การเปรียบเทียบอัลกอริทึมจำแนกมะเร็งมดลูกและมะเร็งปากมดลูก ในการดำเนินงานธุรกิจของโรงพยาบาล

Comparison of Classification Algorithms for Uterine Cancer and Cervical Cancer in Hospital Operations

จีรนันท์	เฉียบแหลม	6201700001
ยศวดี	อุปเนตร	6301700001
เบญจกัลยาณี	แสงมาลา	6301700002

ภาคนิพนธ์เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาบริหารธุรกิจบัณฑิต

รายวิชาโครงงานด้านคอมพิวเตอร์ธุรกิจ

ภาควิชาธุรกิจดิจิทัล คณะเทคโนโลยีสารสนเทศ

มหาวิทยาลัยสยาม

พ.ศ. 2567

หัวข้อภาคนิพนธ์	การเปรียบเทียบอัลกอริทึมจำแนกมะเร็งมดลูกและมะเร็งปากมดลูก ในการดำเนินงานธุรกิจของโรงพยาบาล Comparison of Classification Algorithms for Uterine Cancer and Cervical Cancer in Hospital Operations
หน่วยกิตของภาคนิพนธ์	3 หน่วยกิต
คณะผู้จัดทำ	นางสาวจีรนันท์ เฉียบแหลม 6201700001 นางสาวยศวดี อุปเนตร 6301700001 นางสาวเบญจกัลยาณี แสงมาลา 6301700002
อาจารย์ที่ปรึกษา	ผศ.ดร.พิชญากร เลค
ระดับการศึกษา	บริหารธุรกิจบัณฑิต
สาขาวิชา	ธุรกิจดิจิทัล
ปีการศึกษา	2567

อนุมัติให้ภาคนิพนธ์นี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาบริหารธุรกิจบัณฑิต สาขาวิชาธุรกิจดิจิทัล

คณะกรรมการสอบภาคนิพนธ์

..... ประธานกรรมการ
(ผศ.วัจนา ขาวฟ้า)

..... กรรมการสอบ
(อาจารย์ศรัญธร มั่งมี)

..... อาจารย์ที่ปรึกษา
(ผศ.ดร.พิชญากร เลค)

หัวข้อภาคนิพนธ์	การเปรียบเทียบอัลกอริทึมจำแนกมะเร็งมดลูกและมะเร็งปากมดลูก ในการดำเนินงานธุรกิจของโรงพยาบาล	
หน่วยกิตของภาคนิพนธ์	3 หน่วยกิต	
คณะผู้จัดทำ	นางสาวจีรนันท์ เฉียบแหลม	6201700001
	นางสาวยศวดี อุปเนตร	6301700001
	นางสาวเบญจกัลยาณี แสงมาลา	6301700001
อาจารย์ที่ปรึกษา	ผศ.ดร.พิชญากร เลค	
ระดับการศึกษา	บริหารธุรกิจบัณฑิต	
สาขาวิชา	ธุรกิจดิจิทัล	
ปีการศึกษา	2567	

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองในการพยากรณ์มะเร็งปากมดลูกจากผู้ป่วยที่เป็นมะเร็งปากมดลูกจำนวน 165 คน และผู้ป่วยที่เป็นมะเร็งมดลูก จำนวน 88 คน รวม 253 คน เดือน พฤศจิกายน 2565 ถึงเดือน มกราคม 2566 เทคนิคการจำแนกข้อมูล (Classification) โดยใช้อัลกอริทึม NaiveBayes SMO IBk และ J48 โดยพิจารณาจากค่าความถูกต้อง (Correctly Instances) และค่าความแม่นยำ (Precision)

จากผลการวิจัยพบว่า J48 มีค่าความถูกต้องสูงสุดเท่ากับ 57.33% (±0.60%) NaiveBayes มีค่าความถูกต้องเท่ากับ 56.67% (±0.59%) SMO มีค่าความถูกต้องเท่ากับ 51.33% (±0.55%) และ IBk มีค่าความถูกต้องเท่ากับ 49.33 % (±0.55%) ตามลำดับ จึงอธิบายได้ว่าการเกิดโรคมะเร็งดังกล่าวที่เกิดขึ้นจากการสำรวจและวิจัยผลที่ทำให้เป็นโรคมะเร็งปากมดลูกเกิดจากพฤติกรรมการใช้ชีวิตของตัวผู้ป่วย

คำสำคัญ: วิธีการจำแนกข้อมูล Classifier/ วิธีต้นไม้ตัดสินใจ/ การทำเหมืองข้อมูล

Title	A Comparison of Classification Algorithms for Uterine Cancer and Cervical Cancer in Hospital Operations		
Credits	3 Credits		
By	Miss Jeeranan	Chablame	6201700001
	Miss Yodsawadee	Uppanant	6301700001
	Miss Benjakunlayanee	Sangmala	6301700002
Advisor	Asst. Prof. Dr. Pitchayakorn Lake		
Degree	Bachelors of Business Administration		
Major	Digital Business		
Faculty	Information Technology		
Academic year	2024		

Abstract

The objective of this research is to develop a model for predicting cervical cancer in patients. There were 165 cases of cervical cancer and 88 cases of CA CERVIX, totaling 253 cases, from November 2022 to January 2023. The study used data classification techniques including NaiveBayes, SMO, IBk, and J48, based on the values of correctly instances and precision.

The research findings indicate that the J48 algorithm achieved the highest accuracy at 57.33% (±0.60%), followed by Naïve Bayes at 56.67% (±0.59%), SMO at 51.33% (±0.55%), and IBk at 49.33% (±0.55%), respectively. These results suggest that cervical cancer occurrence, as examined in this study, is influenced by patients' lifestyle behaviors.

Keywords : cervical cancer, classification, data mining, decision tree

กิตติกรรมประกาศ

ภาคนิพนธ์นี้จัดทำขึ้นเพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาบริหารธุรกิจบัณฑิต สาขาวิชาธุรดิจิทัล มหาวิทยาลัยสยาม โดยสามารถสำเร็จลุล่วงเพราะได้รับความกรุณาจาก ผู้ช่วยศาสตราจารย์ ดร.พิชญากร เลค ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ที่กรุณาสละเวลาอันมีค่าช่วยเหลือให้คำแนะนำข้อคิดเห็นต่างๆ อันเป็นประโยชน์ และชี้แนะถึงสิ่งที่สำคัญในการวิจัยโดยตลอด จนสำเร็จได้ด้วยดี ผู้จัดทำขอกราบขอบพระคุณ อาจารย์ ดร.กันทิมา คงสถิตสุวรรณ กราบขอบพระคุณอาจารย์ศรัญธร มั่งมี ซึ่งกรุณาให้คำปรึกษา วิธีคิดและแก้ไขปัญหา ข้อมูลต่างๆ ข้อเสนอแนะในการปรับปรุงวิทยานิพนธ์ให้มีความถูกต้อง และสร้างสรรค์ ผู้จัดทำจึงใคร่ขอขอบพระคุณท่านคณะอาจารย์ทุกท่านเป็นอย่างสูงไว้ ณ ที่นี้

ข้าพเจ้าขอขอบคุณ คณาจารย์ทุกท่าน ที่ให้การอบรมสั่งสอนในวิชาการด้านต่างๆ เพื่อนำความรู้ความสามารถมาปประยุกต์ใช้ให้เกิดประโยชน์ได้ รวมทั้งเพื่อนนักศึกษาภาควิชาการธุรกิจดิจิทัล มหาวิทยาลัยสยามที่ได้ช่วยผลักดันให้สามารถดำเนินการจัดทำวิทยานิพนธ์ได้ตามเวลาที่กำหนด และขอขอบคุณทุกท่านที่มีส่วนเกี่ยวข้องในการช่วยเหลือในการทำวิทยานิพนธ์ฉบับนี้

คณะผู้จัดทำ

สารบัญ

หน้า

บทคัดย่อ ..... ก

Abstract ..... ข

กิตติกรรมประกาศ ..... ค

สารบัญ ..... ง

สารบัญภาพ ..... ฉ

สารบัญตาราง ..... ช

บทที่

1 บทนำ 1

ความเป็นมาและความสำคัญของปัญหา 1

วัตถุประสงค์ 2

กลุ่มเป้าหมาย 2

ขอบเขตของโครงการ 2

ขั้นตอนการดำเนินงาน 3

ประโยชน์ที่คาดว่าจะได้รับ 4

2 แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง 5

ความรู้เกี่ยวกับมะเร็งปากมดลูก 5

ทฤษฎีที่เกี่ยวข้องกับการพยากรณ์ 8

งานวิจัยที่เกี่ยวข้อง 17

บทสรุปงานวิจัยที่เกี่ยวข้อง 20

3 วิธีการดำเนินวิจัย 28

รูปแบบงานวิจัย 29

ชุดข้อมูล 29

เครื่องมือในการวิจัย 29

วิธีดำเนินการวิจัย 29

สารบัญ (ต่อ)

บทที่ หน้า

4 ผลการวิจัย

ผลการวิจัย....33

5 สรุปผลการดำเนินงานและข้อเสนอแนะ

สรุปผลการดำเนินงาน....38

ปัญหาและอุปสรรค....38

ข้อเสนอแนะ....39

บรรณานุกรม....40

ภาคผนวก

ภาคผนวก ก .คู่มือการติดตั้งโปรแกรม....43

ประวัติคณะผู้จัดทำ....51

สารบัญตาราง

ตารางที่ หน้า

2.1 สรุปงานวิจัยที่เกี่ยวข้อง 20
4.1 สรุปผลการทดลอง 33
4.2 รายละเอียดกฎพื้นฐานที่แปลงจากแผนผังต้นไม้ตัดสินใจ 37

สารบัญภาพ

ภาพที่ หน้า

2.1 อวัยวะสืบพันธุ์สตรี 5
2.2 กระบวนการ Classification 11
2.3 ส่วนประกอบของต้นไม้ตัดสินใจ 12
2.4 ตัวอย่างต้นไม้ตัดสินใจ 12
2.5 กระบวนการมาตรฐานในการทำเหมืองข้อมูล 16
3.1 ต้นไม้ตัดสินใจโดยใช้เทคนิคการจำแนกข้อมูลมะเร็งมดลูกและมะเร็งปากมดลูก ในการดำเนินงานธุรกิจของโรงพยาบาล 28
3.2 ผลการคำนวณขนาดกลุ่มตัวอย่าง Krejcie & Morgan 30
3.3 ตัวอย่างตารางข้อมูลมะเร็งปากมดลูก 31
3.4 การเลือกข้อมูลในพยากรและการจำแนกข้อมูลอัลกอริทึม 32
4.1 แสดงผลลัพธ์การจำแนกโดยใช้อัลกอริทึม Naive Bayes 33
4.2 แสดงผลลัพธ์การจำแนกโดยใช้อัลกอริทึม SMO 34
4.3 แสดงผลลัพธ์การจำแนกโดยใช้อัลกอริทึม IBk 34
4.4 แสดงผลลัพธ์การจำแนกโดยใช้อัลกอริทึม J48 35
4.5 การแสดงผลลัพธ์ต้นไม้ตัดสินใจการใช้เทคนิคการจำแนกข้อมูลเพื่อการพยากรณ์มะเร็งปากมดลูกในการดำเนินงานธุรกิจของโรงพยาบาล 36
ก.1 หน้าจอลงโปรแกรม Weka 43
ก.2 หน้าจอเลือก I Agree 44
ก.3 หน้าจอ Choose Components 45
ก.4 หน้าจอ Choose Install Location 46
ก.5 หน้าจอ Installing 47
ก.6 หน้าจอการติดตั้งเสร็จสิ้นแล้ว Weka 3.8.4 48
ก.7 หน้าจอเลือก Applications Weka 49
ก.8 หน้าจอของ Explorer Weka 50

บทที่ 1
บทนำ

<u>ความเป็นมาและความสำคัญของปัญหา</u>

มะเร็งปากมดลูกเป็นปัญหาที่เกิดขึ้นกับผู้หญิงทั่วโลกโดยเฉพาะในประเทศที่พัฒนาแล้ว พบว่ามะเร็งปากมดลูกเป็นอันดับ 2 รองจากมะเร็งเต้านมในแต่ละปีมีผู้ป่วยที่เป็นมะเร็งปากมดลูกหลายแสนคนและจำนวนผู้ป่วยที่เสียชีวิตจากมะเร็งปากมดลูกประมาณครึ่งนึงของจำนวนผู้ที่เป็นมะเร็งปากมดลูก (จตุพล ศรีสมบูรณ์, 2549) โดยที่ 1 ชั่วโมงมีผู้หญิงทั้งโลกเสียชีวิตจากมะเร็งปากมดลูกสูงสุด 30 คน (กระทรวงสาธารณสุข, 2551) สาเหตุหลักของโรคมะเร็งปากมดลูกเป็นการติดเชื้อ Human Papllloma Virus (HPV) ซึ่งลักษณะทางการระบาดเหมือนโรคติดต่อทางเพศสัมพันธ์ เนื่องจากไม่มีประสิทธิภาพในการตรวจคัดกรองโรคมะเร็งปากมดลูกระยะก่อนเป็นมะเร็งและในการรักษาก่อนที่จะเป็นมะเร็งปากมดลูกระยะลุกลาม (สถาบันมะเร็งแห่งชาติ, 2548) โดยมะเร็งปากมดลูกพบในผู้หญิงตั้งแต่อายุก่อน 30 จนถึง 80 ปี โดยพบมากในช่วงอายุ 35 - 50 ปี พฤติกรรมของผู้หญิงไทยส่วนใหญ่มักคิดว่าตัวเองไม่ใช่กลุ่มเสี่ยง และรู้สึกเขินอายกลัวที่จะพบแพทย์เพื่อตรวจภายในการคัดกรองมะเร็งปากมดลูกดังนั้นกว่าตนเองจะรู้ว่ามีอาการผิดปกติความรุนแรงของโรคก็มักอยู่ในระยะลุกลามยากต่อการรักษา

การทำเหมืองข้อมูล คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภททั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร (birdkritsna, 2562)การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูลจากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล ซึ่งการหากฎความสัมพันธ์ (Association rules) เป็นหนึ่งในเทคนิคการทำเหมืองข้อมูลเป็นการค้นหากฎความสัมพันธ์ของข้อมูล โดยค้นหาความสัมพันธ์ของข้อมูลทั้งสองชุดหรือมากกว่าสองชุดขึ้นไปไว้ด้วยกัน ความสำคัญของกฎทำการวัดโดยใช้ข้อมูลสองตัวด้วยกัน คือค่าสนับสนุน (Support) ซึ่งเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎสามารถนำไปใช้หรือเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎที่ใช้มีความถูกต้อง และข้อมูลตัวที่สองที่นำมาใช้วัดคือค่าความมั่นใจ (Confidence) ซึ่งเป็นจำนวนของกรณีที่กฎถูกต้องโดยสัมพันธ์กับจำนวนของกรณี ที่กฎสามารถนำไปใช้ได้ในการหากฎความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน แต่ขั้นตอนวิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลายคือ อัลกอริทึม Classifier

จากบทความข้างต้นผู้วิจัยศึกษาการจำแนกข้อมูลของการส่งเสริมการจัดโปรแกรมการตรวจสุขภาพกรณีศึกษาเป็นข้อมูลมะเร็งปากมดลูกจากสาขารังสีรักษาและมะเร็งวิทยา ฝ่ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์ มาจำแนกข้อมูลการความเสี่ยงที่ทำให้เกิดโรคมะเร็งได้ การใช้เทคนิคเหมืองข้อมูล และเลือกใช้การจำแนกข้อมูลและ และอัลกอริทึมที่นิยมคือ อัลกอริทึม Classifier เพื่อค้นหารูปแบบและการจำแนกข้อมูลที่ซ่อนอยู่ในข้อมูล ผลการวิจัยนี้จะทำให้ทราบถึงความการจำแนกข้อมูลในการเกิดผลเสี่ยงในการเกิดโรคมะเร็งเพราะมีสารหรืออาหารที่มีสารทำให้ก่อมะเร็งหรือไปเร่งเชื้อให้เกิดเร็วขึ้นทำให้ต้องหมั่นดูแลรักษาสุขภาพและ ตรวจสุขภาพเบื้องต้นกันมากขึ้นเพื่อให้พบว่าตนเองมีร่างกายปกติดีและตรวตรวจสุขภาพทุก 6 เดือน หรือ 1 ปี เพื่อสุขภาพที่ดีปลอดภัยจากโรคร้าย

<u>วัตถุประสงค์</u>

1. เพื่อศึกษาข้อเสนอแนะและแนวทางการปรับปรุงแก้ไขเกี่ยวกับการนำไปใช้ในโรงพยาบาล
2. เพื่อศึกษาปัจจัยของบุคคลที่มีผลความเสี่ยงที่จะเกิดโรค
3. เพื่อศึกษาคุณภาพการจัดบริการส่งเสริมคุณภาพและการจัดการด้านบริการของโรงพยาบาล

<u>กลุ่มเป้าหมาย</u>

กลุ่มสตรี ที่มีอายุตั้งแต่ 20ปี ขึ้นไป เป็นบุคคลทั่วไปที่ได้รับการตรวจมะเร็งปากมดลูก จำนวน 150 คน จาก จำนวนประชากร 253 คน

<u>ขอบเขตของโครงงาน</u>

1.ขอบเขตด้านข้อมูล

ข้อมูลที่ผู้วิจัยนำมาวิเคราะห์เป็นข้อมูลมะเร็งปากมดลูกจากสาขารังสีรักษาและมะเร็งวิทยา ฝ่ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์ ที่นำมาวิจัยใน เดือนพฤศจิกายน 2565 ถึง เดือนมกราคม 2566

2. ขอบเขตด้านการวิจัย

การวิจัยเรื่องมะเร็งปากมดลูกผู้วิจัยเลือกข้อมูลจำแนกประเภทข้อมูลโดยการใช้ Classifier ที่สามารถ แบ่งแยกข้อมูลที่มีลักษณะหรือเป้าหมายตามที่กำหนดไว้ ซึ่งมีตัวแปรอิสระ และตัวแปรตามดังนี้

3.ตัวแปรที่ใช้ในการวิจัย

ตัวแปรอิสระ (Independent Variable) ได้แก่

1. ปัจจัยส่วนบุคคลของกลุ่มตัวอย่าง ได้แก่

1.1อายุ

2. พฤติกรรมที่เสี่ยงต่อการเป็นมะเร็งปากมดลูก

2.1 จำนวนบุตร

2.2 การสูบบุหรี่

2.3 โรคติดต่อทางเพศสัมพันธ์

2.4 ระยะของโรค

2.5 มะเร็งที่เป็นอยู่

<u>คุณสมบัติของฮาร์ดแวร์</u>

1. ฮาร์ดแวร์สำหรับผู้พัฒนาระบบ

1.1 CPU AMD Ryzen 2.60 GHz.

1.2 Hard Disk 200 GB.

1.3 RAM 4.00 GB.

1.4 CD-ROM Drive

<u>คุณสมบัติของซอฟต์แวร์</u>

2. ซอฟต์แวร์สำหรับผู้พัฒนาระบบ

1.1 Microsoft Windows 10

1.2 Weka 3.8.4

1.3 Microsoft Excel 2019

1.4 Microsoft Word 2019

<u>ขั้นตอนการดำเนินงาน</u>

ในงานวิจัยการพยากรณ์หาปัจจัยในที่ทำให้เป็นมะเร็งปากมดลูกด้วยเทคนิคการจำแนกเหมืองข้อมูล กรณีศึกษาการใช้เทคนิคการจำแนกข้อมูลเพื่อการพยากรณ์มะเร็งปากมดลูกในการดำเนินงานของโรงพยาบาล โดยมีขั้นตอนดังนี้

1. ศึกษาและทำความเข้าใจกับปัญหาที่เกิดขึ้นของกลุ่มสตรีที่มีพฤติกรรมการใช้ชีวิตที่มีความเสี่ยงก่อให้เกิดการเป็นมะเร็งปากมดลูกโดยที่ไม่รู้ตัว

2. รวบรวมข้อมูลและเก็บข้อมูล โดยการรวบรวมข้อมูลจากศูนย์บริการสาธารณสุข 62 ข้อมูลที่ได้ คือ ผู้ป่วยที่เป็นมะเร็งปากมดลูก จำนวน 165 คน ผู้ป่วยที่เป็นมะเร็งมดลูก จำนวน 88 คน รวมจำนวนผู้ป่วยทั้งหมด 253 คน

3. การเตรียมข้อมูล และคัดเลือกข้อมูลเพื่อนำมาใช้ในการวิจัย ผู้วิจัยเตรียมข้อมูล และการคัดเลือกข้อมูลที่ได้มาเพื่อที่จะนำไปใช้กับเครื่องมือสำหรับการจำแนกของข้อมูล

4. การวิเคราะห์ข้อมูล ผู้วิจัยเลือกการใช้โปรแกรม WEKA 3.8.4 เพื่อการพยากรณ์และจำแนกข้อมูลของโรคมะเร็งปากมดลูก โดยใช้เทคนิคอัลกอริทึม Classifier

5. การทดลองข้อมูล และเลือกเทคนิคที่เหมาะสม เป็นการนำข้อมูลที่ได้จากการรวบรวมข้อมูลและจำแนกข้อมูลแล้ว จึงนำข้อมูลที่ได้มาทำการทดลองในการใช้โปรแกรม WEKA 3.8.4 ซึ่งเลือกเทคนิคที่เหมาะสมคือ เทคนิคอัลกอริทึม Classifier ในการจำแนกกลุ่มข้อมูล เพื่อนำมาใช้ในการส่งเสริมปัจจัยที่มีความเสี่ยงในการเกิดโรคมะเร็งปากมดลูก

6. ผลการวิเคราะห์ข้อมูล เมื่อได้นำเทคนิคอัลกอริทึม Classifier มาทำการทดลองกับข้อมูลแล้ว จึงมีผลลัพธ์ที่ได้จากการทดลองมาวิเคราะห์ที่ถูกต้องและมีกลุ่มข้อมูลที่สัมพันธ์กัน

7. สรุปผลการวิจัยและข้อเสนอแนะ จากผลการวิจัยพบว่าการจำแนกข้อมูลพฤติกรรมการทำให้เกิดโรคมะเร็งปากมดลูก มีระยะเวลาในการเก็บข้อมูลที่สั้น เพราะมีระยะเวลาในการเก็บข้อมูลเพียง 1 เดือนของการเก็บข้อมูลจึงทำให้มีความถูกต้องของข้อมูลที่น้อยกว่าการเก็บข้อมูลที่มีระยะเวลานานกว่า

8. นำเสนอผลวิเคราะห์ให้แก่หน่วยงานที่เกี่ยวข้อง เพื่อนำไปประยุกต์ใช้ในการปรับเปลี่ยนพฤติกรรมการใช้ชีวิตและนำไปใช้ให้เกิดประโยชน์

<u>ประโยชน์ที่คาดว่าจะได้รับ</u>

1. ได้รู้เกี่ยวกับปัจจัยความเสี่ยงที่ก่อให้เกิดโรคมะเร็งปากมดลูก
2. ได้รวบรวมปัจจัยความเสี่ยงมาเสนอแนะเป็นแนวทางการป้องกันโรคมะเร็งปากมดลูกให้กับบุคคลภายนอกได้ร

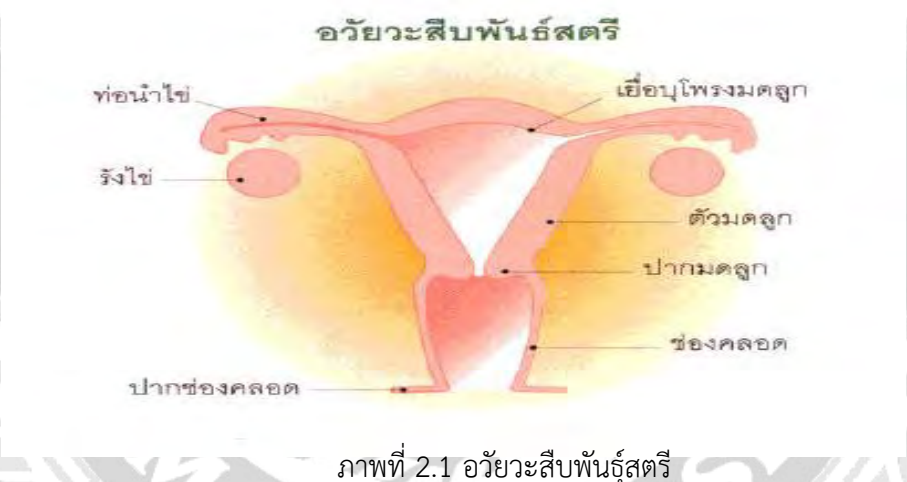
บทที่ 2
แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาค้นคว้าทบทวนวรรณกรรม งานวิจัย และทฤษฎีที่เกี่ยวข้องกับงานวิจัยเรื่องปัจจัยที่เกี่ยวข้องในการก่อให้เกิดมะเร็งปากมดลูก ดังนี้

1 ความรู้เกี่ยวกับมะเร็งปากมดลูก

มะเร็งปากมดลูก เป็นมะเร็งของอวัยวะสืบพันธุ์สตรี เป็นมะเร็งที่พบบ่อยที่สุดในสตรีไทย และทำให้มีผู้ป่วยเสียชีวิตเป็นจำนวนมากจากโรคนี้เราจึงควรมาทำความรู้จักโรคนี้

ปากมดลูกคือ ส่วนของอวัยวะสืบพันธุ์เพศหญิงที่ต่อจากมดลูกในช่องท้องโผล่ยื่นออกในช่องคลอด (ดังรูป)



ภาพที่ 2.1 อวัยวะสืบพันธุ์สตรี

ที่มา: https://med1.healthcare/articles

ปากมดลูกเป็นทางผ่านของเลือดประจำเดือน ซึ่งมาจากมดลูกและไหลออกมาภายนอกผ่านช่องคลอด ขณะมีเพศสัมพันธ์น้ำอสุจิจะไปอยู่ในช่องคลอดเข้าไปสู่มดลูกทางปากมดลูก เมื่อเกิดการปฏิสนธิกับไข่ของคุณผู้หญิงจะทำให้เกิดทารกน้อยๆ อาศัยในโพรงมดลูก และคลอดออกมาทางช่องคลอดเมื่อครบกำหนด สาเหตุสำคัญของโรคนี้เกิดจากเชื้อฮิวแมนแป๊ปปิโลม่าไวรัส (Human papilloma virus = HPV) หรืออีกชื่อหนึ่งว่า เชื้อเอชพีวี (HPV) ซึ่งติดต่อไปยังบุคคลอื่นๆ ได้จากการมีเพศสัมพันธ์ ในชีวิตของผู้หญิงคนหนึ่งอาจจะเคยได้รับเชื้อนี้ แต่ร่างกายสามารถกำจัดไปได้ และมีบางส่วนที่ไม่สามารถกำจัดเชื้อได้ รวมถึงมีปัจจัยเสี่ยงอื่นๆ ที่ทำให้เกิดการเปลี่ยนแปลงของเนื้อเยื่อบริเวณปากมดลูกและทำ

ให้เกิดโรคต่างๆ เช่น หูดหงอนไก่ มะเร็งปากมดลูก และอื่นๆ ปัจจัยเสี่ยงอื่นๆ นอกจากเชื้อ HPV ที่กล่าวมา ได้แก่

1. อายุ ส่วนใหญ่มะเร็งปากมดลูกมักพบในผู้ที่อายุมากกว่า 35 ปีขึ้นไป
2. มีคู่นอนหลายคน ทำให้มีโอกาสได้รับเชื้อ HPV มากขึ้น
3. สูบบุหรี่
4. มีบุตรจำนวนมาก
5. ร่างกายมีภูมิคุ้มกันบกพร่อง (โรคเอดส์)
6. ไม่เคยตรวจภายใน เพื่อค้นหารอยโรคระยะก่อนเป็นมะเร็ง (Precancercous lesion)

ช่วงเวลานี้ผู้ป่วยส่วนใหญ่จะไม่มีอาการ แต่สามารถตรวจพบด้วยการตรวจทางเซลล์วิทยาที่เรียกว่าแป๊ปสเมียร์ (Pap smear) และสามารถรักษาได้ หากตรวจพบรอยโรคในระยะนี้ก็จะสามารถป้องกันการเกิดเป็นมะเร็งปากมดลูกได้การตรวจคัดกรองมะเร็งปากมดลูก หมายถึง การตรวจหาการเปลี่ยนแปลงของเซลล์เนื้อเยื่อบริเวณปากมดลูกก่อนที่จะมีอาการ เป็นการตรวจหารอยโรคระยะก่อนเป็นมะเร็งและมะเร็งระยะเริ่มแรก ทำให้สามารถรักษาและป้องกันไม่ให้เกิดมะเร็งปากมดลูกได้ รวมถึงรักษามะเร็งระยะเริ่มแรกอย่างได้ผลวิธีการตรวจคัดกรองมะเร็งปากมดลูกในปัจจุบันมีหลายวิธี แต่ที่นิยมใช้คือการตรวจเซลล์ปากมดลูกด้วยวิธีทางเซลล์วิทยา หรือแป๊ปสเมียร์ (Pap smear) ซึ่งแพทย์จะทำการตรวจเช็คให้ในขณะที่ทำการตรวจภายใน ซึ่งหลังจากตรวจเสร็จแพทย์จะนัดฟังผลตรวจหรือแจ้งผลให้ทราบในภายหลัง หากมีความผิดปกติก็จะใช้การรักษาตามความผิดปกติของรอยโรค การตรวจคัดกรองมะเร็งปากมดลูกด้วยเซลล์วิทยา ร่วมกับการตรวจหาเชื้อ HPV ชนิดที่มีความเสี่ยงสูง การตรวจวิธีนี้มีข้อดีคือ แพทย์สามารถให้การดูแลรักษาเซลล์ปากมดลูกที่ผิดปกติได้ดีขึ้น แต่มีข้อเสียคือ ผู้ป่วยต้องเสียค่าใช้จ่ายเพิ่มขึ้นค่อนข้างมากในปัจจุบันแนะนำให้ผู้หญิงไทยตรวจภายในพร้อมตรวจแป๊ปสเมียร์ในกลุ่มที่มีเพศสัมพันธ์แล้วประมาณ 3 ปี หรืออายุมากกว่า 30 ปี หรือมีปัจจัยเสี่ยงอื่นๆ

ในระยะรอยโรคก่อนเป็นมะเร็ง และมะเร็งระยะเริ่มแรกผู้ป่วยส่วนใหญ่ไม่มีอาการ แต่เมื่อโรคเป็นมากขึ้น มีอาการดังต่อไปนี้

1. เลือดออกหลังจากมีเพศสัมพันธ์ สวนล้างช่องคลอด ตรวจภายใน
2. เลือดออกทางช่องคลอดที่ไม่ใช่รอบประจำเดือน
3. เลือดออกทางช่องคลอดหลังจากหมดประจำเดือนไปนานแล้ว (วัยทอง)
4. ตกขาวมากขึ้น มีกลิ่นเหม็น
5. ปวดท้องน้อย หรืออุ้งเชิงกราน

6. เจ็บขณะมีเพศสัมพันธ์

หากมีอาการเหล่านี้ควรปรึกษาสูตินรีแพทย์เพื่อตรวจหาความผิดปกติโดยเร็ว ถ้าแพทย์ตรวจแป๊ปสเมียร์พบมีผลเซลล์วิทยาผิดปกติ แพทย์อาจนัดตรวจส่องกล้องขยายดูบริเวณปากมดลูก เพื่อหาความผิดปกติและตัดชิ้นเนื้อไปตรวจ แต่หากแพทย์ตรวจพบความผิดปกติของบริเวณปากมดลูกด้วยตาเปล่า แพทย์จะทำการตัดชิ้นเนื้อไปตรวจเลย ซึ่งทำให้ทราบว่ามีความผิดปกติของเซลล์ปากมดลูกอยู่ในระยะก่อนเป็นมะเร็ง หรืออยู่ในระยะเป็นมะเร็งแล้ว ในกรณีที่เป็นรอยโรคระยะก่อนเป็นมะเร็ง แพทย์จะทำการรักษาด้วยวิธีการรักษาต่างๆ ได้แก่ การตรวจติดตาม, การจี้ความ เย็น, การตัดชิ้นเนื้อปากมดลูกด้วยห่วงไฟฟ้า เป็นต้น หากรอยโรคเป็นระยะมะเร็งปากมดลูกแล้ว แพทย์จะต้องทำการตรวจเพิ่มเติมเพื่อตรวจว่าผู้ป่วยอยู่ในระยะใดของมะเร็ง ซึ่งมะเร็งปากมดลูกมี 4 ระยะ

ระยะที่ 1 เป็นระยะต้น

ระยะที่ 2-3 เป็นระยะกลาง

ระยะที่ 4 เป็นระยะสุดท้าย อาจมีการแพร่กระจายของโรคไปยังอวัยวะอื่นๆ ในร่างกาย เช่น ตับ, ปอด, กระดูก เป็นต้น

การรักษาโรคนี้ขึ้นกับระยะของโรคดังที่กล่าวมาแล้ว หากเป็นระยะก่อนเป็นมะเร็ง หรือมะเร็งปากมดลูกระยะต้น แพทย์อาจใช้การผ่าตัด ซึ่งผลการรักษาดีมากโอกาสหายสูงมาก หากเป็นมะเร็งระยะกลาง การรักษาส่วนใหญ่ใช้การฉายรังสีรักษาร่วมกับให้ยาเคมีบำบัด ซึ่งผลการรักษาดีพอสมควร หากเป็นมะเร็งระยะสุดท้าย การรักษาส่วนใหญ่ทำได้เพียงแค่ประคับประคองอาการ บำบัดอาการเจ็บปวดของผู้ป่วย ผลการรักษาไม่ดี

ผลข้างเคียงจากการรักษา ขึ้นอยู่กับว่าได้รับการรักษาแบบใด โดยทั่วไปหากรักษาด้วยการผ่าตัด ผู้ป่วยมักจะมีภาวะแทรกซ้อนหรือผลข้างเคียง ในขณะผ่าตัด เช่น เสียเลือดมาก หรือหลังผ่าตัดใหม่ๆ เช่น แผลติดเชื้อ เป็นต้น แต่หากเป็นการรักษาด้วยวิธีฉายรังสีรักษาส่วนใหญ่มีผลข้างเคียงมักจะเกิดในช่วงหลังรักษา 2-3 ปีขึ้นไป อาการที่อาจเกิดขึ้น เช่น ปัสสาวะหรืออุจจาระเป็นเลือด อย่างไรก็ตามก่อนทำการรักษาแพทย์จะแจ้งผลข้างเคียงของการรักษาให้ทราบก่อนส่วนการรักษาด้วยยาเคมีบำบัด โดยทั่วไปจะเกิดผลข้างเคียงคือ คลื่นไส้อาเจียน อ่อนเพลีย เม็ดเลือดขาวต่ำติดเชื้อโรคง่ายขึ้น อาจมีผมร่วงในการใช้ยาบางชนิดดังที่กล่าวมาแล้ว คุณผู้หญิงสามารถป้องกันการเกิดมะเร็งปากมดลูกด้วยการรับการตรวจภายในพร้อมตรวจแป๊ปสเมียร์ เพื่อตรวจหารอยโรคระยะก่อนเป็นมะเร็ง ทำให้ป้องกันการเกิดมะเร็งปากมดลูกได้ในปัจจุบันเริ่มมีการผลิตวัคซีนสำหรับป้องกันโรคมะเร็งปากมดลูก ในขณะนี้ยังไม่สามารถป้องกันโรคได้ทั้งหมด หากคุณมีความสนใจอาจปรึกษาสูติ-นรีแพทย์เพื่อขอข้อมูล

เพิ่มเติมได้ (ผศ. นพ.ภุชงค์ ลิขิตธนสมบัติหน่วยมะเร็งวิทยานรีเวชภาควิชาสูติศาสตร์-นรีเวชวิทยา โรงพยาบาลรามาธิบดี)

2. ทฤษฎีที่เกี่ยวข้องกับการพยากรณ์

จากการศึกษาและค้นคว้างานวิจัยที่เกี่ยวกับการพยากรณ์ผู้ที่เป็นมะเร็งปากมดลูก โดยใช้วิธีการจำแนกข้อมูล ผู้วิจัยได้รวบรวมข้อมูล และ ทฤษฎีที่เกี่ยวข้องนำมาสรุปดังนี้

1. ทฤษฎีเหมืองข้อมูลการทำเหมืองข้อมูล (Data Mining Theory) การทำเหมืองข้อมูล (Data Mining) คือ การค้นหาความรู้ที่ถูกซ่อนอยู่ในฐานข้อมูลขนาดใหญ่ โดยการทำการค้นหารูปแบบความสัมพันธ์ของข้อมูลเพื่อให้ค้นพบความรู้ใหม่และนำความรู้ที่ได้มาทำการวิเคราะห์ความน่าเชื่อถือ (สุคนธ์ทิพย์, 2550) ซึ่งขั้นตอนการทำเหมืองข้อมูลมีดังนี้

1.1 การคัดเลือกข้อมูล (Data Selection) เป็นการบอกข้อมูลที่จำเป็นต้องใช้ในเหมืองข้อมูลทั้งหมด

1.2 การเตรียมข้อมูล (Data Preprocessing) เป็นกระบวนการที่ทำให้เกิดความมั่นใจในคุณภาพของข้อมูลที่จะนำมาใช้ โดยมีวิธีการดังนี้

1) การกรองข้อมูล (Data Cleaning)วิเคราะห์ความถูกต้องของข้อมูลโดยการนำข้อมูลที่ไม่ถูกต้องออก เช่น ข้อมูลที่ทำการบันทึกผิด ข้อมูลที่ไม่สอดคล้องกัน

2) การแปลงข้อมูล (Data Transformation)เป็นการแปลงข้อมูลที่เลือกมาให้อยู่ในรูปแบบที่เหมาะสมสำหรับการนำไปใช้วิเคราะห์ตามอัลกอริทึม

3) การทำเหมืองข้อมูลเป็นการใช้เทคนิคของการทำเหมืองข้อมูลต่าง ๆ มาทำการ mining ข้อมูล

4) การวิเคราะห์และประเมินผลลัพธ์ เป็นขั้นตอนการแปลความหมาย และการประเมินผลลัพธ์ที่ได้ว่ามีความเหมาะสมหรือ ตรงกับวัตถุประสงค์ที่ต้องการหรือไม่

ตัวจำแนกประเภทเบย์อย่างง่าย (Naive BayesClassifier) Naive Bayes เป็นเทคนิคที่ถูกตั้งชื่อตามThomas Bayes (1702-1761) และใช้ทฤษฎีของ BayesTheorem ในการคำนวณความน่าจะเป็นซึ่งถูกใช้ในการทำนายผล เมื่อทำการวิเคราะห์กรณีใหม่ การทำนายผลทำได้โดยการรวบรวมผลของตัวแปรอิสระ(independent variable) ที่มีแต่ตัวแปรตาม (dependent variable) Naive Bayes ใช้แก้ปัญหาการจำแนก (บุญเสริม, 2546) ซึ่งจะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นของความสัมพันธ์ที่เกิดขึ้นอีกทั้งยังมีความง่าย ไม่ซับซ้อน ให้ผลลัพธ์ที่ดีและมีความรวดเร็ว แต่อย่างไรก็ตามการใช้ Naive Bayes ไม่สนับสนุนข้อมูล

ต่อเนื่อง (Continuous data) ดังนั้นตัวแปรอิสระหรือตัวแปรตามที่มีค่าเป็นค่าต่อเนื่องจะต้องถูกแบ่งเป็นช่วง ๆ ก่อนจึงสามารถนำมาใช้งานได้(สุคนธ์ทิพย์, 2550)

Bayes' Theorem

$$P(H|E) = [P(E|H)xP(H)] / P(E) \quad (1)$$

โดย

P(H) คือความน่าจะเป็นที่จะเกิดเหตุการณ์ H

P(H|E) คือความน่าจะเป็นที่จะเกิดเหตุการณ์ H เมื่อเกิดเหตุการณ์ E

$$P(c_i \mid \vec{d}_j) = \frac{P(c_i)P(\vec{d}_j \mid c_i)}{P(\vec{d}_j)}$$

2. การตัดสินใจแบบโครงสร้างต้นไม้ (Decision Tree)

2.1 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคหนึ่งของ Classification ซึ่งเป็นวิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูล โดย Classification นั้นเป็นเทคนิคหนึ่งของเหมืองข้อมูล (Data Mining) เทคนิคเหมืองข้อมูล ได้แก่

1) เทคนิค Classification เป็นเทคนิคในการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะต่างๆ ที่ได้มีการกำหนดไว้แล้ว เทคนิคประเภทนี้เหมาะกับการสร้างแบบจำลองเพื่อการพยากรณ์ค่าข้อมูล (Predictive Modeling) ในอนาคตจากการที่ได้จำแนกกลุ่มข้อมูลตัวอย่างไว้แล้ว ซึ่งในลักษณะดังกล่าวนี้เรียกว่า "Supervised Learning" เทคนิคการ Classification มี 2 รูปแบบ ได้แก่ Tree Induction และ Neural Induction และเป็นกระบวนการสร้างแบบจำลองเพื่อจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนด ตัวอย่างเช่น การแบ่งประเภทลูกค้าว่า เชื่อถือได้หรือไม่ซึ่งเป็นการสร้างแบบจำลองโดยการเรียนรู้จากข้อมูลที่ได้กำหนดกลุ่มไว้เรียบร้อยแล้ว

2) เทคนิค Association Rule Discovery เป็นเทคนิคที่ใช้ในการค้นหาความสัมพันธ์ของฐานข้อมูลที่มีขนาดใหญ่ เพื่อที่จะทาการวิเคราะห์ข้อมูลและหาสิ่งที่ซ่อนอยู่ในข้อมูลนั้น เช่นการวิเคราะห์ข้อมูลการซื้อขายในซูเปอร์มาร์เก็ต เพื่อทำการจัดทำการวางแผนเพื่อจัดการส่งเสริมการขาย (Promotion) และเตรียมการวางแผนการเรียงชั้นวางสินค้า (Shelf) เช่น การวางน้ำอัดลมกับข้าวโพดคั่วไว้ใกล้กัน

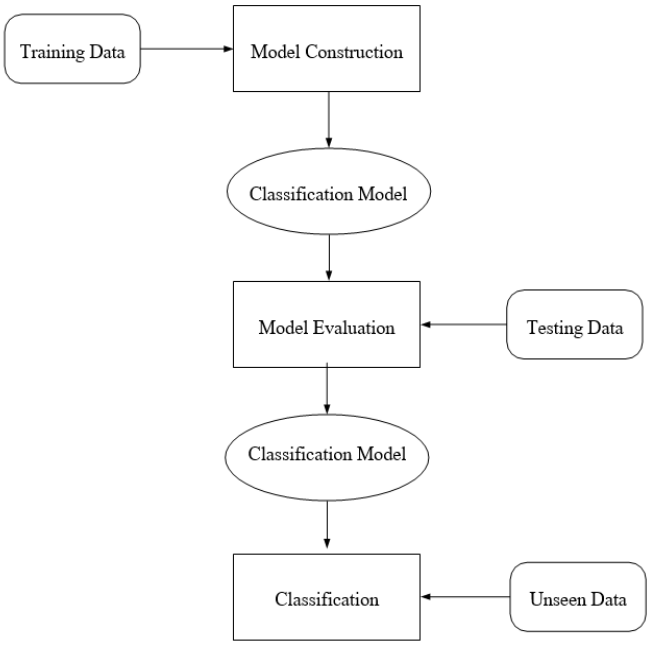
3) เทคนิค Clustering เป็นเทคนิคการลดขนาดของข้อมูลด้วยการรวมกลุ่มตัวแปรที่มีลักษณะเดียวกันไว้ด้วยกัน ทำให้สามารถค้นหาข้อมูลที่ถูกละเลยไปได้ เทคนิคนี้มักถูกใช้เป็นขั้นตอนเบื้องต้นในการทำเหมืองข้อมูล และเหมาะกับข้อมูลที่ยังไม่มีกลุ่มอย่างชัดเจนจึงทำการ Cluster เพื่อหา

กลุ่มต่างๆของข้อมูล โดยจำนวนกลุ่มของข้อมูลมักจะใช้ตัว k แทน ซึ่งผู้ที่ใช้เทคนิคนี้จะเป็นผู้กำหนดจำนวนกลุ่ม วิธีนี้อาจเรียกอีกอย่างว่า K-mean

4) เทคนิค Deviation Detection เป็นเทคนิควิธีในการหาค่าที่แตกต่างไปจากค่ามาตรฐาน หรือค่าที่คาดคิดไว้ว่าต่างไปมากน้อยเพียงใด โดยทั่วไปมักใช้วิธีการทางสถิติ หรือการแสดงให้เห็นภาพ สำหรับเทคนิคนี้ใช้ใน การตรวจสอบลายเซ็นปลอมหรือบัตรเครดิตปลอม เป็นต้น

5) เทคนิค Sequential Analysis เป็นเทคนิคในการวิเคราะห์ลำดับเพื่อค้นพบรูปแบบของการปรากฏของข้อมูล ซึ่งปรากฏในรายการที่แยกออกมา.เช่นถ้าผู้ซื้อ ซื้อสินค้า.A.แล้วเขาจะซื้อสินค้า.B.ในภายหลัง เทคนิคนี้จะแตกต่างจากเทคนิค Association Rule Discovery เพราะคำนึงถึงลำดับการซื้อด้วย

2.2 วิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูล Classification คือ กระบวนการสร้างแบบจำลองเพื่อจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนด เป็นการสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างของข้อมูลที่ได้กำหนดไว้ล่วงหน้า และสามารถพยากรณ์กลุ่มของข้อมูลที่ยังไม่เคยนำมาจัดหมวดหมู่ได้ แบบจำลองที่ได้อาจอยู่ ในรูปแบบการตัดสินใจแบบต้นไม้ตัดสินใจ (Decision Tree) หรือแบบโครงข่ายประสาทเทียม (Neural Network) ในการจัดหมวดหมู่จำเป็นต้องมีกลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) เพื่อให้ข้อมูลเรียนรู้ และสร้างแบบจำลอง (Model Construction) และทดสอบโดยกลุ่มข้อมูลสำหรับการทดสอบ (Testing Data) เพื่อประเมินความถูกต้องของโมเดล (Model Evaluation) อีกทั้งใช้ชุดข้อมูลที่ไม่เคยเห็นมาก่อน (Unseen Data) เพื่อทำการกำหนด Class ให้กับข้อมูลใหม่ที่ได้มา หรือทำนายค่าออกมาตามที่ต้องการ เช่น การจัดหมวดหมู่ของผู้ยื่นขอเครดิต (Credits) เป็นระดับต่ำ ระดับกลาง และระดับสูง ของความเสี่ยงที่จะได้รับ หรือการอนุมัติบุคคลเข้ารับทำงานในลักษณะงานต่างๆ (พยูน, 2548)



ภาพที่ 2.2 กระบวนการ Classification (พยูน, 2548)

2.3 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) คือ แบบจำลองทางคณิตศาสตร์ เพื่อการหาทางเลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning) สามารถสร้างแบบจำลองการจัดหมวดหมู่ (Clustering) ได้จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ล่วงหน้า (Training set) ได้โดยอัตโนมัติและสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วย โดยปกติมักประกอบด้วยกฎในรูปแบบ “ถ้า เงื่อนไข แล้ว ผลลัพธ์” เช่น (พยูน, 2548)

“If Income = High and Married = No THEN Risk = Poor”

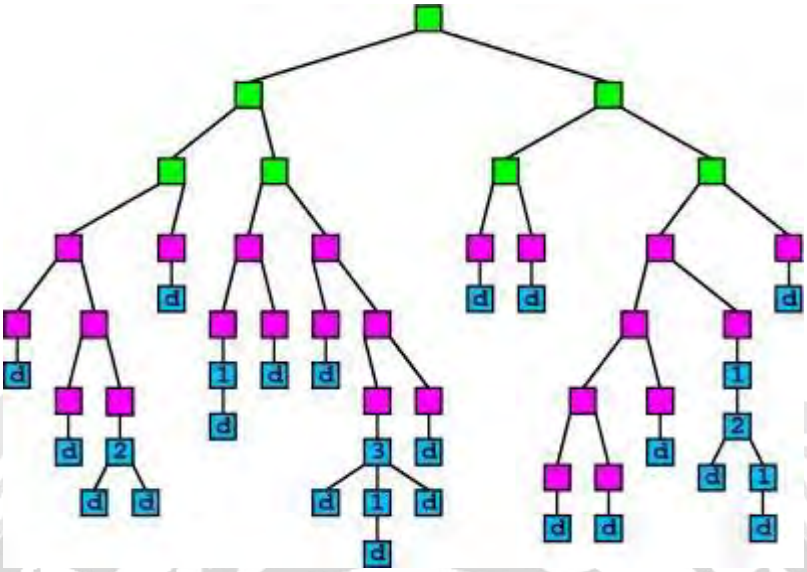
“If Income = High and Married = Yes THEN Risk = Good”

2.3.1 ส่วนประกอบของต้นไม้ตัดสินใจ ประกอบด้วย (จิตตภู, 2550)

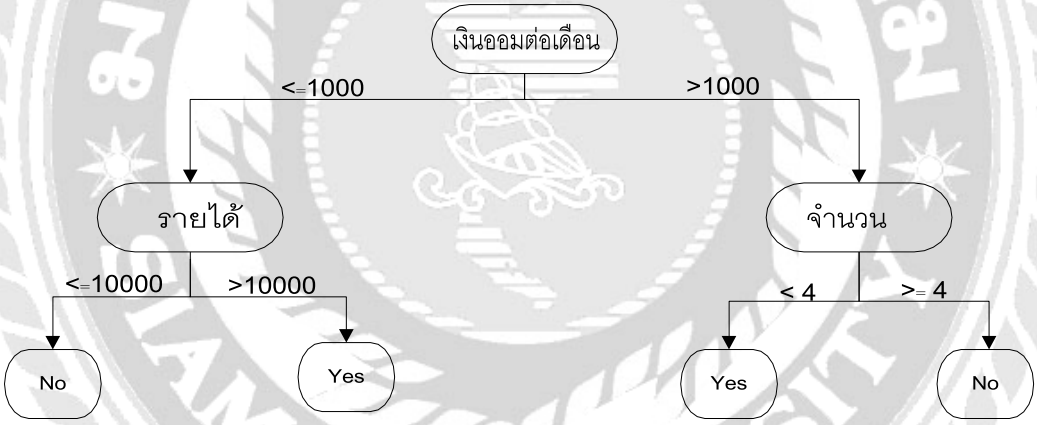
1) โหนด (Node) คือ คุณสมบัติต่างๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใด ซึ่งโหนดที่อยู่สูงสุดเรียกว่า โหนดราก (Root Node)

2) กิ่ง (Branch) คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด

3) ใบ (Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดงส่วนประกอบของต้นไม้ตัดสินใจ ดังภาพที่ 1 และแสดงตัวอย่างของต้นไม้ตัดสินใจได้ดังภาพที่ 2.3



ภาพที่ 2.3 ส่วนประกอบของต้นไม้ตัดสินใจ



ภาพที่ 2.4 ตัวอย่างต้นไม้ตัดสินใจ (ศุภชัย ประคองศิลป์ , 2551)

2.3.2 การสร้างต้นไม้ตัดสินใจ หลักการพื้นฐานของการสร้างต้นไม้ตัดสินใจ เป็นการสร้างในลักษณะจากบนลงล่าง (Top-Down) คือเริ่มจากการสร้างรากของต้นไม้ก่อนแล้วจึงแตกกิ่ง ไปจนถึงใบ โดยแสดงขั้นตอนการสร้างต้นไม้ตัดสินใจได้ดังนี้ (Han and Kamber, 2001)

1. ต้นไม้เริ่มต้นโดยมีโหนดเพียงโหนดเดียวแสดงถึงชุดข้อมูลฝึก (Training Set)
2. ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว ให้โหนดนั้นเป็นใบและตั้งชื่อแยกตามกลุ่มของข้อมูลนั้น

3. ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่ จะต้องวัดค่าเกน (Gain) ของแต่ละแอททริบิวต์เพื่อที่จะใช้เป็นเกณฑ์ในการคัดเลือกแอททริบิวต์ ที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่างๆ ได้ดีที่สุด โดยแอททริบิวต์ที่มีค่าเกนมากที่สุดจะถูกเลือกให้เป็นตัวทดสอบหรือแอททริบิวต์ ใช้ในการตัดสินใจ โดยแสดงในรูปของโหนดบนต้นไม้

4. กิ่งของต้นไม้ ถูกสร้างขึ้นจากค่าต่างๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งออกตามกิ่งต่างๆ ที่สร้างขึ้น

5. ทำการวนซ้ำเพื่อหาแอททริบิวต์ที่มีค่าเกนมากที่สุด สำหรับข้อมูลที่ถูกแบ่งแยกออกมาในแต่ละกิ่งเพื่อนำแอททริบิวต์นี้มาสร้างเป็นโหนดตัดสินใจต่อไป โดยที่แอททริบิวต์ที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีก สำหรับโหนดในระดับต่อๆ ไป

6. ทำการวนซ้ำเพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อยๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งต่อไปนี้เป็นจริง

2.3.3 การคำนวณค่า Information Gain ต้นไม้ตัดสินใจ (Decision Tree) (ขจรศักดิ์ ศรีอ่อน, 2552) เป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล โดยต้นไม้ตัดสินใจจะมีลักษณะคล้ายโครงสร้างต้นไม้ ที่แต่ละโหนดแสดงคุณลักษณะ (Attribute) ในการสร้างต้นไม้ตัดสินใจปัญหาสำคัญที่ต้องพิจารณาคือ ควรจะตัดสินใจเลือกแอททริบิวต์ใดมาทำหน้าที่เป็นโหนดราก ในแต่ละขั้นตอนของการสร้างต้นไม้และต้นไม้ย่อย (Subtree) ของต้นไม้ตัดสินใจ เกณฑ์ที่ใช้ช่วยประกอบการเลือก แอททริบิวต์คือการคำนวณค่ามาตรฐานเกน (Gain Criterion) ซึ่ง เป็นค่าที่บ่งบอกว่าแอททริบิวต์นั้น สามารถจำแนกกลุ่มของข้อมูลได้ดีเพียงใด โดยทดลองเลือก แต่ละแอททริบิวต์ที่เป็นไปได้จากชุดข้อมูลมาทำหน้าที่เป็นโหนดราก ถ้าแอททริบิวต์ใดให้ค่าเกนสูงที่สุด แสดงว่าแอททริบิวต์นั้นสามารถจำแนกกลุ่มของข้อมูลได้ดีที่สุด การใช้ค่า Information Gain จะช่วยลดจำนวนครั้งของการทดสอบ ในการแยกแยะข้อมูลอีกทั้งยังรับประกันว่าต้นไม้ตัดสินใจที่ได้ไม่มีความซับซ้อนมากจนเกินไป
ซึ่งค่า Information Gain นั้นสามารถคำนวณได้จากสมการดังสมการที่ 2-1

$$I(s_1, s_2, \dots, s_n) = -\sum_{i=1}^{n} \frac{s_1}{s} \log_2 \frac{s_1}{s} \qquad (2\text{-}1)$$

เมื่อ

S เป็นเซตของข้อมูลซึ่งประกอบด้วยข้อมูล *s* เรคคอร์ด

n เป็นจำนวนกลุ่มทั้งหมดที่ต่างกันของข้อมูลชุดนั้น

C_i แทนกลุ่มในลำดับ ที่ i โดย ที่ i มีค่าระหว่าง 1 ถึง n

S_i แทนจำนวนข้อมูลที่เป็นสมาชิกของ S และอยู่ในกลุ่ม C_i

S_{ij} แทนจำนวนข้อมูลที่เป็นสมาชิกของ S ในกลุ่ม C_i จากการแบ่งข้อมูลด้วยค่าที่เป็นไปได้ของแอททริบิวต์ A

j ค่าระหว่าง 1 ถึง v

ค่าเอ็นโทรปีของแอททริบิวต์ A ซึ่งมีค่าของแอททริบิวต์เป็น ($a_1, a_2, a_3, \ldots, a_v$) หาได้ในสมการที่ 2-2

$$E(A) = \sum_{j=1}^{v} \frac{s_{1j}+\cdots+s_{nj}}{s} I(s_{1j}, s_{2j}, \ldots, s_{nj}) \quad (2\text{-}2)$$

ดังนั้นจะสามารถพิจารณาค่ามาตรฐานเกนได้ดังสมการที่ 2-3

$$Gain(A) = I(s_{1j}, s_{2j}, \ldots, s_{nj}) - E(A) \quad (2\text{-}3)$$

2.3.4 อัลกอริธึม C4.5 (J48) (พวงทิพย์ แท่นแสง, 2550) เป็นอัลกอริทึมการสร้างกฎจากต้นไม้ตัดสินใจ (Decision Tree) มีพื้นฐานเดียวกับอัลกอริทึม ID3 ถูกออกแบบโดย Quinlan (1992) ซึ่งได้พัฒนาเพิ่มเติมจาก ID3 ดังนี้

1. สามารถหลีกเลี่ยงการสร้างโครงสร้างต้นไม้ที่ใหญ่เกินไป เนื่องจากมีข้อมูล จำนวนมาก อย่างไรก็ตามขึ้นอยู่กับการกำหนดความลึก เมื่อมีการเจริญเติบโตของ Decision Tree
2. ความผิดพลาดลดลง เพราะมีการตัดทอนความผิดพลาดออกไป (Pruning node)
3. มีการสร้างกฎหลังการตัดทอนข้อมูลที่ผิดพลาดออก
4. สามารถใช้กับข้อมูลที่มีความต่อเนื่อง (Continuous Attributes) ที่เป็นตัวเลขได้ เช่น ค่าอุณหภูมิ จำนวนเงิน เป็นต้น
5. การเลือก Attribute ที่วัดการเลือกให้เหมาะสม
6. สามารถใช้กับชุดข้อมูล (Training Data) ที่มีค่าผิดพลาด (Missing Attribute)
7. สามารถใช้สำหรับ Attribute กับ Costs ที่แตกต่างกันได้มีการยอมรับประสิทธิภาพกระบวนการคำนวณที่ปรับปรุง

การสร้างตัวแบบเป็น9 ส่วน และตัวทดสอบ 1 ส่วน แล้วสลับสับเปลี่ยนข้อมูลในการสร้างตัวแบบและทดสอบ ใช้วิธีการวัดค่าความถูกต้อง (Accuracy) จากการสร้างตัวแบบพยากรณ์ สามารถคำนวณได้จากดังสมการที่ (1)

$$Accuracy = \frac{TotalNumberofIns\tan ces}{CorrectyClassifiedInstance} x100 \quad (1)$$

โดย Correctly Classified Instance คือ จำนวนตัวอย่างที่ทำนายถูกต้อง

Total Number of Instances คือ จำนวนตัวอย่างทั้งหมด

ค่ารากที่สองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) คือ การวัดค่าความแตกต่างระหว่างค่าจริงกับค่าที่ได้จากแบบจำลอง หาก RSME มีค่าน้อย แสดงว่าแบบจำลองใกล้เคียงกับค่าจริง สามารถคำนวณได้จาก ดังสมการที่ (2)

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (2)$$

โดย Yi แทน ค่าของข้อมูลที่เกิดขึ้นจริง

Yi แทน ค่าที่ได้จากการพยากรณ์

n แทน จำนวนนำเข้าทั้งหมด n

ที่มา: Hye Rin Kim (2009: 4)

เมื่อสร้างตัวแบบเสร็จแล้วทำการวัดประสิทธิภาพความถูกต้องของตัวแบบโดยพิจารณาจากค่าความถูกต้อง (Accuracy) แสดงดังสมการที่ (1) และค่าความแม่นยำ (Precision) แสดงดังสมการที่ (2) (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)

$$\mathrm{Precision} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

ที่มา : สำนักงานสถิติแห่งชาติ, 2560

กำหนดให้

True Positive (TP) คือจำนวนที่พยากรณ์ถูกว่าเป็นคลาสซึ่งกำลังสนใจอยู่โดยใช้เป็น สมการในวัดตัวแบบ

True Negative (TN) คือ จำนวนที่พยากรณ์ถูกว่าเป็นคลาสซึ่งไม่ได้สนใจอยู่โดยใช้เป็น สมการวัดตัวแบบ

False Positive (FP) คือ จำนวนที่พยากรณ์ผิดมาเป็นคลาสซึ่งกำลังสนใจอยู่โดยใช้เป็น สมการวัดตัวแบบ

False Negative (FN) คือ จำนวนที่พยากรณ์ผิดมาเป็นคลาสซึ่งไม่ได้สนใจอยู่โดยใช้เป็น สมการวัดตัวแบบ

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (2)$$

ที่มา : สำนักงานสถิติแห่งชาติ,2560

กำหนดให้

True Positive (TP) คือ จำนวนที่พยากรณ์ถูกว่าเป็นคลาสซึ่งกำลังสนใจอยู่โดยใช้เป็น สมการในวัดตัวแบบ

False Positive (FP) คือ จำนวนที่พยากรณ์ผิดมาเป็นคลาสซึ่งกำลังสนใจอยู่โดยใช้เป็น สมการวัดตัวแบบ

3. ขั้นตอนการทำเหมืองข้อมูล

ภาพที่ 2.5 กระบวนการมาตรฐานในการทำเหมืองข้อมูล

ที่มา: https://worasupa688.wordpress.com/2018/07/06/data-mining-for-library/

“CRISP-DM” (Shearer, 2000) โดยประกอบด้วย 6 ขั้นตอนทำต่อเนื่องกัน ดังภาพที่ 2.5 คือ

1) Business Understanding เป็นขั้นตอนการเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่ในรูปแบบโจทย์ของการทำเหมืองข้อมูล

2) Data Understanding เป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้เพื่อดูความถูกต้องของข้อมูล และ พิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

3) Data Preparation เป็นข้อมูลที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา Raw Data ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้โดยต้องมีการทำข้อมูลให้ถูกต้อง (Data Cleaning) เช่น การแปลงข้อมูลให้อยู่ในช่วง (Scale) เดียวกัน หรือการเติมข้อมูลที่ขาดหายไป (Missing Value) เป็นต้น

4) Modeling เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทำเหมืองข้อมูล ในที่นี้อาศัยเทคนิคการหากฎความสัมพันธ์ และ จำแนกข้อมูล ในบางครั้งอาจจะต้องมีการย้อนกลับไปขั้นตอน Data Preparation เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิค

5) Evaluation เป็นขั้นตอนที่วัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้กำหนดไว้ในขั้นตอนแรก เพื่อจะได้เปลี่ยนแปลงแก้ไขเพื่อให้ได้ผลลัพธ์ที่ต้องการได้

6) Deployment เป็นการนำเอาองค์ความรู้ที่ได้เหล่านี้ไปใช้ได้จริงในองค์กรหรือบริษัท (กรมส่งเสริมอุตสาหกรรรม กองโลจิสติกส์ : 2019)

3. งานวิจัยที่เกี่ยวข้อง

นันทิดา จันต๊ะวงค์, วท.ม. , ปิยธิดา ตรีเดช, ส.ด. ,สุคนธา ศิริ, ปร.ด. และ ชาญวิทย์ ตรีเดช, Ph.D. (2560) ปัจจัยที่มีความสัมพันธ์กับการเข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูกในสตรีกลุ่มเป้าหมาย อายุ 30-60 ปี อำเภอเมืองจังหวัดปทมุธานี เป็นการวิจัยเชิงสำรวจแบบภาคตัดขวาง เพื่อศึกษาปัจจัยที่มีความสัมพันธ์พบว่าสตรีกลุ่มเป้าหมาย อายุ 30-60 ปี อำเภอเมือง จังหวัดปทุมธานี ส่วนใหญ่เข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูก (ร้อยละ 75.80) และ มีสตรีกลุ่มเป้าหมายที่ไม่เคยเข้ารับบริการตรวจคัดกรอง มะเร็งปากมดลูก (ร้อยละ 24.20) ซึ่งในกลุ่มนี้มีผู้ที่ไม่เคยคิดที่จะเข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูก(ร้อยละ 21.70) ความสัมพันธ์กับการเข้ารับคัดกรองมะเร็งปากมดลูกสตรีกลุ่มเป้าหมาย ได้แก่ อายุ สถานภาพสมรส ประวัติการตั้งครรภ์ การคุมกำเนิด การดูแล ตนเองเรื่องการ

ออกกำลังกาย พบว่าสตรี กลุ่มเป้าหมายที่มีอายุ 50-60 ปี เข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูกมากกว่ากลุ่มที่มีอายุ 30-39 ปี

สุภาภรณ์ พัฒนวงศ์ปราการ (2563) การวิเคราะห์เทคนิคการจำแนกประเภทข้อมูล กรณีศึกษาการทำนายระดับชั้นผู้รับเหมาก่อสร้างสำหรับโครงการก่อสร้างของภาครัฐ วัตถุประสงค์เพื่อจัดทำเหมืองข้อมูลในการนำเสนอการวิเคราะห์เทคนิคการจำแนกประเภทข้อมูล (Classification) ของการทำนายระดับชั้นผู้รับเหมาก่อสร้าง เพื่อให้มีประสิทธิภาพและมีความเหมาะสมมากขึ้น โดยมีข้อมูลโครงการก่อสร้างที่นำมาใช้จำนวน 755 โครงการ โดยการเปรียบเทียบอัลกอริทึม 4 ตัว ได้แก่ อัลกอริทึมแบบ k – Nearest Neighbor Classifier (k-NN), Nave Bayes Classifier, Decision Trees Classifier และ Random Forest Classifier และใช้การประเมินค่า (Evaluation) ในการวัดค่าความถูกต้องและความแม่นยำจากการใช้ อัลกอริทึม พบว่าลำดับค่าความสำคัญในตัวแปรหรือฟีเจอร์(Features) ที่สำคัญมากที่สุด 3 อันดับแรก ได้แก่ 1) RegistCap (ทุนจดทะเบียนผู้รับเหมาก่อสร้าง) 2) 2 (สาขางานก่อสร้างทางและสะพานพิเศษ) และ 3) BPrice (มูลค่างานก่อสร้างที่ชนะการประกวดราคา) ซึ่งเป็นไปตามหลักเกณฑ์และเงื่อนไขการขึ้นทะเบียนการจัดระดับชั้นผู้รับเหมาตามที่กรมบัญชีกลางกำหนด

พรนภา ชุมเชื้อ (2562) การใช้เทคนิคการจัดกลุ่มร่วมกับเทคนิคการจำแนกประเภทสำหรับการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้จากบริบทพฤติกรรม นำเสนอการสร้างตัวจำแนกประเภทสำหรับคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้จากบริบทพฤติกรรม โดยประยุกต์ใช้เทคนิคทางเหมืองข้อมูล ประกอบด้วย เทคนิคการจัดกลุ่ม โดยใช้อัลกอริทึมเคมีน เพื่อจัดกลุ่มผู้เรียนและสร้างคลาสคา ตอบเพื่อหาจำนวนกลุ่มที่เหมาะสม เทคนิคกฎความสัมพันธ์ โดยใช้อัลกอริทึมเอไพรออริเพื่อหาความสัมพันธ์ของคุณลักษณะที่ส่งผลต่อความบกพร่องทางการเรียนรู้ของนักเรียนการนำไปใช้ จากการทดลองจัดกลุ่มผู้เรียนสามารถจัดกลุ่มข้อมูลได้ 2 กลุ่ม คือ กลุ่มปกติ และกลุ่มเสี่ยง เพื่อพัฒนาวิธีการจัดกลุ่มผู้เรียนตามคุณลักษณะที่ส่งผลต่อความบกพร่องทางการเรียนรู้ ค้นหาความสัมพันธ์ของคุณลักษณะที่ส่งผลต่อความบกพร่องทางการเรียนรู้ของนักเรียน สร้างโมเดลจำแนกประเภทเพื่อจำแนกผู้เรียนตามกลุ่มผลลัพธ์ที่ได้จากงานวิจัย ผลลัพธ์ที่ได้จากการทดลองสร้างโมเดลจำแนกประเภทและทดสอบโมเดลด้วยวิธีเคโฟลครอสเวลิเดชันพบว่าโครงข่ายที่เหมาะสมสำหรับจำแนกประเภทผู้เรียนโดยให้ค่าความถูกต้อง 98.30% ค่าความคลาดเคลื่อน RMSE = 0.069 จากนั้นจึงพัฒนาเป็นระบบต้นแบบสำหรับคัดกรองเด็กบกพร่องทางการเรียนรู้จากบริบทพฤติกรรมโดยใช้ภาษาพีเอชพีร่วมกับเวก้าเอฟไอ จากผลการประเมินระดับความคิดเห็นของผู้ใช้งานมีค่าเฉลี่ย 4.34 ซึ่งอยู่ในเกณฑ์มาก

ประเทือง วงษ์ทอง ,วาฤทธิ์ กันแก้ว ,อภิชญา ขวัญแก้ว และ ยุวดี โฉมแดง (2563) การประยุกต์เทคนิคเหมืองข้อมูลค้นหาลักษณะนิสัย ของผู้ประกอบอาชีพด้านคอมพิวเตอร์ วัตถุประสงค์เพื่อสร้างและเปรียบเทียบประสิทธิภาพตัวแบบความสัมพันธ์ของข้อมูล จากเทคนิค Neural Networks, Naive Bayes และ Decision Tree เพื่อค้นหาลักษณะนิสัยของ ผู้ประกอบอาชีพด้านคอมพิวเตอร์ตามทฤษฎีการเลือกอาชีพของฮอลแลนด์ โดยพิจารณาจากตัวแบบ จากค่าความถูกต้อง (Accuracy) และความแม่นยำ (Precision)ดำเนินการเก็บรวบรวมข้อมูลจากผู้ประกอบอาชีพด้านคอมพิวเตอร ใช้การสุ่มตัวอย่างจาก จำนวนผู้ที่สำเร็จการศึกษาระดับปริญญาตรี จำนวนกลุ่มตัวอย่าง 395 คน ผลลัพธ์ที่ได้ที่มีประสิทธิภาพมากที่สุด คือ ตัวแบบที่ได้จากเทคนิค NaiveBayes ที่ค่าความถูกต้อง (Accuracy) เท่ากับ 79.64% (± 3.59%) และ ค่าความแม่นยำ (Precision) เท่ากับ 74.04% (86. 7 ± %) รองลงมา คือ เทคนิค Neural Network มีค่าความถูกต้อง เท่ากับ77.82% (± 5.06%) และค่าความแม่นยำ 73.60% (± 9.07%) และเทคนิค Decision Tree มีค่า ความถูกต้องเท่ากับ 54.69% (66. 1 ± %) และค่าความแม่นยำเท่ากับ 38.72% (62 . 4 ± %) ตามลำดับ

จักรกฤษณ์ หงส์เวียงจันทร์ นิติมา ลักขณานุรักษ์ และไก้รุ่ง เฮงพระพรหม (2019) การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลโรคออทิสติกด้วยเทคนิคเหมืองข้อมูล วัตถุประสงค์เพื่อศึกษาและเปรียบเทียบเทคนิคการจำแนกกลุ่มข้อมูลที่เหมาะสมกับการจำแนกข้อมูลโรคออทิสติก ในการวิจัยครั้งนี้ได้ทำการศึกษาคัดเลือกเทคนิคสำหรับใช้ในการวิจัยนี้ทั้งหมด 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม และเทคนิคเพื่อนบ้านใกล้เคียงที่สุด เพื่อหาประสิทธิภาพของการจำแนกข้อมูลที่ดีที่สุดในกลุ่มข้อมูลโรคออทิสติก ผลการวิจัยพบว่า เทคนิคที่ให้ค่าความแม่นยำมากที่สุด ซึ่งให้ประสิทธิภาพที่ดีที่สุด คือ เทคนิคโครงข่ายประสาทเทียม โดยให้ค่าความแม่นยำอยู่ที่ 99.65 %

จากการศึกษาทฤษฎี และงานวิจัยที่เกี่ยวข้องตามที่ได้กล่าวมาในข้างต้นแล้ว ทำให้ผู้วิจัยมีแนวคิดในการจำแนกข้อมูลของการใช้เทคนิคการจำแนกข้อมูลเพื่อการพยากรณ์มะเร็งปากมดลูก ในการดำเนินงานของโรงพยาบาล โดยใช้เทคนิคอัลกอริทึม Classifier เพื่อจำแนกข้อมูลการพยากรณ์มะเร็งปากมดลูก

<u>บทสรุปงานวิจัยที่เกี่ยวข้อง</u>

จากที่ได้ศึกษางานวิจัยที่เกี่ยวข้องของนักวิจัยหลายๆท่าน สามารถนำมาสรุปได้ดังนี้

ตารางที่ 2.1 สรุปงานวิจัยที่เกี่ยวข้อง

ลำดับ	ชื่อเรื่อง	วัตถุประสงค์	วิธีวิจัย	สรุปผล
1	ปัจจัยที่มีความสัมพันธ์กับการเข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูกในสตรีกลุ่มเป้าหมาย อายุ 30-60 ปี อำเภอเมืองจังหวัดปทุมธานี	1. เพื่อศึกษาการเข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูกในสตรีกลุ่มเป้าหมาย อายุ 30-60 ปี อำเภอเมืองจังหวัดปทุมธานี 2. เพื่อศึกษาปัจจัยส่วนบุคคล ปัจจัยนำ ปัจจัยเอื้อและปัจจัยเสริม ของการเข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูกในสตรีกลุ่มเป้าหมาย อายุ 30-60 ปี อำเภอเมืองจังหวัดปทุมธานี 3. เพื่อศึกษาความสัมพันธ์ระหว่างปัจจัยส่วนบุคคล ปัจจัยนำ ปัจจัยเอื้อและปัจจัยเสริม กับ	การวิจัยครั้งนี้เป็นการวิจัยเชิงสำรวจแบบภาคตัดขวาง (cross-sectional survey research) โดยงานวิจัยได้รับการพิจารณาเห็นชอบจากคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยมหิดล (เอกสารรับรองเลขที่ MUPH 2014-217)	ผลการวิจัยพบว่าสตรีกลุ่มเป้าหมาย อายุ 30-60 ปี อำเภอเมืองจังหวัดปทุมธานี ส่วนใหญ่เข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูก (ร้อยละ 75.80) และมีสตรีกลุ่มเป้าหมายที่ไม่เคยเข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูก (ร้อยละ 24.20) ซึ่งในกลุ่มนี้มีผู้ที่ไม่เคยคิดที่จะเข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูก (ร้อยละ 21.70)

ตารางที่ 2.1 (ต่อ) สรุปงานวิจัยที่เกี่ยวข้อง

ลำดับ	ชื่อเรื่อง	วัตถุประสงค์	วิธีวิจัย	สรุปผล
		การเข้ารับบริการตรวจคัดกรองมะเร็งปากมดลูกในสตรีกลุ่มเป้าหมาย อายุ 30-60 ปี อำเภอเมือง จังหวัดปทุมธานี		
2	การวิเคราะห์เทคนิคการจำแนกประเภทข้อมูลกรณีศึกษาการทำนายระดับชั้นผู้รับเหมาก่อสร้างสำหรับโครงการก่อสร้างของภาครัฐ	1. เพื่อให้ผู้รับเหมาก่อสร้างรับรู้ถึงตัวแปรที่สำคัญที่มีผลต่อการจัดระดับชั้นของผู้รับเหมาก่อสร้าง 2. เพื่อช่วยให้ได้รับการจัดระดับชั้นของผู้รับเหมาก่อสร้างให้เหมาะสมกับหลักเกณฑ์และเงื่อนไขพร้อมทั้งฐานะทางการเงินของผู้รับเหมาก่อสร้างตามที่กรมบัญชีกลางกำหนดได้ 3. เพื่อให้ผู้รับเหมา	งานวิจัยเชิงปริมาณเพื่อจำแนกประเภทข้อมูลทั้ง 2 ประเภท ได้แก่ 1) การจัดระดับชั้นของผู้รับเหมาก่อสร้าง ให้เหมาะสมกับโครงการก่อสร้างที่ผู้รับเหมาเข้าร่วมการประกวดราคาจากเว็บไซต์กรมบัญชีกลาง และ 2) ช่วงของรายได้ผู้รับเหมาก่อสร้างซึ่งสามารถวิเคราะห์ถึงความ	การวิจัยเรื่องการทำนายการจัดระดับชั้นผู้รับเหมาก่อสร้าง พบว่าลำดับค่าความสำคัญในตัวแปรหรือฟีเจอร์ (Features) ที่สำคัญมากที่สุด 3 อันดับแรก ได้แก่ 1) Regist Cap หรือทุนจดทะเบียนผู้รับเหมาก่อสร้าง 2) สาขางานก่อสร้างทางและสะพานพิเศษ และ 3) BPrice หรือมูลค่างานก่อสร้างที่

ตารางที่ 2.1 (ต่อ) สรุปงานวิจัยที่เกี่ยวข้อง

ลำดับ	ชื่อเรื่อง	วัตถุประสงค์	วิธีวิจัย	สรุปผล
		ก่อสร้างรับรู้ถึงตัวแปรที่สำคัญที่มีผลต่อการทำนายในช่วงของรายได้ผู้รับเหมาก่อสร้าง	เสี่ยงที่ผู้ รับเหมาก่อสร้างมีโอกาสที่จะทิ้งงานก่อนกำหนด หรือ มีโอกาสที่จะส่งงานไม่ทันกำหนดได้	ชนะการประกวดราคาซึ่งเป็นไปตามหลักเกณฑ์และเงื่อนไขการขึ้นทะเบียนทุกสาขางานก่อสร้าง ทุกการจัด
3	การใช้เทคนิคการจัดกลุ่มร่วมกับเทคนิคการจำแนกประเภทสำหรับการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้จากบริบทพฤติกรรม	1. เพื่อพัฒนาวิธีการจัดกลุ่มผู้เรียนตามคุณลักษณะที่ส่งผลต่อความบกพร่องทางการเรียนรู้ 2. เพื่อค้นหาความสัมพันธ์ของคุณลักษณะที่ส่งผลต่อความบกพร่องทางการเรียนรู้ของนักเรียน 3. เพื่อสร้างโมเดลจำแนกประเภทเพื่อจำแนกผู้เรียนตามกลุ่มผลลัพธ์ที่ได้จากงานวิจัย	วิธีการดำเนินการวิจัยที่มีการกำหนดขั้นตอนการดำเนินงานวิจัยตามกรอบแนวคิดงานวิจัยโดยยึดหลักกระบวนการมาตรฐาน CRISP-DM ร่วมกับทฤษฎีการทำเหมืองข้อมูลเพื่อสร้างโมเดลสำหรับคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ดังนี้	วิธีการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ในขั้นตอนแรกใช้วิธีการคัดเลือกคุณลักษณะข้อมูลโดยใช้อินฟอร์เมชันเกนเพื่อให้ได้คุณลักษณะที่สำคัญมาใช้งาน สามารถคัดเลือกคุณลักษณะได้ทั้งหมด 22 คุณลักษณะ ซึ่งสามารถจัดกลุ่มข้อมูลได้จำนวน 2 กลุ่ม ซึ่งพบว่าโครงข่ายที่

ตารางที่ 2.1 (ต่อ) สรุปงานวิจัยที่เกี่ยวข้อง

ลำดับ	ชื่อเรื่อง	วัตถุประสงค์	วิธีวิจัย	สรุปผล
			1. กรอบแนวคิดและกระบวนการทำงาน 2. เตรียมข้อมูลและวิเคราะห์ข้อมูล 3. ศึกษาและทดสอบข้อมูลด้วยเทคนิคทางเหมืองข้อมูล 4. สร้างโมเดลจำแนกประเภท 5. การประเมินผลโมเดล 6. การนำไปใช้	เหมาะสมสำหรับจำแนกประเภทผู้เรียน คือโครงข่ายที่มีโครงสร้าง [22-13-6] โดยให้ค่าความถูกต้อง 98.30% ค่าความคลาดเคลื่อน RMSE = 0.069
4	การประยุกต์เทคนิคเหมืองข้อมูลค้นหาลักษณะนิสัยของผู้ประกอบอาชีพด้านคอมพิวเตอร์		1. การจัดเตรียมข้อมูล แบ่งออกเป็น 2 ขั้นตอน โดยแต่ละขั้นตอนมีวิธีการดังนี้ 1) การศึกษาทฤษฎีและพัฒนา	ผลการดำเนินการสร้างตัวแบบและวัดค่าประสิทธิภาพตัวแบบ พบว่าตัวแบบที่มีประสิทธิภาพ มากที่สุด คือ Naive Bayes ที่ค่าความถูกต้อง (Accuracy) เท่ากับ

ตารางที่ 2.1 (ต่อ) สรุปงานวิจัยที่เกี่ยวข้อง

ลำดับ	ชื่อเรื่อง	วัตถุประสงค์	วิธีวิจัย	สรุปผล
		1. สร้างและวัดประสิทธิภาพของตัวแบบความสัมพันธ์ข้อมูลของตัวแบบ Decision Tree, Naïve bay's และ Neural Network 2. เปรียบเทียบประสิทธิภาพของตัวแบบ Decision Tree, Naïve bay's	แบบสอบถามผู้ศึกษาได้ดำเนินการศึกษาทฤษฎีการเลือกอาชีพของฮอลแลนด์ การแบ่งหมวดอาชีพด้านคอมพิวเตอร์ และการพัฒนาแบบสอบถาม 2) การจำแนกหมวดอาชีพด้านคอมพิวเตอร์ และสัดส่วนของผู้ประกอบอาชีพแต่ละด้านและจำแนกเป็น 6 หมวดอาชีพ 2. การวิเคราะห์ข้อมูล ได้แก่ พัฒนาตัวแบบโดยใช้โปรแกรม RapidMiner	79.64% (± 3.59 %) และค่าความแม่นยำ (Precision) เท่ากับ 74.04% (± 7.86%) รองลงมา คือ เทคนิค Neural Network มีค่าความถูกต้องเท่ากับ 77.82% (± 5.06%) และค่าความแม่นยำ 73.60% (± 9.07%) และ เทคนิค Decision Tree มีค่าความถูกต้องเท่ากับ 54.69% (± 1.66%) และค่าความแม่นยำเท่ากับ 38.72% (± 4.62%) ตามลำดับ Naive Bayes และ Neural Networks เมื่อสร้างตัวแบบเสร็จแล้วทำการวัดประสิทธิภาพความถูกต้อง

ตารางที่ 2.1 (ต่อ) สรุปงานวิจัยที่เกี่ยวข้อง

ลำดับ	ชื่อเรื่อง	วัตถุประสงค์	วิธีวิจัย	สรุปผล
			Studio ทำการแบ่งข้อมูลโดยแบ่งเป็นชุดฝึกสอน และชุดทดสอบใช้เทคนิคการแบ่งข้อมูลด้วยวิธี 5 Fold-Cross-Validation เพื่อแบ่งข้อมูลออกเป็นส่วนและทำการวนรอบทดสอบทีละส่วนแล้วใช้เทคนิคในการสร้างตัวแบบได้แก่ Decision Tree เป็นต้น	
5	การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลโรคออทิสติกด้วยเทคนิคเหมืองข้อมูล	เพื่อเปรียบเทียบประสิทธิภาพเทคนิคการจำแนกข้อมูลสำหรับโรคออทิสติกทั้งหมด 3 วิธีได้แก่	1. ข้อมูลสำหรับใช้ในการทดลองผู้วิจัยได้ทำการรวบรวมข้อมูลเด็กที่เป็นโรคออทิสติกและไม่เป็นโรคออทิสติก โดย	งานวิจัยนี้ได้ศึกษาประสิทธิภาพของวิธีการจำแนกกลุ่มข้อมูลโรคออทิสติกด้วยเทคนิคการจำแนกกลุ่ม

ตารางที่ 2.1 (ต่อ) สรุปงานวิจัยที่เกี่ยวข้อง

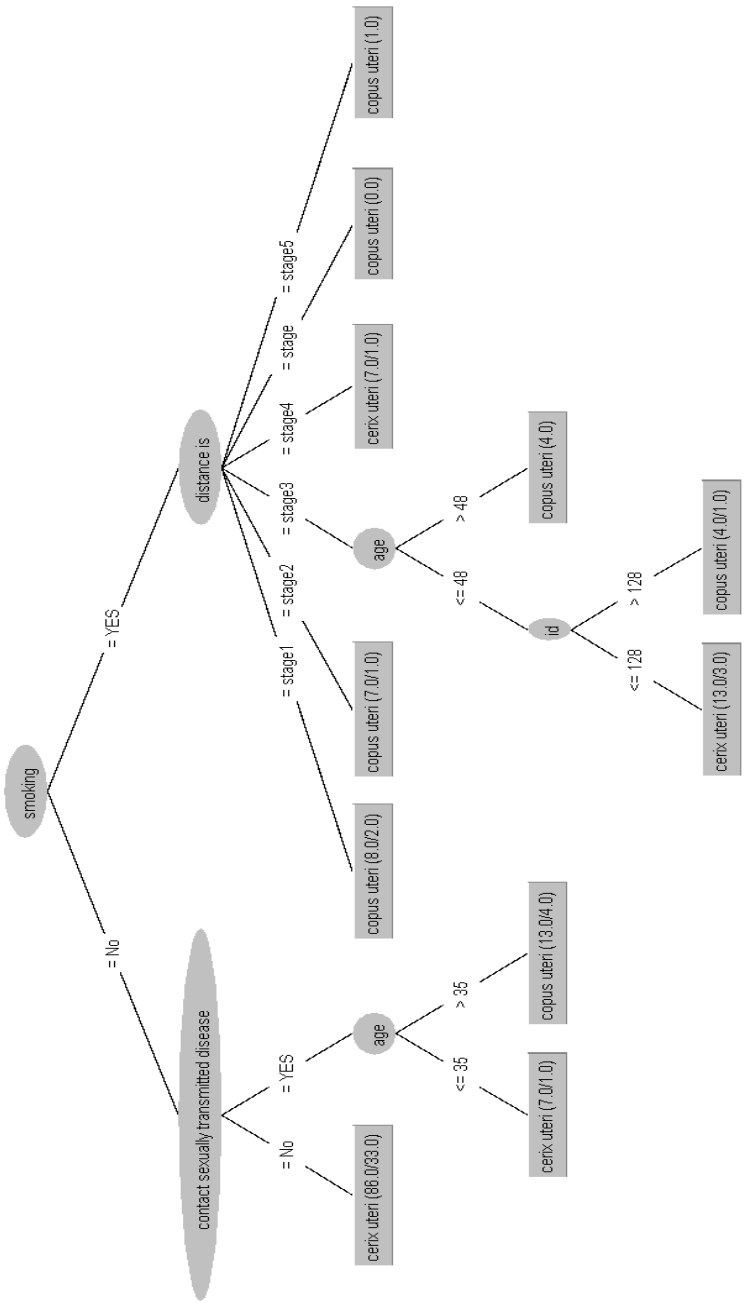
ลำดับ	ชื่อเรื่อง	วัตถุประสงค์	วิธีวิจัย	สรุปผล
		ต้นไม้ตัดสินใจ เครือข่ายประสาทเทียม และ เทคนิคเพื่อนบ้านใกล้เคียงที่สุด	นำข้อมูลดังกล่าวมาจาก เว็บไซต์ Autism Screening Adult Data Set จากนั้นนำข้อมูลมาทำความสะอาด โดยตัดข้อมูลที่มีค่าขาดหาย (Missing Values) ออก 2. การดำเนินการทดลอง โดย 1) นำข้อมูลที่ได้จัดเตรียมไว้ มาเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลโรคออทิสติกด้วยเทคนิคการจำแนกข้อมูล จำนวน 3 เทคนิค	ข้อมูลจำนวน 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม เทคนิคเพื่อนบ้าน ผลการวิจัยพบว่าเทคนิคทั้ง 3 เทคนิค เทคนิคที่ให้ค่าความแม่นยำมากที่สุด จะมีประสิทธิภาพดีที่สุด ดังนั้นนเทคนิคที่ให้ค่าความแม่นยำสูงที่สุด คือ เทคนิคโครงข่ายประสาทเทียม ให้ค่าความแม่นยำอยู่ที่ 99.65 %

ตารางที่ 2.1 (ต่อ) สรุปงานวิจัยที่เกี่ยวข้อง

ลำดับ	ชื่อเรื่อง	วัตถุประสงค์	วิธีวิจัย	สรุปผล
			ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม เทคนิคเพื่อนบ้านใกล้เคียงที่สุด 2) นำผลลัพธ์จากการจำแนกกลุ่มที่ได้จากการทดลองมาหาประสิทธิภาพของการจำแนกกลุ่มโรคออทิสติก	

บทที่ 3

วิธีการดำเนินการวิจัย



ภาพที่3.1 ต้นไม้ตัดสินใจโดยใช้เทคนิคการจำแนกข้อมูลมะเร็งมดลูกและมะเร็งปากมดลูก

รูปแบบงานวิจัย

งานวิจัยนี้เป็นการจำแนกข้อมูลเพื่อวิเคราะห์หาปัจจัยเสี่ยงที่มีผลทำให้เกิดมะเร็งปากมดลูก กรณีศึกษาจากศูนย์บริการสาธารณสุข 62 ในการวิเคราะห์หาปัจจัยผู้วิจัยได้ใช้วิธีการจำแนกข้อมูลแบบแยกข้อมูลโดยใช้อัลกอริทึม Classifier มาช่วยในการจำแนกข้อมูลทำให้สามารถวิเคราะห์และแยะข้อมูลเพื่อให้รู้ถึงปัจจัยเสี่ยงต่อการเกิดโรคมะเร็งปากมดลูก

ชุดข้อมูล

ทางผู้วิจัยได้ทำการสืบค้นหาข้อมูลและเก็บรวบรวมข้อมูลโดยการสืบค้นทางอินเทอร์เน็ตของโรงพยาบาลจุฬาลงกรณ์ สาขารังสีรักษาและมะเร็งวิทยา ฝ่ายรังสีวิทยา ตั้งแต่ เดือน พฤศจิกายน 2565 ถึง เดือน มกราคม 2566 เป็นเป้าหมายในการวิจัยครั้งนี้

เครื่องมือในการวิจัย

ทางผู้วิจัยได้จัดเก็บชุดข้อมูลโดยใช้ Microsoft Excel และจัดเก็บข้อมูลเป็นไฟล์ csv เพื่อที่สามารถนำข้อมูลไปใช้หากฎความสัมพันธ์ โดยใช้โปรแกรม Weka 3.8.4

วิธีดำเนินการวิจัย

ในการวิจัยการจำแนกข้อมูลเพื่อวิเคราะห์หาปัจจัยเสี่ยงของมะเร็งปากมดลูกโดยใช้เทคนิคอัลกอริทึม Classifier มีขั้นตอนการดำเนินการวิจัยด้วยกระบวนการ CRIS-DM มีดังนี้

1 ทำความเข้าใจกับปัญหา (Business Understanding)

2 ทำความเข้าใจข้อกับข้อมูล (Data Understanding)

3 การเตรียมข้อมูล (Data Preparation)

4 การสร้างแบบจำลองข้อมูล (Modelling)

5 การประเมินผล (Evaluation)

6 การนำไปใช้ (Deployment)

ขั้นตอนที่ 1 ทำความเข้าใจธุรกิจ (Business Understanding)

ลักษณะของปัญหาคือกลุ่มสตรีทั่วไปได้มีการละเลยในการดูแลสุขภาพการใช้ชีวิตในประจำวัน การกินความเป็นอยู่ของผู้คนทำให้มีปัจจัยเสี่ยงทำให้เกิดมะเร็งปากมดลูกจึงต้องการศึกษาเกี่ยวกับพฤติกรรมต่างๆของกลุ่มสตรี เพื่อทำให้ทราบพฤติกรรมต่างๆในการเกิดมะเร็งปากมดลูก ทางผู้วิจัยจึงทำการรวบรวมข้อมูลพฤติกรรมที่มีผลต่อการก่อการเกิดมะเร็งปากมดลูก ซึ่งจะใช้ข้อมูลเพียงแค่กลุ่มตัวอย่างเท่านั้น จึงจะต้องหาจำนวนกลุ่มตัวอย่างที่จะเก็บข้อมูล โดยการหากลุ่มตัวอย่างจะใช้เป็นเว็บไซต์คำนวณกลุ่มตัวอย่างช่วยในการคำนวณ จากกลุ่มสตรีมะเร็งวิทยา ฝ่ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์ 253 คน ใช้วิธีคำนวณกลุ่มตัวอย่างแบบ (Krejcie & Morgan) จะได้กลุ่มตัวอย่าง 152.77 ประมาณเป็น 153 คน

ผลผลการคำนวณขนาดกลุ่มตัวอย่าง (Krejcie&Morgan)

(N) จำนวนประชากรที่ศึกษา	253
(e) ความคลาดเคลื่อนที่ยอมรับได้ไม่เกิน (%)	0.05
(X2) ระดับความเชื่อมั่น (%) (ค่าไคสแควร์)	3.841
(P) สัดส่วนของลักษณะที่สนใจในประชากร (ถ้าไม่ทราบให้กำหนด 0.5)	0.5
(s) ขนาดของกลุ่มตัวอย่าง	152.77

ภาพที่ 3.2 ผลการคำนวณขนาดกลุ่มตัวอย่าง Krejcie & Morgan

ขั้นตอนที่ 2 ทำความเข้าใจข้อมูล (Data Understanding)

ข้อมูลที่นำมาใช้วิเคราะห์เป็นการรวบรวมข้อมูลจากโรงพยาบาลจุฬาลงกรณ์ เพื่อให้ทราบสาเหตุการเกิดมะเร็งปากมดลูกโดยการค้นหาและรวบรวมข้อมูลผ่านทางเว็บไซต์ของทาง สาขารังสีรักษาและมะเร็งวิทยา ฝ่ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์ ตั้งแต่ใน เดือนพฤศจิกายน 2565 ถึง เดือนมกราคม 2566 โดยการรวบรวมปัจจัยความเสี่ยงในแต่ละบุคคลเพื่อนำมาอธิบายและวิเคราะห์

ขั้นตอนที่ 3 การเตรียมข้อมูล (Data Preparation)

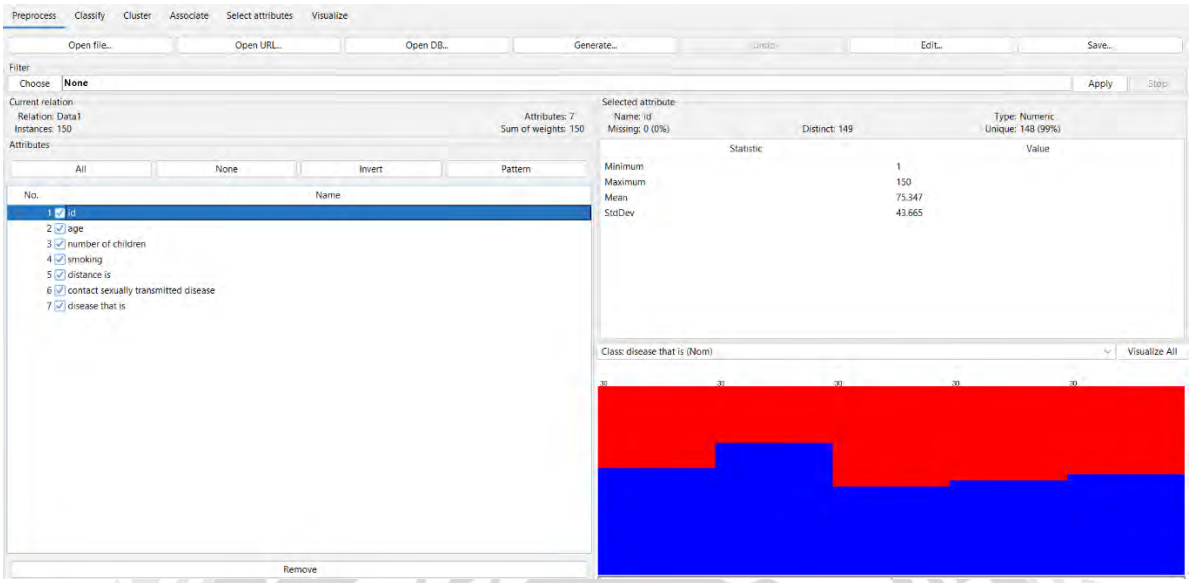
เมื่อรวบรวมข้อมูลของกลุ่มสตรีเสร็จเรียบร้อยแล้ว ต้องเตรียมข้อมูล คัดกรองข้อมูล ซึ่งข้อมูลของกลุ่มสตรีนั้นเป็นข้อมูลของมะเร็งวิทยา ฝ่ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์ ดังนั้นข้อมูลจึงมีรายละเอียดของข้อมูลที่มีจำนวนมาก จึงเลือกนำข้อมูลบางส่วนมาทำการจำแนกการทำให้เกิดโรคมะเร็ง และทำความสะอาดข้อมูลเพื่อให้ข้อมูลมีความถูกต้อง ลดความซ้ำซ้อนของข้อมูล

id	age	number of	smoking	distance is	contact se	disease that is
1	36	5	No	stage1	No	cerix uteri
2	33	3	No	stage2	YES	cerix uteri
3	27	3	No	stage3	No	copus uteri
4	29	4	No	stage1	No	cerix uteri
5	31	4	No	stage2	No	copus uteri
6	41	5	YES	stage3	YES	cerix uteri
7	53	6	No	stage4	YES	copus uteri

ภาพที่ 3.3 ตัวอย่างตารางข้อมูลมะเร็งปากมดลูก

ขั้นตอนที่ 4 การพัฒนาแบบจำลอง (**Model**)

การนำข้อมูลที่ได้จากการแปลงและทำความสะอาดข้อมูลเพื่อให้เหมาะสมแก่การใช้งานมาเข้าสู่กระบวนการจำแนกข้อมูล โดยใช้อัลกอริทึม Classifier มาใช้ในการพยากรณ์และจำแนกข้อมูลเป็นการใช้ในรูปแบบฟังก์ชัน Lazy เป็นการเก็บตัวอย่าง การตัดสินใจโดยใช้เทคนิคอัลกอริทึม NaiveBayes SMO IBK J48 มาใช้จำแนกโดย มีค่าความถูกต้อง (Correctly Classified Instances) และ ค่าความผิดพลาด (Incorrectly Classified Instances) การจำแนกจะอยู่ในรูปแบบของตัวเลขและต้นไม้ตัดสินใจ เช่น ถ้าบุคคลสูบบุหรี่ จะอยู่ในระยะที่3ของการเป็นมะเร็ง นี้มีอายุ 35 ปี จะเป็นมะเร็งปากมดลูก



ภาพที่ 3.4 การเลือกข้อมูลในพยาการและการจำแนกข้อมูลอัลกอริทึม

ขั้นตอนที่ 5 การประเมินแบบจำลอง (Evaluation)

การจำแนกค่าความถูกต้องสามารถทำได้โดยใช้ มีค่าความถูกต้อง (Correctly Classified Instances) คือค่าที่มีประสิทธิภาพของการจำแนกข้อมูลในการพยากรณ์ และ ค่าความผิดพลาด (Incorrectly Classified Instances) คือค่าที่ไม่มีประสิทธิภาพของการจำแนกข้อมูล

ขั้นตอนที่ 6 การนำไปใช้ (Deployment)

การนำผลการวิจัยไปประยุกต์ใช้จะต้องนำผลลัพธ์การจำแนกที่ได้ไปทำการส่งเสริมทางการแพทย์ในการวางแผนในการตรวจสุขภาพประจำปี สามารถแนะนำการตรวจสุขภาพที่เป็นโปรแกรมการตรวจภายในเพื่อหาเชื้อการเกิดมะเร็งก่อนที่จะมาเจอในระยะที่รักษาไม่หายแล้ว

ในด้านของธุรกิจยังสามารถประยุกต์นำไปใช้ในการทำธุรกิจทางการแพทย์ต่างๆที่มีความเกี่ยวข้องกับการพยากรณ์การจำแนกข้อมูลที่วิเคราะห์อีกด้วย

บทที่ 4

ผลการวิจัย

จากการทดลองทำให้ได้ผลลัพธ์ที่ออกมาในรูปแบบของข้อมูลที่แบ่ง Class และค่าเปอร์เซ็นต์ความถูกต้องและความผิดพลาด โดยใช้อัลกอริทึม Classifier เพื่อนำมาจำแนกข้อมูล กรณีศึกษาสาขารังสีรักษาและมะเร็งวิทยา ฝ่ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์ จำนวน 150 คน

ตารางที่ 4.1 สรุปผลการทดลอง

อัลกอริทึม	ค่าความถูกต้อง (%)	ค่าความผิดพลาด (%)	Precision (%)
J48	57.33%	43.33%	±0.60%
Naive Bayes	57.33%	43.33%	±0.59%
SMO	51.33%	48%	±0.55%
IBK	41.33%	49.33%	±0.55%

```
=== Summary ===

Correctly Classified Instances          85               56.6667 %
Incorrectly Classified Instances        65               43.3333 %
Kappa statistic                          0.0961
Mean absolute error                      0.4922
Root mean squared error                  0.5045
Relative absolute error                 99.5145 %
Root relative squared error            101.4385 %
Total Number of Instances              150

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.735    0.642    0.587      0.735    0.652      0.100    0.489     0.534     cerix uteri
                 0.358    0.265    0.522      0.358    0.425      0.100    0.489     0.465     copus uteri
Weighted Avg.    0.567    0.474    0.558      0.567    0.551      0.100    0.489     0.503

=== Confusion Matrix ===

  a  b   <-- classified as
 61 22 |  a = cerix uteri
 43 24 |  b = copus uteri
```

ภาพที่ 4.1 แสดงผลลัพธ์การจำแนกโดยใช้อัลกอริทึม Naive Bayes

```
=== Summary ===

Correctly Classified Instances          78               52      %
Incorrectly Classified Instances        72               48      %
Kappa statistic                         -0.021
Mean absolute error                      0.48
Root mean squared error                  0.6928
Relative absolute error                 97.0484 %
Root relative squared error            139.2982 %
Total Number of Instances              150

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.771    0.791    0.547      0.771    0.640      -0.024   0.490     0.548     cerix uteri
                 0.209    0.229    0.424      0.209    0.280      -0.024   0.490     0.442     copus uteri
Weighted Avg.    0.520    0.540    0.492      0.520    0.479      -0.024   0.490     0.501

=== Confusion Matrix ===

  a  b   <-- classified as
 64 19 |  a = cerix uteri
 53 14 |  b = copus uteri
```

ภาพที่ 4.2 แสดงผลลัพธ์การจำแนกโดยใช้อัลกอริทึม SMO

ผลจากการทดลองโดยใช้อัลกอริทึม SMO ผลที่ได้คือ ค่าความถูกต้องอยู่ที่ 78 หรือ 52 % และค่าความผิดพลาดอยู่ที่ 72 หรือ 48 %

```
=== Summary ===

Correctly Classified Instances          76               50.6667 %
Incorrectly Classified Instances        74               49.3333 %
Kappa statistic                         -0.0009
Mean absolute error                      0.4934
Root mean squared error                  0.6973
Relative absolute error                 99.7639 %
Root relative squared error            140.1928 %
Total Number of Instances              150

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.566    0.567    0.553      0.566    0.560      -0.001   0.500     0.553     cerix uteri
                 0.433    0.434    0.446      0.433    0.439      -0.001   0.500     0.446     copus uteri
Weighted Avg.    0.507    0.508    0.505      0.507    0.506      -0.001   0.500     0.505

=== Confusion Matrix ===

  a  b   <-- classified as
 47 36 |  a = cerix uteri
 38 29 |  b = copus uteri
```

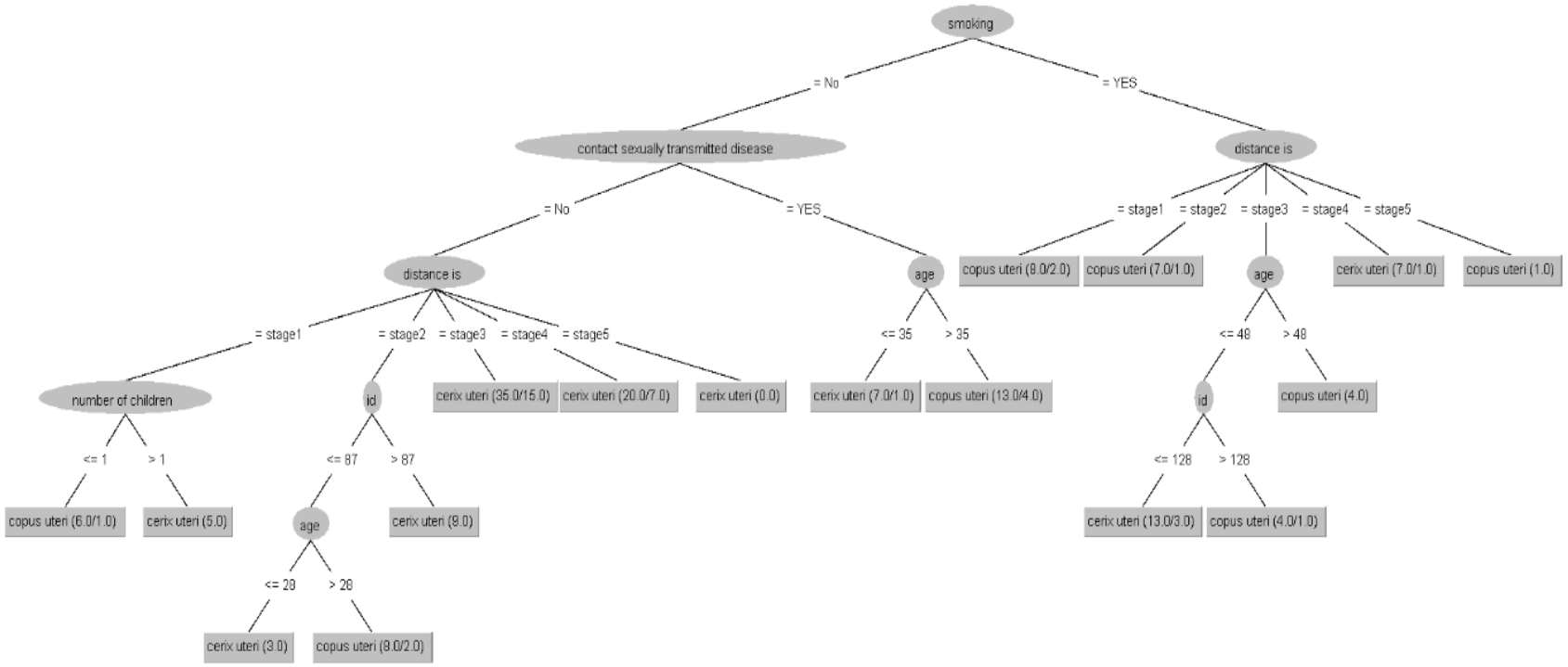
ภาพที่ 4.3 แสดงผลลัพธ์การจำแนกโดยใช้อัลกอริทึม IBk

ผลจากการทดลองโดยใช้อัลกอริทึม IBk ผลที่ได้คือ ค่าความถูกต้องอยู่ที่ 76 หรือ 50.67 % และค่าความผิดพลาดอยู่ที่ 74 หรือ 49.33 %

```
=== Summary ===

Correctly Classified Instances          85               56.6667 %
Incorrectly Classified Instances        65               43.3333 %
Kappa statistic                          0.1093
Mean absolute error                      0.4524
Root mean squared error                  0.5525
Relative absolute error                 91.464  %
Root relative squared error            111.0819 %
Total Number of Instances              150

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.675    0.567    0.596      0.675    0.633      0.111    0.577     0.631     cerix uteri
                 0.433    0.325    0.518      0.433    0.472      0.111    0.577     0.496     copus uteri
Weighted Avg.    0.567    0.459    0.561      0.567    0.561      0.111    0.577     0.571

=== Confusion Matrix ===

  a  b   <-- classified as
 56 27 |  a = cerix uteri
 38 29 |  b = copus uteri
```

ภาพที่ 4.4 แสดงผลลัพธ์การจำแนกโดยใช้อัลกอริทึม J48

ผลจากการทดลองโดยใช้อัลกอริทึม J48 ผลที่ได้คือ ค่าความถูกต้องอยู่ที่ 85 หรือ 56.67 % และค่าความผิดพลาดอยู่ที่ 65 หรือ 43.33 %



ภาพที่4.5 การแสดงผลลัพธ์ต้นไม้ตัดสินใจการใช้เทคนิคการจำแนกข้อมูลเพื่อการพยากรณ์มะเร็งปากมดลูก ในการดำเนินงานธุรกิจของโรงพยาบาล

จากแผนผังต้นไม้ตัดสินใจที่ได้ตามภาพดังกล่าวนั้น สามารถนำมาแปลงเป็นกฎพื้นฐานได้ทั้งหมด 25 กฎพื้นฐาน ซึ่งเป็นเงื่อนไขในการพยากรณ์ความเสี่ยงการเกิดโรคมะเร็งปากมดลูกโดยนำกฎพื้นฐานเหล่านี้ไปใช้ในต้นไม้ตัดสินใจ มีรายละเอียดดังแสดงในตารางที่ 4.2

ตารางที่ 4.2 รายละเอียดกฎพื้นฐานที่แปลงจากแผนผังต้นไม้ตัดสินใจ

ลำดับ	กฎพื้นฐาน	จำนวนข้อมูลตัวอย่าง
1	If smoking = yes and period = stage 1 then copus uteri	8/2
2	If smoking = yes and period = stage 2 then copus uteri	7/1
3	If smoking = yes and distance = stage 3 and age <=48 and id <=128 then cerix uteri	13/3
4	If smoking = yes and period = stage 3 and age <=48 and id >128 then cerix uteri	4/1
5	If smoking = yes and period = stage 3 and age >48 and then copus uteri	4/0
6	If smoking = yes and period = stage 4 then cerix uteri	7/1
7	If smoking = yes and period = stage 5 then copus uteri	1/0

บทที่ 5
สรุปผลการดำเนินงานและข้อเสนอแนะ

สรุปผลการดำเนินงาน

จากการวิจัยผลการจำแนกข้อมูล โดยการทำเหมืองข้อมูลการจำแนกประเภทข้อมูลโดยการใช้ Classifier ที่สามารถแบ่งแยกข้อมูลที่มีลักษณะหรือเป้าหมายตามที่กำหนดได้ ตัวแบบที่ใช้ในการวิจัย Bayers เป็นการ ใช้หลักความน่าจะเป็น Functions เป็นการใช้ในรูปแบบฟังก์ชัน Lazy เป็นการเก็บตัวอย่าง การตัดสินใจเกิดใหม่เมื่อตัวอย่างใหม่ถูกนำเข้ามาเท่านั้น Trees เป็นการสร้างตัวแบบโดยการใช้ต้นไม้ตัดสินใจ และ ผลลัพธ์ที่ได้จากการจำแนกข้อมูล โดยใช้เทคนิคอัลกอริทึม NaiveBayes ผลลัพธ์ค่าความแม่นยำเท่ากับ 85 หรือ คิดเป็นเปอร์เซ็นต์ คือ 56.67 เปอร์เซ็นต์ ค่าความผิดพลาดเท่ากับ 65 หรือ คิดเป็นเปอร์เซ็นต์ คือ 43.33 เปอร์เซ็นต์ การจำแนกข้อมูล โดยใช้เทคนิคอัลกอริทึม SMO ผลลัพธ์ค่าความแม่นยำเท่ากับ 77 หรือ คิดเป็นเปอร์เซ็นต์ คือ 51.33 เปอร์เซ็นต์ ค่าความผิดพลาดเท่ากับ 73 หรือ คิดเป็นเปอร์เซ็นต์ คือ 48.67 เปอร์เซ็นต์ การจำแนกข้อมูล โดยใช้เทคนิคอัลกอริทึม IBk ผลลัพธ์ค่าความแม่นยำเท่ากับ 71 หรือ คิดเป็นเปอร์เซ็นต์ คือ 47.33 เปอร์เซ็นต์ ค่าความผิดพลาดเท่ากับ 79 หรือ คิดเป็นเปอร์เซ็นต์ คือ 52.67 เปอร์เซ็นต์ การจำแนกข้อมูล โดยใช้เทคนิคอัลกอริทึม J48 ผลลัพธ์ค่าความแม่นยำเท่ากับ 86 หรือ คิดเป็นเปอร์เซ็นต์ คือ 57.33 เปอร์เซ็นต์ ค่าความผิดพลาดเท่ากับ 64 หรือ คิดเป็นเปอร์เซ็นต์ คือ 42.66 เปอร์เซ็นต์ จึงอธิบายได้ว่าการเกิดโรคมะเร็งดังกล่าวที่เกิดขึ้นจากการสำรวจและวิจัยผลที่ทำให้เป็นโรคมะเร็งปากมดลูกเกิดจากพฤติกรรมการใช้ชีวิตขอตัวผู้ป่วย

ปัญหาและอุปสรรค

1. เนื่องจากสภาพของโรคระบาดโควิด 19 เกิดขึ้น ณ ช่วยเวลาที่เก็บข้อมูล ทำให้มีอุปสรรคในการสอบถามและเก็บข้อมูล ทำให้ระยะเวลาและข้อมูลที่ได้จากการเก็บและสอบถามน้อยกว่าปกติ

2. เนื่องจากหัวข้อวิจัยที่ทำขึ้นทำให้มีปัญหาในการสอบถามข้อมูลกับผู้ที่ป่วย เพราะเป็นข้อมูลส่วนตัวอาจจะยากในการยินยอมที่จะให้ข้อมูลและอาจจะมีการปฏิเสธในการให้ข้อมูล

ข้อเสนอแนะ

จากผลการวิจัยพบว่าการจำแนกข้อมูลพฤติกรรมการทำให้เกิดโรคมะเร็งปากมดลูก มีระยะเวลาในการเก็บข้อมูลที่สั้น เพราะมีระยะเวลาในการเก็บข้อมูลเพียง 1 เดือนของการเก็บข้อมูลจึงทำให้มีความถูกต้องของข้อมูลที่น้อยกว่าการเก็บข้อมูลที่มีระยะเวลานานกว่า

บรรณานุกรม

กรมส่งเสริมอุตสาหกรรม กองโลจิสติกส์. (2562). *การทำเหมืองข้อมูล (Data Mining)*. https://dol.dip.go.th/th/category/2019-02-08-08-57-30/2019-03-15-08-49-57

กระทรวงสาธารณสุข. (2551). *สธ.ชี้หญิงไทยวัย 35 ปีขึ้นไป กว่า 8 ล้านคน เสี่ยงมะเร็งปากมดลูก เพราะอายหมอ ไม่กล้าไปตรวจมดลูก*. https://pr.moph.go.th/?url=pr/detail/2/04/20810

ขจรศักดิ์ ศรีอ่อน. (2552). *การทำนายสาเหตุของเหตุการณ์กระแสไฟฟ้าขัดข้อง โดยใช้เทคนิคการทำเหมืองข้อมูลในระบบจำหน่ายของการไฟฟ้าส่วนภูมิภาค เขต 1 ภาคกลาง*. (วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิตที่ไม่มีการตีพิมพ์).มหาวิทยาลัยเกษตรศาสตร์.

บุญเสริม กิจศิริกุล. (2548). *ปัญญาประดิษฐ์ (Artificial Intelligence)*. ภาควิชาวิศวกรรมคอมพิวเตอร์, คณะวิศวกรรมศาสตร์. จุฬาลงกรณ์มหาวิทยาลัย.

พยูน พาณิชย์กุล. (2548). *การพัฒนาระบบดาต้าไมน์นิ่ง โดยใช้ Decision Tree*. สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง.

พวงทิพย์ แท่นแสง. (2550). *การทดสอบประสิทธิภาพการทำงานของอัลกอริทึมการไมนิ่งกฎสำหรับจำแนก*. สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

รัตนทัต. (2562, 15 มีนาคม). *การทำเหมืองข้อมูล (Data Mining)*. กองโลจิสติกส์. https://dol.dip.go.th/th/category/2019-02-08-08-57-30/2019-03-15-08-49-57

ศูนย์สูติ-นรีเวช ชั้น 3 โรงพยาบาลพญาไทศรีราชา. (2561). *การดูแลตนเองให้ห่างไกลมะเร็งปากมดลูก*. https://www.phyathai.com/article_detail/1884/th

สถาบันมะเร็งแห่งชาติ. (2567). *ข่าวและกิจกรรม*. https://www.nci.go.th/th/New_web2024/ข่าวและกิจกรรม.html

สำนักงานสถิติแห่งชาติ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม. (2560). *สรุปผลที่สำคัญผู้ทำงานด้านเทคโนโลยีสารสนเทศและการสื่อสาร พ.ศ. 2559*. ผู้แต่ง

เอกสิทธิ์ พัชรวงศ์ศักดา. (2557). *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิงเบื้องต้น*. เอเชีย ดิจิตอล การพิมพ์.

Han, J. & Kamber, M. (2001). *Data mining: Concepts and techniques*. Morgan Kaufmann.

Pregibons, D. (1996). *Data mining*. https://www.nupress.grad.nu.ac.th/data-mining/

TNIC. (2024). *Data mining คืออะไร สำคัญอย่างไรกับธุรกิจ*. https://tnic.tni.ac.th/data-mining/

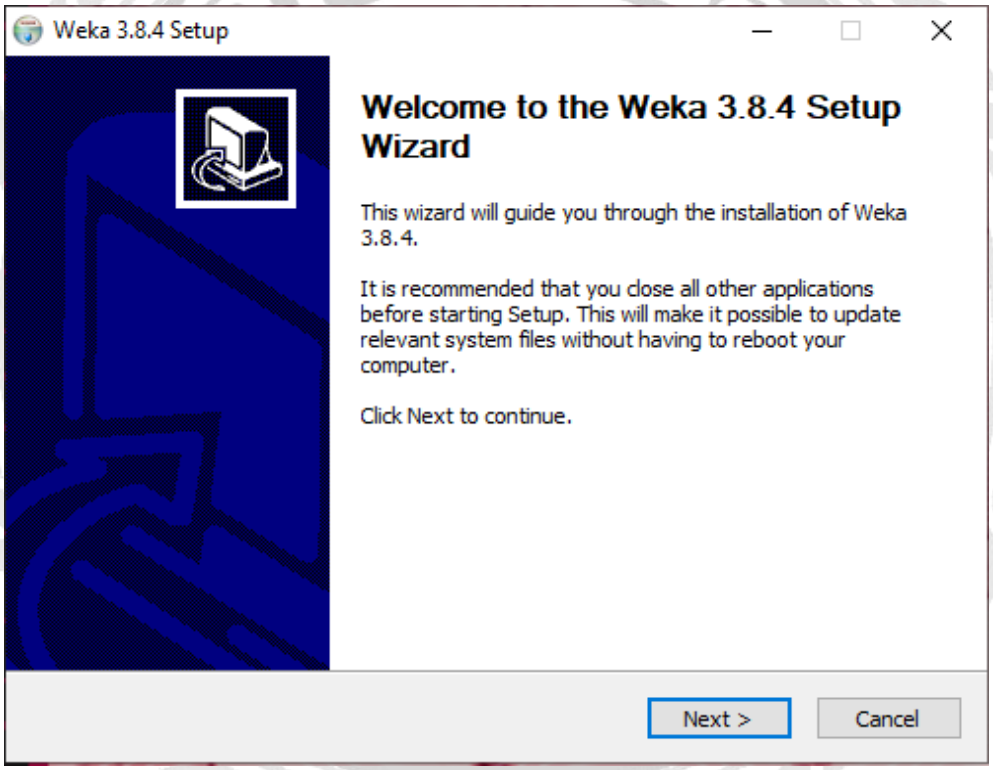
ภาคผนวก

ภาคผนวก ก

คู่มือการติดตั้งโปรแกรม

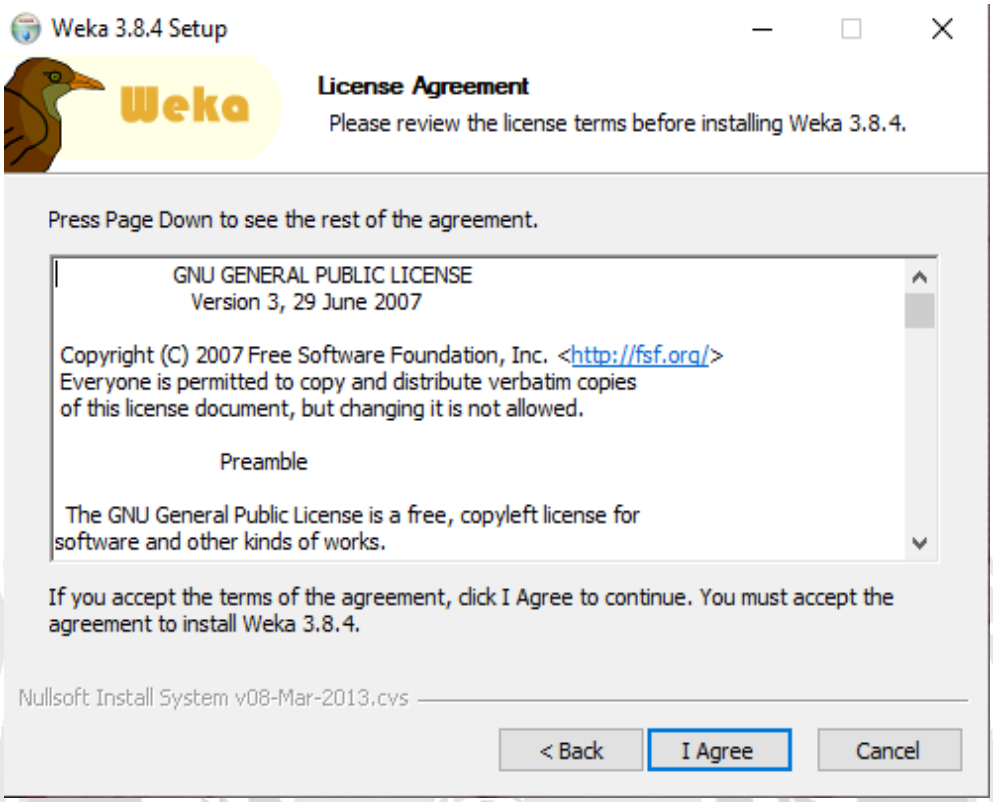
Install Program

1. ดับเบิลคลิก



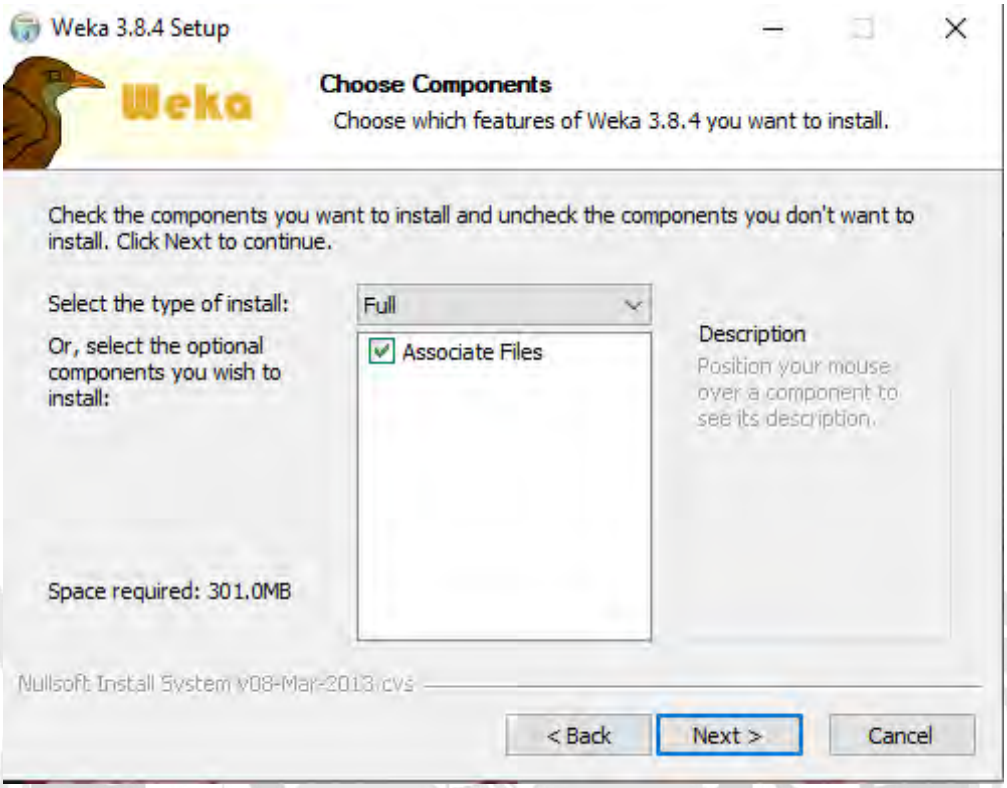
ภาพที่ ก.1 หน้าจอลงโปรแกรม Weka

2. กดปุ่ม Next > เพื่อไปยังหน้าต่อไป



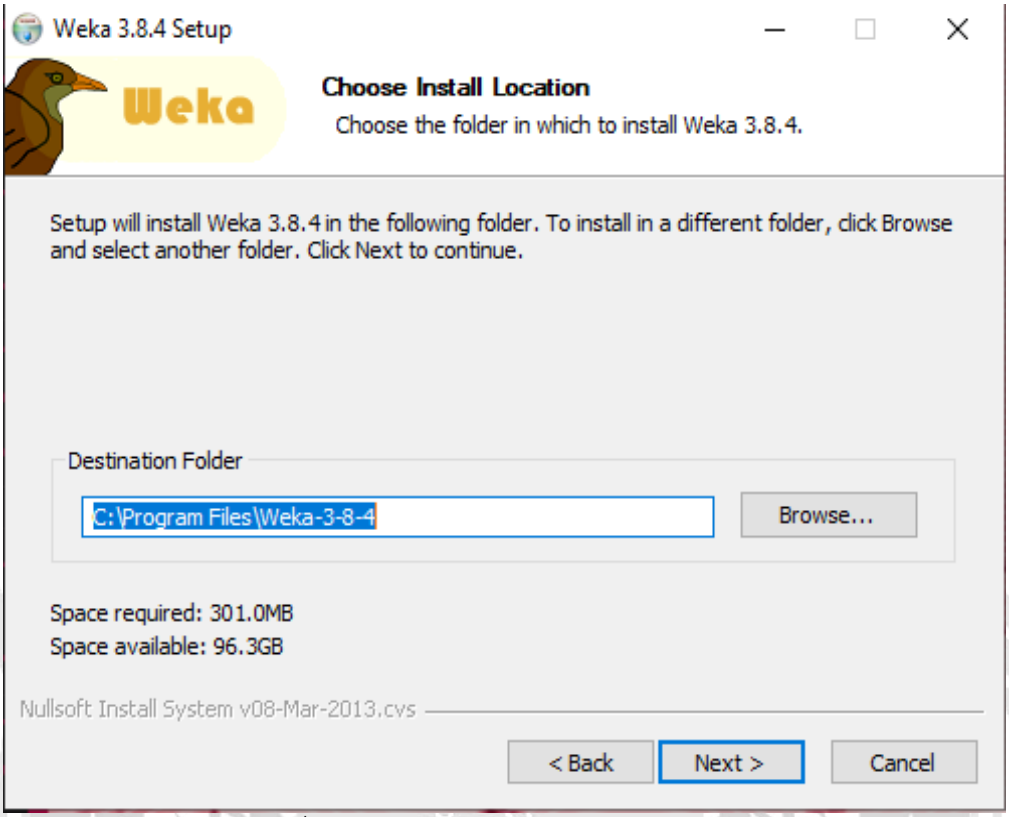
ภาพที่ 2 หน้าจอเลือก I Agree

3. กดปุ่ม I Agree เพื่อไปยังหน้าต่อไป



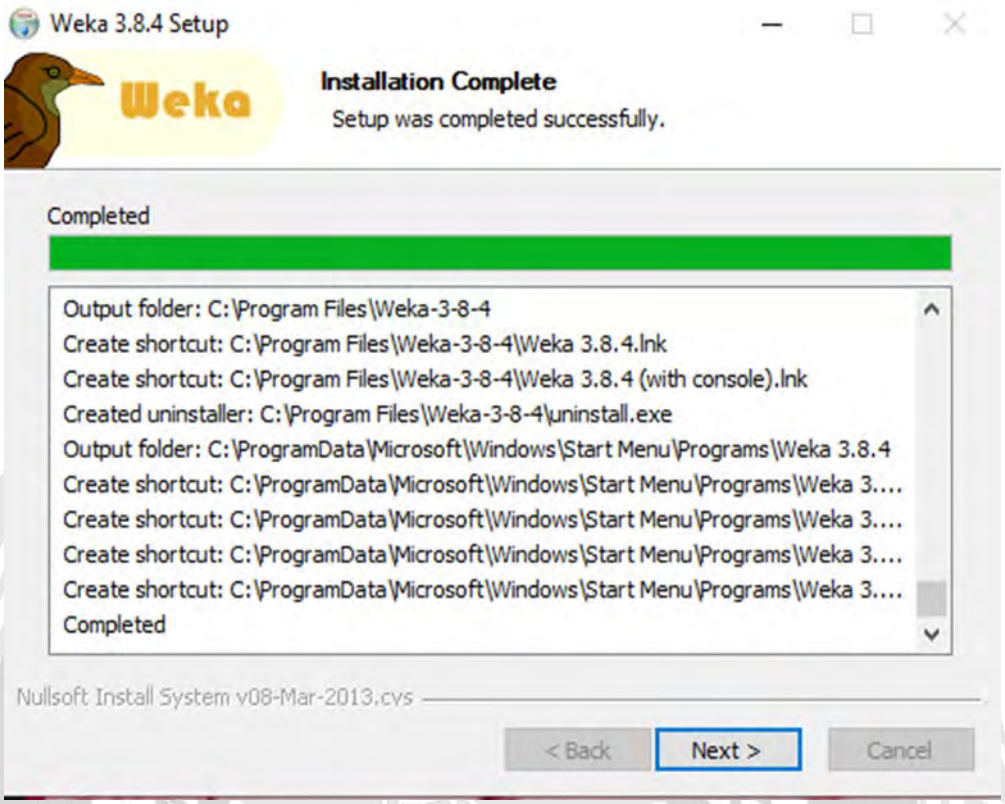
ภาพที่ ก.3 หน้าจอ Choose Components

4. กดปุ่ม Next > เพื่อไปยังหน้าต่อไป



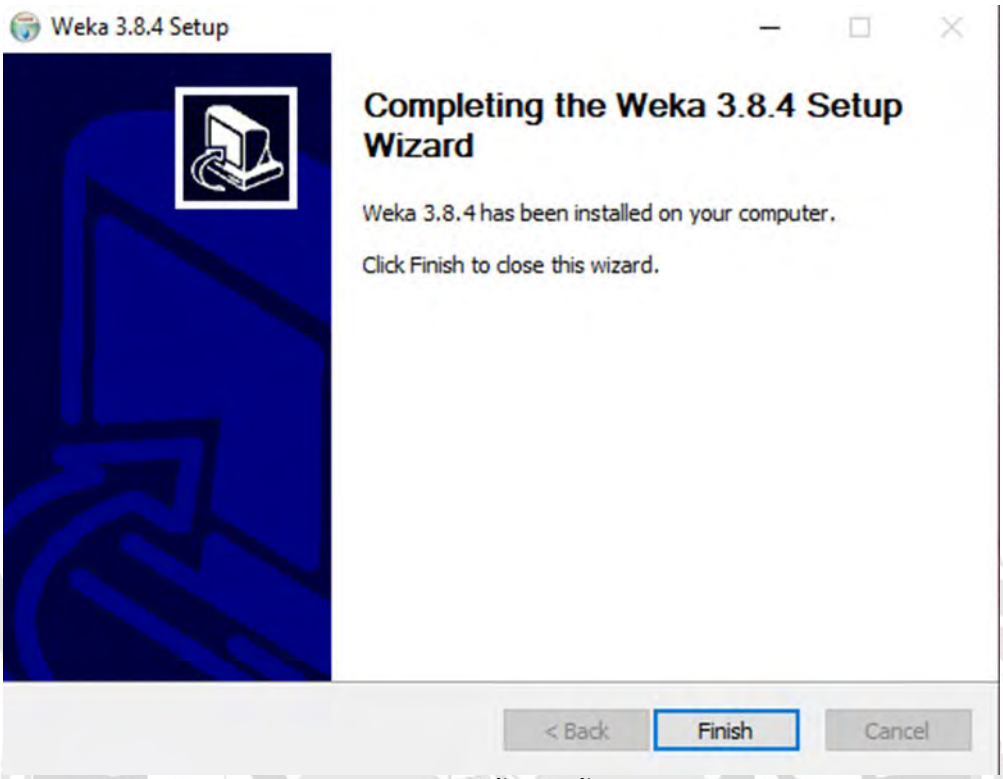
ภาพที่ ก.4 หน้าจอ Choose Install Location

5. กดปุ่ม Next > เพื่อไปยังหน้าต่อไป



ภาพที่ ก.5 หน้าจอ Installing

6. กดปุ่ม Next > เพื่อไปยังหน้าต่อไป



ภาพที่ ก.6 หน้าจอการติดตั้งเสร็จสิ้นแล้ว Weka 3.8.4

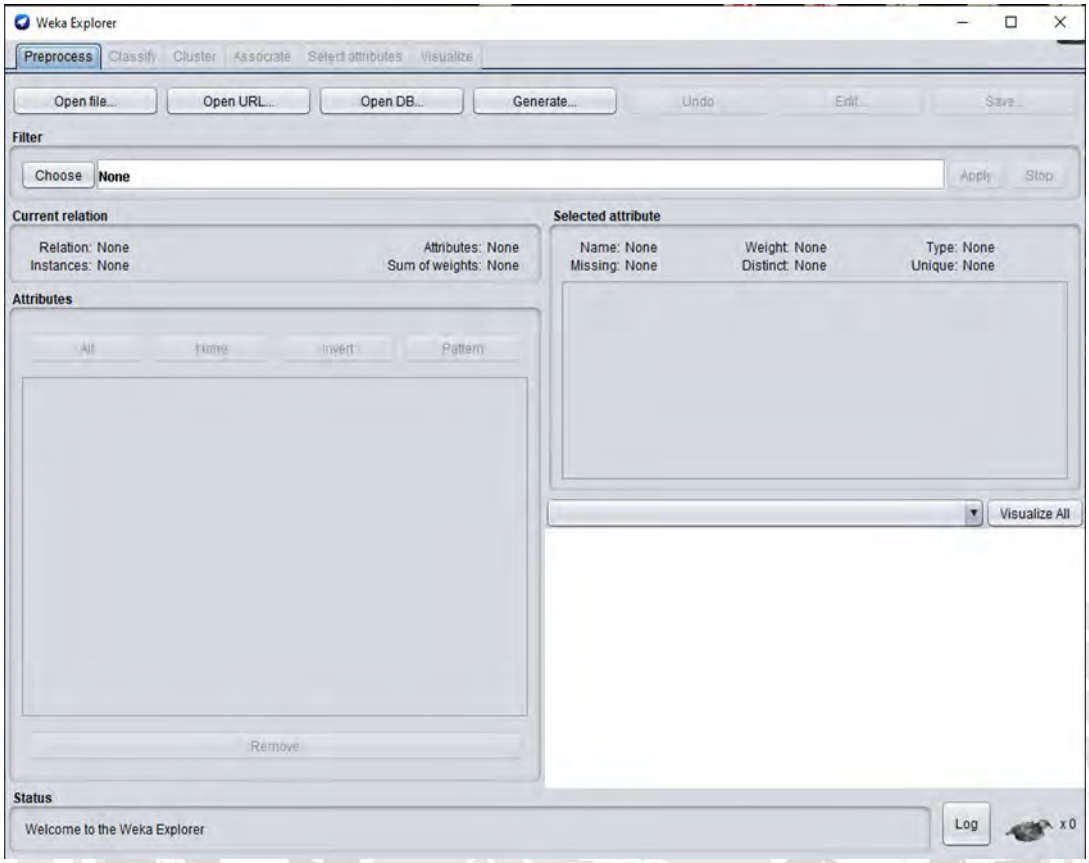
7. กดปุ่ม Finish เพื่อปิดหน้าต่าง

8. ดับเบิลคลิกเพื่อเปิด

โปรแกรม Weka 3.8.4

ภาพที่ ก.7 หน้าจอเลือก Applications Weka

9. กดที่ปุ่ม Explorer เพื่อไปสู่หน้าถัดไป



ภาพที่ ก.8 หน้าจอของ Explorer Weka

10. กดที่ปุ่ม Open File... เพื่อทำการเลือกชุดข้อมูลที่ต้องการ

ประวัติคณะผู้จัดทำ

รหัสนักศึกษา	6201700001
ชื่อ-นามสกุล	นางสาวจีรนันท์ เฉียบแหลม
ที่อยู่	47 ซ.บางแวก124 แขวงบางไผ่ เขตบางแค ถนนบางแวก กรุงเทพมหานคร 10160
เบอร์โทรศัพท์	09-7268-1778
E-Mail	Jeeranan.cha@siam.edu
ระดับประกาศนียบัตรวิชาชีพชั้นสูง	วิทยาลัยพาณิชการธนบุรี
ระดับปริญญาตรี	มหาวิทยาลัยสยาม
รหัสนักศึกษา	6301700001
ชื่อ-นามสกุล	นางสาวยศวดี อุปเนตร
ที่อยู่	503/60 ซ.จรัญฯ37 แขวงบางขุนศรี เขตบางกอกน้อย ถนนจรัญสนิทวงศ์ กรุงเทพมหานคร 10700
เบอร์โทรศัพท์	08-2965-3496
E-Mail	Yodsawadeee.upp@siam.edu
ระดับประกาศนียบัตรวิชาชีพชั้นสูง	วิทยาลัยเทคโนโลยีสยาม(สยามเทค)
ระดับปริญญาตรี	มหาวิทยาลัยสยาม
รหัสนักศึกษา	6301700002
ชื่อ-นามสกุล	นางสาวเบญจกัลยาณี แสงมาลา
ที่อยู่	116/3 ม.5 ต.พันท้ายนรสิงห์ อ.เมืองสมุทรสาคร จ.สมุทรสาคร 74000
เบอร์โทรศัพท์	09-4781-5995
E-Mail	Benjakunlayanee.san@siam.edu
ระดับประกาศนียบัตรวิชาชีพชั้นสูง	วิทยาลัยเทคโนโลยีจรัลสนิทวงศ์
ระดับปริญญาตรี	มหาวิทยาลัยสยาม