประเภท ตัวแบบที่สร้างขึ้นจะเรียนรู้และจำแนกข้อมูลตามกลุ่มหรือประเภทที่มีอยู่ในข้อมูลฝึกฝน (training data) และจากนั้นสามารถใช้ในการจำแนกข้อมูลใหม่ที่เข้ามาโดยไม่รู้ล่วงหน้า ประเภทของข้อมูลนั้น ๆ ตัวอย่างเช่น การจำแนกความเสี่ยงของลูกค้าในการกู้ยืมเงิน เป็นต้น

2.1.3.2.2 การจัดกลุ่มข้อมูล (Clustering) เป็นกระบวนการที่นำข้อมูลที่มีอยู่และแบ่งแยกออกเป็นกลุ่มๆ โดยที่ข้อมูลในกลุ่มเดียวกันจะมีลักษณะที่คล้ายคลึงกันและข้อมูลในกลุ่มต่างกลุ่มจะมีลักษณะที่แตกต่างกัน

2.1.3.2.3 การสร้างกฎความสัมพันธ์ (Association Rules) เป็นกระบวนการที่นำข้อมูลการเกิดขึ้นพร้อม ๆ กันอยู่เสมอ (frequent itemsets) เพื่อหากฎที่บอกรถึงความสัมพันธ์ระหว่างคุณสมบัติต่าง ๆ ในข้อมูล ปัญหาลักษณะนี้เรียกว่า "การวิเคราะห์ตระกล้าสินค้า" (market

basket analysis) และมักนำมาใช้ในร้านค้าเพื่อทำนายการซื้อสินค้าหรือความสัมพันธ์ระหว่างสินค้าต่าง ๆ

2.1.3.2.4 การคาดคะเน (Estimation/Regression) มีลักษณะคล้ายกับการจำแนกประเภทข้อมูล คือ การคาดคะเนเน้นการประมาณค่าต่อเนื่องหรือตัวเลข และการจำแนกประเภทข้อมูลเน้นการจำแนกข้อมูลลงในกลุ่มหรือประเภทที่มีอยู่แล้วในข้อมูล

2.1.3.2.5 การค้นหาข้อมูลที่มีความผิดปกติ (Data Anomaly Detection) เป็นกระบวนการที่ใช้ในการค้นหาข้อมูลที่มีลักษณะผิดปกติหรือความแตกต่างจากข้อมูลส่วนใหญ่ (Outlier Data) โดยข้อมูลที่มีความผิดปกติมักมีจำนวนน้อยและไม่สม่ำเสมอในข้อมูล (สัญญา พันธุ์แพง, 2563)

2.1.4 ลักษณะเฉพาะของข้อมูลที่มีการใช้ทำ Data Mining

2.1.4.1 ข้อมูลขนาดใหญ่มักมีปริมาณข้อมูลที่มาเกินกว่าที่จะสามารถพิจารณาความสัมพันธ์ที่ซ่อนอยู่ภายในข้อมูลได้ด้วยตาเปล่า โดยใช้วิธีการ Database Management System (DBMS) เพื่อจัดการฐานข้อมูลขนาดใหญ่ให้อยู่ในรูปแบบที่เหมาะสมและสามารถเข้าถึงข้อมูลได้อย่างมีประสิทธิภาพ

2.1.4.2 ข้อมูลที่มาจากหลายแหล่งนั้นอาจเป็นข้อมูลที่ถูกรวบรวมมาจากหลายระบบปฏิบัติการหรือหลายระบบจัดการฐานข้อมูล (DBMS) ต่าง ๆ ซึ่งอาจรวมถึงระบบ DBMS ที่แตกต่างกันเช่น Oracle, DB2, MS SQL, MS Access เป็นต้น

2.1.4.3 ข้อมูลที่ไม่มีการเปลี่ยนแปลงตลอดช่วงเวลาที่ทำการ Data Mining หมายถึงข้อมูลที่คงที่และไม่มีการอัปเดตหรือเปลี่ยนแปลงข้อมูลเกี่ยวกับคุณสมบัติหรือค่าข้อมูลตลอดระยะเวลาที่กำหนดสำหรับกระบวนการ Data Mining หรือการวิเคราะห์ข้อมูลที่ไม่มีการเปลี่ยนแปลงมักจะเหมาะสำหรับการทำ Data Mining ในบางกรณี เนื่องจากการแปลงข้อมูลที่ไม่เปลี่ยนแปลงมีความเสถียรและสามารถใช้ในการวิเคราะห์ข้อมูลในรูปแบบที่เหมือนกันตลอดเวลา

2.1.4.4 ข้อมูลที่มีโครงสร้างซับซ้อน เช่น ข้อมูลรูปภาพ ข้อมูลมัลติมีเดีย เป็นข้อมูลที่มีความซับซ้อนในการประมวลผลและการวิเคราะห์ เนื่องจากข้อมูลเหล่านี้มีลักษณะที่ไม่ได้เป็นตัวอักษรหรือตัวเลขแบบที่สามารถนำเข้ามาในโมเดลวิเคราะห์ได้อย่างตรงไปตรงมาการทำ Data Mining กับข้อมูลที่มีโครงสร้างซับซ้อนนี้ต้องใช้เทคนิคและเครื่องมือที่ทันสมัย เพื่อให้สามารถแยกแยะและวิเคราะห์ข้อมูลได้ถูกต้องและมีประสิทธิภาพ

2.1.5 ประโยชน์ของ Data Mining

2.1.5.1 การคาดการณ์ผลลัพธ์และการชี้แนะทางการตัดสินใจมีบทบาทสำคัญในการจัดการและการตัดสินใจในหลายสาขาอาชีพ เช่น ธุรกิจ, วิทยาการคอมพิวเตอร์, และวิทยาศาสตร์ข้อมูล เพื่อช่วยให้การตัดสินใจมีความระมัดระวังและมีเสถียรภาพมากยิ่งขึ้นในสถานการณ์ที่มีความไม่แน่นอนและการเปลี่ยนแปลงตลอดเวลา

2.1.5.2 การเพิ่มความเร็วในการวิเคราะห์ฐานข้อมูลขนาดใหญ่เป็นแนวทางสำคัญในการจัดการข้อมูลขนาดใหญ่เพื่อให้สามารถดำเนินการวิเคราะห์และนำข้อมูลไปสู่การตัดสินใจได้อย่างมีประสิทธิภาพ

2.1.5.3 การค้นหาส่วนประกอบที่ซ่อนอยู่ในเอกสารเป็นกระบวนการที่ใช้ในการวิเคราะห์ข้อมูลเพื่อค้นหาโครงสร้างและความสัมพันธ์ของส่วนประกอบต่าง ๆ ของเอกสาร

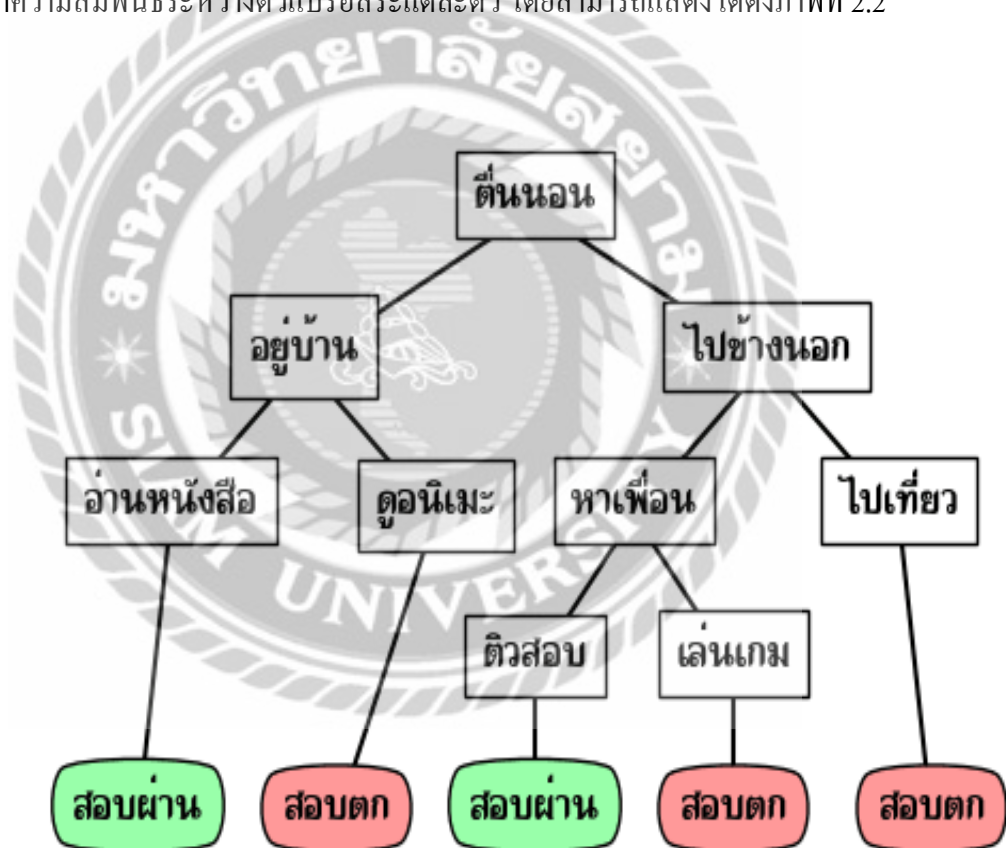
2.1.5.4 การเชื่อมโยงหน่วยงานต่าง ๆ ภายในองค์กรเป็นกระบวนการที่สำคัญในการสนับสนุนความร่วมมือและการทำงานร่วมกันในองค์กรเพื่อบรรลุวัตถุประสงค์ขององค์กรอย่างมีประสิทธิภาพ

2.1.5.5 การจัดกลุ่มข้อมูล เช่น ในการจัดการลูกค้าหรือข้อมูลในองค์กร ที่ช่วยให้สามารถปรับเปลี่ยนแผนการบริหารและการบริการตามความต้องการของลูกค้าได้อย่างมีประสิทธิภาพ (ศศิ วณิช, 2558)

2.1.6 การจำแนกประเภท (Classification)

เทคนิคการจำแนกประเภทข้อมูล (Data Classification) เป็นกระบวนการที่สำคัญในการสืบค้นความรู้พื้นฐานข้อมูลขนาดใหญ่ (Knowledge Discovery from very large Database: KDD) หรือ ดาต้าไมน์นิง (Data Mining) เทคนิคการจำแนกประเภทข้อมูลนี้มีความสำคัญในการดึงความรู้และข้อมูลอย่างมีประสิทธิภาพจากข้อมูลขนาดใหญ่ เป็นเทคนิคกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้จากกลุ่มตัวอย่างข้อมูลที่เรียกว่าข้อมูลสอนระบบ (training data) และมีการประยุกต์ใช้กว้างขวางในหลายด้านของสาขาวิชาต่าง ๆ เพื่อช่วยในการตัดสินใจและการวิเคราะห์ข้อมูลในสถานการณ์ต่าง ๆ อย่างมีประสิทธิภาพ เทคนิคการจำแนกประเภทข้อมูลนี้ได้นำไปประยุกต์ใช้ในหลายด้าน เช่น การจัดกลุ่มลูกค้าทางการตลาด, การตรวจสอบความผิดปกติ และการวิเคราะห์ทางการแพทย์ เป็นต้น (ชนวิวัฒน์ ปัญญาวงศ์, 2555)

2.1.6.1 อัลกอริทึม J48 Decision Tree, หรืออัลกอริทึมของ C4.5, เป็นอัลกอริทึมที่ถูกพัฒนาโดย Ross Quinlan และเป็นการปรับปรุงของอัลกอริทึม ID3 Decision Tree โดยอัลกอริทึมนี้เป็นเทคนิคที่ใช้ในการตัดสินใจโดยการสร้างโครงสร้างต้นไม้ที่ช่วยในการแบ่งข้อมูลไปยังกลุ่มย่อย โดยใช้หลักการของ Information Entropy และคำนวณ Normalized Information Gain เพื่อเลือกคุณลักษณะข้อมูลที่เหมาะสมสำหรับการแบ่งข้อมูล อัลกอริทึมที่ใช้ใน Decision Tree ได้แก่ J48 และ ID3. NBTree เป็นการผสมผสานระหว่างการใช้เทคนิคการเรียนรู้แบบต้นไม้ตัดสินใจในการสร้างโหนดและการใช้เทคนิคการเรียนรู้แบบ Naïve Bayes ในโหนด นอกจากนี้ Decision Tree และ Naïve Bayes ยังมี Bayesian Network (BayesNet) ที่ใช้ตัวแบบแบบชุดความน่าจะเป็นในการจัดกลุ่มข้อมูลโดยพิจารณาความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละตัว โดยสามารถแสดงได้ดังภาพที่ 2.2

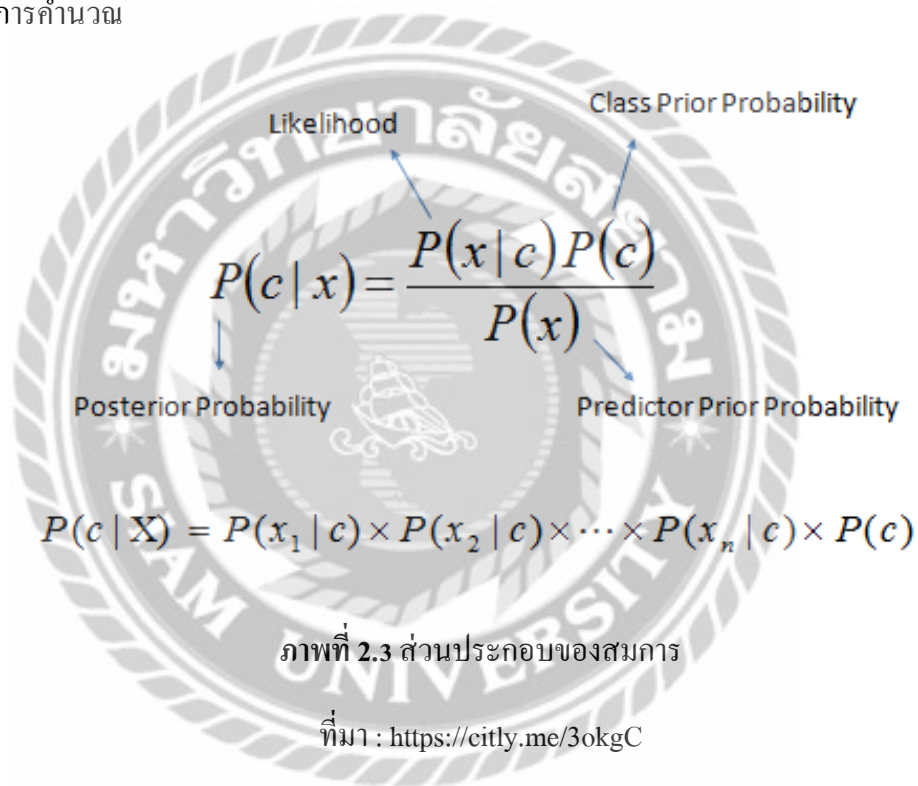


ภาพที่ 2.2 การตัดสินใจเลือกว่าจะอยู่บ้านหรือออกไปข้างนอก

ที่มา: <https://phyblas.hinaboshi.com/20171105>

ผลลัพธ์ของการตัดสินใจขั้นตอนการสร้างต้นไม้อย่างง่ายสามารถสร้างได้จากคุณลักษณะที่เป็นไปได้จะอธิบายตัวอย่างที่ง่ายๆ ที่หลายๆคนอาจจะคงเคยเห็นมากันบ้าง คือการตัดสินใจในการเลือกว่าจะอยู่บ้านหรือออกไปข้างนอก อยู่บ้านจะอ่านหนังสือหรือเล่นเกม ออกไปข้างนอกแล้วไปเจอเพื่อนหรือเปล่าเจอเพื่อนแล้วจะชวนกันไปเล่นเกมหรือไปอ่านหนังสือสุดท้ายการกระทำเหล่านั้นจะเป็นตัวตัดสินใจว่าจะสอบปลายภาคผ่านหรือเปล่า (อรนุช พันโท และ มนต์ชัย เทียนทอง ,2558)

2.1.6.2 Naive Bayes Classification เป็นการจัดหมวดหมู่ที่ใช้หลักการความน่าจะเป็นเข้ามาช่วยในการคำนวณ



ความหมายของสมการ แทนตัวแปร 3 ตัว คือ

c หมายถึง Class

x หมายถึง Attribute

P หมายถึง Probability (ความน่าจะเป็น)

P(c|x) Posterior probability หมายถึง ความน่าจะเป็น ข้อมูลที่มีแอททริบิวต์ x จะเป็นคลาส C

$P(x|c)$ Likelihood หมายถึง ความน่าจะเป็น ข้อมูลที่มีคลาส C มีแอททริบิวต์เป็น x

$P(c)$ Prior probability หมายถึง จำนวน Class ที่จะเกิดขึ้น หรือจำนวน Class ทั้งหมด คือความน่าจะเป็นของ Class C

$P(x)$ Predictor Prior probability หมายถึง จำนวน Attribute ทั้งหมด
ตัวอย่าง การทำนายข้อมูลบุคคลจาก Dataset

A	B	C
ลำดับ	เพศ	สีผิว
1	ชาย	ขาว
2	ชาย	ขาว
3	ชาย	ขาว
4	หญิง	ขาว
5	หญิง	ขาว
6	หญิง	ขาว
7	หญิง	ขาว
8	หญิง	ขาว
9	หญิง	ขาว
10	หญิง	ขาว
11	หญิง	ขาว
12	หญิง	ขาว
13	หญิง	น้ำตาล
14	ชาย	น้ำตาล
15	ชาย	น้ำตาล

ภาพที่ 2.4 การทำนายระหว่างบุคคล

ที่มา : <https://citly.me/3okgC>

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability

Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

ภาพที่ 2.5 ส่วนประกอบของสมการ

ที่มา : <https://citly.me/3okgC>

การกำหนดค่าของตัวแปรในสมการมี ดังนี้

c หมายถึง Class (เพศชาย , เพศหญิง)

x หมายถึง Attribute (สีผิว , ส่วนสูง , น้ำหนัก)

P หมายถึง Probability (ความน่าจะเป็น)

จำนวนเพศหญิงใน Dataset เท่ากับ 10

จำนวนเพศชายใน Dataset เท่ากับ 5

$P(x|c)$ Likelihood หมายถึง ความน่าจะเป็น ข้อมูลที่มีคลาส C และมีแอททริบิวต์ เป็น x

ความน่าจะเป็นที่ข้อมูลชายและหญิงจะมีคนที่ มีสีผิว , น้ำหนัก และส่วนสูงตามเกณฑ์ที่กำหนด เช่น

- จำนวนผู้หญิง 10 คน จะมีผิวสีขาว 9 คน ค่า $P(x|c)$ เท่ากับเพศหญิง) เท่ากับ 9/10

- จำนวนผู้ชาย 5 คน จะมีผิวสีขาว 3 คน ค่า $P(x|c)$ เท่ากับเพศชาย) เท่ากับ 3/5

$P(c)$ Prior probability หมายถึง ความน่าจะเป็นของ Class คือจำนวน Class / Class ทั้งหมด เช่น

- ถ้าคิดค่า $P(c)$ ของผู้หญิงเท่ากับ 10/15

- ถ้าค่า $P(c)$ ของผู้ชายเท่ากับ 5/15

b, c หมายถึง จำนวนข้อมูลที่มีในการทำนายผิด $P(x)$ Predictor Prior probability เท่ากับ $12/15$ มีคนผิวสีขาวจำนวน 12 คนจากทั้งหมด 15 คน

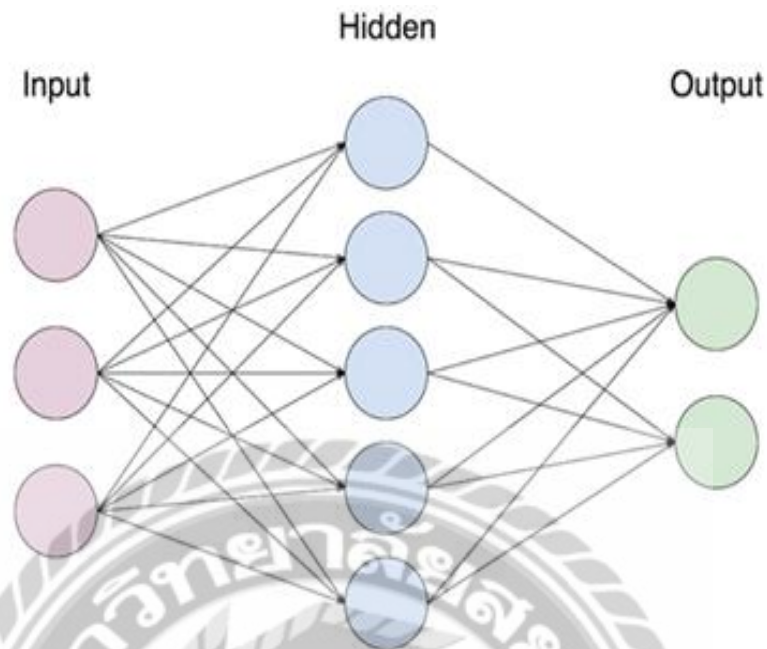
ดังนั้นการแทนค่าในสูตรจะได้

ค่าของ $P(c \text{ เท่ากับเพศหญิง} | x)$ เท่ากับ 0.75

ค่าของ $P(c \text{ เท่ากับเพศชาย} | x)$ เท่ากับ 0.25

ดังนั้นจากข้อมูลใน Dataset จะมีคนผิวสีขาวเป็นเพศหญิง 75% และเป็นเพศชาย 25% (สรุป Machine Learning (EP.6) - การจัดหมวดหมู่ด้วย Naive Bayes, 2020)

2.1.6.3 โครงข่ายประสาทเทียมหรือ Neural Networks เป็นเทคนิคที่นำมาใช้ในการทำเหมืองข้อมูล (Data Mining) และประมวลผลสารสนเทศโดยใช้การคำนวณแบบคอนเนกชันนิสต์ (Connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ โดยหมายถึงการสร้างโมเดลทางคณิตศาสตร์ที่ประกอบด้วยหน่วยประมวลผลที่เรียกว่า "นิวรอน" (Neurons) และการเชื่อมต่อระหว่างนิวรอนที่เรียกว่า "จุดประสานประสาท" (Synapses) โดยแต่ละนิวรอนมีส่วนหนึ่งที่ได้รับข้อมูลเข้ามาเรียกว่า "เดนไดรต์" (Dendrite) และส่วนหนึ่งที่ส่งข้อมูลออกไปเรียกว่า "แอกซอน" (Axon) โดยนิวรอนจะปฏิกิริยาด้วยกระแสไฟฟ้าเคมีเมื่อมีการกระตุ้นด้วยสิ่งต่าง ๆ ที่มาจากภายนอกหรือจากนิวรอนอื่น ๆ กระแสไฟฟ้าเคมีนี้จะถูกส่งผ่านเดนไดรต์และถ้ามีพอต่อการกระตุ้น นิวเคลียสซึ่งเป็นส่วนประมวลผลหลักของนิวรอนจะกระตุ้นนิวรอนอื่น ๆ ต่อไปผ่านทางแอกซอน วัตถุประสงค์ของโครงข่ายประสาท คือ การสร้างเครื่องมือที่สามารถเรียนรู้การจดจำรูปแบบ (Pattern Recognition) และการสร้างความรู้ใหม่ (Knowledge Extraction) อย่างมีความคล้ายคลึงกับการทำงานของสมองมนุษย์ โดยได้แรงบันดาลใจจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric Network) ในสมองที่ประกอบด้วยนิวรอนและจุดประสานประสาท ที่มีหน้าที่รับและส่งข้อมูลผ่านกระแสไฟฟ้าเคมี เพื่อตัดสินใจและประมวลผลข้อมูลในลักษณะที่มีความคล้ายคลึงกับการทำงานของโครงข่ายประสาทในสมองมนุษย์ (วิทยา พรพัชรพงศ์, 2012)



ภาพที่ 2.6 ส่วนประกอบของโครงข่ายประสาทเทียม

ที่มา : <https://hackernoon.com/artificial-neural-network-a843ff870338>

โครงข่ายประสาทเทียมมี 3 ส่วน ดังภาพด้านบน

1. Input Layer เป็นข้อมูลนำเข้า input จำนวน node จะเท่ากับจำนวน input
2. Hidden Layer เป็นชั้นตรงกลาง จะมีกี่ชั้นก็ได้ และมีผลต่อประสิทธิภาพการเรียนรู้ของ model ทุก node ต้องมี nonlinearity function
3. Output Layer เป็นชั้นที่นำข้อมูลไปใช้ โดยจำนวน node ขึ้นอยู่กับ output ที่นำไปใช้

2.1.7 Confusion Matrix เป็นเครื่องมือสำคัญในการประเมินผลลัพท์ของการทำนายหรือ Prediction ที่เกิดจาก Model ที่เราสร้างขึ้นใน Machine Learning. มีโอเคเดียวมาจากการวัดและประเมินความแม่นยำของการทำนายโดยเปรียบเทียบระหว่างสิ่งที่ Model ทำนายกับสิ่งที่เกิดขึ้นจริงในรูปแบบ Confusion Matrix ซึ่งเป็นตารางที่นำมาคำนวณค่าต่าง ๆ เพื่อประเมินประสิทธิภาพของการทำนายด้วย Model โดยที่ (ปกรณ์ กัชชาติ, 2019)

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

ภาพที่ 2.7 Confusion Matrix

ที่มา : <https://citly.me/noqEN>

True Positives (TP) จำนวนข้อมูลที่ถูกทำนายให้เป็น Positive และเป็นจริงในความเป็นจริง (Actual Positive)

True Negatives (TN): จำนวนข้อมูลที่ถูกทำนายให้เป็น Negative และเป็นจริงในความเป็นจริง (Actual Negative)

False Positives (FP): จำนวนข้อมูลที่ถูกทำนายให้เป็น Positive แต่เป็นจริงในความเป็นจริงคือ Negative

False Negatives (FN): จำนวนข้อมูลที่ถูกทำนายให้เป็น Negative แต่เป็นจริงในความเป็นจริงคือ Positive

ด้วยข้อมูลดังกล่าว สามารถคำนวณหลายตัวชี้วัดความสำเร็จของระบบได้ เช่น

ความแม่นยำ (Accuracy): สัดส่วนของข้อมูลทั้งหมดที่ถูกทำนายถูกต้อง (TP + TN) ต่อจำนวนข้อมูลทั้งหมด (TP + TN + FP + FN)

ความแม่นยำบนแต่ละกลุ่ม (Precision): สัดส่วนของ True Positives ต่อทั้งหมดที่ระบบทำนายว่าเป็น Positive (TP / (TP + FP))

ความแม่นยำของกลุ่ม Positive (Recall หรือ Sensitivity): สัดส่วนของ True Positives ต่อทั้งหมดของความเป็นจริงที่เป็น Positive (TP / (TP + FN))

ค่า F1-Score: ค่าความสมดุลระหว่าง Precision และ Recall ที่คำนวณจากสูตร $F1-Score = 2 * ((Precision * Recall) / (Precision + Recall))$

ความแม่นยำบนกลุ่ม Negative (Specificity): สัดส่วนของ True Negatives ต่อทั้งหมดของความเป็นจริงที่เป็น Negative (TN / (TN + FP))

Confusion Matrix เป็นเครื่องมือที่ช่วยให้เราเข้าใจการทำนายและประสิทธิภาพของระบบจำแนกข้อมูลอย่างชัดเจน และช่วยในการปรับปรุงและประสิทธิภาพของโมเดล Machine Learning ของเราด้วยการวิเคราะห์ผลลัพธ์ที่ได้จากการทดสอบ

2.2 แนวความคิดที่เกี่ยวข้องกับการสูบบุหรี่

การให้คำจำกัดความและชี้แจงส่วนประกอบในการสูบบุหรี่เป็นข้อมูลที่จะนำไปพยากรณ์เพื่อให้เห็นถึงความอันตรายและพิษภัยของบุหรี่ ดังต่อไปนี้

2.2.1 สารประกอบของบุหรี่

2.2.1.1 นิโคติน เป็นสารพิษที่กระตุ้นสมองและระบบประสาทส่วนกลาง ส่งผลให้ความดันโลหิตสูงขึ้นและเกิดผลกระทบทางหัวใจและชีพจร เป็นสาเหตุของโรคหัวใจและความดันโลหิตสูง

2.2.1.2 ทาร์ เป็นสารที่เกี่ยวข้องกับการสูบบุหรี่ และมีผลกระทบที่เสียหายต่อสุขภาพปอดและระบบทางเดินหายใจ โดยเฉพาะการเป็นสาเหตุของมะเร็งปอดและโรคถุงลมโป่งพอง

2.2.1.3 คาร์บอนมอนอกไซด์ เป็นก๊าซที่เกิดจากการเผาไหม้ เมื่อถูกดูดเข้าไปจะทำให้เลือดข้นและหนัก ส่งผลให้หัวใจทำงานหนักขึ้น

2.2.1.4 ไฮโดรเจนไซยาไนด์ สารนี้ส่งผลให้เกิดอาการไอ มีเสมหะ ปวดศีรษะ และคลื่นไส้อาเจียน มีความพิษในการทำสงคราม

2.2.3.5 ไนโตรเจนไดออกไซด์ เป็นสารที่มีความสามารถทำลายเยื่อหุ้มหลอดเลือดส่วนปลายและ
 ถูกลดในปอด ซึ่งเป็นสาเหตุหนึ่งของโรคถุงลมปอดโป่งพอง

2.2.3.6 แอมโมเนีย ใช้ในการปรุงแต่งรสชาติและช่วยให้ไนโคตินดูดซึมเข้าสู่สมองและ
 ประสาทเร็วขึ้น แอมโมเนียมีหลายลักษณะการใช้งานและมีผลกระทบต่อสุขภาพที่ต่างกัน

2.2.3.7 ไนโตรเจนออกไซด์ ทำให้หัวใจเต้นเร็วขึ้นและเป็นอันตราย

2.2.3.8 โซดาไนต์ ใช้เป็นยาเบื่อหนู การได้รับสารนี้มากจะทำให้หัวใจเป็นอัมพาตและหยุด
 หายใจได้ (สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ, 2560)

2.2.2 โรคของการสูบบุหรี่

บุหรี่นั้นทำให้เกิดความเสียหายหลายอย่างเช่น ทำให้เสื่อมสมรรถภาพทางเพศ เป็นมะเร็งใน
 ช่องปาก มะเร็งหลอดลมและหลอดอาหาร มะเร็งกล่องเสียง มะเร็งปอด ถุงลมโป่งพอง โรคตับแข็ง
 ทำให้มีกลิ่นปากและตัวรุนแรง สูญเสียเงินโดยไร้เหตุ ฟันเหลือง ตาแดง เล็บเขียว และยังส่งผลร้ายต่อ
 ผู้ที่อยู่รอบข้างผู้สูบบุหรี่ (สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ, 2561)

2.2.3 ปัจจัยเสี่ยงที่ทำให้ติดบุหรี่

2.2.3.1 ความอยากลองเป็นเรื่องธรรมชาติของวัยรุ่นที่ทำทายเป็นสนุกสนาน ทำให้ร้อยละ
 38.4 ของเยาวชนเริ่มสูบบุหรี่เพราะอยากลอง แม้ทราบว่ามันไม่ดีต่อสุขภาพ

2.2.3.2 ตามอย่างเพื่อน เยาวชนบางคนอาจเริ่มสูบบุหรี่เพราะได้รับแรงจูงใจจากเพื่อน การ
 กระตุ้นจากกลุ่มเพื่อนสามารถทำให้บางคนลองสูบบุหรี่และติดตามท่าทางของเพื่อนที่สูบบุหรี่

2.2.3.3 ตามอย่างคนในบ้าน การมีคนในบ้านที่สูบบุหรี่อาจส่งผลให้เยาวชนลองสูบบุหรี่
 เพราะพบการสูบบุหรี่ในสภาพแวดล้อมใกล้ชิด การมีบิดา มารดา หรือญาติที่สูบบุหรี่สามารถกระตุ้น
 ให้เยาวชนลองสูบบุหรี่มากขึ้น

2.2.3.4 เพื่อเข้าสังคม บางคนอาจเริ่มสูบบุหรี่เพราะจำเป็นต้องทำเช่นในงานเลี้ยงสังสรรค์ งาน
 สังสรรค์ หรืองานเลี้ยงรับรอง ซึ่งมักมีการสูบบุหรี่ในสังคมนั้น การเริ่มสูบบุหรี่เพื่อเข้าสังคมเป็น
 ตัวกระตุ้นให้บางคนติดตามมาสูบบุหรี่

2.2.3.5 ลดความเครียด บางคนสูบบุหรี่เพื่อลดความเครียด สารนิโคตินในบุหรี่สามารถสร้าง
 ความผ่อนคลายในร่างกายและสมอง แต่เมื่อปริมาณนิโคตินลดลง อาจทำให้เกิดความเครียดและความ
 หงุดหงิดในระยะต่อมา ทำให้บุคคลต้องสูบบุหรี่อย่างต่อเนื่อง

2.2.3.6 เชื้อกระแสสื่อโฆษณา สื่อโฆษณามีอิทธิพลในการส่งเสริมการสูบบุหรี่ แม้ว่าประเทศ
 ไทยจะมีกฎหมายห้ามโฆษณาบุหรี่ แต่บางประเทศอาจไม่มีกฎหมายห้าม ซึ่งสื่อโฆษณานูหรีสามารถ

เข้าถึงวัยรุ่นผ่านหลายช่องทาง การโฆษณาแฝงและการให้ความสนับสนุนในรูปแบบต่าง ๆ สามารถกระตุ้นให้เยาวชนสนใจและลองสูบบุหรี่ได้ (สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ, 2561)

2.3 พฤติกรรมการสูบบุหรี่

พฤติกรรมการสูบบุหรี่ของคนสามารถแบ่งออกเป็น 4 ประเภท ตามที่ซัลลิแวน ทอมกินส์ (Sullivan Tomkins) ได้กล่าวถึงไว้ ได้แก่

2.3.1 พฤติกรรมการสูบบุหรี่ที่เป็นนิสัยเคยชิน (Habitual smoking) มักจะทำตามนิสัยเป็นประจำ โดยสูบบุหรี่ตลอดเวลาและมีบุหรี่ใกล้ตัวเสมอ การสูบบุหรี่ทำให้รู้สึกดีขึ้นหรือมีความรู้สึกว่าจะเพิ่มความมั่นใจ และเป็นพฤติกรรมที่อาจกลายเป็นนิสัยอัตโนมัติ ถ้าต้องการเลิกสูบบุหรี่จะต้องเข้าใจลักษณะของพฤติกรรมการสูบบุหรี่ของตนเองเพื่อนำมาใช้ในการเลิกสูบบุหรี่ได้เป็นอย่างดี

2.3.2 พฤติกรรมการสูบบุหรี่ที่มองโลกในแง่ดี (Positive affect smoking) คือการสูบบุหรี่เพื่อกระตุ้นความพอใจและความสุข โดยเรียกร้องให้เกิดความสุขและความตื่นเต้น เพื่อผ่อนคลายความเครียด เช่น การสูบบุหรี่หลังอาหาร หรือเพื่อแสดงตัวตนเหมือนคนที่ตนเองชื่นชอบ เช่น การจำลองคาราหรือนำแบบฉบับมาตามความชอบ กลุ่มนี้บางคนมีความสุขในการสูบบุหรี่และมีความสุขในการสัมผัสควันบุหรี่ คนที่มีพฤติกรรมการสูบบุหรี่แบบนี้ต้องใช้ความพยายามมากเพื่อเลิกสูบบุหรี่.

2.3.3 พฤติกรรมการสูบบุหรี่เนื่องจากมองโลกในแง่ลบ (Negative affect smoking) คือการสูบบุหรี่เมื่อมีความเครียดหรือความกดดัน เช่น เมื่อรู้สึกตื่นเต้น หวาดกลัว หรือมีปัญหา พฤติกรรมการสูบบุหรี่ในกลุ่มนี้ไม่ได้ติดต่อกันตลอดเวลา แต่เมื่อเกิดความเครียดหรืออารมณ์ลบก็จะสูบบุหรี่เป็นบางครั้ง เช่น เมื่อรู้สึกสุขหรือเศร้า หรือเงินอายุ หรือในสถานการณ์ที่ต้องอยู่คนเดียว เพื่อลดความเครียด พยายามหาวิธีทดแทนการสูบบุหรี่อย่างมีประสิทธิภาพอยู่ตลอดเวลา

2.3.4 พฤติกรรมขาดการสูบบุหรี่ไม่ได้ (Addiction smoking) เกิดเมื่อผู้สูบบุหรี่รู้สึกว่าบุหรี่คือสิ่งที่จำเป็นและเป็นสิ่งที่ขาดไม่ได้ การสูบบุหรี่ช่วยให้รู้สึกสบายใจและพอใจเพื่อลดความรู้สึกกดดัน บุคคลในกลุ่มนี้อาจมีความยากลำบากในการเลิกสูบบุหรี่และต้องใช้ความพยายามและการตั้งใจมากเพื่อเลิกสูบบุหรี่ การสนับสนุนและการส่งเสริมกำลังใจในการเลิกสูบบุหรี่เป็นสิ่งสำคัญ ข้อเสนอแนะสำหรับนักติดยุติประเภทนี้คือ ควรพยายามหยุดสูบบุหรี่เท่าที่จะเป็นไปได้ และไม่ให้ตนเองสูบบุหรี่มาก ๆ เพื่อให้เกิดความไม่สบายใจ เช่น อาจจะมีมือ ไอ หรือเสียงแหบหลังสูบบุหรี่ เพื่อให้รู้สึกถึงความไม่ดีและโทษของการสูบบุหรี่ได้และเริ่มคิดอย่างระมัดระวังก่อนที่จะสูบบุหรี่อีกครั้ง นอกจากนี้

ยังสามารถใช้วิธีต่าง ๆ เพื่อช่วยในการเลิกสูบบุหรี่ได้ (วิลาวัณย์ ประทีปแก้ว, พรเทพ ล้อมพรม และ วัลภา พรหมชัย, 2552)

2.4 กฎหมายที่เกี่ยวข้องกับบุหรี่

กฎหมายเกี่ยวกับบุหรี่

พ.ร.บ.ควบคุมผลิตภัณฑ์ยาสูบ พ.ศ. 2560 มีมาตรการสำคัญที่ประชาชนต้องรับทราบเพื่อการปฏิบัติตามกฎหมายที่ถูกต้อง ดังนี้

1. กำหนดห้ามขายหรือให้ผลิตภัณฑ์ยาสูบแก่บุคคลที่มีอายุต่ำกว่า 20 ปี เพื่อรักษาและส่งเสริมสุขภาพและป้องกันการใช้ยาสูบในวัยที่ยังไม่เหมาะสม ซึ่งเป็นมาตรการสำคัญในการลดความเสี่ยงต่อสุขภาพและป้องกันปัญหาที่เกี่ยวข้องกับการสูบบุหรี่ในวัยเยาว์

2. ห้ามให้บุคคลที่มีอายุต่ำกว่า 18 ปี ใช้งานหรือเข้าถึงสิ่งบริการที่อาจมีข้อจำกัดความเสี่ยงต่อการพัฒนาและสุขภาพของวัยรุ่นนั้นได้

3. ห้ามขายผลิตภัณฑ์ยาสูบใน 4 กลุ่มสถานที่ต่อไปนี้ วัดหรือสถานที่ปฏิบัติพิธีกรรมทางศาสนา, สถานพยาบาลและร้านขายยา, สถานศึกษาทุกระดับ, สวนสาธารณะ, สวนสัตว์ และสวนสนุก เป็นต้น การห้ามขายในสถานที่เหล่านี้เป็นมาตรการที่เกี่ยวข้องกับการควบคุมการใช้ยาสูบและส่งเสริมสุขภาพของประชาชน เพื่อป้องกันผู้ที่มีอายุต่ำกว่า 18 ปี จากการเข้าถึงผลิตภัณฑ์ยาสูบและลดความเสี่ยงต่อสุขภาพและสวัสดิการของเด็กและวัยรุ่น

4. กำหนดห้ามโฆษณาและสื่อสารการตลาดผลิตภัณฑ์ยาสูบในทุกรูปแบบ รวมถึงพรีดตี้ที่ส่งเสริมการขายในงานคอนเสิร์ตและอีเวนต์ทางบันเทิงต่าง ๆ เพื่อควบคุมการสื่อสารและโฆษณาเกี่ยวกับผลิตภัณฑ์ยาสูบและลดการกระตุ้นให้บุคคลบริโภคสนใจผลิตภัณฑ์นี้ และเพื่อลดความเสี่ยงต่อสุขภาพและป้องกันปัญหาที่เกี่ยวข้องกับการสูบบุหรี่

5. ห้ามผู้ประกอบการธุรกิจผลิตภัณฑ์ยาสูบทำกิจกรรม CSR (Corporate Social Responsibility) อุปถัมภ์สนับสนุนบุคคลหรือองค์กรที่เป็นการสร้างภาพลักษณ์ของผลิตภัณฑ์ยาสูบ การห้ามนี้เกี่ยวข้องกับการป้องกันการใช้กิจกรรม CSR ในการโฆษณาหรือสร้างภาพบวกของผลิตภัณฑ์ยาสูบ ซึ่งอาจส่งผลต่อการเสริมสร้างภาพของผลิตภัณฑ์นั้นให้ดูดีและเป็นมิตรต่อสังคมอย่างไม่เหมาะสม

6. ห้ามตั้งวางโชว์ผลิตภัณฑ์ยาสูบหรือซองบุหรี่ที่จุดขายปลีกที่ทำให้ผู้บริโภคหรือประชาชนมองเห็น โดยมีเลข 7 เป็นตัวเลขบนซองบุหรี่ ซึ่งเป็นมาตรการที่จัดทำขึ้นเพื่อควบคุมการแสดงผลและ

โฆษณาผลิตภัณฑ์ยาสูบที่อาจสร้างสิ่งกระตุ้นให้ผู้บริโภคมองเห็นและติดตัวรูปสัญลักษณ์หรือข้อมูลที่เป็นไปในทางที่ไม่เหมาะสม มาตรการนี้เน้นการควบคุมการสื่อสารและการโฆษณาของผลิตภัณฑ์ยาสูบเพื่อลดผลกระทบต่อ การสูบบุหรี่และสุขภาพของประชาชนในสถานที่ขายปลีกและสาธารณะทั่วไป

7. ห้ามแบ่งซองขายบุหรี่เป็นรายมวน หมายถึงการห้ามแยกบรรจุซองบุหรี่เพื่อขายเป็นรายมวนหรือเป็นชิ้นส่วนเล็ก ๆ ที่สามารถซื้อเป็นบรรจุหนึ่งเดียว การห้ามนี้มีเพื่อควบคุมการขายบุหรี่และลดการเข้าถึงบุหรี่โดยการแยกขายเป็นรายมวน ซึ่งช่วยลดความน่าสนใจและการเข้าถึงบุหรี่โดยเฉพาะในกลุ่มที่มีความเสี่ยงสูง เช่น วัยรุ่น และส่งเสริมสุขภาพของประชาชนในระยะยาว

8. เพิ่มโทษผู้ฝ่าฝืนสูบบุหรี่ในเขตปลอดบุหรี่ให้มีปรับไม่เกิน 5,000 บาท เป็นมาตรการที่กำหนดโทษในกรณีที่บุคคลฝ่าฝืนกฎหมายหรือข้อกำหนดที่ห้ามสูบบุหรี่ในพื้นที่เขตที่ระบุว่าเป็น "เขตปลอดบุหรี่"

9. กำหนดหน้าที่ของเจ้าของสถานที่สาธารณะในการรักษาและสนับสนุนการปฏิบัติตามข้อกำหนดห้ามการสูบบุหรี่ในพื้นที่ที่เป็น "เขตปลอดบุหรี่" โดยรวมถึงการแจ้งเตือนและประชาสัมพันธ์กฎหมายที่เกี่ยวข้อง และหากไม่ปฏิบัติตามมาตรการนี้เจ้าของสถานที่สาธารณะอาจถูกปรับตามกฎหมายไม่เกิน 3,000 บาท เพื่อส่งเสริมความปลอดภัยและป้องกันการสูบบุหรี่ในพื้นที่เขตปลอดบุหรี่ให้ดำเนินอย่างมีความเสี่ยงต่อสุขภาพและสวัสดิการของประชาชนในสถานที่สาธารณะ (เทศบาลตำบลขุนทะเล, 2564)

2.5 งานวิจัยที่เกี่ยวข้อง

งานวิจัยเรื่อง “การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์ไทรอยด์ด้วยเทคนิคเหมืองข้อมูล” โดยนักวิจัย ณัฐวดี หงษ์บุญมี และประภาศิริ ตรีพาณิชย์กุล (2562) งานวิจัยนี้มีวัตถุประสงค์เปรียบเทียบประสิทธิภาพของการจำแนกข้อมูลด้วยอัลกอริทึมเหมืองข้อมูลสามแบบ : โครงข่ายประสาทเทียม, การเรียนรู้แบบเบย์, และต้นไม้ตัดสินใจ เพื่อหาอัลกอริทึมที่มีประสิทธิภาพสูงสุดในการพยากรณ์ความเสี่ยงของโรคไฮเปอร์ไทรอยด์ โดยลดการนำเข้าข้อมูลที่ละปัจจัย ใช้ข้อมูลจากโรงพยาบาลในจังหวัดพิษณุโลก 323 ชุดข้อมูล มีจำนวน 12 ปัจจัย ผลการเปรียบเทียบพบว่าโครงข่ายประสาทเทียมให้ความถูกต้องสูงสุดที่ 82.97% มากกว่าต้นไม้ตัดสินใจและการเรียนรู้แบบเบย์ ตามลำดับ ปัจจัยที่ส่งผลต่อความเสี่ยงโรคไฮเปอร์ไทรอยด์มีความสำคัญคือ อารมณ์แปรปรวนและเหนื่อยง่าย และปัจจัยส่วนบุคคลที่มีความสำคัญคือ

เพศ นอกจากนี้ยังสามารถพัฒนาระบบการพยากรณ์ความเสี่ยงโรคไฮเปอร์ไทรอยด์บนสมาร์ทโฟน เพื่อช่วยสนับสนุนการตัดสินใจในการวิเคราะห์ความเสี่ยงโรคไฮเปอร์ไทรอยด์และแนะแนวทางการรักษาของแพทย์และผู้ป่วยในอนาคต

งานวิจัยเรื่อง “การเปรียบเทียบปัจจัยโรคประจำตัวผู้สูงอายุโดยใช้อัลกอริทึมการจัดกลุ่ม J48 และ NaiveBayes : กรณีศึกษาสาธารณสุขโพธิ์กลางนครราชสีมา” โดยนักวิจัย เบญจกัศ จงหมื่น ไวย์ (2558) จากข้อมูลที่น่ามาวิเคราะห์พบว่ามีผู้สูงอายุทั้งหมด 594 ราย แบ่งเป็น 6 กลุ่มอายุต่างกัน โดยค้นหาช่วงอายุระหว่าง 60-90 ปี พบว่าความน่าจะเป็นสูงสุดอยู่ที่อายุ 85-90 ปี คือ 0.445 โดยการอธิบายความสัมพันธ์กล่าวว่าถ้าอายุน้อยกว่าหรือเท่ากับ 86 ปี ส่วนใหญ่มีโรคประจำตัวคือความดันโลหิตสูงและโรคพาร์กินสัน ระดับความน่าจะเป็นคือ 2.0 แต่ถ้าอายุมากกว่า 86 ปี พบว่าผู้สูงอายุมีโรคประจำตัวคือโรคเบาหวานและความดันโลหิตสูง และมีเพศชายมากกว่าเพศหญิง มีตัวแบบที่สองในการทำนายผู้สูงอายุ คือ Decision Tree ที่มีความถูกต้องของการทำนายเท่ากับ 0.856% และ 91.43% และ Naive Bayes ที่มีความถูกต้องของการทำนายเท่ากับ 10.27% และ 89.72% ดังนั้น Decision Tree เหมาะสำหรับการทำนายผู้สูงอายุในกลุ่มนี้เพราะมีความถูกต้องสูงกว่า Decision Tree

งานวิจัยเรื่อง “การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย” โดยนักวิจัย สุรวัชร ศรีเปารยะ และ สายชล สีนสมบูรณ์ทอง (2560) การวิจัยนี้มุ่งหวังเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มผู้ป่วยโรคไตเรื้อรังด้วยวิธีการทำเหมืองข้อมูลแต่ละวิธี พบว่า วิธีความใกล้เคียงกันมากที่สุดมีความถูกต้องถึง 94.17% และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย คือ 0.0579 วิธีต้นไม้ตัดสินใจมีความถูกต้องถึง 100% และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย คือ 0.0059 วิธีโครงข่ายประสาทเทียมมีความถูกต้องถึง 98.33% และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย คือ 0.0115 วิธีซัพพอร์ตเวกเตอร์แมชชีนมีความถูกต้องถึง 98.33% และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย คือ 0.0167 วิธีฐานภูมิมีความถูกต้องถึง 95% และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย คือ 0.0454 วิธีการถดถอยลอจิสติกมีความถูกต้องถึง 93.33% และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย คือ 0.0660 วิธีนาอูฟเบย์มีความถูกต้องถึง 95% และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย คือ 0.0409 จากผลการวิจัย วิธีต้นไม้ตัดสินใจ (J48) เป็นวิธีการจำแนกกลุ่มที่มีประสิทธิภาพในการจำแนกข้อมูลโรคไตเรื้อรังที่ดีที่สุด โดยมีความถูกต้องถึง 100%

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดคือ 0.0059 ซึ่งเหมาะสำหรับการใช้ในการจำแนกกลุ่มผู้ป่วยโรคไตเรื้อรังตามข้อมูลที่มีอยู่ในการวิจัยนี้

งานวิจัยเรื่อง “การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุล โดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดเชื้อในเยื่อหุ้มสมอง” โดยนักวิจัย ภรณ์ญา ปาลวิสุทธิ (2559) การวิจัยนี้มุ่งหวังพัฒนาตัวแบบเพื่อพยากรณ์การเป็นโรคติดเชื้อในเยื่อหุ้มสมองที่มีการใช้อินเทอร์เน็ตมากขึ้นไป โดยใช้เทคนิคการสุ่มเพิ่มตัวอย่างกลุ่มน้อย (SMOTE) เพื่อปรับความสมดุลของข้อมูล และใช้วิธีต้นไม้ตัดสินใจ J48, ID3, LMT, CART, และ Random Forest เป็นตัวแบบในการพยากรณ์ ผลการวิจัยพบว่า วิธี Random Forest มีประสิทธิภาพในการพยากรณ์การเป็นโรคติดเชื้อในเยื่อหุ้มสมองที่ดีที่สุดโดยมีค่าความแม่นยำถึง 87.15% วิธี J48, ID3, LMT, และ CART มีประสิทธิภาพในการพยากรณ์ต่ำกว่า Random Forest การวิจัยนี้ช่วยให้เราเข้าใจเพิ่มเติมเกี่ยวกับปัญหาโรคติดเชื้อในเยื่อหุ้มสมองและวิธีการพยากรณ์การเป็นโรคนี้นี้ด้วยโมเดลแบบ Random Forest ที่มีประสิทธิภาพสูงที่สุดในการพยากรณ์ข้อมูลในงานวิจัยนี้โดยมีค่าความแม่นยำถึง 87.15% ร่วมกับค่าความไวและความจำเพาะที่มีค่าสูงอีกด้วย การพัฒนาตัวแบบนี้จะช่วยในการตรวจสอบและดูแลเยื่อหุ้มสมองที่มีความเสี่ยงต่อโรคติดเชื้อในเยื่อหุ้มสมองได้เป็นอย่างดี

งานวิจัยเรื่อง “การเปรียบเทียบประสิทธิภาพในการทำนายการติดเกมของเด็กและวัยรุ่นในกรุงเทพมหานคร” โดยนักวิจัย สายชล สิ้นสมบูรณ์ทอง (2561) ในการศึกษานี้ได้ทำการศึกษาและเปรียบเทียบประสิทธิภาพในการทำนายการติดเกมของเด็กและวัยรุ่นในกรุงเทพมหานคร โดยใช้วิธีการจำแนกกลุ่ม 7 วิธีที่แตกต่างกัน และใช้ข้อมูลจากการสำรวจแบบสอบถาม เรียกว่าวิธีต้นไม้การตัดสินใจมีประสิทธิภาพที่ดีที่สุดในการทำนายการติดเกม มีค่าความแม่นยำถึง 92.17% และค่าความระลึกของชั้นไม้ได้ติดเกมมากที่สุดเท่ากับ 0.976 อันแสดงถึงความเหมาะสมในการทำนายและความแม่นยำในการระบุผู้ที่ไมติดเกม ในขณะที่วิธีซัพพอร์ตเวกเตอร์แมชชีนมีค่าความระลึกของชั้นไม้ได้ติดเกมมากที่สุดเท่ากับ 0.976 และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยน้อยที่สุดเท่ากับ 0.0870 วิธีฐานภูมิมีค่าความเที่ยงของชั้นไม้ได้ติดเกมและความระลึกของชั้นไม้ได้ติดเกมมากที่สุดเท่ากับ 0.930 และ 0.806 ตามลำดับ การเลือกใช้วิธีการจำแนกกลุ่มที่เหมาะสมสำหรับการทำนายการติดเกมของเด็กและวัยรุ่น

ในกรุงเทพมหานครจึงขึ้นอยู่กับความต้องการและความสำคัญของแต่ละปัจจัยที่สำคัญต่อการวิเคราะห์และการพยากรณ์ที่ถูกต้องในสถานการณ์ที่กำลังสนใจ

งานวิจัยเรื่อง “การเปรียบเทียบประสิทธิภาพในการพยากรณ์การเป็นโรคเบาหวานของโรงพยาบาลแห่งหนึ่ง” สายชล สตินสมบูรณ์ทอง (2561) การศึกษานี้ได้สำรวจข้อมูลตั้งแต่เดือนมกราคมถึงเมษายน พ.ศ. 2559 จำนวน 1,233 ชุด โดยใช้เทคนิคการจำแนกประเภทข้อมูล 6 เทคนิคที่แตกต่างกัน พบว่าเทคนิคโครงข่ายประสาทเทียมมีประสิทธิภาพในการพยากรณ์ที่ดีที่สุด โดยให้ค่าความถูกต้อง ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) มีค่าเท่ากับ 95.94%, 0.0491, และ 0.0396 ตามลำดับ สรุปได้ว่าเทคนิคโครงข่ายประสาทเทียมมีประสิทธิภาพสูงสุดในการทำนายข้อมูลในการศึกษานี้ การศึกษานี้ได้สำรวจข้อมูลตั้งแต่เดือนมกราคมถึงเมษายน พ.ศ. 2559 จำนวน 1,233 ชุด โดยใช้เทคนิคการจำแนกประเภทข้อมูล 6 เทคนิคที่แตกต่างกัน พบว่าเทคนิคโครงข่ายประสาทเทียมมีประสิทธิภาพในการพยากรณ์ที่ดีที่สุด โดยให้ค่าความถูกต้อง ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) มีค่าเท่ากับ 95.94%, 0.0491, และ 0.0396 ตามลำดับ สรุปได้ว่าเทคนิคโครงข่ายประสาทเทียมมีประสิทธิภาพสูงสุดในการทำนายข้อมูลในการศึกษานี้

งานวิจัยเรื่อง “การจำแนกข้อมูลเพื่อวินิจฉัยความเสี่ยงการเป็นโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูล” นพรัตน์ นนท์ศิริ, ราตรี มนต์ศิลา และ กริช สมกันธา (2565) งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างแบบจำลองการจำแนกข้อมูลเพื่อวินิจฉัยความเสี่ยงในการเป็นโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูล 4 วิธี คือ วิธีนาอิวเบย์ (Naive Bayes) วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) วิธีความใกล้เคียงกันที่สุด (K-Nearest Neighbor) และวิธีต้นไม้ตัดสินใจ (Decision Tree) โดยใช้ข้อมูลของผู้ป่วยโรคเบาหวานที่ได้รับการทบทวนจากเวชระเบียนย้อนหลัง จำนวน 1,435 ชุด ข้อมูล 16 คุณลักษณะ จากนั้นทำการวัดความถูกต้องของแบบจำลอง (Accuracy) โดยใช้วิธี 10-Fold cross validation พบว่า วิธีต้นไม้ตัดสินใจให้ค่าประสิทธิภาพสูงสุดโดยมีค่าความถูกต้อง 93.73% ส่วนวิธีนาอิวเบย์ค่าความถูกต้อง 88.92% วิธีความใกล้เคียงกันที่สุดและวิธีซัพพอร์ตเวกเตอร์แมชชีนค่าความถูกต้อง 86.97% และ 86.13% ตามลำดับ จะเห็นว่าวิธีต้นไม้ตัดสินใจมีประสิทธิภาพสูงสุดในการสร้างแบบจำลอง ด้วยเหตุผลว่าเป็นวิธีที่ไม่มีการแจกแจงหรือไม่ใช้พารามิเตอร์ทางสถิติที่ต้องปรับ

เหมาะสมสำหรับการจัดการข้อมูลที่มีมิติสูงและสามารถนำไปพัฒนาระบบวินิจฉัยความเสี่ยงโรคเบาหวานในอนาคต และเป็นแนวทางในการสนับสนุนการตัดสินใจทางการแพทย์ในการวินิจฉัยความเสี่ยงโรคเบาหวานของผู้ป่วยในอนาคต

งานวิจัยเรื่อง “การเปรียบเทียบโมเดลการเรียนรู้ของเครื่องสำหรับคัดกรองผู้ป่วยโรคเบาหวานที่มีภาวะชาปลายเท้า” อรุณรักษ์ ตันพานิช, คุณฉวีศุภวรร ษนะกุล, พิเชฐ บัญญัติ และ จริญญา จันทน (2562) งานวิจัยนี้ใช้ข้อมูลจากโรงพยาบาลส่งเสริมสุขภาพประจำตำบลท่าจีน จังหวัดสงขลา จำนวน 300 ชุด แบ่งเป็น 2 กลุ่ม คือ กลุ่มผู้ป่วยโรคเบาหวานชนิดที่ 2 ที่มีภาวะชาปลายเท้า และกลุ่มผู้ป่วยปกติ ในการวิเคราะห์ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคการเรียนรู้แบบอัตโนมัติด้วยการทำงานของโครงข่ายประสาทของมนุษย์ (Deep Learning) และเทคนิคป่าสุ่ม (Random Forest) โดยเปรียบเทียบประสิทธิภาพด้วยค่าความถูกต้อง ส่วนเบี่ยงเบนมาตรฐาน (SD) และระยะเวลาในการประมวลผล พบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพที่ดีที่สุด ด้วยค่าความถูกต้อง 93.3% ส่วนเบี่ยงเบนมาตรฐาน 8.2% และระยะเวลาในการประมวลผล 46 วินาที ตามลำดับ

งานวิจัยเรื่อง “การพัฒนาตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาด้วยเทคนิคการเหมืองข้อมูล” เพชรรัตน์ ม่วงน้อย และคณะ (2021) งานวิจัยนี้ใช้เครื่องมือที่ประกอบด้วยแบบสอบถามข้อมูลทั่วไปและแบบประเมินแบบทดสอบภาวะซึมเศร้า PHQ-9 เพื่อวิเคราะห์ข้อมูลโดยใช้เทคนิคเหมืองข้อมูลด้วยอัลกอริทึม Random, Tree, LMT, PART J48, และ JRIP และทำการเปรียบเทียบประสิทธิภาพด้วยการทดสอบ Cross-validation พบว่าอัลกอริทึม Random Tree มีประสิทธิภาพสูงสุดโดยมีค่าความถูกต้อง 96.00% รองลงมาคือ อัลกอริทึม LMT มีค่า 95.90% อัลกอริทึม PART มีค่า 95.10% อัลกอริทึม J48 มีค่า 94.80% และอัลกอริทึม JRIP มีค่า 94.00% ตามลำดับ

งานวิจัยเรื่อง "การวิเคราะห์ความรู้สึกแบบผสมสำหรับความคิดเห็นต่อโรงแรมในประเทศไทย" วาติศย์ คำพรมมา (2562) งานวิจัยนี้มีวัตถุประสงค์เพื่อทำการวิเคราะห์ความรู้สึกเกี่ยวกับความคิดเห็นต่อโรงแรมในประเทศไทยโดยใช้เทคนิคการวิเคราะห์ความรู้สึกในรูปแบบเหมืองความคิดเห็น (opinion mining) เพื่อให้ความเข้าใจเกี่ยวกับความคิดเห็นของผู้ใช้บริการโรงแรมเมื่อเข้ารับบริการ งานวิจัยได้รวบรวมความคิดเห็นจากเว็บไซต์ APT TUBE จำนวน 10,000 ประโยค การวิเคราะห์ความ

คิดเห็น (Sentiment Analysis) : ใช้เทคนิค K-means ในการแบ่งความคิดเห็นเป็น 5 กลุ่ม ได้แก่ การเข้าถึง (accessibility), กิจกรรมและความบันเทิง (activities & entertainment), อาหารและเครื่องดื่ม (food & beverage), พนักงานผู้ให้บริการ (staff), สถานที่ (place) การแยกประเภท (Classification) ใช้ K-Nearest Neighbors (K-NN) เพื่อแยกประเภทข้อความในแต่ละกลุ่มการวัดประสิทธิภาพด้วยค่าความเที่ยง (Precision), ค่าความระลึก (Recall), ค่าความถูกต้อง (Accuracy), ค่า F-Measure ร่วมกับ 10 Fold Cross Validation. และเปรียบเทียบ Decision tree, Support Vector Machine (SVM), K-Nearest Neighbors (K-NN) ผลลัพธ์ที่ได้ระบุว่าเทคนิคที่ให้ค่าความถูกต้องสูงที่สุดคือเทคนิค K-means ร่วมกับ K-NN ที่ให้ค่าความถูกต้องถึง 94.80 %



ตารางที่ 2.1 ตารางเปรียบเทียบงานวิจัย

ชื่อผู้แต่ง/ปี พ.ศ./ชื่อเรื่อง	กลุ่มตัวอย่าง	อัลกอริทึม	ผลการวิจัย
การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์ไทรอยด์ด้วยเทคนิคเหมืองข้อมูล โดยนักวิจัย ฉัฐวดี หงษ์บุญมี และประภาสรี ศรีพานิชกุล (2562)	ข้อมูลผู้ป่วยโรคไฮเปอร์ไทรอยด์จำนวน 323 ชุด	1. โครงข่ายประสาทเทียม 2. Decision Tree (J48) 3. Naïve Bayes	โครงข่ายประสาทเทียมให้ค่าประสิทธิภาพสูงสุดโดยมีค่าความถูกต้อง 82.97%
การเปรียบเทียบปัจจัยโรคประจำตัวผู้สูงอายุโดยใช้อัลกอริทึม การจัดกลุ่ม J48 และ Naïve Bayes : กรณีศึกษาสาธารณสุขสูงโพธิ์กลาง นครราชสีมา โดยนักวิจัยเบญจภักดิ์ จงหมื่นไวย์ (2558)	ข้อมูลผู้สูงอายุทั้งหมด 594 คน	1. Decision Tree (J48) 2. Naïve Bayes	ผลการวิจัยพบว่า Decision Tree ร้อยละค่าความถูกต้องของการทำนายเท่ากับ 91.43%
การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย โดยนักวิจัยสุรวัชร ศรีเปารยะ และสายชล สินสมบูรณ์ทอง (2560)	ข้อมูลผู้ป่วยโรคไตเรื้อรังทั้งหมด 400 ชุด	1. โครงข่ายประสาทเทียม 2. J48 3. ซัพพอร์ตเวกเตอร์แมชชีน	ผลการวิจัยพบว่าโครงข่ายประสาทเทียมมีค่าความถูกต้อง คือ 98.33%

ชื่อผู้แต่ง/ปี พ.ศ./ชื่อเรื่อง	กลุ่มตัวอย่าง	อัลกอริทึมที่ใช้	ผลการวิจัย
		4.ฐานกฎ 5.การทอดยอลอ จิสติก 6. Naive Bayes	
การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุล โดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดเชื้อในเนื้อ โดยนักวิจัย ภรณ์ยา ปาลวิสุทธิ (2559)	ใช้ข้อมูลกลุ่มเยาวชนทั้งหมด 892 ชุด	1. Random Forest 2. J48 3. ID3 4. LMT 5. CART	ผลการวิจัยพบว่าตัวแบบการพยากรณ์จากเทคนิค Random Forest เป็นตัวแบบที่มีประสิทธิภาพในการพยากรณ์สูงที่สุด โดยมีค่า 87.53%
การเปรียบเทียบประสิทธิภาพในการทำนายการติดเชื้อของเด็กและวัยรุ่นในกรุงเทพมหานคร โดยนักวิจัย สายชล สิ้นสมบูรณ์ทอง (2561)	กลุ่มตัวอย่างในเขตลาดกระบัง มีนบุรีและบางเขน เขตละ 500 ชุด (คน) รวม 1,500 ชุด	1. KNN 2. IBk 3. J48 4. MLP 5. SMO 6. Binary logistic regression 7. Naive Bayes 8.ฐานกฎ	J48 เป็นวิธีที่สามารถทำนายการติดเชื้อของเด็กและวัยรุ่นในกรุงเทพมหานครที่มีประสิทธิภาพมากที่สุดคือ 92.17%
ทำการศึกษาเปรียบเทียบประสิทธิภาพในการพยากรณ์การเป็นโรคเบาหวานของ	ข้อมูลผู้ป่วยโรคเบาหวาน 1,024 ชุด	1. KNN (Ibk) 2. Decision Tree (J48)	จากผลการศึกษา พบว่าเทคนิค ANN (Multi perceptron)

ชื่อผู้แต่ง/ปี พ.ศ./ชื่อเรื่อง	กลุ่มตัวอย่าง	อัลกอริทึมที่ใช้	ผลการวิจัย
โรงพยาบาลแห่งหนึ่ง โดย นักวิจัย สายชล สินสมบูรณ์ ทอง (2561)		3. ANN (Multi perceptron) 4. SVM (SMO) 5. Binary Logistic Regression 6. Naïve Baye	มีประสิทธิภาพในการ พยากรณ์ดีที่สุด โดยให้ ค่าความถูกต้อง 95.94%
งานวิจัยเรื่อง “การจำแนก ข้อมูลเพื่อวินิจฉัยความเสี่ยง การเป็นโรคเบาหวาน โดยใช้ เทคนิคเหมืองข้อมูล” โดย นักวิจัย นพรัตน์ นนทศิริ, ราตรี มนัสศิลา และ กริช สมกันธา (2565)	ข้อมูลผู้ป่วย โรคเบาหวาน จำนวน 1,435 ชุด	1. Support Vector Machine 2. Naive Bayes 3. Decision Tree 4. K-Nearest Neighbor	ผลการทดสอบพบว่า Decision Tree ให้ค่า ประสิทธิภาพสูงสุดโดย มีค่าความถูกต้อง 93.73%
การเปรียบเทียบโมเดลการ เรียนรู้ของเครื่องสำหรับคัด กรองผู้ป่วยโรคเบาหวานที่มี ภาวะช้ำปลายเท้า โดยนักวิจัย อรุณรักษ์ ตันพานิช, ดุชนิศุ ภววรรณกุล, พิเชฐ บัญญัติ และ จริญญา จันทน (2562)	ชุดข้อมูล ผู้ป่วย โรคเบาหวาน และ ผู้ป่วยที่มี ภาวะช้ำปลาย เท้า ทั้งหมด 300 ชุด	1. SVM 2. Deep Learning 3. Random Forest	จากผลการศึกษาพบว่า เทคนิค SVM พยากรณ์ ได้ผลลัพธ์ดีที่สุด โดยค่า ความถูกต้อง เท่ากับ 93.3%
ศึกษาและพัฒนาตัวแบบ ประเมินภาวะความเสี่ยง การ เป็นโรคซึมเศร้าของนักศึกษา ด้วยเทคนิคการเหมืองข้อมูล	นักศึกษา ระดับปริญญา ตรี มหาวิทยาลัย	1. Random Tree 2. LMT 3. PART	การเปรียบเทียบ ประสิทธิภาพด้วยการ ทดสอบ

ชื่อผู้แต่ง/ปี พ.ศ./ชื่อเรื่อง	กลุ่มตัวอย่าง	อัลกอริทึมที่ใช้	ผลการวิจัย
โดยนักวิจัย เพชรรัตน์ ม่วงน้อย และคณะ (2021)	ราชภัฏนครปฐม จำนวน 600 คน	4. J48 5. JRIP	Cross- validation ผลการวิจัยพบว่า Random Tree มีประสิทธิภาพสูงสุดโดยมีค่าความถูกต้อง 96.00%
การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์ โดยนักวิจัย วิไลลักษณ์ ตรีพีช (2562)	ข้อมูลความคิดเห็นจากเว็บไซต์ APT TUBE จำนวน 10,000 ประโยค	1.K-means 2. K-NN 3. Decision tree 4. Support Vector Machine (SVM)	ผลการวิจัยพบว่าเทคนิคที่ให้ค่าความถูกต้องมากที่สุดคือ เทคนิค K-means ร่วมกับ K-NN ให้ค่า ความถูกต้องสูงที่สุดที่ 94.80%

บทที่ 3

วิธีการดำเนินงานวิจัย

วิธีการดำเนินงานวิจัยของการเปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูญบุหรืของวัยรุ่นด้วยเทคนิคเหมืองข้อมูลผู้วิจัยได้กำหนดขั้นตอนของการดำเนินงาน ดังนี้

- 3.1 การศึกษาข้อมูลที่เกี่ยวข้อง
- 3.2 รวบรวมคัดเลือก และจัดเตรียมข้อมูล
- 3.3 เครื่องมือที่ใช้ในการเก็บข้อมูล
- 3.4 ทดสอบอัลกอริทึม

3.1 การศึกษาข้อมูลที่เกี่ยวข้อง

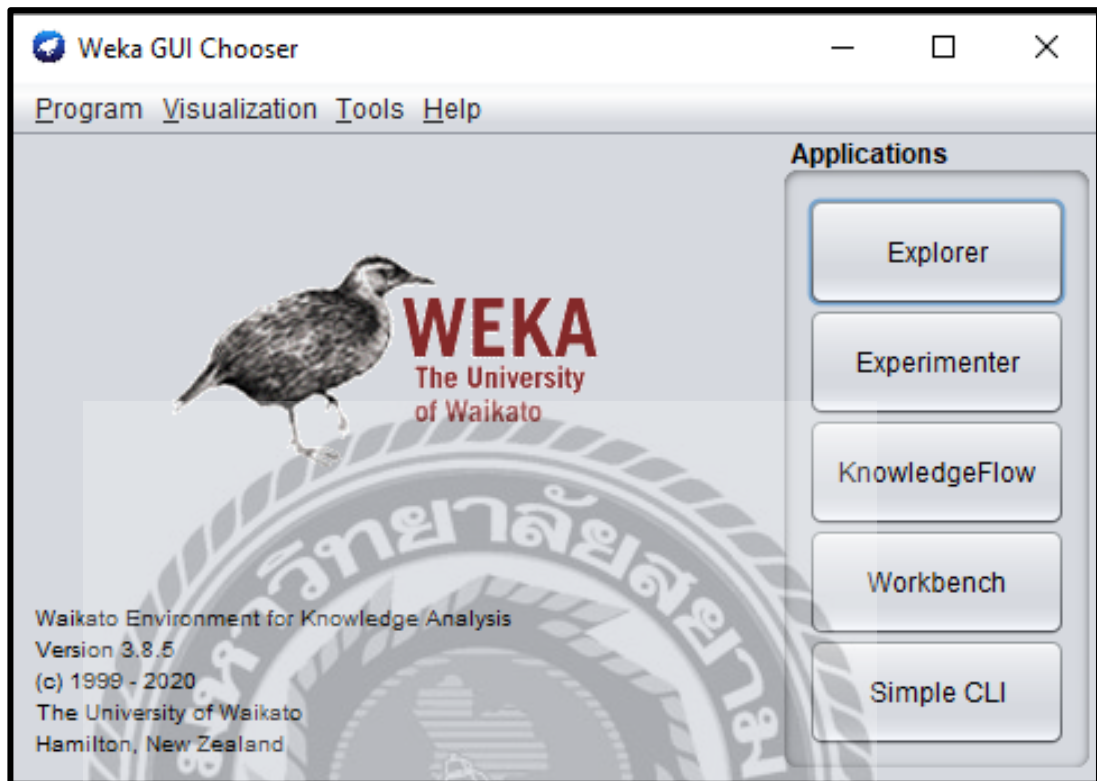
ในการทำวิจัยการเปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูญบุหรืของวัยรุ่นด้วยเทคนิคเหมืองข้อมูลผู้วิจัยต้องศึกษาค้นคว้าข้อมูลและความรู้ที่จะนำมาใช้ในงานวิจัย ดังนี้

3.1.1 ศึกษาและทบทวนวรรณกรรมหรือผลงานวิจัยที่เกี่ยวข้อง (Review Literature) เพื่อเป็นแนวทางในการดำเนินงานวิจัย

3.1.2 ศึกษาขั้นตอนการวิเคราะห์ข้อมูลโดยใช้การจำแนกประเภทข้อมูล (Classification) และอัลกอริทึมของเทคนิคการจำแนกประเภทข้อมูลแต่ละตัว

3.1.2.1 ศึกษาอัลกอริทึมของเทคนิคการจำแนกประเภทข้อมูล (Classification) ที่จะนำมาทดสอบ เพื่อหาอัลกอริทึมที่เหมาะสมในการสร้างโมเดลมากที่สุด จากการศึกษาผู้วิจัยได้เลือกอัลกอริทึมที่จะนำมาทดสอบ ดังนี้ J48, Naive Bayes และ โครงข่ายประสาทเทียม

3.1.2.2 ศึกษาถึงเครื่องมือ หรือซอฟต์แวร์ที่นำมาใช้ในการสร้างโมเดล และทดสอบอัลกอริทึม การวิจัยครั้งนี้ผู้วิจัยได้เลือกใช้เครื่องมือ คือซอฟต์แวร์โปรแกรม Weka ซึ่งเป็นโปรแกรมที่สามารถนำมาทดสอบอัลกอริทึมของการจำแนกประเภทข้อมูล (Classification) ได้เนื่องจากมีอัลกอริทึมที่ได้ระบุไว้ให้เลือกใช้ในโปรแกรมครบตามที่กำหนด และสามารถแสดงค่าความแม่นยำของอัลกอริทึมจาก Confusion Matrix ในโปรแกรม Weka ได้



ภาพที่ 3.1 หน้าจอแสดงแท็บเมนู

3.2 รวบรวม คัดเลือกและจัดเตรียมข้อมูล

ทำการจัดเตรียมข้อมูลเพื่อใช้ในการสร้างโมเดลและทดสอบอัลกอริทึม มีรายละเอียด ดังนี้

3.2.1 กำหนดกลุ่มประชากรที่ใช้ในการวิจัยการเปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูญบุหรืของวัยรุ่น คือนักศึกษามหาวิทยาลัยสยามจำนวนมากกว่า 12,000 คน จากการเก็บข้อมูลจากแบบสอบถามจำนวน 543 คน จากทั้งหมด

Timestamp	L/M/H	วันที่ตอบข้อที่ 1	วันที่ตอบข้อที่ 2	วันที่ตอบข้อที่ 3	วันที่ตอบข้อที่ 4	วันที่ตอบข้อที่ 5	วันที่ตอบข้อที่ 6	วันที่ตอบข้อที่ 7	วันที่ตอบข้อที่ 8	วันที่ตอบข้อที่ 9
9/18/2022 14:24:48	พฤ		22 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 5	2.00-2.49	2.00-2.49		มากกว่า 3.50 ขึ้นไป	ปานกลาง	ปาน
9/18/2022 15:06:47	พฤ		22 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 4	3.00-3.49	3.00-3.49		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/18/2022 15:12:24	ชาว		24 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 5	2.00-2.49	2.50-2.99		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/18/2022 15:15:21	ชาว		25 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 4	3.00-3.49	3.00-3.49		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/18/2022 22:46:31	พฤ		21 คณะบริหารธุรกิจ ระดับปริญญาตรี	ข้อที่ 3	2.50-2.99	2.50-2.99		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/20/2022 13:08:14	ชาว		25 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 1	2.50-2.99	3.00-3.49		มากกว่า 3.50 ขึ้นไป	พียงข้าง	ค่อนข้างดีเยี่ยม
9/23/2022 10:22:02	ชาว		22 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 2	2.50-2.99	2.50-2.99		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/23/2022 10:22:04	ชาว		19 คณะศิลปการศึกษ ระดับปริญญาตรี	ข้อที่ 2	3.00-3.49	3.00-3.49		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/23/2022 10:23:54	ชาว		20 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 1	มากกว่า 3.50 ขึ้นไป	มากกว่า 3.50 ขึ้นไป		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/23/2022 10:25:51	พฤ		19 คณะบริหารธุรกิจ ระดับปริญญาตรี	ข้อที่ 2	2.50-2.99	3.00-3.49		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/23/2022 10:26:21	พฤ		20 คณะบริหารธุรกิจ ระดับปริญญาตรี	ข้อที่ 2	2.50-2.99	3.00-3.49		มากกว่า 3.50 ขึ้นไป	มากกว่า	พอใช้
9/23/2022 10:26:40	พฤ		20 คณะบริหารธุรกิจ ระดับปริญญาตรี	ข้อที่ 2	2.50-2.99	มากกว่า 3.50 ขึ้นไป		มากกว่า 3.50 ขึ้นไป	มากกว่า	พอใช้
9/23/2022 10:31:59	พฤ		20 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 3	3.00-3.49	3.00-3.49		มากกว่า 3.50 ขึ้นไป	มากกว่า	ค่อนข้างดีเยี่ยม
9/23/2022 10:33:44	พฤ		20 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 1	2.50-2.99	3.00-3.49		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/23/2022 10:34:09	ชาว		21 คณะวิศวกรรมศาสตร์ ระดับปริญญาตรี	ข้อที่ 1	มากกว่า 3.50 ขึ้นไป	2.50-2.99		มากกว่า 3.50 ขึ้นไป	มากกว่า	ปาน
9/23/2022 10:35:05	พฤ		20 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 3	2.50-2.99	3.00-3.49		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/23/2022 10:37:37	ชาว		21 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 1	2.00-2.49	2.00-2.49		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/23/2022 10:38:53	ชาว		24 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 3	3.00-3.49	3.00-3.49		มากกว่า 3.50 ขึ้นไป	พียงข้าง	ปาน
9/23/2022 10:45:47	ชาว		18 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 1	3.00-3.49	3.00-3.49		มากกว่า 3.50 ขึ้นไป	พียงข้าง	ปาน
9/23/2022 10:47:00	พฤ		18 คณะบริหารธุรกิจ ระดับปริญญาตรี	ข้อที่ 1	มากกว่า 3.50 ขึ้นไป	2.50-2.99		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/23/2022 10:47:31	ชาว		22 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 1	มากกว่า 3.50 ขึ้นไป	2.00-2.49		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/23/2022 10:47:49	ชาว		19 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 1	2.00-2.49	ค่าที่ 2.00		มากกว่า 3.50 ขึ้นไป	มากกว่า	ปาน
9/23/2022 10:52:54	พฤ		19 คณะบริหารธุรกิจ ระดับปริญญาตรี	ข้อที่ 1	มากกว่า 3.50 ขึ้นไป	มากกว่า 3.50 ขึ้นไป		มากกว่า 3.50 ขึ้นไป	ค่อนข้างดีเยี่ยม	
9/23/2022 10:54:26	ชาว		18 คณะบริหารธุรกิจ ระดับปริญญาตรี	ข้อที่ 1	มากกว่า 3.50 ขึ้นไป	3.00-3.49		มากกว่า 3.50 ขึ้นไป	อยู่ต่ำกว่า	ปาน
9/23/2022 11:16:37	ชาว		19 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 1	2.50-2.99	2.50-2.99		มากกว่า 3.50 ขึ้นไป	มากกว่า	ปาน
9/23/2022 11:16:56	ชาว		18 คณะเทคโนโลยีสารสนเทศ ระดับปริญญาตรี	ข้อที่ 1	2.50-2.99	2.50-2.99		มากกว่า 3.50 ขึ้นไป	พียงข้าง	พอใช้

ภาพที่ 3.2 ข้อมูลที่ยังไม่ผ่านการ Data Cleaning

ข้อมูลที่ยังไม่ผ่านการ Data Cleaning คือ ข้อมูลที่ได้มาหลังจากการตอบแบบสอบถาม

id	age	credit	year	Grade before	Grade after	sex	state	address	With who	Work f	Work m	Smoking f	Smoking m	Income f	Income m	where	house	preference	amount	id
2	1	22	10	5	2	2	3	1	1	1	4	2 no	no	2	2	1	3	3	3	
3	1	22	10	4	4	4	1	1	1	1	1	1 no	no	4	3	1	50	3	3	
4	2	24	10	5	2	3	1	1	1	1	1	1 yes	no	2	2	2	3	3	3	
5	2	25	10	4	4	4	1	1	1	2	2 no	no	5	5	1	3	2	2	2	
6	1	21	4	3	3	3	1	2	1	2	9 no	no	4	4	1	6	62	2	2	
7	2	25	10	1	3	4	2	3	4	1	1 yes	yes	1	1	6	18	1	1	1	
8	2	22	10	2	3	3	1	1	1	6	6 no	no	4	1	1	3	2	2	2	
9	2	19	1	2	4	4	1	1	1	4	5	1 yes	no	2	2	2	3	4	4	
10	2	20	10	1	5	5	1	1	1	7	7 no	no	5	4	1	3	3	3	3	
11	1	19	4	2	3	4	1	1	1	2	2 no	no	5	5	1	3	2	2	2	
12	1	20	4	2	3	4	3	4	2	1	10 yes	no	2	1	1	3	1	1	1	
13	1	20	4	2	3	5	3	1	1	1	1 no	yes	1	1	2	3	2	2	2	
14	1	20	10	3	4	4	3	3	1	1	3 no	yes	1	1	1	3	2	2	2	
15	1	20	10	4	3	4	1	1	1	1	1 no	no	2	2	1	3	2	2	2	
16	2	21	5	1	5	3	3	1	1	4	6 yes	no	5	5	1	99	3	3	3	
17	1	20	10	3	3	4	1	1	1	2	2 no	no	5	1	2	50	4	4	4	
18	2	21	10	1	2	2	1	1	1	5	1 no	no	5	5	1	3	3	3	3	
19	2	24	10	3	4	4	2	1	1	3	1 no	no	5	1	1	22	2	2	2	
20	2	18	10	1	4	4	2	1	1	6	6 yes	no	2	2	1	3	2	2	2	
21	1	18	4	1	5	3	1	1	1	6	1 no	no	2	2	1	3	4	4	4	
22	2	22	10	1	5	2	1	1	1	6	2 no	no	5	4	1	3	4	4	4	
23	2	19	10	1	2	1	3	1	1	10	1 yes	no	1	5	2	3	2	2	2	
24	1	10	4	1	5	5	1	2	1	6	6 no	no	2	1	1	3	2	2	2	

ภาพที่ 3.3 ข้อมูลการสูบบุหรี่ที่ผ่านการ Data Cleaning แล้วแปลงเป็นไฟล์ .csv

คณะผู้วิจัยทำการแปลงข้อมูลไฟล์ Excel ให้เป็นไฟล์ .csv แล้วนำข้อมูลเข้าใน โปรแกรม Weka เพื่อประมวลผลหาค่าการพยากรณ์ความถูกต้อง ในแต่ละอัลกอริทึมที่ใช้

3.2.2 การเก็บรวบรวมข้อมูล ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลโดยใช้แบบสอบถาม โดยแบบสอบถามที่ได้ ได้มาจากการศึกษาข้อมูลจากงานวิจัยต่าง ๆ แบบสอบถามในส่วนที่ 1 ข้อมูลทั่วไป และส่วนที่ 2 โทษและพิษภัยของบุหรี่ ได้ศึกษาค้นคว้าข้อมูลจากงานวิจัยที่สนคดี และความตั้งใจในการเลิกสูบบุหรี่ของนักศึกษา วิทยาลัยราชพฤกษ์ (นนทรี สัจจาธรรม, 2555) และในส่วนที่ 3 เจตคติและค่านิยมต่อการสูบบุหรี่ ได้ศึกษาค้นคว้าข้อมูลจากพฤติกรรมกรรมการสูบบุหรี่ของนักเรียนในช่วงชั้นที่ 3 โรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒ ปทุมวัน (ใจชื่น ตะเกาพงษ์, 2550) พฤติกรรมการสูบบุหรี่ของนักศึกษามหาวิทยาลัยสยามเป็นจำนวน 543 ชุด ตามจำนวนของกลุ่มตัวอย่างที่ได้ระบุไว้ซึ่งคณะผู้วิจัยรวบรวมข้อมูล โดยแจกแบบสอบถามให้กลุ่มตัวอย่างที่สูบบุหรี่และไม่สูบบุหรี่โดยเก็บรวบรวมข้อมูลในมหาวิทยาลัยสยาม

3.3 เครื่องมือที่ใช้ในการเก็บข้อมูล

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลประกอบไปด้วย แบบสอบถามพฤติกรรมกรรมการสูบบุหรี่ภายในแบบสอบถามแบ่งออกเป็น 3 ส่วน ดังนี้

ส่วนที่ 1 ข้อมูลทั่วไป

ส่วนที่ 2 โทษและพิษภัยของบุหรี่

ส่วนที่ 3 เจตคติและค่านิยมต่อการสูบบุหรี่

ตารางที่ 3.1 ข้อมูลทั่วไป ตัวแปรสำหรับการประมวลผลข้อมูลด้วยเทคนิคเหมืองข้อมูล

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
1	Sex	เพศ	1 2	หญิง ชาย
2	Age	อายุ	18 19 20 21 22 23 24 25 26 27 29	อายุ 18 ปี อายุ 19 ปี อายุ 20 ปี อายุ 21 ปี อายุ 22 ปี อายุ 23 ปี อายุ 24 ปี อายุ 25 ปี อายุ 26 ปี อายุ 27 ปี อายุ 29 ปี
3	Faculty	คณะ	1 2 3 4 5 6 7 8 9 10 11	คณะศิลปศาสตร์ คณะนิติศาสตร์ คณะรัฐศาสตร์ คณะบริหารธุรกิจ คณะวิศวกรรมศาสตร์ คณะพยาบาลศาสตร์ คณะเกษตรศาสตร์ คณะสาธารณสุขศาสตร์ คณะวิทยาศาสตร์ คณะเทคโนโลยีสารสนเทศ คณะนิติศาสตร์

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
3	Faculty	คณะ	12 13 14	คณะแพทยศาสตร์ วิทยาลัยดนตรีและ ศิลปะการแสดงวิทยาลัยนานาชาติ
4	EI	ระดับการศึกษา	1 2 3	ระดับปริญญาตรี ระดับปริญญาโท ระดับปริญญาเอก
5	Year	ชั้นปี	1 2 3 4 5 6	ชั้นปีที่1 ชั้นปีที่2 ชั้นปีที่3 ชั้นปีที่4 ชั้นปีที่5 ชั้นปีที่6
6	Grade before	ก่อนเข้าเรียนใน สาขานี้เกรดเฉลี่ย สะสมของท่าน เท่าไร	1 2 3 4 5	ต่ำกว่า 2.00 2.00-2.49 2.50-2.99 3.00-3.49 มากกว่า 3.50 ขึ้นไป
7	Grade after	ระดับคะแนน เฉลี่ยสะสมของ ท่านที่ศึกษาใน ปัจจุบันได้เกรด เฉลี่ยสะสม เท่าไร	1 2 3 4 5	ต่ำกว่า 2.00 2.00-2.49 2.50-2.99 3.00-3.49 มากกว่า 3.50 ขึ้นไป
8	marital status	สถานภาพสมรส ของบิดา-มารดา	1 2 3	อยู่ด้วยกัน หย่าร้าง แยกกันอยู่

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
8	marital status	สถานภาพสมรส ของบิดา-มารดา	4	หม้าย
9	address	สถานที่พักอาศัย ปัจจุบัน	1 2 3 4	บ้าน อพาร์ทเมนท์ คอนโดมิเนียม หอพัก
10	With who	พักอาศัยอยู่กับ ใคร	1 2 3 4 5 6	ครอบครัว บิดา มารดา แฟน เพื่อน อยู่คนเดียว
11	Work f	อาชีพของบิดา	1 2 3 4 5 6 7 8 9 10	อาชีพอิสระ เจ้าของธุรกิจ รัฐวิสาหกิจ ราชการ พนักงานบริษัท เอกชน เล่นหุ้น ชวานา ไม่ได้ทำงาน เสียชีวิต
12	Work m	อาชีพของมารดา	1 2 3 4	อาชีพอิสระ เจ้าของธุรกิจ รัฐวิสาหกิจ ราชการ

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
12	Work m	อาชีพของมารดา	5	พนักงานบริษัท
			6	เอกชน
			7	เล่นหุ้น
			8	ชวานา
			9	พนักงานทำความสะอาด
			10	ไม่ได้ทำงาน
			11	เสียชีวิต
13	Smoking f	บิดาสูบบุหรี่ หรือไม่	Yes	สูบบุหรี่
			No	ไม่สูบบุหรี่
14	Smoking m	มารดาสูบบุหรี่ หรือไม่	Yes	สูบบุหรี่
			No	ไม่สูบบุหรี่
15	Income f	รายได้ของบิดา	1	ต่ำกว่า 15,000 บาท
			2	15,000-20,000 บาท
			3	2,0001- 25,000 บาท
			4	25,001-30,000 บาท
			5	มากกว่า 30,000 บาทขึ้นไป
16	Income m	รายได้ของมารดา	1	ต่ำกว่า 15,000 บาท
			2	15,000-20,000 บาท
			3	2,0001- 25,000 บาท
			4	25,001-30,000 บาท
			5	มากกว่า 30,000 บาทขึ้นไป
17	whose house	บ้านที่ท่านพักอยู่ เป็นของใคร	1	บิดา – มารดา
			2	ญาติ
			3	แฟน
			4	เพื่อน
			5	คนเดียว

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
17	whose house	บ้านที่ท่านพักอยู่ เป็นของใคร	6	บ้านเช่า
18	province	ภูมิลำเนา	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24	จังหวัดอ่างทอง จังหวัดอำนาจเจริญ กรุงเทพมหานคร จังหวัดบึงกาฬ จังหวัดบุรีรัมย์ จังหวัดฉะเชิงเทรา จังหวัดชัยนาท จังหวัดชัยภูมิ จังหวัดจันทบุรี จังหวัดเชียงใหม่ จังหวัดเชียงราย จังหวัดชลบุรี จังหวัดชุมพร จังหวัดกาฬสินธุ์ จังหวัดกาญจนบุรี จังหวัดขอนแก่น จังหวัดเลย จังหวัดลพบุรี จังหวัดมหาสารคาม จังหวัดมุกดาหาร จังหวัดนครนายก จังหวัดนครปฐม จังหวัดนครราชสีมา จังหวัดนครสวรรค์

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัส ข้อมูล	รายละเอียด
18	Province	ภูมิภาค	25	จังหวัดนครศรีธรรมราช
			26	จังหวัดน่าน
			27	จังหวัดหนองบัวลำภู
			28	จังหวัดหนองคาย
			29	จังหวัดนนทบุรี
			30	จังหวัดปัตตานี
			31	จังหวัดปทุมธานี
			32	จังหวัดพังงา
			33	จังหวัดพัทลุง
			34	จังหวัดพะเยา
			35	จังหวัดเพชรบูรณ์
			36	จังหวัดเพชรบุรี
			37	จังหวัดพิจิตร
			38	จังหวัดพิษณุโลก
			39	จังหวัดแพร่
			40	จังหวัดพระนครศรีอยุธยา
			41	จังหวัดปราจีนบุรี
			42	จังหวัดประจวบคีรีขันธ์
			43	จังหวัดภูเก็ต
			44	จังหวัดระนอง
			45	จังหวัดราชบุรี
			46	จังหวัดระยอง
			47	จังหวัดร้อยเอ็ด
			48	จังหวัดสระแก้ว
			49	จังหวัดสกลนคร
			50	จังหวัดสมุทรปราการ

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
18	province	ภูมิลำเนา	51	จังหวัดสมุทรสาคร
			52	จังหวัดสระบุรี
			53	จังหวัดสตูล
			54	จังหวัดสิงห์บุรี
			55	จังหวัดศรีสะเกษ
			56	จังหวัดสงขลา
			57	จังหวัดสุโขทัย
			58	จังหวัดสุพรรณบุรี
			59	จังหวัดสุราษฎร์ธานี
			60	จังหวัดสุรินทร์
			61	จังหวัดตรัง
			62	จังหวัดอุบลราชธานี
			63	จังหวัดอุดรธานี
			64	จังหวัดอุตรดิตถ์
			65	จังหวัดยะลา
			66	จังหวัดยโสธร
			67	จังหวัดกระบี่
			68	จังหวัดกำแพงเพชร
			69	จังหวัดตราด
			70	จังหวัดตาก
			71	จังหวัดนครพนม
			72	จังหวัดนครราชสีมา
			73	จังหวัดแม่ฮ่องสอน
			74	จังหวัดลำปาง
			75	จังหวัดลำพูน
			76	จังหวัดสมุทรสงคราม

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
18	Province	ภูมิลำเนา	77	จังหวัดอุทัยธานี
19	amount	จำนวนสมาชิก ภายในครอบครัว	1 2 3 4	1-2 คน 3-4 คน 5-6 คน มากกว่า 6 คนขึ้นไป
20	stay with	พักอาศัยอยู่กับ ใครมากที่สุด	1 2 3 4 5	ครอบครัว พี่น้อง ญาติ คนที่รู้จัก เพื่อน
21	Smoking friend	เพื่อนสูบบุหรี่	m n	สูบบุหรี่ ไม่สูบบุหรี่
22	Smoking you	คุณสูบบุหรี่	yes no	สูบบุหรี่ ไม่สูบบุหรี่

ตารางที่ 3.2 โทษและพิษภัยของบุหรี่

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
1	Smoking causes cancer	การสูบบุหรี่เป็นสาเหตุทำให้ เกิดโรคมะเร็งที่อวัยวะ ต่าง ๆ ของร่างกาย	2 3 4	เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย
2	Smoking causes blood vessels in the brain to rupture.	การสูบบุหรี่เป็นสาเหตุทำให้ เส้นเลือดในสมองแตก และเป็นอัมพาตได้	2 3 4	เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
3	Nicotine	สารพิษในควันบุหรี่ที่ทำให้ผู้สูบมีอาการติดยาหรือคือสารนิโคติน	2 3 4	เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย
4	emphysema	โรคถุงลมโป่งพองเกิดจากการสูบบุหรี่ทำให้ถุงลมปอดถูกทำลายและมีอาการหืดเรื้อรังของหลอดลม	2 3 4	เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย
5	cigarette butt	ก้นกรองของบุหรี่ไม่ได้ทำให้ปริมาณของสารนิโคตินในบุหรีลดลง	2 3 4	เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย
6	Lead	ในควันบุหรีมีสารตะกั่วซึ่งเป็นสารที่ยับยั้งการเจริญเติบโตของร่างกายและทำลายเนื้อเยื่อสมอง	2 3 4	เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย
7	secondhand smoke	ควันบุหรีที่ออกจากผู้สูบได้ทำร้ายคนข้างเคียงเรียกว่า “ควันบุหรีมือสอง”	2 3 4	เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย
8	Non-smoking area	ตามประกาศกระทรวงสาธารณสุข พ.ศ. 2551 กำหนดให้แสดงเครื่องหมายของเขตสูบบุหรี และเขตปลอดบุหรีในที่เปิดเผย และสามารถมองเห็นได้ ชัดเจน	2 3 4	เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย
9	violator	ผู้ที่ฝ่าฝืนสูบบุหรีในเขตพื้นที่ปลอดบุหรีมีโทษปรับไม่เกิน 2,000 บาท	2 3 4	เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
10	place owner	เจ้าของสถานที่ ถ้าไม่จัดให้ มีเครื่องหมายเขตปลอด บุหรี่ มีโทษปรับไม่เกิน 2,000 บาท	2 3 4	เห็นด้วย ไม่เห็นด้วย ไม่เห็นด้วย

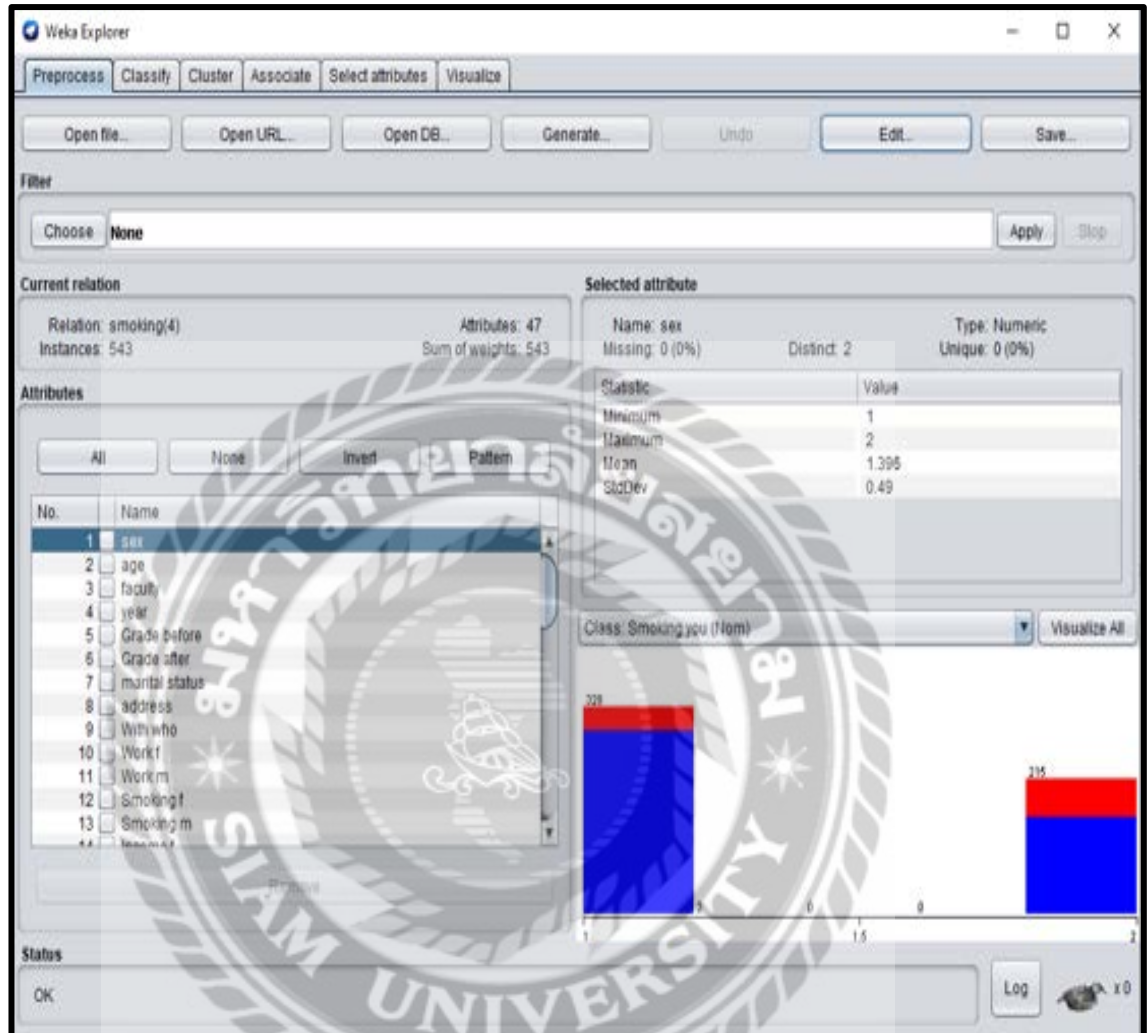
ตารางที่ 3. 3 เจตคติและค่านิยมต่อการสูบบุหรี่

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
1	unusual behavior	การสูบบุหรี่เป็นพฤติกรรม ไม่ปกติในสังคมไทย	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่เห็นด้วย ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
2	man smoking	ผู้ชายสูบบุหรี่เป็นพฤติกรรม ไม่ปกติ	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่เห็นด้วย ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
3	smoking woman	ผู้หญิงสูบบุหรี่เป็นพฤติกรรม ไม่เหมาะสม	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่เห็นด้วย ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
4	adult smoking	ผู้ใหญ่สูบบุหรี่เป็นพฤติกรรมที่ไม่ควรยอมรับ	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
5	smoking teen	วัยรุ่นสูบบุหรี่เป็นพฤติกรรมที่ไม่เหมาะสม	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
6	not modern	การสูบบุหรี่ไม่แสดงถึงการเป็นคนทันสมัยเท่าที่	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
7	not accepted	การสูบบุหรี่ไม่ทำให้คุณเป็นที่ยอมรับของเพื่อนๆ	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
8	inimitable	การสูบบุหรี่เป็นพฤติกรรมที่ไม่น่าเลียนแบบ	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
9	Drugs	บุหรี่เป็นสิ่งเสพติด ทำลายสุขภาพ	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
10	dangerous drug	บุหรี่เป็นสิ่งเสพติดที่อันตราย เพราะจะนำไปสู่สิ่งเสพติดอื่น ๆ	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
11	hate friends	ท่านรังเกียจเพื่อนหรือบุคคลที่สูบบุหรี่	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
12	disgusting friend	ถ้าท่านสูบบุหรี่เพื่อนหรือคนรอบข้างรังเกียจ	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
13	smoking group	ถ้าท่านอยู่ในกลุ่มที่สูบบุหรี่ ท่านจำเป็นต้องสูบด้วย	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง

ลำดับ	ชนิดข้อมูล	ชื่อค่าข้อมูล	รหัสข้อมูล	รายละเอียด
14	Waste	การสูบบุหรี่ทำให้สิ้นเปลือง ค่าใช้จ่ายในการรักษาโรคที่มี สาเหตุจากการสูบบุหรี่โดย ไม่จำเป็น	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง
15	lose personality	การเลิกบุหรี่ทำให้เสีย บุคลิกภาพของตนเอง	1 2 3 4 5	เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง



ภาพที่ 3.4 หน้าจอแสดงการนำข้อมูลเข้า Weka

นำไฟล์ข้อมูลมาประมวลผลด้วยโปรแกรม Weka โดยใช้ Smoking you เป็น Attributes as Class ทำการเปลี่ยนข้อมูล Attributes Age ให้เป็น Numeric และทำการจัดการเปลี่ยนข้อมูล Attributes ทั้งหมดให้เป็น Nominal ยกเว้น Smoking you กับ Age ที่ทำการเปลี่ยนข้อมูลไปแล้ว ดังภาพที่ 3.5

Viewer

Relation: smoking(4)-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,4...

No.	1: sex	2: age	3: faculty	4: year	5: Grade before	6: Grade after	7: marital status	8: address	9: With who	10: Work f	11: Work m	12: Smoking f	13: Smoking m	14: Income
	Nominal	Numeric	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	1	4.0	10	5	2	2	3	1	1	4	2	no	no	2
2	1	4.0	10	4	4	4	1	1	1	1	1	no	no	4
3	2	6.0	10	5	2	3	1	1	1	1	1	yes	no	2
4	2	7.0	10	4	4	4	1	1	1	2	2	no	no	5
5	1	3.0	4	3	3	3	1	2	1	2	9	no	no	4
6	2	7.0	10	1	3	4	2	3	4	1	1	yes	yes	1
7	2	4.0	10	2	3	3	1	1	1	6	6	no	no	4
8	2	1.0	1	2	4	4	1	1	1	5	1	yes	no	2
9	2	2.0	10	1	5	5	1	1	1	7	7	no	no	5
10	1	1.0	4	2	3	4	1	1	1	2	2	no	no	5
11	1	2.0	4	2	3	4	3	4	2	1	10	yes	no	2
12	1	2.0	4	2	3	5	3	1	1	1	1	no	yes	1
13	1	2.0	10	3	4	4	3	3	1	1	3	no	yes	1
14	1	2.0	10	4	3	4	1	1	1	1	1	no	no	2
15	2	3.0	5	1	5	3	3	1	1	4	6	yes	no	5
16	1	2.0	10	3	3	4	1	1	1	2	2	no	no	5
17	2	3.0	10	1	2	2	1	1	1	5	1	no	no	5
18	2	6.0	10	3	4	4	2	1	1	3	1	no	no	5
19	2	0.0	10	1	4	4	2	1	1	6	6	yes	no	2
20	1	0.0	4	1	5	3	1	1	1	6	1	no	no	2
21	2	4.0	10	1	5	2	1	1	1	6	2	no	no	5
22	2	1.0	10	1	2	1	3	1	1	10	1	yes	no	1
23	1	1.0	4	1	5	5	1	3	1	6	6	no	no	2
24	2	0.0	4	1	5	4	1	1	1	2	9	no	no	5
25	2	1.0	10	1	3	3	3	1	1	1	1	no	no	3
26	2	0.0	10	1	3	3	2	4	3	4	4	no	no	5
27	2	0.0	10	1	3	4	2	4	1	6	6	yes	no	2
28	2	0.0	10	1	3	4	2	4	1	6	6	yes	no	2
29	2	0.0	10	1	5	5	1	1	1	1	1	no	no	2

Add instance Undo OK Cancel

ภาพที่ 3.5 หน้าจอแสดงการประเภทข้อมูล

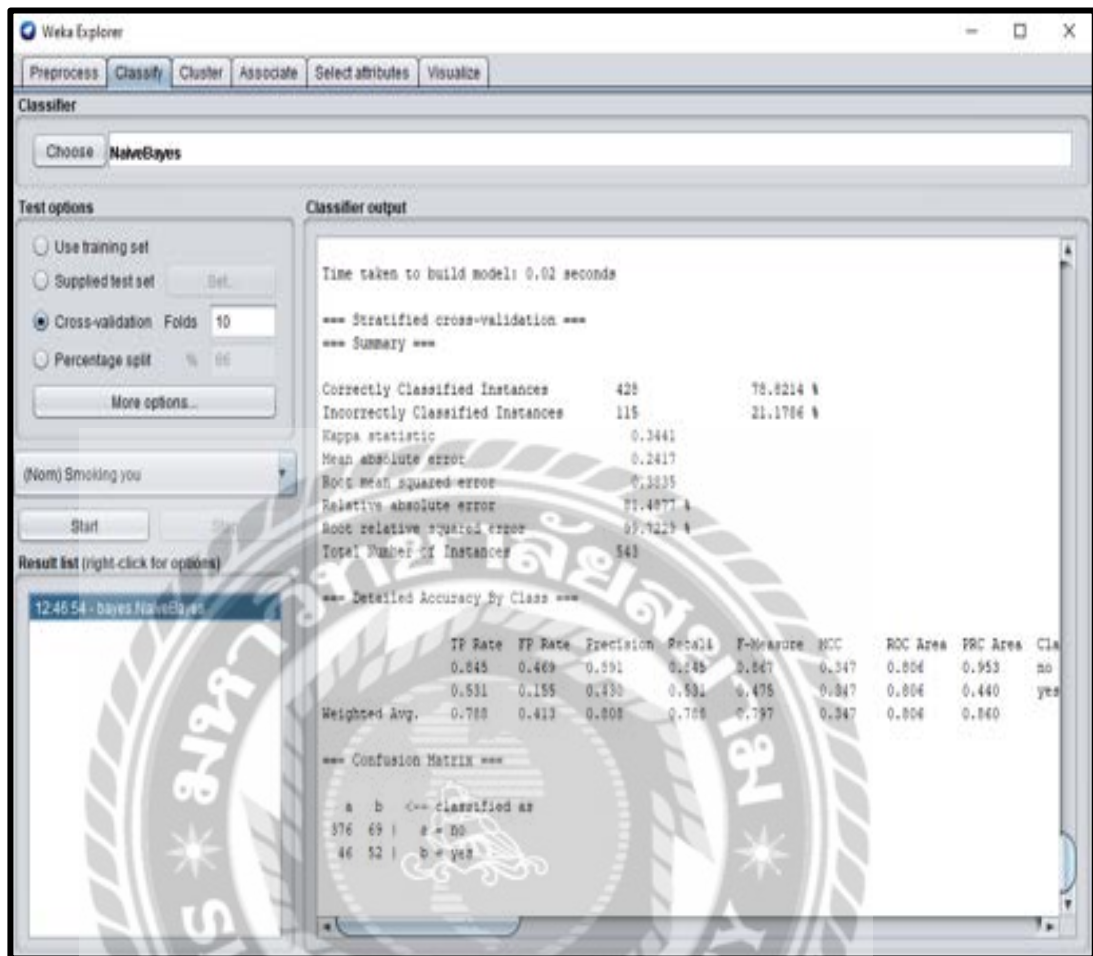
3.4 ทดสอบอัลกอริทึม

หลังจากขั้นตอนการรวบรวมทดสอบและจัดเตรียมข้อมูลเสร็จสิ้นจะได้ข้อมูลซึ่งอยู่ในรูปแบบฐานข้อมูล จากนั้นจะเข้าสู่ขั้นตอนการทดสอบหาอัลกอริทึม โดยมีวิธีการ ดังนี้

3.4.1. การเตรียมโมเดลจะประกอบด้วย การทดลองใช้โมเดล ซึ่งในการทดสอบจะใช้โปรแกรม โดยผู้วิจัยได้ทำการเลือกอัลกอริทึมของการจำแนกประเภทข้อมูลไว้ซึ่งได้แก่ J48, Naive Bayes และ โครงข่ายประสาทเทียม เพื่อนำมาทดสอบ และใช้วิธี 10-fold cross validation เพื่อใช้ในการทดสอบประสิทธิภาพของโมเดล Cross – validation นี้จะแบ่งข้อมูลเป็น 10 ส่วน โดยแต่ละส่วนมีจำนวนข้อมูลเท่ากันหลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดลทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้

3.4.2. การกำหนดพื้นที่การทดลองใช้โมเดล โดยใช้กลุ่มตัวอย่าง คือ นักศึกษามหาวิทยาลัยสยาม ที่สูบบุหรี่ และไม่สูบบุหรี่

3.4.3. สร้างแบบจำลอง (Model Building) นำข้อมูลในฐานข้อมูลมาสร้างการแบ่งประเภท โดยใช้เหมือนข้อมูลนำกลุ่มตัวอย่างที่ใช้ในการทดสอบทำแบบสอบถาม จำนวน 543 ชุด มาทำการประมวลผลเพื่อหาค่าความแม่นยำและถูกต้อง การทดสอบอัลกอริทึมแต่ละอัลกอริทึมจะทำการทดสอบ 3 ครั้ง โดยแบ่งข้อมูลในการสร้างโมเดลเป็น 3 ชุด ชุดที่ 1 ประกอบด้วยข้อมูลทั่วไป ชุดที่ 2 ประกอบด้วยข้อมูลทั่วไป, ข้อมูลโทษและพิษของบุหรี่ และชุดที่ 3 ประกอบด้วยข้อมูลทั่วไป, ข้อมูลโทษและพิษของบุหรี่, ข้อมูลเจตคติและค่านิยมต่อการสูบบุหรี่ มีขั้นตอน ดังนี้



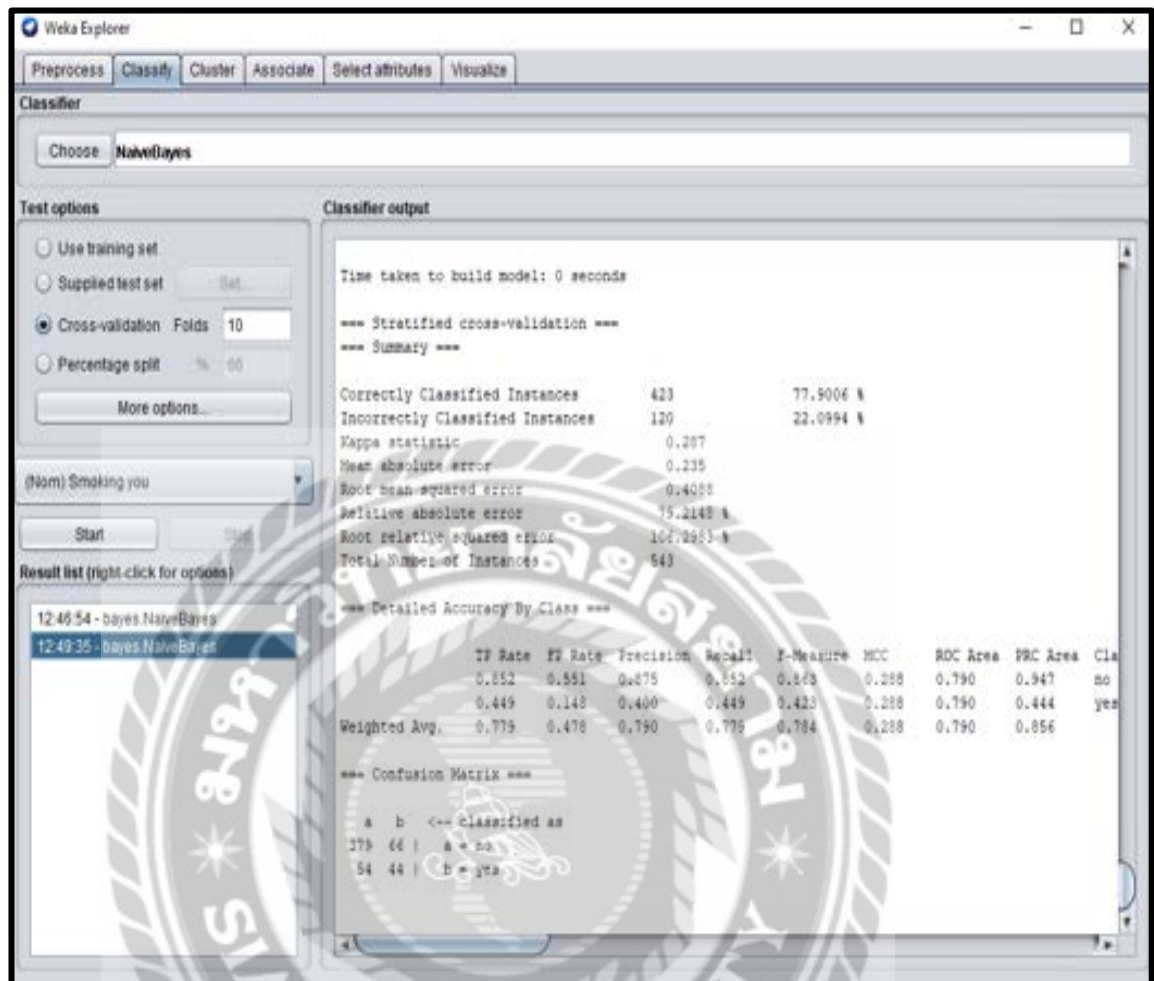
ภาพที่ 3.6 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 1 โดยใช้อัลกอริทึม Naïve Bayes

3.4.3.1 ส่วนวัดประสิทธิภาพของโมเดล (Evaluation on training set) สามารถพิจารณาผลได้ดังนี้

1) Correctly Classified Instances จากข้อมูล 543 Instances โมเดลสามารถทำนายข้อมูลถูกต้อง 428 Instances หรือคิดเป็น 78.82% ของทั้งหมด

2) Incorrectly Classified Instances จากข้อมูล 543 Instances โมเดลทำนายข้อมูลไม่ถูกต้อง 115 Instances หรือคิดเป็น 21.18% ของทั้งหมด

3) Root Mean Squared Error (RMSE) ค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.3835



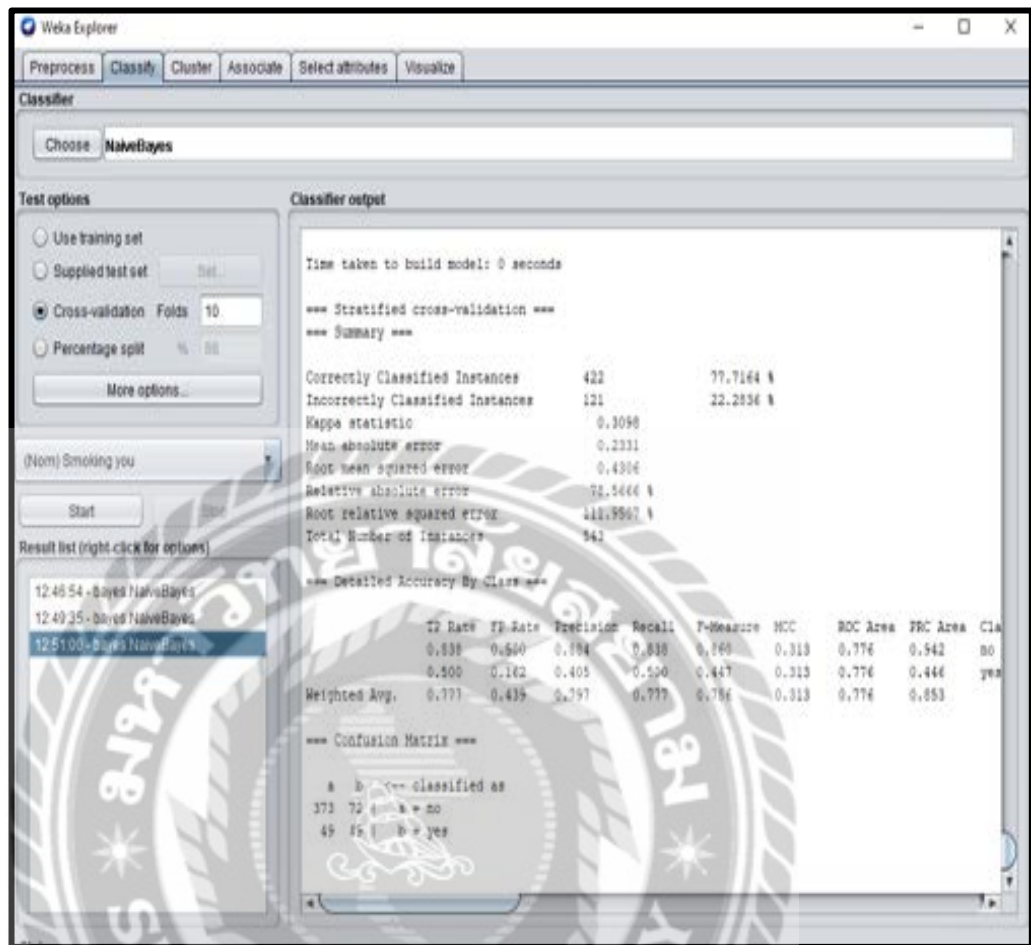
ภาพที่ 3.7 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 2 โดยใช้อัลกอริทึม Naïve Bayes

3.4.3.2 ส่วนวัดประสิทธิภาพของโมเดล (Evaluation on training set) สามารถพิจารณาผลได้ดังนี้

1) Correctly Classified Instances จากข้อมูล 543 Instances โมเดลสามารถทำนายข้อมูลถูกต้อง 423 Instances หรือคิดเป็น 77.90% ของทั้งหมด

2) Incorrectly Classified Instances จากข้อมูล 543 Instances โมเดลทำนายข้อมูลไม่ถูกต้อง 120 Instances หรือคิดเป็น 22.10% ของทั้งหมด

3) Root Mean Squared Error (RMSE) ค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.4088



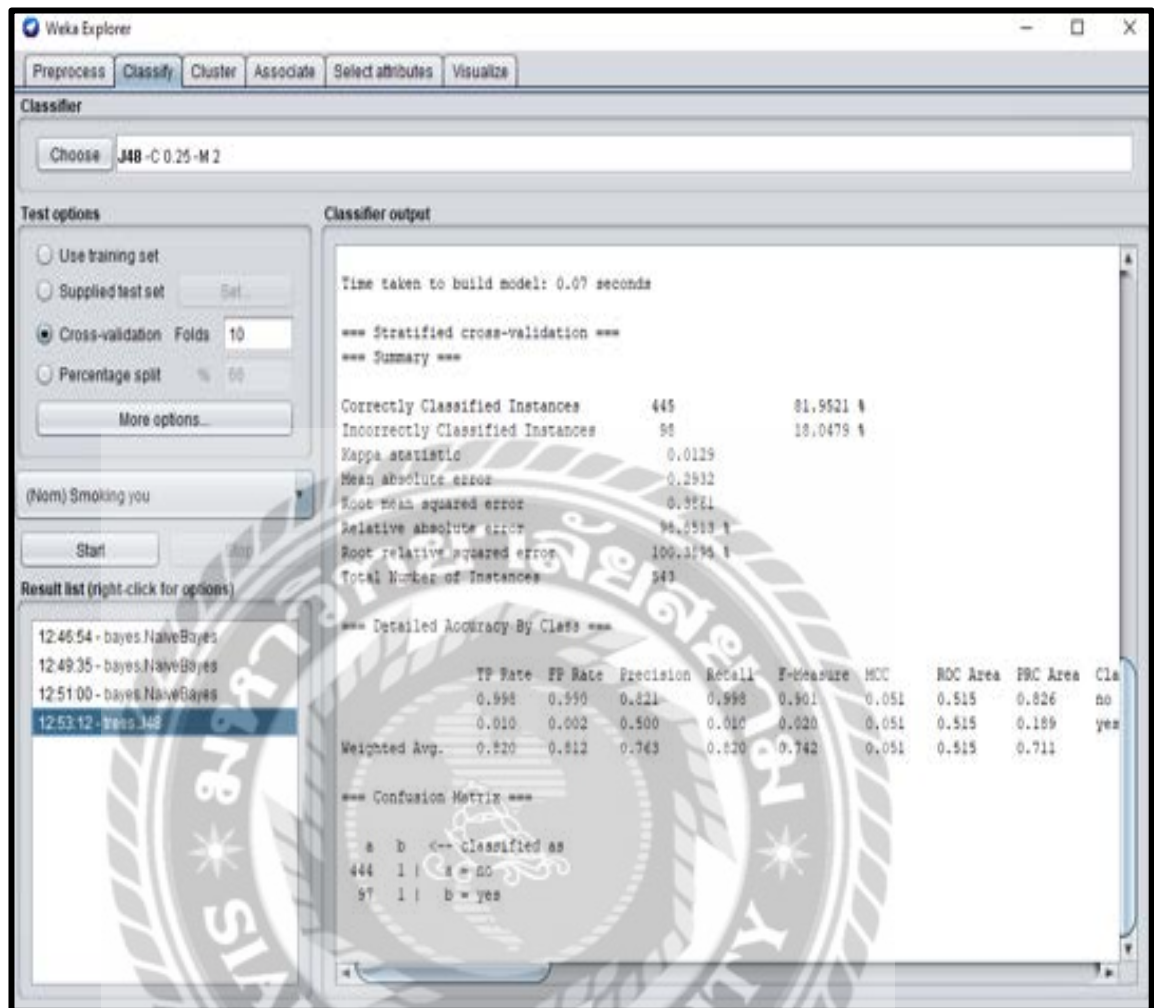
ภาพที่ 3.8 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 3 โดยใช้อัลกอริทึม Naive Bayes

3.4.3.3 ส่วนวัดประสิทธิภาพของโมเดล (Evaluation on training set) สามารถพิจารณาผลได้ดังนี้

1) Correctly Classified Instances จากข้อมูล 543 Instances โมเดลสามารถทำนายข้อมูลถูกต้อง 422 Instances หรือคิดเป็น 77.71% ของทั้งหมด

2) Incorrectly Classified Instances จากข้อมูล 543 Instances โมเดลทำนายข้อมูลไม่ถูกต้อง 121 Instances หรือคิดเป็น 22.28% ของทั้งหมด

3) Root Mean Squared Error (RMSE) ค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.4306



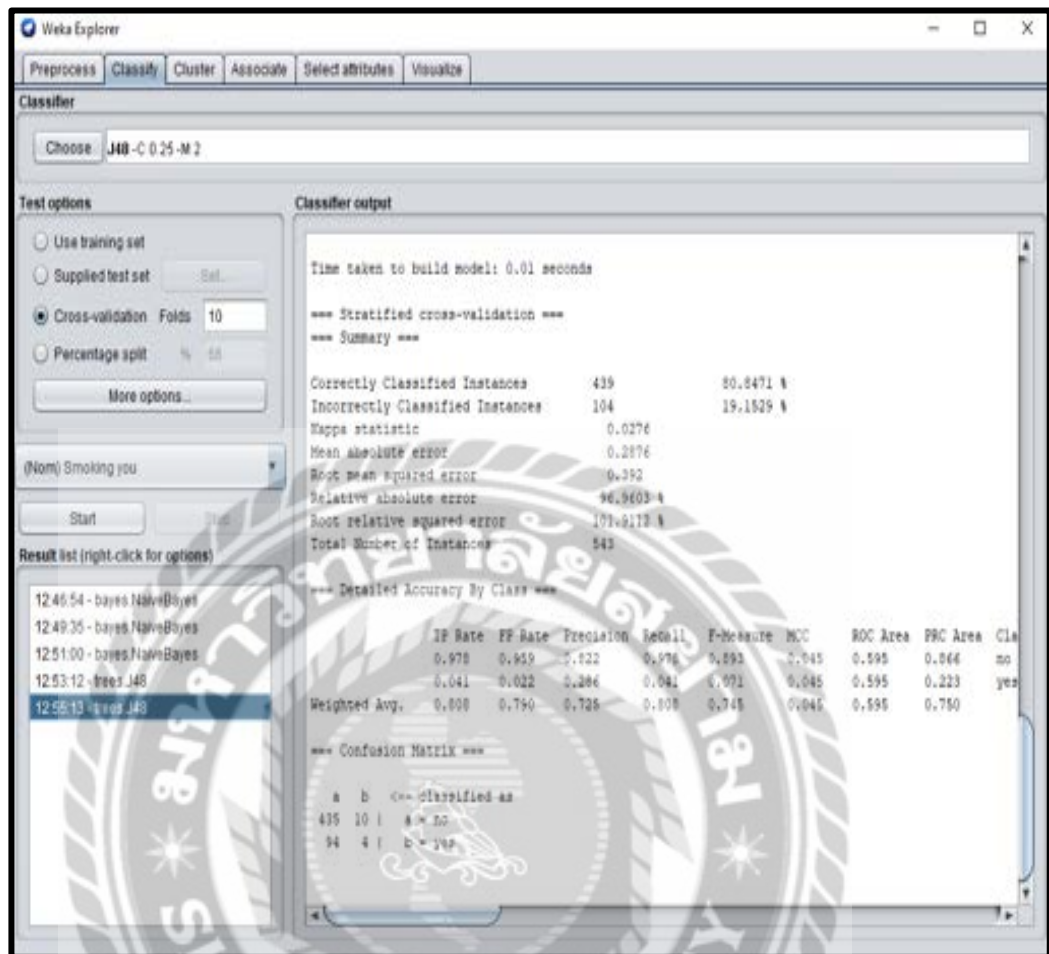
ภาพที่ 3.9 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 1 โดยใช้อัลกอริทึม J48

3.4.3.4 ส่วนวัดประสิทธิภาพของโมเดล (Evaluation on training set) สามารถพิจารณาผลได้ดังนี้

1) Correctly Classified Instances จากข้อมูล 543 Instances โมเดลสามารถทำนายข้อมูลถูกต้อง 445 Instances หรือคิดเป็น 81.95% ของทั้งหมด

2) Incorrectly Classified Instances จากข้อมูล 543 Instances โมเดลทำนายข้อมูลไม่ถูกต้อง 98 Instances หรือคิดเป็น 18.05% ของทั้งหมด

3) Root Mean Squared Error (RMSE) ค่าความคลาดเคลื่อนระหว่างค่าจริงและค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.3861



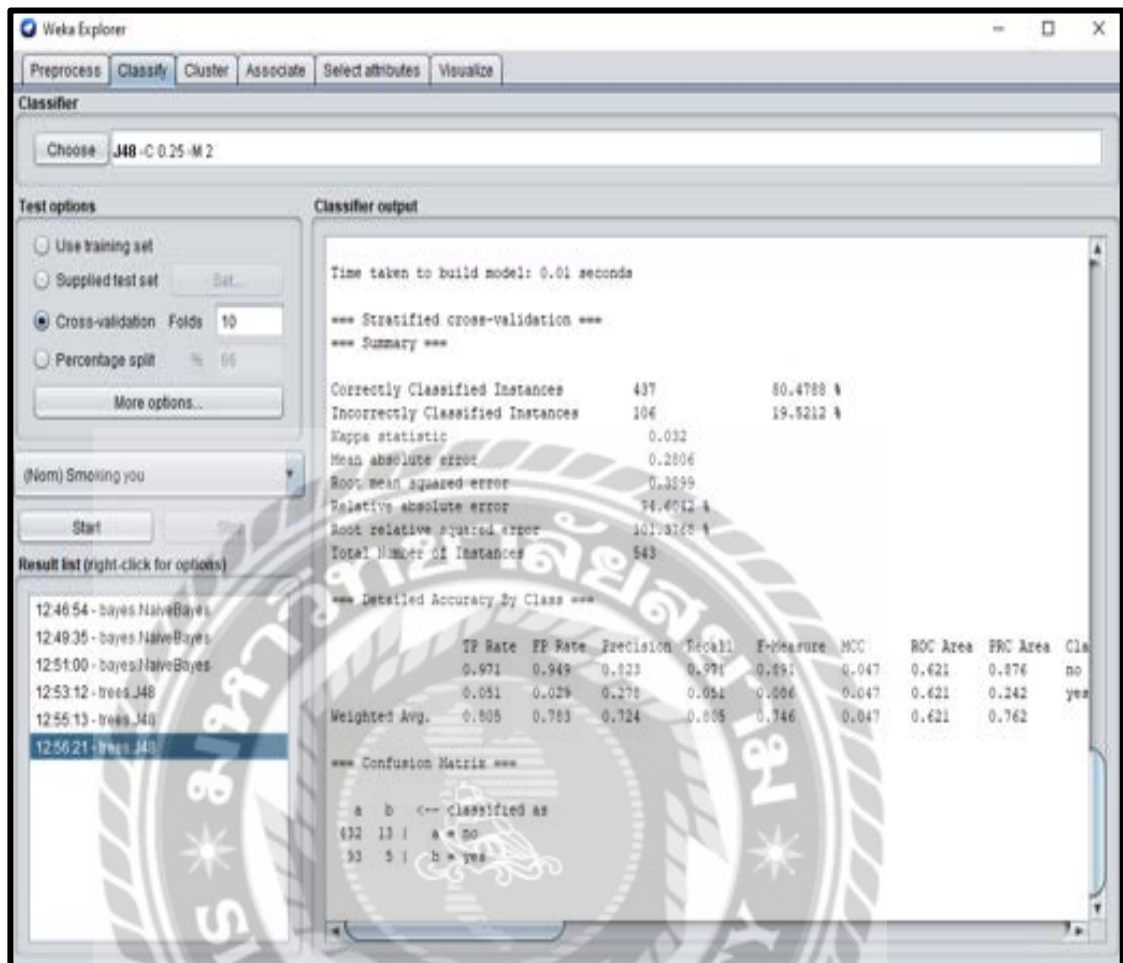
ภาพที่ 3.10 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 2 โดยใช้อัลกอริทึม J48

3.4.3.5 ส่วนวัดประสิทธิภาพของโมเดล (Evaluation on training set) สามารถพิจารณาผลได้ดังนี้

1) Correctly Classified Instances จากข้อมูล 543 Instances โมเดลสามารถทำนายข้อมูลถูกต้อง 439 Instances หรือคิดเป็น 80.85% ของทั้งหมด

2) Incorrectly Classified Instances จากข้อมูล 543 Instances โมเดลทำนายข้อมูลไม่ถูกต้อง 104 Instances หรือคิดเป็น 19.15% ของทั้งหมด

3) Root Mean Squared Error (RMSE) ค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.392



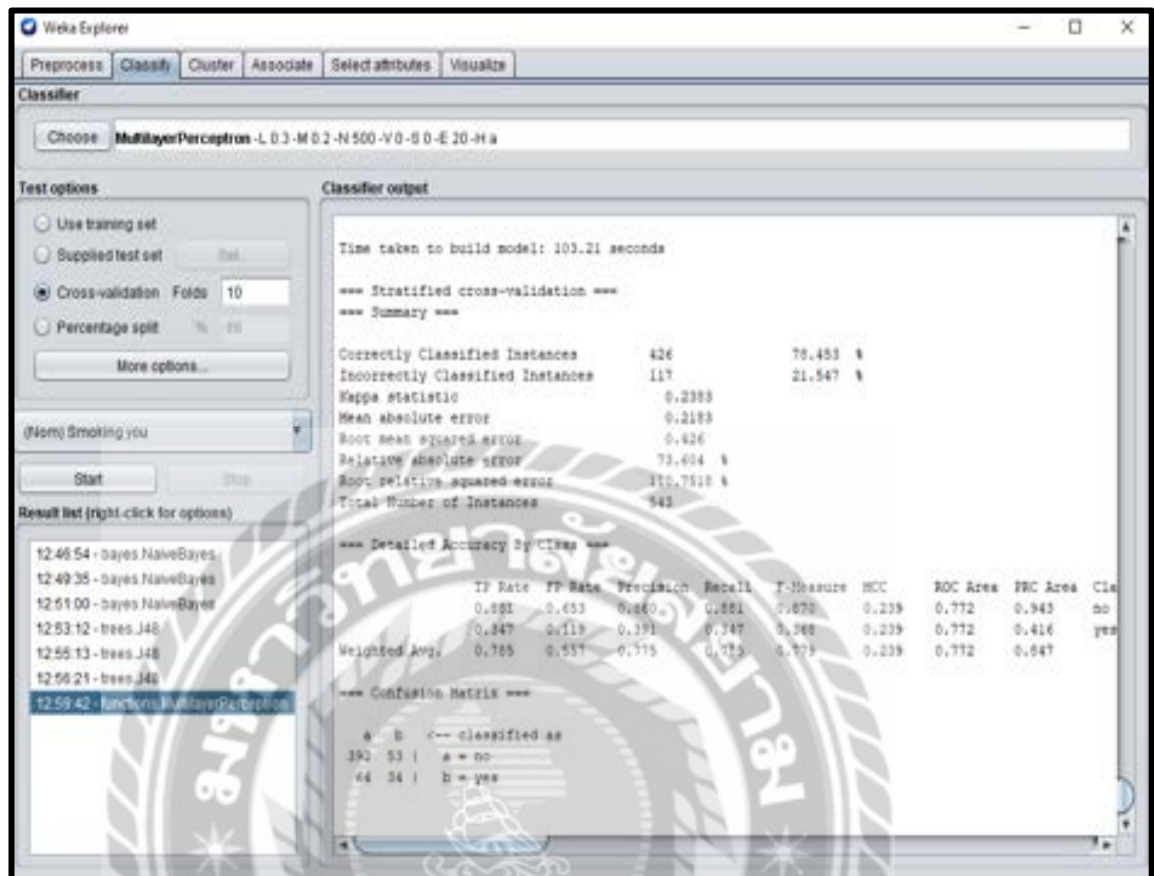
ภาพที่ 3.11 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 3 โดยใช้อัลกอริทึม J48

3.4.3.6 ส่วนวัดประสิทธิภาพของโมเดล (Evaluation on training set) สามารถพิจารณาผลได้ดังนี้

1) Correctly Classified Instances จากข้อมูล 543 Instances โมเดลสามารถทำนายข้อมูลถูกต้อง 437 Instances หรือคิดเป็น 80.48% ของทั้งหมด

2) Incorrectly Classified Instances จากข้อมูล 543 Instances โมเดลทำนายข้อมูลไม่ถูกต้อง 106 Instances หรือคิดเป็น 19.52% ของทั้งหมด

3) Root Mean Squared Error (RMSE) ค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.3899



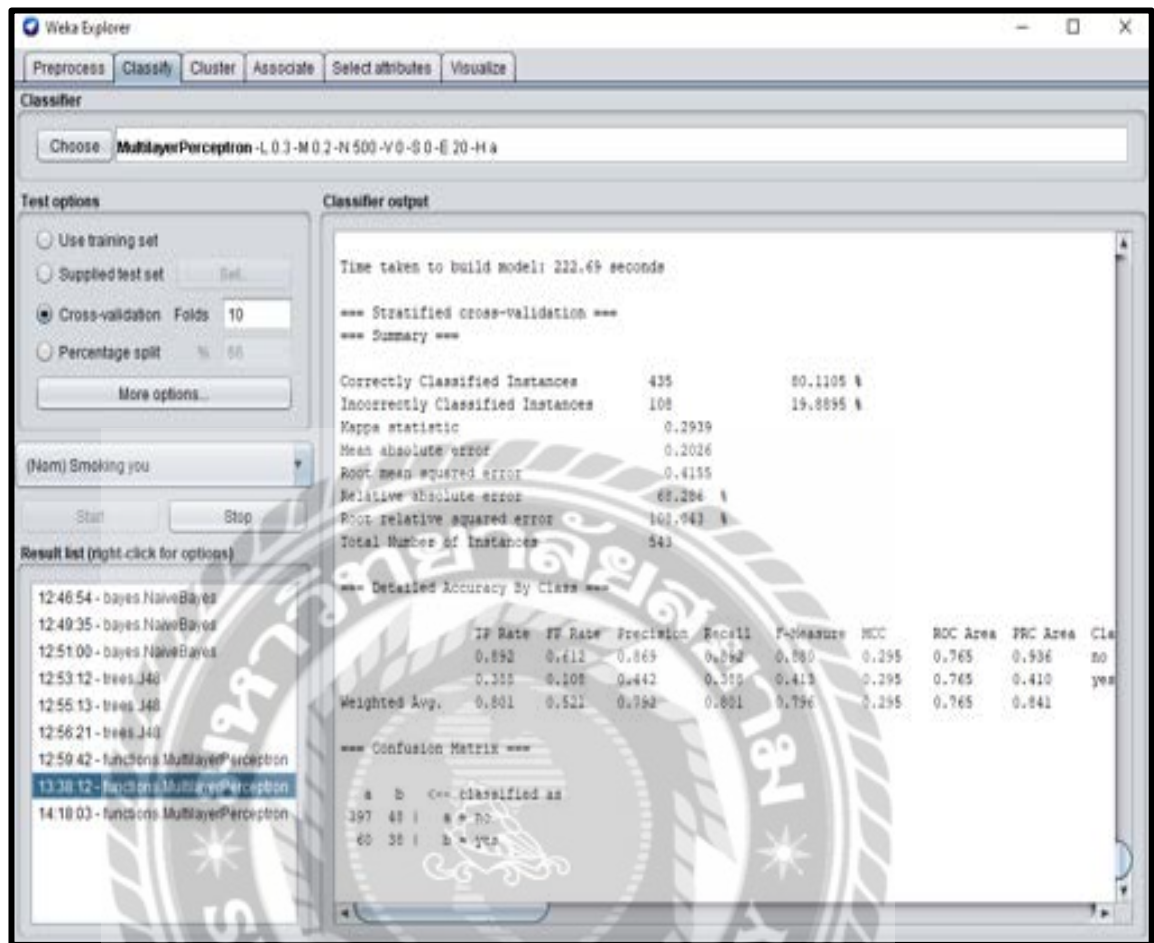
ภาพที่ 3.12 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 1 โดยใช้อัลกอริทึมโครงข่ายประสาทเทียม

3.4.3.7 ส่วนวัดประสิทธิภาพของโมเดล (Evaluation on training set) สามารถพิจารณาผลได้ดังนี้

1) Correctly Classified Instances จากข้อมูล 543 Instances โมเดลสามารถทำนายข้อมูลถูกต้อง 426 Instances หรือคิดเป็น 78.45% ของทั้งหมด

2) Incorrectly Classified Instances จากข้อมูล 543 Instances โมเดลทำนายข้อมูลไม่ถูกต้อง 117 Instances หรือคิดเป็น 21.55% ของทั้งหมด

3) Root Mean Squared Error (RMSE) ค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.426



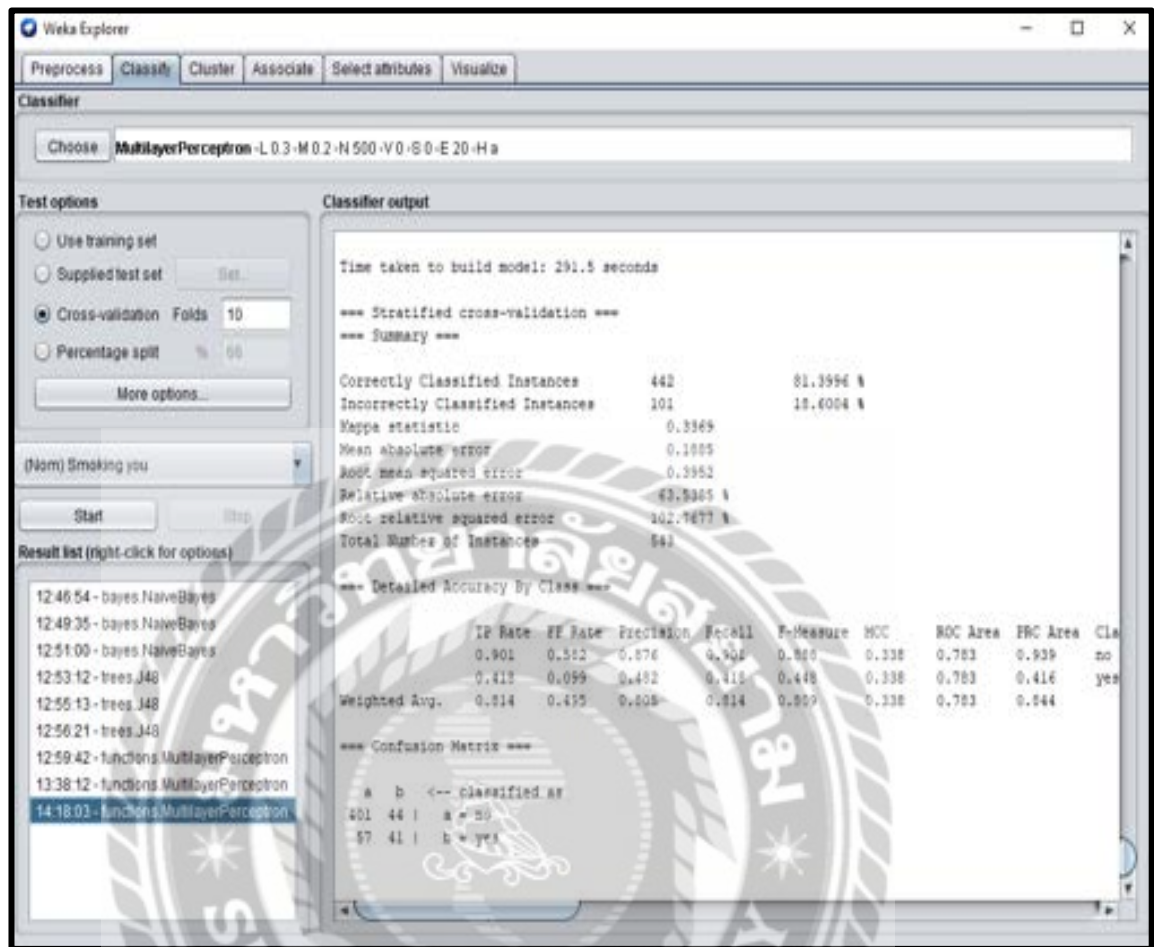
ภาพที่ 3.13 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 2 โดยใช้อัลกอริทึมโครงข่ายประสาทเทียม

3.4.3.8 ส่วนวัดประสิทธิภาพของโมเดล (Evaluation on training set) สามารถพิจารณาผลได้ดังนี้

1) Correctly Classified Instances จากข้อมูล 543 Instances โมเดลสามารถทำนายข้อมูลถูกต้อง 435 Instances หรือคิดเป็น 80.11% ของทั้งหมด

2) Incorrectly Classified Instances จากข้อมูล 543 Instances โมเดลทำนาย ข้อมูลไม่ถูกต้อง 108 Instances หรือคิดเป็น 19.89% ของทั้งหมด

3) Root Mean Squared Error (RMSE) ค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.4155



ภาพที่ 3.14 ส่วนวัดประสิทธิภาพโมเดลด้วยข้อมูลชุดที่ 3 โดยใช้อัลกอริทึมโครงข่ายประสาทเทียม

3.4.3.9 ส่วนวัดประสิทธิภาพของโมเดล (Evaluation on training set) สามารถพิจารณาผลได้ดังนี้

1) Correctly Classified Instances จากข้อมูล 543 Instances โมเดลสามารถทำนายข้อมูลถูกต้อง 435 Instances หรือคิดเป็น 81.40% ของทั้งหมด

2) Incorrectly Classified Instances จากข้อมูล 543 Instances โมเดลทำนายข้อมูลไม่ถูกต้อง 108 Instances หรือคิดเป็น 18.60% ของทั้งหมด

3) Root Mean Squared Error (RMSE) ค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่พยากรณ์ได้ มีค่าเท่ากับ 0.3952

บทที่ 4

ผลของการวิจัย

งานวิจัยในครั้งนี้ผู้วิจัยได้ใช้เทคนิค J48, Naive Bayes และ โครงข่ายประสาทเทียม ในการ จำแนกเพื่อทำนายข้อมูลการสูบบุหรี่ของวัยรุ่นในมหาวิทยาลัยสยาม กลุ่มตัวอย่างทั้งหมด 543 คน ซึ่ง จะกล่าวถึงผลของการวิจัย ดังนี้

4.1 ผลการวิเคราะห์ข้อมูล ผลการวิเคราะห์เกี่ยวกับข้อมูลส่วนบุคคลและข้อมูลทั่วไปเกี่ยวกับ โทษและค่านิยมต่อการสูบบุหรี่ของผู้ตอบแบบสอบถามการวิเคราะห์ข้อมูล แบ่งออกเป็น 3 ส่วน ดังนี้

ส่วนที่ 1 ข้อมูลทั่วไป

ส่วนที่ 2 โทษและพิษภัยของบุหรี่

ส่วนที่ 3 เจตคติและค่านิยมต่อการสูบบุหรี่

ส่วนที่ 1 คุณลักษณะกลุ่มตัวอย่าง เป็นข้อมูลทั่วไปของผู้ตอบแบบสอบถาม ซึ่งเป็นนักศึกษา ของมหาวิทยาลัยสยาม จำนวน 543 คน จำแนกตาม เพศ อายุ คณะที่ศึกษา การมีบุคคลในครอบครัวที่ สูบบุหรี่ การคบเพื่อน ฯ ดำเนินการวิเคราะห์ข้อมูลด้วยสถิติ คะแนนร้อยละ และคะแนนเฉลี่ย รายละเอียด ดังตารางที่ 4.1

ตารางที่ 4.1 จำนวนและร้อยละของคุณลักษณะของกลุ่มตัวอย่าง

N = 543

ลำดับ	ข้อมูล	คุณลักษณะของกลุ่มตัวอย่าง	จำนวน	ร้อยละ
1	เพศ	ชาย	328	60.40
		หญิง	215	39.59
2	อายุ	อายุ 18 ปี	58	10.68
		อายุ 19 ปี	91	16.75
		อายุ 20 ปี	110	20.25
		อายุ 21 ปี	112	20.62
		อายุ 22 ปี	86	15.83
		อายุ 23 ปี	35	6.44
		อายุ 24 ปี	20	3.68
		อายุ 25 ปี	13	2.39
		อายุ 26 ปี	5	0.92
		อายุ 27 ปี	10	1.84
3	คณะ	คณะศิลปศาสตร์	77	14.18
		คณะนิเทศศาสตร์	47	8.656
		คณะรัฐศาสตร์	1	0.184
		คณะบริหารธุรกิจ	91	16.75
		คณะวิศวกรรมศาสตร์	75	13.81
		คณะพยาบาลศาสตร์	38	6.99
		คณะเกษตรศาสตร์	65	11.97
		คณะสาธารณสุขศาสตร์	13	11.97
		คณะวิทยาศาสตร์	17	2.39

ลำดับ	ข้อมูล	คุณลักษณะของกลุ่มตัวอย่าง	จำนวน	ร้อยละ
3	คณะ	คณะเทคโนโลยีสารสนเทศ	86	15.83
		คณะนิติศาสตร์	6	1.10
		คณะแพทยศาสตร์	20	3.68
		วิทยาลัยดนตรีและ ศิลปะการแสดง	2	0.36
		วิทยาลัยนานาชาติ	5	0.92
4	ระดับการศึกษา	ระดับปริญญาตรี	543	100
		ระดับปริญญาโท	0	0
		ระดับปริญญาเอก	0	0
5	ชั้นปี	ชั้นปีที่ 1	185	34.07
		ชั้นปีที่ 2	128	23.57
		ชั้นปีที่ 3	111	20.44
		ชั้นปีที่ 4	100	18.41
		ชั้นปีที่ 5	16	2.94
		ชั้นปีที่ 6	3	0.55
6	ก่อนเข้าเรียนใน สาขานี้เกรดเฉลี่ย สะสมของท่าน เท่าไร	ต่ำกว่า 2.00	10	1.84
		2.00-2.49	75	13.81
		2.50-2.99	113	20.81
		3.00-3.49	179	32.96
		มากกว่า 3.50 ขึ้นไป	166	30.57
7	ระดับคะแนนเฉลี่ย สะสมของท่านที่ ศึกษาในปัจจุบันได้ เกรดเฉลี่ยสะสม เท่าไร	ต่ำกว่า 2.00	10	1.84
		2.00-2.49	71	13.07
		2.50-2.99	122	22.46
		3.00-3.49	177	32.59
		มากกว่า 3.50 ขึ้นไป	163	30.01

ลำดับ	ข้อมูล	คุณลักษณะของกลุ่มตัวอย่าง	จำนวน	ร้อยละ
8	สถานภาพสมรสของ บิดา-มารดา	อยู่ด้วยกัน	359	66.11
		หย่าร้าง	79	14.54
		แยกกันอยู่	81	14.91
		หม้าย	24	4.42
9	สถานที่พักอาศัย ปัจจุบัน	บ้าน	347	63.90
		อพาร์ทเมนท์	8	1.47
		คอนโดมิเนียม	49	9.02
		หอพัก	139	25.59
10	พักอาศัยอยู่กับใคร	ครอบครัว	322	59.30
		บิดา	8	1.47
		มารดา	51	9.39
		แฟน	24	4.42
		เพื่อน	49	9.02
		อยู่คนเดียว	89	16.39
11	อาชีพของบิดา	อาชีพอิสระ	199	36.64
		เจ้าของธุรกิจ	119	21.91
		รัฐวิสาหกิจ	14	2.57
		ราชการ	50	9.20
		พนักงานบริษัท	19	3.49
		เอกชน	83	15.28
		เล่นหุ้น	11	2.06
		ชานา	8	1.47
		ไม่ได้ทำงาน	23	4.23
		เสียชีวิต	17	3.13
12	อาชีพของมารดา	อาชีพอิสระ	222	40.88
		เจ้าของธุรกิจ	118	21.73
		รัฐวิสาหกิจ	7	1.28

ลำดับ	ข้อมูล	คุณลักษณะของกลุ่มตัวอย่าง	จำนวน	ร้อยละ
12	อาชีพของมารดา	ราชการ	43	7.91
		เอกชน	92	16.94
		เล่นหุ้น	12	2.21
		ชานา	8	1.47
		พนักงานทำความสะอาด	19	3.50
		ไม่ได้ทำงาน	16	2.94
		เสียชีวิต	6	1.10
13	บิดาสืบบุหรีหรือไม่	สืบบุหรี	153	28.17
		ไม่สืบบุหรี	390	71.82
14	มารดาสืบบุหรีหรือไม่	สืบบุหรี	21	3.86
		ไม่สืบบุหรี	522	96.13
15	รายได้ของบิดา	ต่ำกว่า 15,000 บาท	105	19.33
		15,000-20,000 บาท	156	28.72
		2,0001- 25,000 บาท	58	10.68
		25,001-30,000 บาท	61	11.23
		มากกว่า 30,000 บาทขึ้นไป	163	30.01
16	รายได้ของมารดา	ต่ำกว่า 15,000 บาท	142	26.15
		15,000-20,000 บาท	147	27.07
		2,0001- 25,000 บาท	61	11.23
		25,001-30,000 บาท	55	10.12
		มากกว่า 30,000 บาทขึ้นไป	138	25.41
17	บ้านที่พักอยู่เป็นของใคร	บิดา – มารดา	357	65.74
		ญาติ	72	13.24
		แฟน	2	0.36
		เพื่อน	2	0.36
		คนเดียว	11	2.02
		บ้านเช่า	99	18.23

ลำดับ	ข้อมูล	คุณลักษณะของกลุ่มตัวอย่าง	จำนวน	ร้อยละ
18	ภูมิลาเนา	จังหวัดอ่างทอง	3	0.55
		จังหวัดอำนาจเจริญ	2	0.36
		จังหวัดกรุงเทพมหานคร	282	51.93
		จังหวัดบึงกาฬ	2	0.36
		จังหวัดบุรีรัมย์	6	1.10
		จังหวัดฉะเชิงเทรา	2	0.36
		จังหวัดชัยนาท	1	0.18
		จังหวัดชัยภูมิ	5	0.92
		จังหวัดจันทบุรี	3	0.55
		จังหวัดเชียงใหม่	2	0.36
		จังหวัดเชียงราย	7	1.28
		จังหวัดชลบุรี	6	1.10
		จังหวัดชุมพร	3	0.55
		จังหวัดกาฬสินธุ์	4	0.73
		จังหวัดกาญจนบุรี	6	1.28
		จังหวัดขอนแก่น	5	1.10
		จังหวัดเลย	4	0.92
		จังหวัดลพบุรี	2	0.73
		จังหวัดมหาสารคาม	7	0.36
		จังหวัดมุกดาหาร	2	1.28
		จังหวัดนครนายก	1	0.36
		จังหวัดนครปฐม	15	0.18
		จังหวัดนครราชสีมา	9	2.76
		จังหวัดนครสวรรค์	5	1.65
		จังหวัดนครศรีธรรมราช	2	0.92
		จังหวัดน่าน	1	0.36
		จังหวัดหนองบัวลำภู	1	0.18

ลำดับ	ข้อมูล	คุณลักษณะของกลุ่มตัวอย่าง	จำนวน	ร้อยละ
18	ภูมิลาเนา	จังหวัดหนองคาย	2	0.18
		จังหวัดนนทบุรี	11	0.36
		จังหวัดปัตตานี	1	2.02
		จังหวัดปทุมธานี	1	0.18
		จังหวัดพังงา	3	0.18
		จังหวัดพัทลุง	2	0.55
		จังหวัดพะเยา	3	0.36
		จังหวัดเพชรบูรณ์	5	0.55
		จังหวัดเพชรบุรี	2	0.92
		จังหวัดพิจิตร	1	0.36
		จังหวัดพิษณุโลก	5	0.18
		จังหวัดแพร่	3	0.92
		จังหวัดพระนครศรีอยุธยา	3	0.55
		จังหวัดปราจีนบุรี	1	0.55
		จังหวัดประจวบคีรีขันธ์	3	0.18
		จังหวัดภูเก็ต	1	0.55
		จังหวัดระนอง	1	0.18
		จังหวัดร้อยเอ็ด	7	0.36
		จังหวัดสระแก้ว	1	1.28
		จังหวัดสกลนคร	2	0.18
		จังหวัดสมุทรปราการ	15	0.36
		จังหวัดสมุทรสาคร	26	2.76
		จังหวัดสระบุรี	1	4.78
จังหวัดสตูล	1	0.18		
จังหวัดสิงห์บุรี	2	0.18		
จังหวัดศรีสะเกษ	7	0.36		
จังหวัดสงขลา	5	1.28		

ลำดับ	ข้อมูล	คุณลักษณะของกลุ่มตัวอย่าง	จำนวน	ร้อยละ
18	ภูมิสำเนา	จังหวัดสุโขทัย	1	0.92
		จังหวัดสุพรรณบุรี	3	0.18
		จังหวัดสุราษฎร์ธานี	7	0.55
		จังหวัดสุรินทร์	6	1.28
		จังหวัดตรัง	3	1.10
		จังหวัดอุบลราชธานี	7	0.55
		จังหวัดอุดรธานี	3	1.28
		จังหวัดอุดรดิษฐ์	2	0.55
		จังหวัดยะลา	1	0.36
		จังหวัดยโสธร	2	0.18
19	จำนวนสมาชิก ภายในครอบครัว	1-2 คน	48	8.84
		3-4 คน	278	51.19
		5-6 คน	135	15.10
		มากกว่า 6 คนขึ้นไป	82	24.86
20	พักอาศัยอยู่กับใคร มากที่สุด	ครอบครัว	485	89.31
		พี่น้อง	9	1.65
		ญาติ	36	6.63
		คนที่รู้จัก	13	2.39
		เพื่อน	0	0
21	เพื่อนสูบบุหรี่	สูบบุหรี่	401	73.84
		ไม่สูบบุหรี่	142	26.15
22	คุณสูบบุหรี่	สูบบุหรี่	98	81.95
		ไม่สูบบุหรี่	445	18.04

ตารางที่ 4.1 จากการวิเคราะห์ข้อมูลการตอบแบบสอบถามนักศึกษามหาวิทยาลัยสยามจำนวน 543 คน เป็นเพศชายจำนวนร้อยละ 60.40 และเป็นเพศหญิงจำนวนร้อยละ 39.59 มีอายุโดยเฉลี่ยอยู่ที่ 18-27 ปี จำนวนร้อยละ 99.45 และมีอายุโดยเฉลี่ยอยู่ที่ 29 ปี จำนวนร้อยละ 0.55 เป็นนักศึกษาคณะ

บริหารธุรกิจจำนวนร้อยละ 16.75 เป็นนักศึกษาคณะเทคโนโลยีสารสนเทศจำนวนร้อยละ 15.83 เป็นนักศึกษาคณะศิลปศาสตรจำนวนร้อยละ 14.18 เป็นนักศึกษาคณะวิศวกรรมศาสตร์จำนวนร้อยละ 13.81 เป็นนักศึกษาคณะเภสัชศาสตร์จำนวนร้อยละ 11.97 เป็นนักศึกษาคณะนิเทศศาสตร์จำนวนร้อยละ 8.65 เป็นนักศึกษาคณะพยาบาลศาสตร์จำนวนร้อยละ 6.99 เป็นนักศึกษาคณะแพทยศาสตร์จำนวนร้อยละ 3.68 เป็นนักศึกษาคณะวิทยาศาสตร์จำนวนร้อยละ 3.13 เป็นนักศึกษาคณะสาธารณสุขศาสตร์จำนวนร้อยละ 2.39 เป็นนักศึกษาคณะนิติศาสตร์จำนวนร้อยละ 1.10 เป็นนักศึกษาวิทยาลัยดนตรีและศิลปะการแสดงวิทยาลัยนานาชาติจำนวนร้อยละ 0.36 และเป็นนักศึกษาคณะรัฐศาสตร์ศิลปศาสตรจำนวนร้อยละ 0.18 เป็นนักศึกษาชั้นปีที่ 1-5 จำนวนร้อยละ 95.80 และเป็นนักศึกษาชั้นปีที่ 6 จำนวนร้อยละ 0.55 มีคะแนนสอบก่อนเข้าเรียนเฉลี่ยต่ำกว่า 2.00 จำนวนร้อยละ 1.84 มีคะแนนสอบก่อนเข้าเรียนเฉลี่ย 2.00-2.49 จำนวนร้อยละ 13.81 มีคะแนนสอบก่อนเข้าเรียนเฉลี่ย 2.50-2.99 จำนวนร้อยละ 20.81 มีคะแนนสอบก่อนเข้าเรียนเฉลี่ย 3.00-3.49 จำนวนร้อยละ 32.96 และมีคะแนนสอบก่อนเข้าเรียนเฉลี่ยมากกว่า 3.50 จำนวนร้อยละ 30.57 มีคะแนนสอบระหว่างเรียนเฉลี่ยต่ำกว่า 2.00 จำนวนร้อยละ 1.84 มีคะแนนสอบระหว่างเรียนเฉลี่ย 2.00-2.49 จำนวนร้อยละ 13.07 มีคะแนนสอบระหว่างเรียนเฉลี่ย 2.50-2.99 จำนวนร้อยละ 22.46 คะแนนสอบระหว่างเรียนเฉลี่ย 3.00-3.49 จำนวนร้อยละ 32.59 และมีคะแนนสอบระหว่างเรียนเฉลี่ยมากกว่า 3.50 จำนวนร้อยละ 30.01 บิดามารดาอาศัยอยู่ด้วยกันจำนวนร้อยละ 66.11 บิดามารดาแยกกันอยู่อาศัยจำนวนร้อยละ 14.91 บิดามารดาได้ทำการหย่าร้างจำนวนร้อยละ 14.54 และบิดาหรือมารดาได้เสียชีวิตจำนวนร้อยละ 4.42 พักอาศัยอยู่ที่บ้านเป็นจำนวนร้อยละ 63.90 พักอาศัยอยู่ที่หอพักเป็นจำนวนร้อยละ 25.59 พักอาศัยอยู่ที่คอนโดมิเนียมเป็นจำนวนร้อยละ 9.02 และพักอาศัยอยู่ที่อพาร์ทเมนท์เป็นจำนวนร้อยละ 1.47 พักอาศัยอยู่กับครอบครัวจำนวนร้อยละ 59.30 พักอาศัยอยู่กับบิดาจำนวนร้อยละ 1.47 พักอาศัยอยู่กับมารดาจำนวนร้อยละ 9.39 พักอาศัยอยู่กับแม่แฟนจำนวนร้อยละ 4.42 พักอาศัยอยู่กับเพื่อนจำนวนร้อยละ 9.02 และพักอาศัยอยู่กับอยู่คนเดียวจำนวนร้อยละ 16.39 บิดาทำอาชีพอิสระจำนวนร้อยละ 36.64 บิดาทำอาชีพเจ้าของธุรกิจจำนวนร้อยละ 21.91 บิดาทำอาชีพเอกชนจำนวนร้อยละ 15.28 บิดาทำอาชีพราชการจำนวนร้อยละ 9.20 บิดาทำอาชีพพนักงานบริษัทจำนวนร้อยละ 3.49 บิดาทำอาชีพรัฐวิสาหกิจจำนวนร้อยละ 2.57 บิดาทำอาชีพเล่นหุ้นจำนวนร้อยละ 2.06 บิดาทำอาชีพขานาจำนวนร้อยละ 1.47 บิดาเสียชีวิตจำนวนร้อยละ 3.13 และบิดาไม่ได้ทำงานจำนวนร้อยละ 4.23 มารดาทำอาชีพอิสระจำนวนร้อยละ 40.88 มารดาทำอาชีพเจ้าของธุรกิจจำนวนร้อยละ 21.73 และมารดาทำอาชีพราชการจำนวนร้อยละ 1.28 พักอาศัยอยู่ในจังหวัดกรุงเทพมหานครจำนวนร้อยละ 51.93 ซึ่งมากที่สุด และพักอยู่ในจังหวัด

หนองบัวลำภูจำนวนร้อยละ 0.18 ซึ่งน้อยที่สุด มีสมาชิกในครอบครัว 1-2 คนจำนวนร้อยละ 8.84 มีสมาชิกในครอบครัว 3-4 คนจำนวนร้อยละ 51.19 มีสมาชิกในครอบครัว 5-6 คนจำนวนร้อยละ 15.10 และมีสมาชิกในครอบครัวมากกว่า 6 คนขึ้นไปจำนวนร้อยละ 24.86 ที่พักอาศัยเป็นของครอบครัว จำนวนร้อยละ 89.31 ที่พักอาศัยเป็นของญาติจำนวนร้อยละ 6.63 ที่พักอาศัยเป็นของคนที่ย้ายมาจำนวน ร้อยละ 2.39 และที่พักอาศัยเป็นของพี่น้องจำนวนร้อยละ 1.65 มีเพื่อนที่สูบบุหรี่จำนวนร้อยละ 73.84 และมีเพื่อนที่ไม่สูบบุหรี่จำนวนร้อยละ 26.15 ผู้ตอบแบบสอบถามที่สูบบุหรี่จำนวนร้อยละ 81.95 และ ผู้ตอบแบบสอบถามที่ไม่สูบบุหรี่จำนวนร้อยละ 18.04

ส่วนที่ 2 การวิเคราะห์ความคิดเห็นเกี่ยวกับพิษภัยและโทษของบุหรี่

2.1 พิษภัยของบุหรี่ โดยการวิเคราะห์ความคิดเห็นเกี่ยวกับการสูบบุหรี่มีผลต่อร่างกายทำให้ ร่างกายได้รับสารพิษ เป็นสาเหตุของการเจ็บป่วยด้วยโรคต่าง ๆ

2.2 โทษของการสูบบุหรี่ โดยการวิเคราะห์ความคิดเห็นเกี่ยวกับกฎหมายคุ้มครองสุขภาพ ของผู้ไม่สูบบุหรี่ และการควบคุมผลิตภัณฑ์ยาสูบ ซึ่งกำหนดโทษสำหรับผู้ฝ่าฝืนสูบบุหรี่ใน บริเวณ ที่ห้ามสูบ หรือไม่จัดเขตปลอดบุหรี่ตามประกาศของกระทรวงสาธารณสุข ดังแสดงใน ตารางที่ 4.2

ตารางที่ 4.2 การวิเคราะห์ความคิดเห็นเกี่ยวกับพิษภัยและโทษของบุหรี่

ข้อ ที่	ข้อความ	เห็น ด้วย	ไม่เห็น ใจ	ไม่เห็น ด้วย
1	การสูบบุหรี่เป็นสาเหตุทำให้เกิด โรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย	92.08	6.63	1.28
2	การสูบบุหรี่เป็นสาเหตุทำให้เส้นเลือดในสมองแตก และ เป็นอัมพาตได้	80.66	17.49	1.84
3	สารพิษในควันบุหรี่ที่ทำให้ผู้สูบบมีอาการติดบุหรี่คือ สารนิโคติน (Nicotine)	89.13	8.65	2.21
4	โรคถุงลมโป่งพองเกิดจากการสูบบุหรี่ทำให้ถุงลมปอดฉีก ทำลายและมีอาการหืดเกร็งของหลอดลม	88.39	10.12	1.47
5	ก้นกรองของบุหรี่ไม่ได้ทำให้ปริมาณของสารนิโคติน ในบุหรี่ลดลง	73.84	22.65	3.49

ข้อ ที่	ข้อความ	เห็น ด้วย	ไม่เห็นใจ	ไม่เห็น ด้วย
6	ในวันบวหฺรึมีสารตะกั่วซึ่งเป็นสารที่ยับยั้งการเจริญเติบโตของร่างกาย และทำลายเนื้อเยื่อสมอง	81.58	15.83	2.57
7	ลวันบวหฺรึที่ออกจากผู้สูบได้ทำร้ายคนข้างเคียงเรียกว่า “ลวันบวหฺรึมือสอง”	88.39	9.94	1.65
8	ตามประกาศกระทรวงสาธารณสุข พ.ศ. 2551 กำหนดให้แสดงเครื่องหมายของเขตสูบบวหฺรึ และเขตปลอดบวหฺรึในที่เปิดเผย และสามารถมองเห็นได้ชัดเจน	83.05	14.54	2.39
9	ผู้ที่ฝึบสูบบวหฺรึในเขตพื้นที่ปลอดบวหฺรึมีโทษปรับไม่เกิน 2,000 บาท	82.50	13.99	3.49
10	เจ้าของสถานที่ ถ้าไม่จัดให้มีเครื่องหมายเขตปลอดบวหฺรึ หรือเขตสูบบวหฺรึ มีโทษปรับไม่เกิน 2,000 บาท	79.37	17.68	2.94

จากตารางที่ 4.2 ความคิดเห็นเกี่ยวกับพิษภัย และโทษของการสูบบวหฺรึ การสูบบวหฺรึทำให้ผู้สูบบวหฺรึได้รับสารพิษจากนิโคติน และสาร ตะกั่วซึ่งจะทำลายสมอง และยับยั้งการเจริญเติบโตของร่างกาย เป็นสาเหตุให้เกิดโรคมะเร็งที่ อวัยวะต่าง ๆ ของร่างกายและมีความเสี่ยงต่อการเสียชีวิตจากโรคหัวใจ ทั้งเป็นสาเหตุให้เกิดโรคถุงลมโป่งพองซึ่งจะทำให้ได้รับความทรมานจากถุงลมปอดถูกทำลาย ความคิดเห็นเกี่ยวกับก้นกรองของบวหฺรึ เห็นด้วยร้อยละ 73.84 ไม่เห็นใจร้อยละ 22.65 และไม่เห็นว่าร้อยละ 3.49 ไม่ได้ทำให้ปริมาณของสาร นิโคตินลดลง ด้านความคิดเห็นเกี่ยวกับโทษของการสูบบวหฺรึ นักศึกษากลุ่มตัวอย่างมีความคิดเห็นว่า ใน บริเวณที่ติดป้ายเขตปลอดบวหฺรึ หากมีการฝึบสูบบวหฺรึในพื้นที่เขตปลอดบวหฺรึจะมีโทษปรับไม่เกิน 2,000 บาท ซึ่งเจ้าของสถานที่ต้องให้มีเครื่องหมายเขตปลอดบวหฺรึ หรือเขตที่ใ้สูบบวหฺรึได้ โดยติดป้ายให้เห็นชัดเจนในที่เปิดเผย หากไม่จัดให้มีเครื่องหมายดังกล่าวจะมีโทษปรับไม่เกิน 2,000 บาท และหากไม่จัดพื้นที่ให้เป็นเขตปลอดบวหฺรึ และเขตบวหฺรึก็จะมีโทษปรับไม่เกิน 20,000 บาท ซึ่งความคิดเห็นส่วนมากจะเห็นด้วยร้อยละ 79.37 ไม่เห็นใจร้อยละ 17.68 และไม่เห็นว่าร้อยละ 2.94

ส่วนที่ 3 การวิเคราะห์เจตคติต่อการสูบบุหรี่ ดังตารางที่ 4.3
 ตารางที่ 4.3 เจตคติต่อการสูบบุหรี่

ข้อ ที่	ข้อความ	เห็น ด้วย อย่าง ยิ่ง	เห็นด้วย	ไม่ แน่ใจ	ไม่เห็น ด้วย	ไม่เห็น ด้วย ยิ่ง
1	การสูบบุหรี่เป็นพฤติกรรมไม่ปกติในสังคมไทย	26.15	15.10	33.51	20.44	4.78
2	ผู้ชายสูบบุหรี่เป็นพฤติกรรมไม่ปกติ	21.91	15.10	32.96	23.57	6.44
3	ผู้หญิงสูบบุหรี่เป็นพฤติกรรมไม่เหมาะสม	26.70	18.78	27.44	20.81	6.26
4	ผู้ใหญ่สูบบุหรี่เป็นพฤติกรรมที่ไม่ควรยอมรับ	26.70	16.57	29.28	21.73	5.70
5	วัยรุ่นสูบบุหรี่เป็นพฤติกรรมที่ไม่เหมาะสม	30.93	24.49	25.96	15.10	3.49
6	การสูบบุหรี่ไม่แสดงถึงการเป็นคนทันสมัยเท่าที่	45.30	17.12	16.57	0	20.99
7	การสูบบุหรี่ไม่ทำให้คุณเป็นที่ยอมรับของเพื่อนๆ	36.83	16.39	22.83	14.73	9.20
8	การสูบบุหรี่เป็นพฤติกรรมที่ไม่น่าเลียนแบบ	56.53	22.46	14.73	3.86	2.39
9	บุหรี่เป็นสิ่งเสพติด ทำลายสุขภาพ	64.64	19.15	11.60	3.13	1.47

ข้อ ที่	ข้อความ	เห็น ด้วย อย่าง ยิ่ง	เห็นด้วย	ไม่ แน่ใจ	ไม่เห็น ด้วย	ไม่เห็น ด้วย อย่าง ยิ่ง
10	บุหรี่เป็นสิ่งเสพติดที่อันตรายเพราะจะนำไปสู่สิ่งเสพติดอื่น ๆ	50.64	21.91	18.04	6.63	2.76
11	ทำนรังเกียจเพื่อนหรือบุคคลที่สูบบุหรี่	24.67	13.62	23.75	26.33	11.60
12	ถ้าทำนสูบบุหรี่เพื่อนหรือคนรอบข้างรังเกียจ	27.80	15.65	32.22	16.20	8.10
13	ถ้าทำนอยู่ในกลุ่มที่สูบบุหรี่ ทำนจำเป็นต้องสูบด้วย	21.54	9.02	15.83	26.33	27.25
14	การสูบบุหรี่ทำให้สิ้นเปลืองค่าใช้จ่ายในการรักษาโรคที่มีสาเหตุจากการสูบบุหรี่โดยไม่จำเป็น	59.48	19.52	14.73	4.60	1.65
15	การเลิกบุหรี่ทำให้เสียบุคลิกภาพของตนเอง	29.46	12.89	16.94	19.70	20.99

จากตารางที่ 4.3 ผลจากการทดสอบความคิดเห็นและทัศนคติของประชากรในสังคมไทยต่อการสูบบุหรี่ ผู้วิจัยได้ทำการสำรวจโดยใช้ข้อมูลจากการทำแบบสอบถามเกี่ยวกับแนวคิดเห็นต่าง ๆ โดยสรุปข้อมูลดังนี้

การสูบบุหรี่ของนักศึกษากลุ่มตัวอย่าง พบว่า ร้อยละข้อมูลสูงสุด คือ 64.64 ซึ่งเป็น เจนคติที่รู้สึกว่าการสูบบุหรี่เป็นสิ่งเสพติดที่ทำลายสุขภาพ ลำดับต่อมาคือ การสูบบุหรี่เป็นพฤติกรรมที่ไม่น่าเลียนแบบ มีความเห็นด้วยอย่างยิ่งร้อยละ 56.53 และไม่เห็นด้วยอย่างยิ่ง 2.39 เจนคติต่อการสูบบุหรี่ซึ่งทำให้สิ้นเปลืองค่าใช้จ่ายในการรักษาโรคที่มีสาเหตุจากการสูบบุหรี่โดยไม่จำเป็น มีความเห็นด้วยอย่างยิ่งร้อยละ 59.48 และไม่เห็นด้วยอย่างยิ่งร้อยละ 1.65 เจนคติต่อการสูบบุหรี่เป็นพฤติกรรมที่ไม่เหมาะสมสำหรับผู้หญิงมีความเห็นด้วยอย่างยิ่งร้อยละ 30.93 และรู้สึกว่าการสูบบุหรี่เป็นอันตรายซึ่งจะเป็นต้นเหตุนำไปสู่การเสพติดสิ่งเสพติดอื่น ๆ ที่รุนแรงมากยิ่งขึ้นมีความเห็นด้วยอย่างยิ่งร้อยละ 50.64

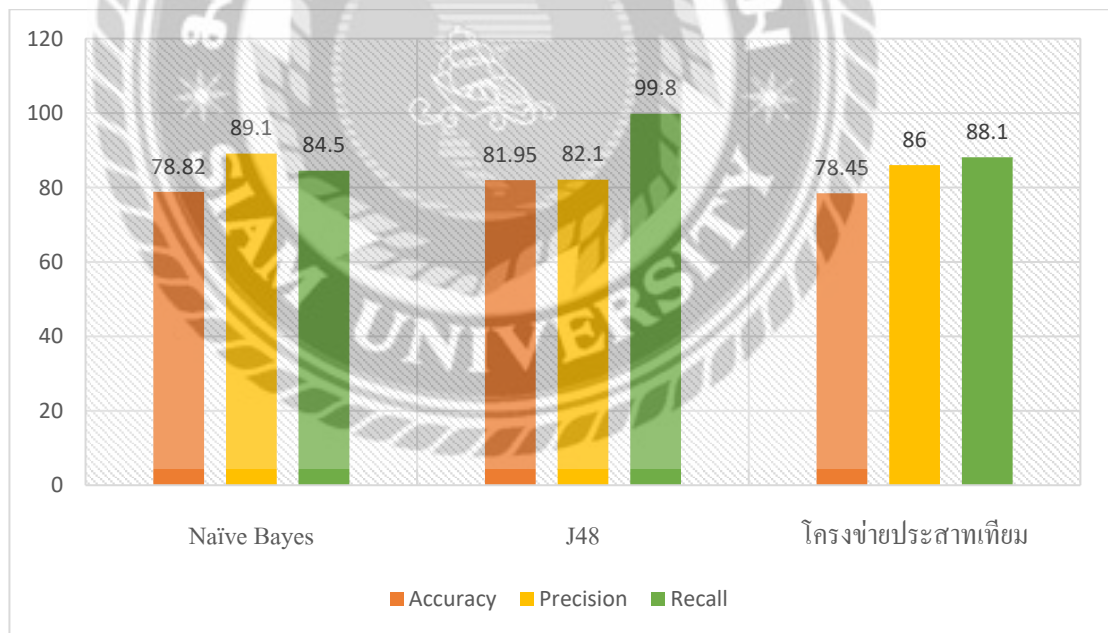
และไม่เห็นอย่างย้ง 2.76 เจนคติที่รู้สึกว่าการสูบบุหรี่ไม่ได้แสดงว่าเป็นคนทันสมัยหรือเท่แต่อย่างใด ซึ่งมีความเห็นด้วยอย่างยิ่งร้อยละ 45.30 และไม่เห็นด้วยอย่างยิ่งร้อยละ 20.99 เจนคติที่นักศึกษา รู้สึกว่าการที่บุคคลจะเลิกสูบบุหรี่ไม่ทำให้บุคคลนั้นเสียบุคลิกภาพ

4.2 ผลการเปรียบเทียบประสิทธิภาพของข้อมูล

4.2.1 สรุปผลการเปรียบเทียบประสิทธิภาพของข้อมูลชุดที่ 1 โดยมีผลลัพธ์ดังนี้

ตารางที่ 4.4 การเปรียบเทียบประสิทธิภาพความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) เฉลี่ย และ ค่าความระลึก (Recall) เฉลี่ย ชุดข้อมูลที่ 1 จำนวน 22 Attribute

Algorithms	Accuracy (%)	Precision (%)	Recall (%)
Naïve Bayes	78.82	89.10	84.50
J48	81.95	82.10	99.80
โครงข่ายประสาทเทียม	78.45	86.00	88.10



ภาพที่ 4.1 การเปรียบเทียบประสิทธิภาพวิธี Naïve Bayes, J48 และ โครงข่ายประสาทเทียม ชุดข้อมูลที่ 1 จำนวน 22 Attribute

จากตารางที่ 4.4 และภาพที่ 4.1 สรุปได้ว่า การนำชุดข้อมูลที่ 1 จำนวน 22 Attribute ที่ได้จากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องมาสร้างโมเดลเพื่อทำนายการสูญบุหรืของวัยรุ่น พบว่าประสิทธิภาพทำนายวิธี J48 มีค่าความถูกต้อง (Accuracy) มากที่สุดคือ 81.95% รองลงมาวิธี Naïve Bayes มีค่าความถูกต้องเท่ากับ 78.82% ลำดับสุดท้ายวิธี โครงข่ายประสาทเทียมมีค่าความถูกต้องเท่ากับ 78.45%

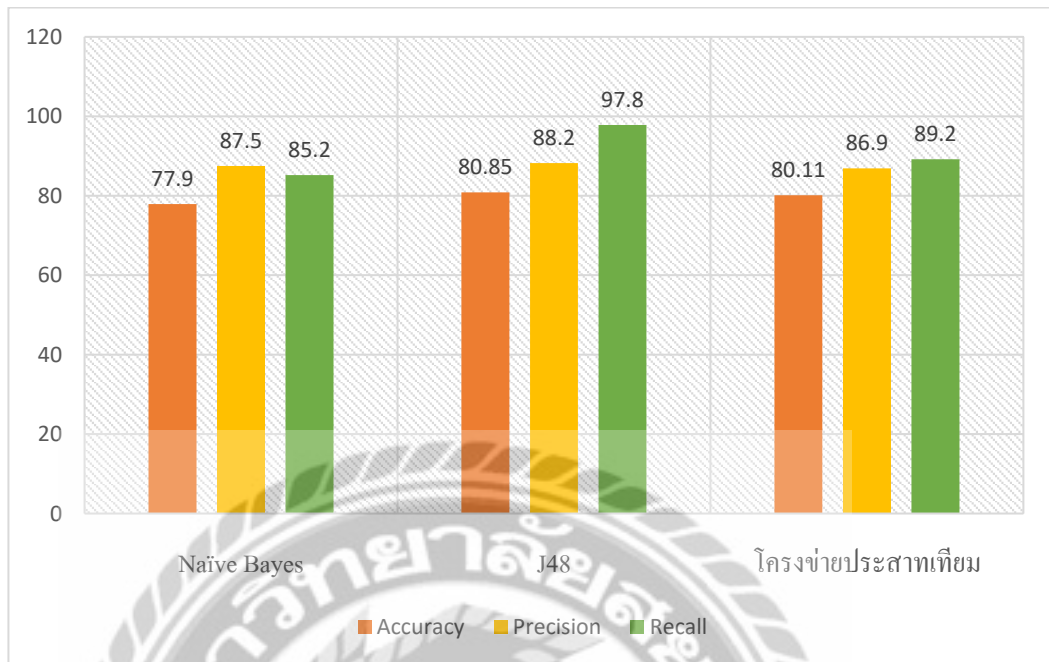
ด้านการเปรียบเทียบค่าความแม่นยำ (Precision) เฉลี่ย พบว่าวิธี Naïve Bayes มากที่สุด 89.10% รองลงมาวิธี โครงข่ายประสาทเทียม 86.00% ลำดับสุดท้ายวิธี J48 82.10%

ด้านการเปรียบเทียบค่าความระลึก (Recall) เฉลี่ย พบว่าวิธี J48 มากที่สุด 99.80% รองลงมาวิธี โครงข่ายประสาทเทียม 88.10% ลำดับสุดท้ายวิธี Naïve Bayes 84.50%

4.2.2 สรุปผลการเปรียบเทียบประสิทธิภาพของข้อมูลชุดที่ 2 โดยมีผลลัพธ์ดังนี้

ตารางที่ 4.5 การเปรียบเทียบประสิทธิภาพความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) เฉลี่ย และ ค่าความระลึก (Recall) เฉลี่ย ชุดข้อมูลที่ 2 จำนวน 32 Attribute

Algorithms	Accuracy (%)	Precision (%)	Recall (%)
Naïve Bayes	77.90	87.50	85.20
J48	80.85	88.20	97.80
โครงข่ายประสาทเทียม	80.11	86.90	89.20



ภาพที่ 4.2 การเปรียบเทียบประสิทธิภาพวิธี Naïve Bayes, J48 และ โครงข่ายประสาทเทียม ชุดข้อมูลที่ 2 จำนวน 32 Attribute

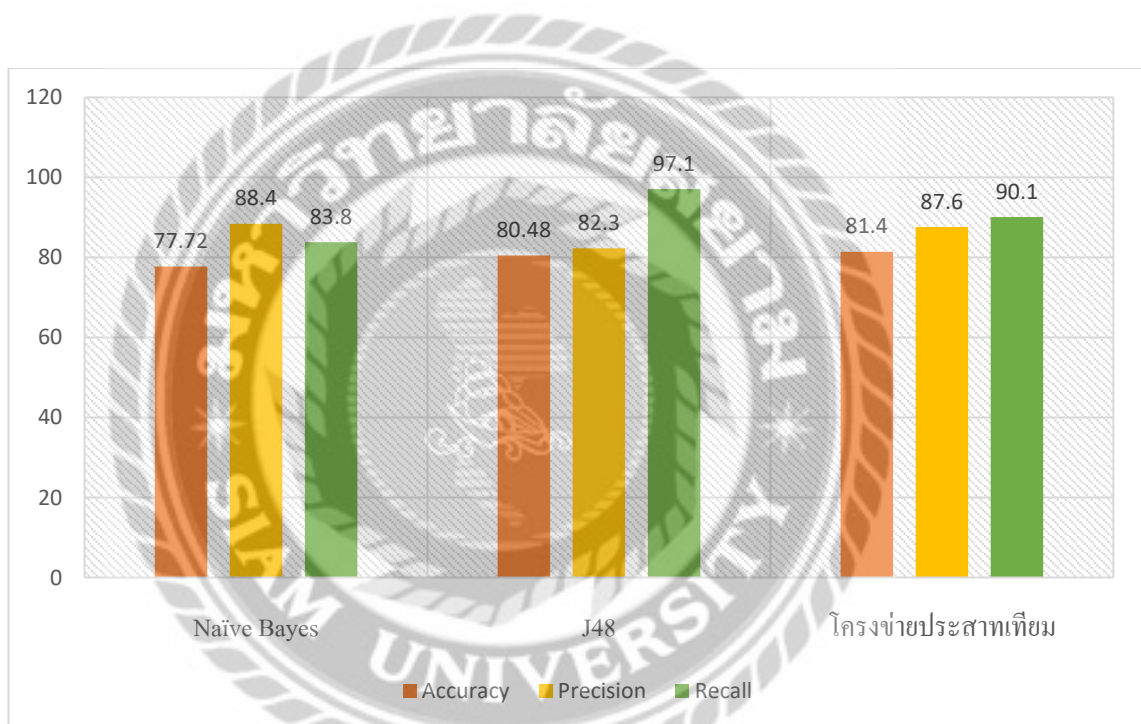
จากรายที่ 4.5 และภาพที่ 4.2 สรุปได้ว่า การนำชุดข้อมูลที่ 2 จำนวน 32 Attribute ที่ได้จากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องมาสร้าง โมเดลเพื่อทำนายการสูญบุหรืของวัยรุ่น พบว่า ประสิทธิภาพทำนายวิธี J48 มีค่าความถูกต้อง (Accuracy) มากที่สุดคือ 80.85% รองลงมาวิธี โครงข่ายประสาทเทียมมีค่าความถูกต้องเท่ากับ 80.11% ลำดับสุดท้ายวิธี Naïve Bayes มีค่าความถูกต้องเท่ากับ 77.90%

ด้านการเปรียบเทียบค่าความแม่นยำ (Precision) เฉลี่ย พบว่าวิธี J48 มากที่สุด 88.20% รองลงมาวิธี Naïve Bayes 87.50% ลำดับสุดท้ายวิธี โครงข่ายประสาทเทียม 86.90%

ด้านการเปรียบเทียบค่าความระลึก (Recall) เฉลี่ย พบว่าวิธี J48 มากที่สุด 97.80% รองลงมาวิธี โครงข่ายประสาทเทียม 89.20% ลำดับสุดท้ายวิธี Naïve Bayes 85.20%

4.2.3 สรุปผลการเปรียบเทียบประสิทธิภาพของข้อมูลชุดที่ 3 โดยมีผลลัพธ์ดังนี้
 ตารางที่ 4.6 การเปรียบเทียบประสิทธิภาพความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) เฉลี่ย และ ค่าความระลึก (Recall) เฉลี่ย ชุดข้อมูลที่ 3 จำนวน 47 Attribute

Algorithms	Accuracy (%)	Precision (%)	Recall (%)
Naïve Bayes	77.72	88.40	83.80
J48	80.48	82.30	97.10
โครงข่ายประสาทเทียม	81.40	87.60	90.10



ภาพที่ 4.3 การเปรียบเทียบประสิทธิภาพวิธี Naïve Bayes, J48 และ โครงข่ายประสาทเทียม ชุดข้อมูลที่ 3 จำนวน 47 Attribute

จากตารางที่ 4.6 และภาพที่ 4.3 สรุปได้ว่า การนำชุดข้อมูลที่ 3 จำนวน 47 Attribute ที่ได้จากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องมาสร้าง โมเดลเพื่อทำนายการสูญบุหรืงของวัยรุ่น พบว่า ประสิทธิภาพทำนายวิธี โครงข่ายประสาทเทียม มีค่าความถูกต้อง (Accuracy) มากที่สุดคือ 81.40% รองลงมาวิธี J48 มีค่าความถูกต้องเท่ากับ 80.48% ถ้าับสุดท้ายวิธี Naïve Bayes มีค่าความถูกต้องเท่ากับ 77.72%

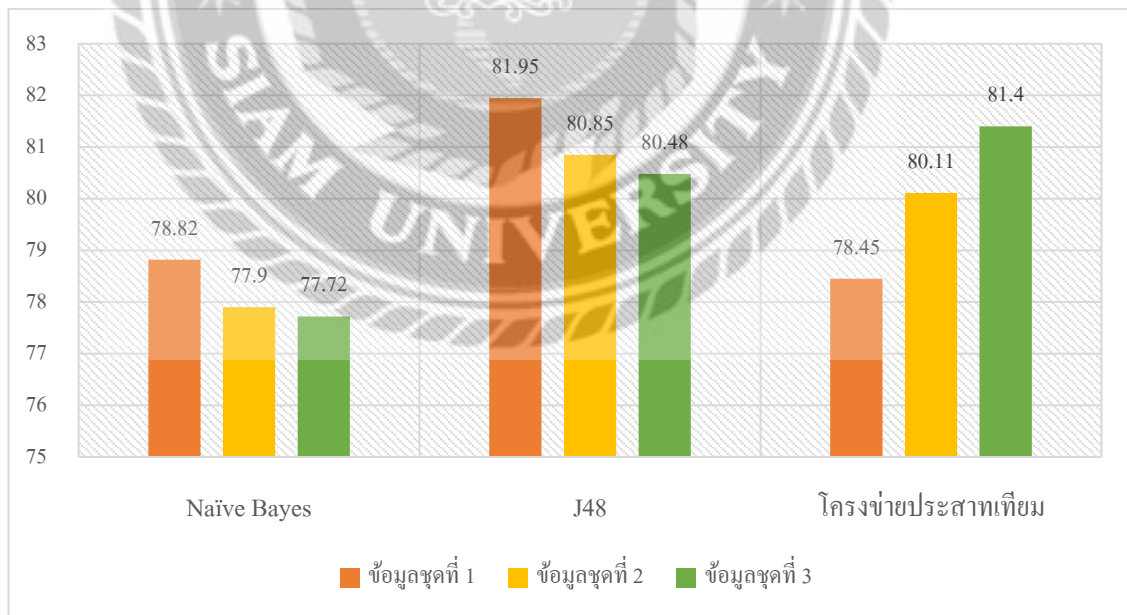
ด้านการเปรียบเทียบค่าความแม่นยำ (Precision) เฉลี่ย พบว่าวิธี Naïve Bayes มากที่สุด 88.40% รองลงมาวิธี โครงข่ายประสาทเทียม 87.60% ลำดับสุดท้ายวิธี J48 82.30%

ด้านการเปรียบเทียบค่าความระลึก (Recall) เฉลี่ย พบว่าวิธี J48 มากที่สุด 97.10% รองลงมาวิธี โครงข่ายประสาทเทียม 90.10% ลำดับสุดท้ายวิธี Naïve Bayes 83.80%

4.2.4 สรุปผลการเปรียบเทียบประสิทธิภาพของข้อมูลทั้ง 3 ชุด ที่ให้ค่าความถูกต้อง (Accuracy) มากที่สุด โดยมีผลลัพธ์ดังนี้

ตารางที่ 4.7 ค่าความถูกต้อง (Accuracy) การวิเคราะห์ข้อมูลชุดที่ 1, ข้อมูลชุดที่ 2 และ ข้อมูลชุดที่ 3

Algorithms	ค่าความถูกต้อง (Accuracy) ข้อมูลชุดที่ 1	ค่าความถูกต้อง (Accuracy) ข้อมูลชุดที่ 2	ค่าความถูกต้อง (Accuracy) ข้อมูลชุดที่ 3
Naïve Bayes	78.82	77.90	77.72
J48	81.95	80.85	80.48
โครงข่ายประสาทเทียม	78.45	80.11	81.40



ภาพที่ 4.4 การเปรียบเทียบค่าความถูกต้อง (Accuracy) วิธี Naïve Bayes, J48 และ โครงข่ายประสาทเทียม ชุดข้อมูลที่ 1, ชุดข้อมูลที่ 2 และ ชุดข้อมูลที่ 3

จากตารางที่ 4.7 และภาพที่ 4.4 สรุปได้ว่า จากการเปรียบเทียบค่าความถูกต้อง (Accuracy) ของการใช้ข้อมูลชุดที่ 1, ข้อมูลชุดที่ 2 และ ข้อมูลชุดที่ 3 ทำนายผลการสูบบุหรี่ของวัยรุ่น พบว่า โมเดลทำนายของชุดข้อมูลทั้ง 3 ชุด มีค่าความถูกต้อง (Accuracy) เพิ่มขึ้นและลดลงแตกต่างกัน พิจารณารายวิธีพบว่า วิธี Naïve Bayes ลดลง 0.92% และ 0.18% , J48 ลดลง 1.1% และ 0.37% , โครงข่ายประสาทเทียม เพิ่มขึ้น 1.66% และ 1.29% สรุปได้ว่า โมเดลการทำนายผลการสูบบุหรี่ของวัยรุ่นด้วยวิธี J48 โดยใช้ข้อมูลชุดที่ 1 มีค่าความถูกต้องมากที่สุดคือ 81.95%

4.3 สรุปกฎของอัลกอริทึม J48

4.3.1 สรุปกฎอัลกอริทึม J48 จำนวน 24 ข้อ ของชุดข้อมูลที่ 1,2,3 โดยข้อมูลชุดที่ 1 ไม่มี Tree

4.3.2 กฎอัลกอริทึมของ J48 คือ

1. เมื่อ Smoking friend มีค่าเท่ากับ n คือค่าที่ไม่มีเพื่อนสูบบุหรี่ ซึ่งทำนายได้ว่าจะไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
2. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 2 คือค่าที่เห็นด้วย ซึ่งทำนายได้ว่าจะไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
3. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 4 คือค่าที่ไม่เห็นด้วย ซึ่งทำนายได้ว่าจะไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
4. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ ควันบุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 2 คือค่าที่เห็นด้วย ซึ่งทำนายได้ว่าจะไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
5. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ ควันบุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 4 คือค่าที่ไม่เห็นด้วย ซึ่งทำนายได้ว่าจะมีพฤติกรรมที่สูบบุหรี่ คือค่า yes
6. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่และบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ

- ควันบุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 1 คืออาชีพอิสระและรายได้ของมารดา Income m มีค่าเท่ากับ 1 ซึ่งมีรายได้ต่ำกว่า 15,000 บาท จะทำนายได้ว่าไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
7. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ ควันบุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 1 คืออาชีพอิสระและรายได้ของมารดา Income m มีค่าเท่ากับ 2 ซึ่งมีรายได้ตั้งแต่ 15,000-20,000 บาท จะทำนายได้ว่าไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
8. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ ควันบุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 1 คืออาชีพอิสระและรายได้ของมารดา Income m มีค่าเท่ากับ 3 ซึ่งมีรายได้ตั้งแต่ 2,001-25,000 บาท จะทำนายได้ว่ามีพฤติกรรมที่สูบบุหรี่ คือค่า yes
9. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ ควันบุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 1 คืออาชีพอิสระและรายได้ของมารดา Income m มีค่าเท่ากับ 4 ซึ่งมีรายได้ตั้งแต่ 25,001-30,000 บาท จะทำนายได้ว่าไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
10. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ ควันบุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 1 คืออาชีพอิสระและรายได้ของมารดา Income m มีค่าเท่ากับ 5 ซึ่งมีรายได้มากกว่า 30,000 บาทขึ้นไปจะทำนายได้ว่าไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no

แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 8 คือชานา ซึ่งทำนายได้ว่าจะมีพฤติกรรมที่สูบบุหรี่ คือค่า yes

17. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ คำนับุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 9 คือพนักงานทำความสะอาด ซึ่งทำนายได้ว่าจะไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
18. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ คำนับุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 10 คือไม่ได้ทำงานและพักอยู่กับใคร With who มีค่าเท่ากับ 1 คืออยู่กับครอบครัว ซึ่งทำนายได้ว่าจะมีพฤติกรรมที่สูบบุหรี่ คือค่า yes
19. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ คำนับุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 10 คือไม่ได้ทำงาน และพักอยู่กับใคร With who มีค่าเท่ากับ 2 คืออยู่กับบิดา ซึ่งทำนายได้ว่าจะไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
20. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ คำนับุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 10 คือไม่ได้ทำงาน และพักอยู่กับใคร With who มีค่าเท่ากับ 3 คืออยู่กับมารดา ซึ่งทำนายได้ว่าจะไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
21. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ คำนับุหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 10 คือไม่ได้ทำงาน และพักอยู่กับใคร With who มีค่าเท่ากับ 4 คืออยู่กับแฟนซึ่งทำนายได้ว่าจะไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no

22. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ คำนับหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 10 คือไม่ได้ทำงาน และพักอยู่กับใคร With who มีค่าเท่ากับ 5 คืออยู่กับเพื่อน ซึ่งทำนายได้ว่าจะไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
23. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ คำนับหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 10 คือไม่ได้ทำงาน และพักอยู่กับใคร With who มีค่าเท่ากับ 6 คืออยู่คนเดียว ซึ่งทำนายได้ว่าจะไม่มีพฤติกรรมที่สูบบุหรี่ คือค่า no
24. เมื่อ Smoking friend มีค่าเท่ากับ m คือค่าที่มีเพื่อนสูบบุหรี่ ซึ่งบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย Smoking causes cancer มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ คำนับหรี่ที่ออกจากผู้สูบบุหรี่ทำร้ายคนรอบข้าง secondhand smoke มีค่าเท่ากับ 3 คือค่าที่ไม่แน่ใจ อาชีพของมารดา Work m มีค่าเท่ากับ 11 คือเสียชีวิต ซึ่งทำนายได้ว่าจะมีพฤติกรรมที่สูบบุหรี่ คือค่า yes

บทที่ 5

สรุปผลและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้ ผู้วิจัยได้เปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูญบุหรีของวัยรุ่นด้วยเทคนิคเหมืองข้อมูล เพื่อที่จะค้นหาตัวแบบหรือโมเดลที่ใช้ในการจำแนกประเภทข้อมูลที่ดีที่สุด ขั้นตอนแรกผู้วิจัยได้ทำการศึกษาและทบทวนวรรณกรรมหรือผลงานวิจัยที่เกี่ยวข้องเพื่อให้เป็นแนวทางในการดำเนินงานวิจัย นอกจากนี้ยังศึกษาขั้นตอนการวิเคราะห์ข้อมูลโดยใช้การจำแนกประเภทข้อมูล (Classification) และเลือกใช้อัลกอริทึมที่เหมาะสมในการสร้างโมเดล กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้คือ นักศึกษามหาวิทยาลัยสยาม ซึ่งเป็นนักศึกษาชั้นปีที่ 1 ถึง 6 ทุกสาขาวิชา จำนวน 543 คน ดำเนินการเก็บรวบรวมข้อมูลตั้งแต่วันที่ 13 ถึง วันที่ 31 ธันวาคม 2565 เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลคือ แบบสอบถามที่ใช้กับกลุ่มตัวอย่างจำนวน 543 คน และนำข้อมูลที่ได้จากแบบสอบถามนำมาวิเคราะห์ โดยใช้โปรแกรม Weka นำข้อมูลเข้าโมเดล มี Attribute จำนวนทั้งหมด 49 Attribute และแบ่งชุดข้อมูลเป็น 3 ชุด ทำการเปรียบเทียบประสิทธิภาพการจำแนกประเภทข้อมูลโดยใช้เทคนิคเหมืองข้อมูล 3 วิธี คือ วิธีด้วยต้นไม้ตัดสินใจ คือ 1. อัลกอริทึม J48 2. วิธีทฤษฎีความน่าจะเป็นแบบเบย์ คือ อัลกอริทึม Naive Bayes และ 3. วิธีระบบคอมพิวเตอร์จากโมเดลทางคณิตศาสตร์ คือ อัลกอริทึม โครงข่ายประสาทเทียม จากผลการเปรียบเทียบเทคนิคเหมืองข้อมูล ทั้ง 3 วิธี สรุปได้ว่า ชุดข้อมูลที่ 1 จำนวน 22 Attribute ที่ได้จากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องมาสร้างโมเดลเพื่อทำนายการสูญบุหรีของวัยรุ่น พบว่า ประสิทธิภาพทำนายวิธี J48 มีค่าความถูกต้อง (Accuracy) มากที่สุดคือ 81.95% รองลงมาวิธี Naive Bayes มีค่าความถูกต้องเท่ากับ 78.82% ลำดับสุดท้ายวิธี โครงข่ายประสาทเทียมมีค่าความถูกต้องเท่ากับ 78.45% , ชุดข้อมูลที่ 2 จำนวน 32 Attribute ที่ได้จากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องมาสร้างโมเดลเพื่อทำนายการสูญบุหรีของวัยรุ่น พบว่า ประสิทธิภาพทำนายวิธี J48 มีค่าความถูกต้อง (Accuracy) มากที่สุดคือ 80.85% รองลงมาวิธี โครงข่ายประสาทเทียมมีค่าความถูกต้องเท่ากับ 80.11% ลำดับสุดท้ายวิธี Naive Bayes มีค่าความถูกต้องเท่ากับ 77.90% และชุดข้อมูลที่ 3 จำนวน 47 Attribute ที่ได้จากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องมาสร้างโมเดลเพื่อทำนายการสูญบุหรีของวัยรุ่น พบว่า ประสิทธิภาพ

ทำนายวิธี โครงข่ายประสาทเทียม มีค่าความถูกต้อง (Accuracy) มากที่สุดคือ 81.40% รองลงมาวิธี J48 มีค่าความถูกต้องเท่ากับ 80.48% ลำดับสุดท้ายวิธี Naïve Bayes มีค่าความถูกต้องเท่ากับ 77.72% จากการเปรียบเทียบค่าความถูกต้อง (Accuracy) ของการใช้ข้อมูลชุดที่ 1, ข้อมูลชุดที่ 2 และ ข้อมูลชุดที่ 3 ทำนายผลการสูบบุหรี่ของวัยรุ่น พบว่า โมเดลทำนายของชุดข้อมูลทั้ง 3 ชุด มีค่าความถูกต้อง (Accuracy) เพิ่มขึ้นและลดลงแตกต่างกัน พิจารณารายวิธีพบว่า วิธี Naïve Bayes ลดลง 0.92% และ 0.18% , J48 ลดลง 1.1% และ 0.37% , โครงข่ายประสาทเทียม เพิ่มขึ้น 1.66% และ 1.29% สรุปได้ว่า โมเดลการทำนายผลการสูบบุหรี่ของวัยรุ่นด้วยวิธี J48 โดยใช้ข้อมูลชุดที่ 1 มีค่าความถูกต้องมากที่สุด คือ (Accuracy) 81.95%

5.2 ข้อเสนอแนะ

จากการวิจัยครั้งนี้ เพื่อให้เป็นแนวทางในการพัฒนางานวิจัยให้มีประสิทธิภาพมากขึ้นมีข้อเสนอแนะดังนี้

5.2.1 สามารถนำวิธีการวิเคราะห์ J48 ไปเป็นแนวทางในการคัดกรองปัจจัยที่ทำให้วัยรุ่นสูบบุหรี่ เนื่องจากเป็นวิธีที่มีร้อยละในการพยากรณ์ถูกต้องค่อนข้างสูง

5.2.2 เพื่อนำข้อมูลที่ได้ไปปรับใช้เป็นข้อมูลพื้นฐานในการจัดโครงการอบรมเยาวชนปรับเปลี่ยนพฤติกรรมการสูบบุหรี่ของกลุ่มผู้สูบบุหรี่ ใช้เป็นแนวทางสู่การเป็นมหาวิทยาลัยปลอดบุหรี่ตามแนวทางการเข้าร่วมนโยบายของเครือข่ายวิชาชีพสุขภาพเพื่อสังคมไทยปลอดบุหรี่

บรรณานุกรม

- ใจชื่น ตะเกาพงษ์. (2550). พฤติกรรมการสูบบุหรี่ของนักเรียนในช่วงชั้นที่ 3 โรงเรียน สาธิต
มหาวิทยาลัยศรีนครินทรวิโรฒ ปทุมวัน (วิทยานิพนธ์ปริญญาโทมหาบัณฑิตที่ไม่มีการตีพิมพ์).
มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ณัฐวดี หงษ์บุญมี และ ประภาสศิริ ตรีพานิชกุล. (2562). การเปรียบเทียบประสิทธิภาพการจำแนก
ข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์ไทรอยด์ด้วยเทคนิค
เหมืองข้อมูล. วารสารโครงการวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ,
1(9), 41-51.
- เทศบาลตำบลขุนทะเล. (2564). พระราชบัญญัติควบคุมผลิตภัณฑ์ยาสูบ พ.ศ. 2560 มีผลบังคับใช้ตาม
กฎหมายในวันที่ 4 กรกฎาคม 2560. <http://www.khuntalae.go.th/networknews/detail/75954>
- ชนวัฒน์ ปัญญาวงศ์. (2555). เทคนิคการจำแนกข้อมูล (Classification).
<http://tanawatpan.blogspot.com/2012/06/classification.html>
- ชนวัฒน์ เสริฐสุวรรณกุล. (2564). การศึกษาพฤติกรรมผู้ใช้เว็บไซต์สำนักหอสมุดกลางมหาวิทยาลัย
ศรีนครินทรวิโรฒ (รายงานการวิจัย). สำนักหอสมุดกลาง มหาวิทยาลัยศรีนครินทรวิโรฒ.
- นนทรี สัจจาธรรม. (2555). ทักษะคิด และความตั้งใจในการเลิกสูบบุหรี่ของนักศึกษาวิทยาลัย
ราชพฤกษ์ (รายงานการวิจัย). วิทยาลัยราชพฤกษ์.
- นพรัตน์ นนท์ศิริ, ราตรี มนต์ศิลา และ กริช สมกันธา. (2561). การจำแนกข้อมูลเพื่อวินิจฉัยความเสี่ยง
การเป็นโรคเบาหวาน โดยใช้เทคนิคเหมืองข้อมูล. วารสารวิชาการพระจอมเกล้าพระนคร
เหนือ, 33(2), 538-547.
- นิเวศ จิระวิจิตรชัย. (2553). การค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ
(รายงานการวิจัย). มหาวิทยาลัยราชภัฏสวนสุนันทา.
- บริษัท คิวบ์ อนาคติกส์ คอนซัลตติ้ง จำกัด. (2561). กระบวนการเหมืองข้อมูลของ CRISP-DM.
<https://datacubeth.ai/crisp-dm/>

- เบญจกัศ จงหมื่นไวย์. (2558). การเปรียบเทียบปัจจัยโรคประจำตัวผู้สูงอายุโดยใช้อัลกอริทึม การจัดกลุ่ม J48 และ Naïve Bayes : กรณีศึกษาสาธารณสุข โพรซีกกลางนครราชสีมา. *วารสารวิชาการ การประยุกต์ใช้เทคโนโลยีสารสนเทศ, 1*(2), 43-51. <https://ph02.tci-thaijo.org/index.php/project-journal/article/view/152660>
- ปกรณ์ กัชชาติ . (2562) . *Confusion Matrix เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนายใน Machine learning*. <https://citly.me/noqEN>
- ปริศนา คำเงิน. (2553). การวิจัยเรื่องความคิดเห็นต่อปัจจัยจูงใจที่ส่งผลต่อการสูบบุหรี่ของนักเรียนระดับปริญญาตรี (วิทยานิพนธ์ปริญญาโทที่ไม่มีใครตีพิมพ์). สาขาวิชาบริหารการ พัฒนาสำนักบริหารและพัฒนาวิชาการ คณะพัฒนาการท่องเที่ยว มหาวิทยาลัยแม่โจ้.
- เพชรรัตน์ ม่วงน้อย และคณะ. (2021). ศึกษาและพัฒนาตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาด้วยเทคนิคการเหมืองข้อมูล. *วารสารวิชาการ การประยุกต์ใช้เทคโนโลยีสารสนเทศ, 7*(1), 54-63.
- ภรณ์ยา ปาลวิสุทธิ. (2559). การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดเชื้อในเนื้อ. *วารสารเทคโนโลยีสารสนเทศ, 12*(1), 10.
- วาทิตย์ คำพรมมา. (2562). *แบบจำลองการวิเคราะห์ ความรู้สึกแบบผสมสำหรับความคิดเห็นต่อโรงแรมในประเทศไทยโดยใช้ K-means และ K-NN* (รายงานการวิจัย). ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.
- วิทยา พรพัชรพงศ์. (2012). *โครงข่ายประสาทเทียม (Artificial Neural Networks - ANN)*. <https://www.gotoknow.org/posts/163433>
- วิลาวัณย์ ประทีปแก้ว, พรเทพ ล้อมพรม และ วัดภา พรหมชัย. (2552). *พฤติกรรมกาสูบบุหรี่ของเยาวชนในจังหวัดนนทบุรี* (รายงานการวิจัย). สำนักงานสาธารณสุขจังหวัดนนทบุรี สำนักงานปลัดกระทรวงสาธารณสุข กระทรวงสาธารณสุข.
- ศจี วานิช. (2558). *Data Mining (เหมืองข้อมูล)*. <http://sajeegm301.blogspot.com/2015/11/data-mining.html>
- ศิริพร คุรุบา. (2553). *ทฤษฎีการทำเหมืองข้อมูล*. <http://siripornk.blogspot.com/2010/08/data-mining.html?m=1>

- สรุป Machine Learning(EP.6)-การจัดหมวดหมู่ด้วย Naive Bayes. (2563) . <https://citly.me/3okgC>
- สัญญา พันธุ์แพง. (2563). การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลสำหรับการพยากรณ์การศึกษาต่อ
นักศึกษาระดับปริญญาตรี ในมหาวิทยาลัยราชภัฏเชียงใหม่. (รายงานการวิจัย). วิทยาลัย
แม่ฮ่องสอน มหาวิทยาลัยราชภัฏเชียงใหม่.
- สายชล สิ้นสมบูรณ์ทอง. (2561 ก). การเปรียบเทียบประสิทธิภาพในการทำนายการติดเกมของเด็ก
และวัยรุ่นในกรุงเทพมหานคร. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 26(3), 405-416.
- สายชล สิ้นสมบูรณ์ทอง. (2561 ข). การเปรียบเทียบประสิทธิภาพในการทำนายผลการเป็น
โรคเบาหวาน. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 26(2), 195-207.
- สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ. (2560). รู้จักกับสารพิษสำคัญในบุหรี่.
<https://shorturl.asia/7eNLQ>
- สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ. (2561 ก). 19 โรค จากการสูบบุหรี่.
<https://shorturl.asia/e4qJN>
- สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ. (2561 ข). ปัจจัยเสี่ยงที่ทำให้ติดบุหรี่.
<https://shorturl.asia/fDVbQ>
- สุพัตรา ปัญญาคุณ, ธัญลักษณ์ คล้ายสงคราม และ ชิปไต่ย พงษ์ศาสตร์. (2557). การศึกษาข่ายงานเบย์
และการประยุกต์ *PARTs and Applications*.
http://sc2.kku.ac.th/stat/statweb/images/Eventpic/60/Seminar/02_1_Bayesian-networks-and-Applications.pdf
- สุรวัชร ศรีเปารยะ และ สายชล สิ้นสมบูรณ์ทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนก
กลุ่มการเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย, *วารสาร
วิทยาศาสตร์และเทคโนโลยี*, 25(5), 839-853.
- อรนุช พันโท และ มนต์ชัย เทียนทอง. (2558). การเปรียบเทียบประสิทธิภาพการจำแนกรูปแบบการ
เรียนรู้ VARK ด้วยเทคนิคเหมืองข้อมูล. *วารสารเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏ
อุบลราชธานี*, 4(1), 1-11.

อรุณรักษ์ ตันพานิช, คุณฉวี สุภววรรณกุล, พิเชฐ บัญญัติ และ จรุง จันแทน. (2562). การเปรียบเทียบ
โมเดลการเรียนรู้ของเครื่องสำหรับคัดกรองผู้ป่วยเบาหวานที่มีภาวะช้ำปลายเท้า. ใน *การประชุม
หาดใหญ่วิชาการระดับชาติและนานาชาติ ครั้งที่ 10* (น. 736-749). มหาวิทยาลัย
หาดใหญ่.





ภาคผนวก

ภาคผนวก ก

แบบสอบถามงานวิจัย

เรื่อง การเปรียบเทียบประสิทธิภาพการจำแนกเพื่อทำนายข้อมูลการสูญบุหรีของวัยรุ่นด้วยเทคนิค
เหมืองข้อมูล

คำชี้แจง

แบบสอบถามเรื่องนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการสร้างแบบจำลองในการทำนายข้อมูล โดยใช้เทคนิคเหมืองข้อมูล ดังนั้นผู้วิจัยจึงใคร่ขอความอนุเคราะห์และความร่วมมือในการตอบแบบสอบถามครั้งนี้ตามความเป็นจริง ผู้วิจัยขอรับรองว่าข้อมูลที่ท่านให้ถือเป็นความลับและจะนำเสนอการวิเคราะห์ในภาพรวมเท่านั้น

โดยแบบสอบถามชุดนี้แบ่งออกเป็น 3 ตอน ดังนี้

ตอนที่ 1 ข้อมูลทั่วไป จำนวน 24 ข้อ

ตอนที่ 2 โทษและพิษภัยของบุหรี จำนวน 10 ข้อ

ตอนที่ 3 เจตคติและค่านิยมต่อการสูญบุหรี จำนวน 15 ข้อ

ขอขอบคุณในความร่วมมือ

นางสาวศุภนิดา เปี่ยมาลัย

นางสาวนภัสกร นำพรชวลรุ่ง

นางสาวลลิตา เกิดสันเทียะ

นักศึกษาหลักสูตร วิทยาศาสตร์บัณฑิต สาขาเทคโนโลยีสารสนเทศ

มหาวิทยาลัยสยาม

ตอนที่ 1 ข้อมูลทั่วไป

คำชี้แจง : โปรดทำเครื่องหมาย ✓ ลงในช่อง () ที่ตรงกับข้อมูลของท่านเพียงช่องเดียว

1. เพศ

() หญิง () ชาย

2. ท่านอายุเท่าไร.....

3. ท่านเรียนอยู่คณะอะไร

() คณะสาธารณสุขศาสตร์ () คณะแพทยศาสตร์
 () คณะพยาบาลศาสตร์ () คณะเภสัชศาสตร์
 () คณะวิทยาศาสตร์ () คณะบริหารธุรกิจ
 () คณะวิศวกรรมศาสตร์ () วิทยาลัยดนตรีและศิลปการแสดง
 () คณะนิเทศศาสตร์ () คณะศิลปศาสตร์
 () คณะรัฐศาสตร์ () คณะนิติศาสตร์
 () คณะเทคโนโลยีสารสนเทศ () คณะรัฐศาสตร์
 () วิทยาลัยนานาชาติ

4. ระดับการศึกษา

() ระดับปริญญาตรี () ระดับปริญญาโท () ระดับปริญญาเอก

5. ท่านอยู่ชั้นปีที่เท่าไร

() ชั้นปีที่ 1 () ชั้นปีที่ 2
 () ชั้นปีที่ 3 () ชั้นปีที่ 4
 () ชั้นปีที่ 5 () ชั้นปีที่ 6

6. ก่อนเข้าเรียนในสาขานี้เกรดเฉลี่ยสะสมของท่านเท่าไร

() ต่ำกว่า 2.00 () 2.00 - 2.49
 () 2.50 - 2.99 () 3.00 - 3.49
 () มากกว่า 3.50 ขึ้นไป

7. ระดับคะแนนเฉลี่ยสะสมของท่านที่ศึกษาในปัจจุบันได้เกรดเฉลี่ยสะสมเท่าไร

() ต่ำกว่า 2.00 () 2.00 - 2.49
 () 2.50 - 2.99 () 3.00 - 3.49
 () มากกว่า 3.50 ขึ้นไป

8.สถานภาพสมรสของบิดา-มารดา

- () อยู่ด้วยกัน () หม้าย
() แยกกันอยู่ () หย่าร้าง

9.สถานที่พักอาศัย (ระหว่างกำลังเรียนอยู่)

- () หอพัก () บ้าน
() คอนโดมิเนียม () อื่น ๆ (โปรดระบุ)

10.ปัจจุบันท่านพักอาศัยอยู่กับใคร

- () ครอบครัว () เพื่อน
() แฟน () คนเดียว
() บิดา () มารดา
() อื่น ๆ (โปรดระบุ)

11.อาชีพของบิดา ของท่านทำอาชีพอะไร

- () เอกชน () ราชการ
() รัฐวิสาหกิจ () อาชีพอิสระ
() เจ้าของธุรกิจ () อื่น ๆ (โปรดระบุ)

12.อาชีพของมารดา ของท่านทำอาชีพอะไร

- () เอกชน () ราชการ
() รัฐวิสาหกิจ () อาชีพอิสระ
() เจ้าของธุรกิจ () อื่น ๆ (โปรดระบุ)

13.ปัจจุบันบิดาของท่าน สุขุมหรือหรือไม่

- () สุขุม () ไม่สุขุม

14.ปัจจุบันมารดาของท่าน สุขุมหรือหรือไม่

- () สุขุม () ไม่สุขุม

15.มีรายได้ของบิดาเฉลี่ยต่อเดือนเท่าไร

- () ต่ำกว่า 15,000 บาท () 15,000 - 20,000 บาท
() 20,001 - 25,000 บาท () 25,001 - 30,000 บาท
() มากกว่า 30,000 บาทขึ้นไป

16. มีรายได้ของมารดาเฉลี่ยต่อเดือนเท่าไร

- () ต่ำกว่า 15,000 บาท () 15,000 - 20,000 บาท
 () 2,0001 - 25,000 บาท () 25,001 - 30,000 บาท
 () มากกว่า 30,000 บาทขึ้นไป

17. บ้านที่ท่านพักอยู่เป็นของใคร

- () บิดา มารดา () เซ้า
 () เพื่อน () ญาติ
 () อื่น ๆ (โปรดระบุ)

18. ภูมิลำเนาเดิมของคุณอยู่จังหวัดอะไร

19. สมาชิกภายในครอบครัวของท่านมีด้วยกันกี่คน

- () 1 - 2 คน () 3 - 4 คน
 () 5 - 6 คน () มากกว่า 6 คนขึ้นไป

20. ตั้งแต่เกิดมาท่านพักอาศัยอยู่กับใครมากที่สุด

- () ครอบครัว () ญาติ
 () พี่น้อง () เพื่อน
 () คนที่รู้จัก () อื่น ๆ (โปรดระบุ)

21. ท่านเล่นอินเทอร์เน็ตเน็ทวันละกี่ชั่วโมง

- () 1 - 3 ชั่วโมง () 4 - 6 ชั่วโมง
 () 7 - 8 ชั่วโมง () มากกว่า 8 ชั่วโมงขึ้นไป

22. ท่านชอบเล่นเกมสล็อตออนไลน์แนวไหน (ตอบได้มากกว่า 1 ข้อ)

- () เล่นแบบหรือการจำลอง () ผจญภัย () RPG
 () ต่อสู้ () วางแผน () ปริศนา
 () กีฬาและการแข่งขัน () การศึกษา () ไม่เล่น

23. ท่านมีเพื่อนสูบบุหรี่หรือไม่

- () มี () ไม่มี

24. ท่านสูบบุหรี่ไม่

- () สูบบุหรี่ () ไม่สูบบุหรี่

ตอนที่ 2 โทษและพิกษณ์ของบุหรี่

คำชี้แจง : โปรดทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับความคิดเห็นของท่านมากที่สุดเพียงช่องเดียว

ข้อ ที่	ข้อความ	เห็น ด้วย	ไม่ แน่ใจ	ไม่เห็น ด้วย
1	การสูบบุหรี่เป็นสาเหตุทำให้เกิดโรคมะเร็งที่อวัยวะต่าง ๆ ของร่างกาย			
2	การสูบบุหรี่เป็นสาเหตุทำให้เส้นเลือดในสมองแตก และเป็นอัมพาตได้			
3	สารพิษในควันบุหรี่ที่ทำให้ผู้สูบบุหรี่มีอาการติดบุหรี่คือ สารนิโคติน (Nicotine)			
4	โรคถุงลมโป่งพองเกิดจากการสูบบุหรี่ทำให้ถุงลมปอดถูกทำลายและมีอาการหดร่งของหลอดลม			
5	ก้นกรองของบุหรี่ไม่ได้ทำให้ปริมาณของสารนิโคตินในบุหรีลดลง			
6	ในควันบุหรีมีสารตะกั่วซึ่งเป็นสารที่ยับยั้งการเจริญเติบโตของร่างกาย และทำลายเนื้อเยื่อสมอง			
7	ควันบุหรีที่ออกจากผู้สูบได้ทำร้ายคนข้างเคียงเรียกว่า “ควันบุหรีมือสอง”			
8	ตามประกาศกระทรวงสาธารณสุข พ.ศ. 2551 กำหนดให้แสดงเครื่องหมายของเขตสูบบุหรี และเขตปลอดบุหรีในที่เปิดเผย และสามารถมองเห็นได้ชัดเจน			
9	ผู้ที่ฝ่าฝืนสูบบุหรีในเขตพื้นที่ปลอดบุหรีมีโทษปรับไม่เกิน 2,000 บาท			
10	เจ้าของสถานที่ ถ้าไม่จัดให้มีเครื่องหมายเขตปลอดบุหรี หรือเขตสูบบุหรี มีโทษปรับไม่เกิน 2,000 บาท			

ตอนที่ 3 เจตคติและค่านิยมต่อการสูบบุหรี่

คำชี้แจง : โปรดทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับความคิดเห็นของท่านมากที่สุดเพียงช่องเดียว

ข้อที่	ข้อความ	เห็น ด้วย อย่าง ยิ่ง	เห็น ด้วย	ไม่ แน่ใจ	ไม่ เห็น ด้วย	ไม่เห็น ด้วย อย่าง ยิ่ง
1	การสูบบุหรี่เป็นพฤติกรรมไม่ปกติในสังคมไทย					
2	ผู้ชายสูบบุหรี่เป็นพฤติกรรมไม่ปกติ					
3	ผู้หญิงสูบบุหรี่เป็นพฤติกรรมไม่เหมาะสม					
4	ผู้ใหญ่สูบบุหรี่เป็นพฤติกรรมที่ไม่ควรยอมรับ					
5	วัยรุ่นสูบบุหรี่เป็นพฤติกรรมที่ไม่เหมาะสม					
6	การสูบบุหรี่ไม่แสดงถึงการเป็นคนทันสมัย เท่าที่					
7	การสูบบุหรี่ไม่ทำให้คุณเป็นที่ยอมรับ ของเพื่อน ๆ					
8	การสูบบุหรี่เป็นพฤติกรรมที่ไม่น่าเลียนแบบ					
9	บุหรี่เป็นสิ่งเสพติด ทำลายสุขภาพ					
10	บุหรี่เป็นสิ่งเสพติดที่อันตรายเพราะจะนำไปสู่สิ่ง เสพติดอื่น ๆ					
11	ท่านรังเกียจเพื่อนหรือบุคคลที่สูบบุหรี่					
12	ถ้าท่านสูบบุหรี่เพื่อนหรือคนรอบข้างรังเกียจ					
13	ถ้าท่านอยู่ในกลุ่มที่สูบบุหรี่ ท่านจำเป็นต้องสูบบุหรี่					
14	การสูบบุหรี่ทำให้สิ้นเปลืองค่าใช้จ่ายในการ รักษาโรคที่มีสาเหตุจากการสูบบุหรี่โดยไม่ จำเป็น					
15	การเลิกบุหรี่ทำให้เสียบุคลิกภาพของตนเอง					

** ขอขอบพระคุณทุกท่านที่กรุณาเสียสละเวลาในการให้ข้อมูลตอบแบบสอบถามในครั้งนี้ **

ประวัติคณะผู้จัดทำ

รหัสนักศึกษา	6105100001
ชื่อ-นามสกุล	นางสาวศุภนิดา เปี่ยมาลัย
ที่อยู่	1269 ซอยอนามัยงามเจริญ 27 แขวงท่าข้าม เขตบางขุนเทียน จังหวัดกรุงเทพฯ 10150
เบอร์โทรศัพท์	092-272-5814
E-Mail	onlyyou2205@gmail.com
ระดับมัธยมศึกษาตอนปลาย	โรงเรียนรัตนโกสินทร์สมโภชบางขุนเทียน
ระดับปริญญาตรี	มหาวิทยาลัยสยาม
รหัสนักศึกษา	6105100005
ชื่อ-นามสกุล	นางสาวนภัสกร นำพรชาวรุ่ง
ที่อยู่	79/203 เขตจอมทอง แขวงบางมด หมู่บ้านสินทวี ซอยอนามัยงามเจริญ 11 ถนนพระราม 2 จังหวัดกรุงเทพฯ 10150
เบอร์โทรศัพท์	081-802-5585
E-Mail	nutsumi34@gmail.com
ระดับมัธยมศึกษาตอนปลาย	โรงเรียนสายปัญญาในพระบรมราชินูปถัมภ์
ระดับปริญญาตรี	มหาวิทยาลัยสยาม
รหัสนักศึกษา	6105100008
ชื่อ-นามสกุล	นางสาวลลิตา เกิดสันเทียะ
ที่อยู่	20/1 หมู่ที่ 1 ตำบล หอมเกร็ด อำเภอ สามพราน จังหวัดนครปฐม 73110
เบอร์โทรศัพท์	061-629-2944
E-Mail	Lalita.Ker@siam.edu
ระดับมัธยมศึกษาตอนปลาย	โรงเรียนตลาดโพธิ์พิทยาคม
ระดับปริญญาตรี	มหาวิทยาลัยสยาม