

# ทำนายปัจจัยที่มีผลต่อการเลือกอาชีพในอนาคตของนักศึกษา สำนักวิทยาศาสตร์และศิลปดิจิทัล

ปัญญาภัส พุทธิรักษา<sup>1</sup> ณัฐชา สุขลัด<sup>2</sup> ภัทราพร พรหมเงิน<sup>3</sup> และ ธุรกิจ ไชยโพธิ์<sup>4</sup>

Panlapas Phuttharaksa<sup>1</sup>, Natchaya Suklad<sup>2</sup>, Pattaraporn Phomngern<sup>3</sup> and Thoorakit Chaiyapo<sup>4</sup>

สำนักวิทยาศาสตร์และศิลปดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี

B6400873@g.sut.ac.th<sup>1</sup>, B6428273@g.sut.ac.th<sup>2</sup>, B6423087@g.sut.ac.th<sup>3</sup> and B6417338@g.sut.ac.th<sup>4</sup>

## บทคัดย่อ

งานวิจัยนี้เป็นการศึกษาเกี่ยวกับเทคนิคการทำเหมืองข้อมูล ปัจจุบันมีการคำนึงถึงการเลือกอาชีพที่ต้องการในหลากหลายมิติมากขึ้น เพื่อให้ทราบว่ามีปัจจัยใดบ้างที่เข้ามาเกี่ยวข้องในการเลือกอาชีพของนักศึกษาในแต่ละชั้นปี ในงานวิจัยนี้จึงได้ศึกษาเปรียบเทียบประสิทธิภาพตัวแบบจำลอง การตัดสินใจปัจจัยที่เกี่ยวข้องกับการเลือกอาชีพของนักศึกษา โดยดูจากแนวโน้มการเลือกตอบคำถามปัจจัยด้านอาชีพ เพื่อเป็นแนวทางในการประกอบการตัดสินใจของนักศึกษาเอง และเป็นแนวทางในการจัดรูปแบบการเรียนการสอนให้สอดคล้องตรงกับความต้องการในการประกอบอาชีพของนักศึกษา ใช้เทคนิคการวิเคราะห์ 2 เทคนิค คือ เทคนิคต้นไม้ตัดสินใจ เทคนิคเพื่อนบ้านใกล้สุด มาทำนายและเปรียบเทียบความถูกต้องของแต่ละเทคนิค ผลการทดสอบพบว่า ตัวแบบจำลองที่สร้างจากเทคนิคต้นไม้ตัดสินใจ ด้วยชุดคำสั่งการสุ่มป่าไม้ ให้ค่าความถูกต้องสูงสุด ซึ่งมีค่า 58.08% ค่าความแม่นยำ 56.30% ค่าความระลึก 58.10% ค่าความถ่วงดุล 56.70% ตัวแบบจำลองที่สร้างจากเทคนิคเพื่อนบ้านใกล้สุด ใช้ชุดคำสั่ง ไอปีเค ให้ค่าความถูกต้อง ซึ่งมีค่า 56.90% ค่าความแม่นยำ 56.60% ค่าความระลึก 56.90% ค่าความถ่วงดุล 56.30% และตัวแบบจำลองที่สร้างจากเทคนิคต้นไม้ตัดสินใจ ด้วยชุดคำสั่งเจ48 ให้ค่าความถูกต้องน้อยที่สุด ซึ่งมีค่า 55.39% ค่าความแม่นยำ 54.60% ค่าความระลึก 55.40% ค่าความถ่วงดุล 54.80% ผลที่ได้จากการวิจัยสามารถนำผลมาประยุกต์ใช้ในการสร้างแผนการเรียนการสอนให้เหมาะสมกับความต้องการ และเป็นแนวทางในการใช้ประกอบการตัดสินใจของนักศึกษาต่อไป

**คำสำคัญ:** การทำเหมืองข้อมูล ทำนายกลุ่มอาชีพ เทคนิคต้นไม้ตัดสินใจ เทคนิคเพื่อนบ้านใกล้สุด

## บทนำ

การประกอบอาชีพเป็นส่วนหนึ่งของชีวิต อาจกล่าวได้ว่า มนุษย์ทุกคนเกิดมาต้องดิ้นรน เพื่อให้มีชีวิตอยู่รอดในโลกต่อไป การประกอบอาชีพเพื่อแสวงหารายได้ จึงเป็นอีกหนทางหนึ่ง การทำงานของมนุษย์ทำเพื่อให้ได้ค่าตอบแทนเป็นกิจกรรมที่ทำให้เกิดรายได้และเกิดการแลกเปลี่ยนทรัพยากรระหว่างผู้ซื้อกับผู้ขายเพื่อนำไปใช้จ่ายสำหรับปัจจัยทั้ง 4 ได้แก่ อาหาร ที่อยู่อาศัย เครื่องนุ่งห่ม และยารักษา ในปัจจุบันการตัดสินใจเลือกอาชีพ นับว่าเป็นเรื่องสำคัญอย่างยิ่งในชีวิตมนุษย์ อาจกล่าวได้ว่า “งานคือชีวิต” (กระทรวงแรงงาน, 2552) ในการเลือกเส้นทางการประกอบอาชีพ จำเป็นต้องมีการเริ่มต้นด้วยการวางแผนชีวิตด้านอาชีพตั้งแต่วัยเรียน ซึ่งเป็นการวางแผนระยะยาวที่ต้องใช้เวลาตัดสินใจอย่างยาวนาน และใช้ความพยายามอย่างมาก เพื่อให้ได้รับผลตอบแทนที่คุ้มค่ามากที่สุดจากการตัดสินใจ ในปัจจุบันมีการเปลี่ยนแปลงทางเศรษฐกิจสังคม และด้านอื่น ๆ อย่างรวดเร็วซึ่งเป็นผลมาจากความเจริญก้าวหน้าทางเทคโนโลยีและวิทยาการ ทำให้มีอาชีพที่หลากหลายมากยิ่งขึ้น การพัฒนาเทคโนโลยีเพื่อตอบสนองความต้องการของมนุษย์ได้เจริญก้าวหน้าอย่างรวดเร็ว เห็นได้จากสิ่งของเครื่องใช้ต่าง ๆ ที่พบเห็นในชีวิตประจำ ซึ่งในปัจจุบันมีการให้ความสำคัญกับเทคโนโลยีที่ใช้เพื่อการทำงานหรือเทคโนโลยีการประกอบอาชีพ เพราะเป็นปัจจัยที่ทำให้การจัดการในกระบวนการทำงานดี และมีประสิทธิภาพมากยิ่งขึ้นเมื่อเทียบกับการใช้วิถีเดิมที่ต้องลงแรงหรืออาศัยแรงจากธรรมชาติ ซึ่งในปัจจุบันเอง ปฏิเสธไม่ได้เลยว่า ความก้าวหน้าทางด้านวิทยาศาสตร์และเทคโนโลยี ส่งผลให้มีการพัฒนาและคิดค้นสิ่งอำนวยความสะดวกต่อการดำรงชีวิตเป็นอย่างมาก โดยเทคโนโลยีได้เข้ามามีบทบาทในชีวิตประจำวันของมนุษย์อย่างเห็นได้ชัด รวมถึงปัจจัยพื้นฐานการดำรงชีวิต ซึ่งผลิตสินค้าและให้บริการต่าง ๆ เพื่อตอบสนองความต้องการของมนุษย์มากขึ้น จะเห็นว่าปัจจุบันมนุษย์เกี่ยวข้องกับเทคโนโลยีเป็นอย่างมาก โดยเฉพาะเทคโนโลยีการประกอบอาชีพ เนื่องจากพื้นฐานของเทคโนโลยีนั้นมีประโยชน์ต่อการพัฒนาสังคม

ขณะเดียวกันสถานการณ์ตลาดแรงงานในปัจจุบันได้รับผลกระทบจากทั้งวิกฤตการแพร่ระบาดของไวรัสโควิด-19 และเศรษฐกิจโลกที่ชะลอตัวอย่างต่อเนื่องจากปีที่ผ่านมา ทำให้ภาคอุตสาหกรรม และภาคธุรกิจต่าง ๆ รวมทั้งแรงงานต้องปรับตัวให้พร้อมรับมือ บุคลากรสายงานไอทีนับเป็นกลุ่มแรงงานที่ตลาดแรงงานมีความต้องการสูง ติดอันดับที่ 4 จาก 10 สายงาน ที่ตลาดต้องการในปัจจุบัน และยังมีแนวโน้มการขยายตัวอย่างต่อเนื่องที่เพิ่มขึ้นทุก ๆ ปีอีกด้วย เนื่องจากเป็นสายงานที่มีความสอดคล้องกับยุคดิจิทัลในปัจจุบัน ส่งผลให้ความต้องการทักษะใหม่ ๆ ในอาชีพเพิ่มขึ้น ทั้งนี้เพื่อขับเคลื่อนองค์กรอย่างมีประสิทธิภาพและบรรลุเป้าหมายตามที่วางไว้ (กรุงเทพธุรกิจ, 2565) ทางสำนักวิทยาศาสตร์และศิลปดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี เล็งเห็นถึงความก้าวหน้าทางเทคโนโลยีดังกล่าว จึงได้มีการจัดกลุ่มวิชาชีพที่นักศึกษาสามารถเลือกศึกษาได้ด้วยตนเอง โดยแบ่งออกเป็น 2 สาขาวิชา ดังนี้ 1) สาขาวิชาเทคโนโลยีดิจิทัล กลุ่มวิชาชีพที่สามารถเลือกศึกษา ประกอบด้วย กลุ่มวิชาชีพด้าน

โปรแกรมประยุกต์บนเว็บ กลุ่มวิชาชีพด้านโปรแกรมประยุกต์บนมือถือ กลุ่มวิชาชีพด้านซอฟต์แวร์วิสาหกิจ กลุ่มวิชาชีพด้านการจัดการข้อมูล กลุ่มวิชาชีพด้านวิทยาศาสตร์ข้อมูล และกลุ่มวิชาชีพด้านเทคโนโลยีคลาวด์

2) สาขาวิชานิเทศศาสตร์ดิจิทัล กลุ่มวิชาชีพที่สามารถเลือกศึกษา ประกอบด้วย กลุ่มวิชาชีพด้านการออกแบบ visuals กลุ่มวิชาชีพด้านเนื้อหาดิจิทัล กลุ่มวิชาชีพด้านแอนิเมชัน กลุ่มวิชาชีพด้านวีดิทัศน์ดิจิทัล กลุ่มวิชาชีพด้านการวางแผนสื่อดิจิทัล กลุ่มวิชาชีพด้านเกมคอมพิวเตอร์ กลุ่มวิชาชีพด้านการรายงานข่าวแบบเรียลไทม์ กลุ่มวิชาชีพด้านโปรแกรมประยุกต์บนมือถือ และกลุ่มวิชาชีพด้านการวิจัยสื่อดิจิทัล (สำนักวิทยาศาสตร์และศิลปดิจิทัล (มหาวิทยาลัยเทคโนโลยีสุรนารี, 2561) เพื่อให้นักศึกษาเลือกศึกษาเรียนรู้และประกอบอาชีพให้เหมาะสมกับความสามารถหรือความถนัดของแต่ละบุคคล

จากการสำรวจพบว่า กลุ่มอาชีพที่มีทักษะทางด้านเทคโนโลยีดิจิทัล เป็นกลุ่มอาชีพที่มีความสำคัญต่อตลาดแรงงานเป็นอย่างมาก (กรมจัดหางาน, 2561) ดังนั้น ความสำคัญของการตัดสินใจเลือกประกอบอาชีพจึงมีความสำคัญ นอกจากการคำนึงถึงการเลือกอาชีพให้สอดคล้องกับความต้องการของตลาดแรงงานในยุคปัจจุบันแล้ว ยังควรเลือกให้เหมาะสมกับตนเอง โดยคำนึงถึงปัจจัยในด้านต่าง ๆ เช่น ด้านความรู้ความสามารถ ด้านความสนใจ ด้านความถนัด รวมไปถึงด้านสังคม เพื่อให้เกิดสภาพแวดล้อมในการทำงานที่มีประสิทธิภาพ ซึ่งส่งผลต่อรูปแบบงานโดยตรง (โสธยา สุภาพผล, 2563)

ด้วยเหตุผลดังกล่าวทำให้ผู้วิจัยเล็งเห็นความสำคัญของการเลือกประกอบอาชีพจากปัจจัยที่เกี่ยวข้องอย่างถี่ถ้วนมากขึ้น เพื่อเป็นแนวทางประกอบการตัดสินใจในการเลือกประกอบอาชีพของนักศึกษาสาขาวิชาเทคโนโลยีดิจิทัล และสาขาวิชานิเทศศาสตร์ดิจิทัล สำนักวิทยาศาสตร์และศิลปดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี ดังนั้นงานวิจัยนี้จึงได้นำเทคนิคเหมืองข้อมูลเข้ามาช่วยในการวิเคราะห์เปรียบเทียบปัจจัยที่มีผลต่อการเลือกประกอบอาชีพในอนาคตของนักศึกษา ซึ่งจะสามารถวิเคราะห์ได้ว่าปัจจัยใดที่ส่งผลต่อการเลือกประกอบอาชีพมากที่สุดในแต่ละกลุ่มอาชีพ และช่วยให้นักศึกษาเลือกประกอบอาชีพที่เหมาะสมกับตนเองในอนาคต

### วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาปัจจัยการเลือกอาชีพของนักศึกษาในแต่ละชั้นปีการศึกษาของสำนักวิทยาศาสตร์และศิลปดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี
2. เพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อการเลือกอาชีพของนักศึกษาสำนักวิทยาศาสตร์และศิลปดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารีด้วยเทคนิคเหมืองข้อมูล

## ทฤษฎีที่เกี่ยวข้อง

### การทำเหมืองข้อมูล (Data Mining)

คำนิยามของเหมืองข้อมูลนั้นได้มีผู้นิยาม หรือความหมายต่าง ๆ ไว้ ยกตัวอย่างนิยามของเหมืองข้อมูล ดังนี้

สุรพงศ์ เอื้อวัฒนามงคล (2557) ได้ให้ความหมายการทำเหมืองข้อมูลว่า เป็นขบวนการวิเคราะห์ข้อมูลที่มีขั้นตอนเพื่อให้ได้มาซึ่งรูปแบบ (Pattern) ซึ่งแสดงความสัมพันธ์ระหว่างข้อมูลโดยผลลัพธ์ความรู้เกี่ยวกับข้อมูลที่ถูกต้องสามารถนำไปใช้ในการตัดสินใจ และดำเนินงานได้โดยไม่ผิดพลาด เพื่อให้ได้ตามวัตถุประสงค์ในการดำเนินงาน

เอกสิทธิ์ พัทธวงศ์ศักดิ์ (2557) ได้ให้ความหมายของเหมืองข้อมูล หมายถึง เป็นคำศัพท์ที่ใช้เปรียบเทียบกับ การขุดเหมืองแร่ ซึ่งในการขุดเหมืองแร่นั้น ทำเพื่อที่จะได้ แร่ ที่มีมูลค่าออกมา เช่น เพชร พลอย ทอง ที่มีมูลค่ามหาศาล ซึ่งในการขุดเหมืองแร่นั้น ต้องมีการระเบิดภูเขา ขุดดิน ขุดหิน เพื่อหาแร่ต่าง ๆ เหล่านั้น โดยสามารถเปรียบเทียบได้กับองค์กรหรือ บริษัทมีภูเขาข้อมูลมากมาย มีทั้งที่มีมูลค่านำไปใช้ได้ และที่ไม่มีมูลค่าอะไรเลย ต้องกำจัดทิ้ง ซึ่งเป็นการค้นหาสิ่งที่มีประโยชน์ มีมูลค่าจากฐานข้อมูลขนาดใหญ่

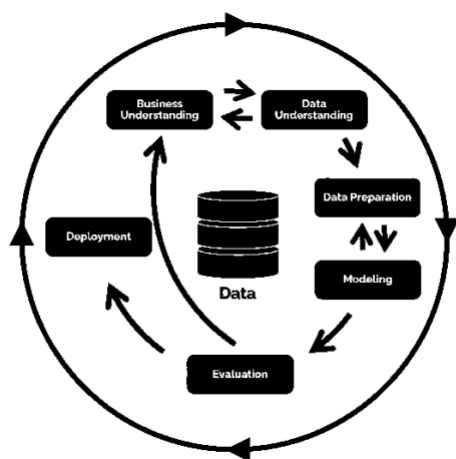
จากนิยามดังกล่าวสามารถสรุปได้ว่า การทำเหมืองข้อมูล (Data Mining) เป็นกระบวนการหาความสัมพันธ์ของข้อมูลที่มีจำนวนมาก ซึ่งจะใช้ขั้นตอนวิธีทางสถิติ ซึ่งจะค้นหาความสัมพันธ์และรูปแบบทั้งหมดที่มีอยู่ในฐานข้อมูล โดยความสัมพันธ์และรูปแบบอาจถูกซ่อนไว้ภายในข้อมูลที่มีจำนวนมาก ในการการทำเหมืองข้อมูลนั้น จะทำการสำรวจและวิเคราะห์ข้อมูลที่มีอยู่ ให้อยู่ในรูปแบบเดียวกันมีความหมายและรูปแบบของกฎ เพื่อให้ได้ความรู้ใหม่ที่มีประโยชน์ในฐานข้อมูล โดยมีเทคนิคในการวิเคราะห์ 3 แบบ ดังนี้ (หนึ่งหทัย ชัยอารกร, 2559)

1) การหากฎความสัมพันธ์ (Association Rule) คือ การค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปหารูปแบบที่เกิดขึ้นบ่อย ๆ (Frequent Pattern) และใช้ในการวิเคราะห์ความสัมพันธ์หรือทำนาย โดยข้อมูลที่น่ามาใช้จะอยู่ในรูปแบบประเภท หรือ ตัวเลขเท่านั้น (ปรีชญา นวนแก้ว, 2563)

2) การแบ่งกลุ่มข้อมูล (Clustering) คือ เป็นเครื่องจักรการเรียนรู้ประเภทหนึ่ง ซึ่งอยู่ในประเภทการเรียนรู้แบบไม่มีผู้สอน หมายถึง การสร้างโมเดลที่ใส่ข้อมูลเข้าไปโดยที่ไม่ได้กำหนดกลุ่มเป้าหมาย แต่ให้โมเดลคอมพิวเตอร์เป็นคนกำหนดเองว่าข้อมูลที่ได้รับมาสามารถแปลงออกมาเป็นโมเดลในลักษณะรูปแบบใดได้บ้าง เป็นการวิเคราะห์ข้อมูล ที่จะแบ่งชุดข้อมูลออกเป็นกลุ่ม นำข้อมูลที่มีลักษณะเหมือนกัน หรือคล้ายกันจัดไว้ในกลุ่มเดียวกัน ซึ่งอาศัยความเหมือน (Similarity) หรือความใกล้ชิด (Proximity) ของข้อมูลในการจัดกลุ่ม ยกตัวอย่างชุดคำสั่ง (Algorithm) การแบ่งกลุ่มข้อมูล เช่น การแบ่งกลุ่มข้อมูลแบบเคมีน (K-Mean Clustering) ฐานข้อมูลแบบลำดับขั้น (Hierarchical Clustering Model) เป็นต้น (Alfred L., 1932)

3) การจำแนกประเภทข้อมูล (Classification) คือ เป็นเครื่องจักรการเรียนรู้ประเภทหนึ่ง ซึ่งอยู่ในประเภทการเรียนรู้แบบมีผู้สอน จะเป็นแบบจำลองในรูปแบบของการจำแนกประเภทข้อมูลเพื่อให้ได้คำตอบที่เป็นตัวเลือก หรือกลุ่มข้อมูลของโมเดล เช่น ใช่กับไม่ใช่ เป็นกับไม่เป็น หรือเป็นกลุ่มคำตอบว่าเป็นกลุ่ม 1 2 หรือ 3 เป็นต้น (สถาบันนวัตกรรมและธรรมาภิบาลข้อมูล, 2565)

กระบวนการวิเคราะห์ข้อมูลด้วยคริสปี้ ดีเอ็ม (Cross Industry Standard Process for Data Mining: CRISP DM)



ภาพที่ 1 กระบวนการวิเคราะห์ข้อมูลด้วยคริสปี้ ดีเอ็ม (CRISP DM)

ที่มา: NICK HOTZ (2023)

จากภาพที่ 1 ได้แสดงกระบวนการวิเคราะห์ข้อมูลด้วยคริสปี้ ดีเอ็ม (Cross Industry Standard Process for Data Mining: CRISP DM) ซึ่งเป็นกระบวนการที่ใช้เป็นพื้นฐานของการวิเคราะห์ข้อมูล โดยมีขั้นตอนทั้งหมด 6 ขั้นตอน ดังนี้ (NICK HOTZ, 2023)

1) การเข้าใจปัญหาทางธุรกิจ (Business Understanding) ในขั้นตอนนี้คือทำความเข้าใจวัตถุประสงค์และข้อกำหนดของโครงการ

2) การทำความเข้าใจข้อมูล (Data Preparation) ในขั้นตอนนี้คือการเพิ่มความเข้าใจในการใช้ข้อมูลธุรกิจ เพื่อที่จะช่วยผลักดัน และมุ่งเน้นในการระบุ รวบรวม และวิเคราะห์ชุดข้อมูลที่สามารถช่วยให้บรรลุเป้าหมายของธุรกิจได้

3) การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่ใช้เวลานานมากที่สุด เพราะจะต้องสกัดข้อมูลให้ออกมาได้ตรงตามความต้องการ และสามารถนำไปวิเคราะห์ได้จริงโดยเกิดข้อผิดพลาดที่น้อยที่สุด โดย

การเตรียมข้อมูลจะมี 3 ขั้นตอนย่อย ดังนี้ การคัดเลือกข้อมูล (Data Selection) การทำความสะอาดข้อมูล (Data Cleaning) และการแปลงข้อมูล (Data Transformation) (เอกสิทธิ์ พัทธวงศ์ศักดิ์, 2557)

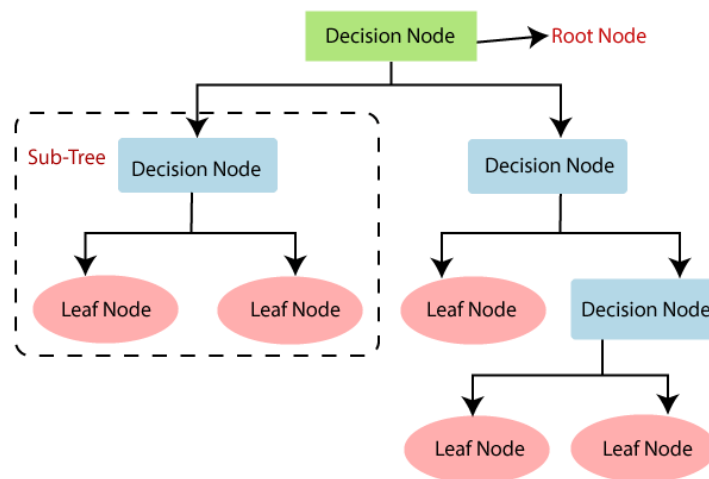
4) การสร้างตัวแบบจำลอง (Modeling) ในขั้นตอนนี้คือการวิเคราะห์ข้อมูลด้วยเทคนิคเหมือนข้อมูลต่าง ๆ โดยในงานวิจัยนี้จะเป็นการใช้เทคนิคต้นไม้ตัดสินใจ และเทคนิคเพื่อนบ้านใกล้สุด

5) การประเมินประสิทธิภาพ (Evaluation) ในขั้นตอนนี้คือการทดสอบตัวแบบจำลองว่ามีความถูกต้องมากน้อยเพียงใด เพื่อที่จะสามารถนำไปประยุกต์ใช้งานได้จริง

6) การนำผลลัพธ์ไปใช้งาน (Deployment) ในขั้นตอนนี้คือการนำองค์ความรู้ที่ได้จากวิเคราะห์ข้อมูลด้วยเทคนิคเหมือนข้อมูลต่าง ๆ และได้ค่าความถูกต้องที่ดีที่สุดไปใช้งาน ให้ตรงกับวัตถุประสงค์ขององค์กรต่อไป

### ต้นไม้ตัดสินใจ (Decision Tree)

เป็นเครื่องมือที่ใช้ในการจำแนกประเภทและการทำนาย ซึ่งเป็นการเรียนรู้จากการคาดคะเนเหตุการณ์ที่เกิดขึ้น โดยอาศัยเงื่อนไขเป็นตัวช่วยในการตัดสินใจ มีรูปแบบของเทคนิคเป็นโครงสร้างแบบแผนผังลำดับเหมือนต้นไม้ ซึ่งจะประกอบไปด้วย 3 ส่วน ส่วนที่เป็นโหนดภายใน (Inner node) เป็นโหนดของต้นไม้ใช้แทนตัวแปรต่าง ๆ ที่ใช้ในการตัดสินใจ ส่วนที่กิ่ง (Branch, Link) ใช้แทนค่าที่เป็นได้ของตัวแปร และส่วนที่เป็นโหนดใบ (Leaf node) แสดงกลุ่มในแต่ละโหนดภายในแสดงถึงการทดสอบของแอตทริบิวต์รวมไปถึงแสดงผลลัพธ์ของการทดสอบศึกษาถึงการเรียนรู้ ดังภาพที่ 2

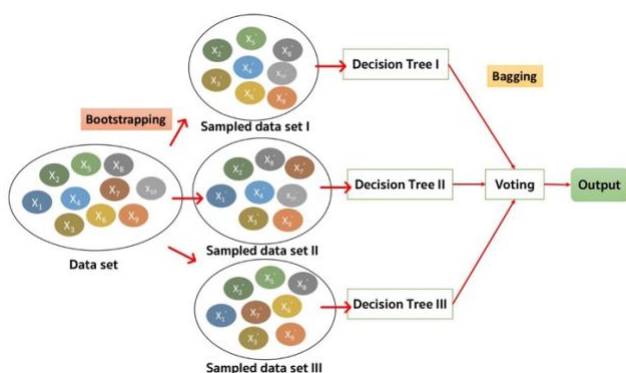


ภาพที่ 2 ต้นไม้ตัดสินใจ (Decision Tree)

ที่มา: Kong Ruksiam (2020)

## การสุ่มป่าไม้ (Random Forest)

เป็นขั้นตอนเทคนิคหนึ่งของการเรียนรู้ที่นิยมใช้มาก โดยการสุ่มป่าไม้ (Random Forest) เป็นขั้นตอนวิธีพัฒนาต่อยอดมาจาก ต้นไม้ตัดสินใจ (Decision Tree) แต่มีความแตกต่างกันที่ การสุ่มป่าไม้จะเป็นการเพิ่มจำนวนต้นไม้ เป็นหลาย ๆ ต้น ทำให้ ประสิทธิภาพการทำงานและพยากรณ์สูงขึ้น โดยมีหลักการทำงาน คือ จะแบ่งข้อมูลออกเป็น ต้นไม้ตัดสินใจ หลาย ๆ ต้น โดยแต่ละต้นจะได้รับคุณลักษณะ (Feature) และข้อมูล (Data) ที่แตกต่างกัน เพื่อให้ได้ต้นไม้ที่มีความหลากหลายและมีความอิสระต่อกันมากขึ้น ดังภาพที่ 3 (PradyaSin, 2019)



ภาพที่ 3 การสุ่มป่าไม้ (Random Forest)

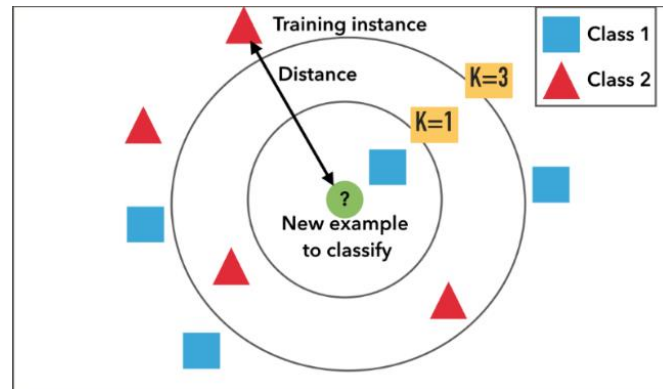
ที่มา: Kong Ruksiam (2020)

## เทคนิคเพื่อนบ้านใกล้สุด (K-Nearest Neighbors)

เป็นเทคนิคที่ใช้ในการจำแนกประเภทในรูปแบบจัดกลุ่มข้อมูล (Classification) โดยใช้หลักการเปรียบเทียบข้อมูลที่สนใจกับข้อมูลชุดอื่นว่ามีความคล้ายคลึงมากน้อยขนาดไหน หากข้อมูลที่กำลังสนใจอยู่ใกล้ข้อมูลใดมากที่สุด ระบบจะให้คำตอบเป็นข้อมูลที่อยู่ใกล้ที่สุด และอยู่ในลักษณะการทำงานแบบไม่ได้ใช้ข้อมูลชุดเรียนรู้ (Training Data) ในการสร้างแบบจำลอง มีลักษณะการทำงานตามรูปภาพที่ 4 สามารถอธิบายเป็นขั้นตอนการทำงานในลักษณะดังนี้ (Kong Ruksiam, 2020)

- 1) ทำการเปรียบเทียบจุดใหม่วงกลมสีเขียว (New Point) กับจุดทั้งหมดใน Training Set เพื่อหาจุดใกล้เคียงกับจุดใหม่
- 2) กำหนดจำนวนจุดใกล้เคียงกับจุดใหม่จำนวน K จุด
- 3) หาก K=1 แสดงว่าให้นำจุดที่ใกล้เคียงกับจุดใหม่ 1 จุดมาพิจารณานั้นคือ Class A มากำหนดค่า Class ให้จุดใหม่เป็น Class A เหมือนกัน

4) หาก  $K=3$  แสดงว่าให้นำจุดที่ใกล้เคียงกับจุดใหม่ 3 จุดมาพิจารณานั้นคือ Class A 1 จุด กับ Class B อีก 2 จุดมาคำนวณผลรวมค่าที่ใกล้เคียง ผลปรากฏว่า Class B มี 2 จุดค่าผลรวมความใกล้เคียงมากกว่า Class A ดังนั้นจุดใหม่จะถูกกำหนดให้เป็น Class B



ภาพที่ 4 รูปแบบการทำงานของเทคนิคเพื่อนบ้านใกล้สุด (K-Nearest Neighbor)

ที่มา: Kong Ruksiam (2020)

### เมทริกซ์ข้อผิดพลาด (Confusion Matrix)

คือตารางสำคัญในการวัดความสามารถของเครื่องจักรการเรียนรู้ ในการแก้ปัญหาการจำแนกประเภทข้อมูล (chengz, 2019)

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

ภาพที่ 5 ตารางConfusion Matrix

ที่มา : (chengz, 2019)

True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่า “จริง” และ มีค่าเป็น “จริง ”

True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่า “ไม่จริง” และ มีค่า “ไม่จริง ”

False Positive (FP) คือ สิ่งที่โปรแกรมทำนายว่า “จริง” แต่ มีค่าเป็น “ไม่จริง”

False Negative (FN) คือ สิ่งที่โปรแกรมทำนายว่า “ไม่จริง” แต่ มีค่าเป็น “จริง”



### ค่าความถูกต้อง (Accuracy)

เป็นค่าที่บ่งบอกถึงความสามารถของเครื่องมือวัด ในการอ่านค่าหรือแสดงค่าที่วัดได้เข้าใกล้ค่าจริง โดยการคำนวณค่าความถูกต้องโดยใช้สมการ (ทวิพล, 2555)

$$\text{ค่าความถูกต้อง} = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

### ค่าความแม่นยำ (Precision)

เป็นค่าที่โมเดลทายเป็นคลาสที่กำลังพิจารณาถูกหรือค่าที่โมเดลทำนายว่าเป็นคลาสที่กำลังพิจารณาทั้งถูกและผิดโดยการคำนวณค่าความแม่นยำสามารถใช้สมการ (Ninenox, 2020)

$$\text{ค่าความแม่นยำ} = \frac{TP}{(TP+FP)}$$

### ค่าความครบถ้วน (Recall)

เป็นการวัดความสามารถของระบบในการดึงเอกสารที่เกี่ยวข้องออกมา โดยค่าความครบถ้วนเป็นอัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและถูกดึงออกมา กับจำนวนเอกสารที่เกี่ยวข้องทั้งหมด คำนวณได้จากสมการ (Thaiall.com, 2558)

$$\text{ค่าความครบถ้วน} = \frac{TP}{(TP + FN)}$$

### ค่าความถ่วงดุล (F-measure)

เป็นค่าที่เกิดจากการเปรียบเทียบระหว่างค่าความแม่นยำและค่าความครบถ้วนของแต่ละคลาสเป้าหมาย คำนวณได้จากสมการ (อนุพงศ์, 2561)

$$\text{ค่าความถ่วงดุล} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

### สโมต (SMOTE)

SMOTE (synthetic minority over- sampling technique) เป็นเทคนิคที่ใช้ในการแก้ปัญหาการจำแนกข้อมูลที่ไม่สมดุล เนื่องจากข้อมูลในแต่ละคลาสมีจำนวนแตกต่างกันมาก ทำให้ผลลัพธ์ของการจำแนกไม่สมเหตุสมผล ดังนั้นวิธีสโมต เป็นการเพิ่มจำนวนข้อมูลกลุ่มน้อยนั้น ให้มีจำนวนเพิ่มขึ้น โดยการเพิ่มข้อมูลในการปรับข้อมูลที่ไม่สมดุล และใช้งานร่วมกับเทคนิคการจำแนกข้อมูลต่าง ๆ ซึ่งพบว่าการใช้งานเทคนิคสโมต ทำให้ผลการทำนายมีความถูกต้องเพิ่มขึ้น โดยการเพิ่มข้อมูลในกลุ่มน้อยนั้นทำให้การกระจายของกลุ่มข้อมูลมีความสมดุลมากยิ่งขึ้น (พุทธิพร, 2562)

## งานวิจัยที่เกี่ยวข้อง

### จากการศึกษางานวิจัยที่เกี่ยวข้อง

สุภลักษณ์ สีสุทอง (2559) ได้เลือกการพัฒนาแบบจำลองการตัดสินใจในการเลือกกลุ่มวิชาสำหรับนักศึกษา สาขาวิชาเทคโนโลยีสารสนเทศปีที่ 2 มีการใช้เทคนิคแบ่งออกเป็น 3 แบบได้แก่ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคการจัดลำดับ (Ranking) เทคนิคกฎการจำแนก (Classification Rules) โดยใช้เทคนิคกระบวนการ ลำดับชั้นเชิงวิเคราะห์ในการประเมินเกณฑ์และทางเลือกในการตัดสินใจการเลือกกลุ่มวิชาของนักศึกษาใน 4 กลุ่ม วิชา ได้แก่ กลุ่มวิชาซอฟต์แวร์วิสาทกิจ กลุ่มวิชานิติศาสตร์ กลุ่มวิชาระบบสารสนเทศเพื่อการจัดการ และกลุ่ม วิชาสารสนเทศศึกษา จำนวน 165 คน พบว่าปัญหาในการเลือกเข้ากลุ่มวิชาของนักศึกษา 3 อันดับ อันดับแรก คือ ด้านผลการเรียน คิดเป็นร้อยละ 84.5% อันดับสอง คือ ด้านการไม่รู้ข้อมูลการเรียนรู้ของแต่ละกลุ่มวิชาดีพอ คิด เป็นร้อยละ 71.4% และอันดับสาม คือ ด้านความไม่ชอบไม่ถนัดและไม่สามารถด้านนั้น ๆ เป็นพื้นฐาน คิด เป็นร้อยละ 59% และผลประเมินออกมาว่าการใช้เทคนิคกฎการจำแนก มาใช้สำหรับวิเคราะห์การเลือกกลุ่มวิชา สำหรับนักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ โดยผลประเมินพบว่าตัวแบบจำลองการตัดสินใจในการเลือกกลุ่ม วิชาให้ค่าระลึกละเอียดสูงถึง 84.5% สรุปได้ว่าแบบจำลองนี้สามารถใช้งานได้และมีความน่าเชื่อถือสูงสุด

รัชฎา เทพประสิทธิ์ และ จริญญา แสนราช (2562) ได้นำเอาแผนการเรียนรู้ก่อนเข้าศึกษาและเพศโดยใช้ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เพื่อวิเคราะห์ปัจจัยที่มีผลต่อการเลือกเรียนสาขาวิชาในคณะครุศาสตร์ มหาวิทยาลัยราชภัฏเชียงรายของนักศึกษาระดับปริญญาตรี โดยมีการนำข้อมูลจากชุดเรียนรู้มาทดสอบด้วยชุด ข้อมูลทดสอบที่ให้ค่าถูกต้อง โดยใช้วิธีการทดสอบแบบไขว้ (Cross-Validation Test) เพื่อหาค่าความถูกต้อง (Accuracy) โดยตัวแบบจำลองที่พัฒนาจากเทคนิคต้นไม้ตัดสินใจที่ถูกทดสอบด้วยข้อมูลที่แบ่งออกเป็น 10 ส่วน ได้ค่าความถูกต้องเท่ากับ 72.5% ซึ่งถือว่าเป็นการทดสอบที่เชื่อถือได้ จากงานวิจัยดังกล่าวพบว่า การใช้เทคนิค ต้นไม้ตัดสินใจให้ค่าความถูกต้องที่น่าเชื่อถือ

สำราญ วานนท์ และ ธรัช แสนราช (2561) ได้นำข้อมูลภาวะการมีงานทำกับข้อมูลระยะเย็นประวัติของ ผู้สำเร็จการศึกษา ในสาขาวิชาทางด้านคอมพิวเตอร์ของมหาวิทยาลัยในประเทศมาศึกษาโดยใช้เทคนิคการจำแนก ประเภทข้อมูล (Classification) การใช้คุณลักษณะทางด้านความสามารถหรือผลการเรียนของผู้สำเร็จการศึกษาที่ ได้รับจากหลักสูตร ด้านความสามารถพิเศษ ด้านอาชีพของบิดามารดา ด้านรายได้ของบิดามารดา ด้านเพศ ด้าน ตำแหน่งงาน ด้านความสอดคล้องสาขา ด้านสาขาวิชา งานวิจัยนี้ได้เสนอเทคนิคการจำแนกประเภทข้อมูล 3 เทคนิคได้แก่ เทคนิคการจำแนกประเภทข้อมูล ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) เทคนิคการจำแนกประเภท ข้อมูลด้วยวิธีการสุ่มป่าไม้ (Random Forest) และเทคนิคการจำแนกประเภทข้อมูลด้วยวิธีแบ็กกิง (Bagging)

สำหรับวิเคราะห์ค่าความแม่นยำการพยากรณ์และนำเทคนิคที่ให้ค่าความแม่นยำมากที่สุดได้แก่ การสุ่มป่าไม้ 84.29% ไปใช้เป็นตัวแบบจำลองในการพัฒนาระบบพยากรณ์เพื่อการแนะแนวอาชีพต่อไป

สมฤทัย กลัดแก้ว อรไท ชั่วเจริญ และ ชำนาญ เจริญรุ่งเรือง (2557) ได้นำข้อมูลเพื่อศึกษาปัจจัยที่มีผลต่อการเลือกตำแหน่งงานให้สอดคล้องกับความสามารถของบัณฑิตด้วยการวิเคราะห์การถดถอยโลจิสติกส์พหุกลุ่ม (Logistic Regression) และการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) เพื่อเปรียบเทียบความถูกต้องของการพยากรณ์ระหว่างการวิเคราะห์การถดถอยโลจิสติกส์พหุกลุ่มและการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจ เพื่อพัฒนาตัวแบบจำลองการตัดสินใจการเลือกตำแหน่งงานให้สอดคล้องกับความสามารถของบัณฑิต สถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ การวิเคราะห์การถดถอยโลจิสติกส์พหุกลุ่ม และการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ผลวิจัยพบว่าเทคนิคการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจมีค่าความถูกต้องมากกว่าการวิเคราะห์การถดถอยโลจิสติกส์พหุกลุ่มเล็กน้อย โดยค่าความถูกต้องของเทคนิคการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจเท่ากับ 57.37% และการวิเคราะห์การถดถอยโลจิสติกส์พหุกลุ่มมีค่าความถูกต้อง 56.3%

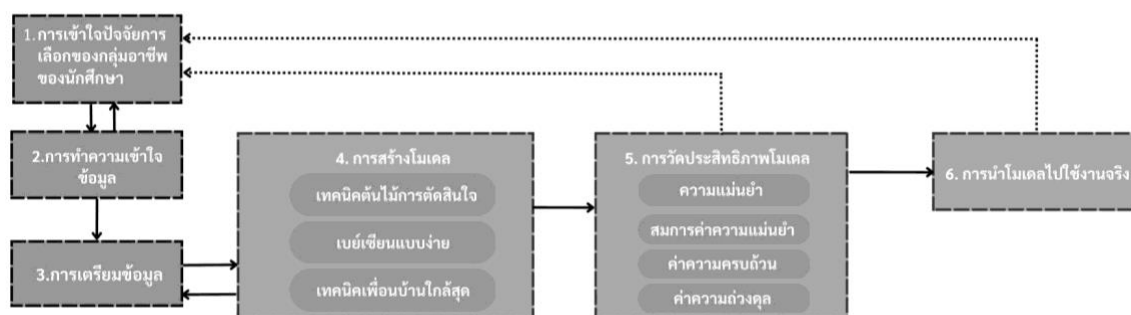
กมลรัตน์ สมใจ (2563) ได้นำข้อมูลของนักศึกษาสาขาเทคโนโลยีสารสนเทศมาวิเคราะห์ 3 ตัวแบบจำลองด้วยกันได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) นาอ็ฟเบย์ (Naive Bayes) และโครงข่ายประสาทเทียม (Neural Network) ผลของการเปรียบเทียบพบว่าโมเดลต้นไม้ตัดสินใจ มีการให้ค่าความถูกต้องมากที่สุด จำแนกเป็นกลุ่มวิชาคือกลุ่มวิชาเทคโนโลยีสารสนเทศ มีค่าความถูกต้องคิดเป็นร้อยละ 94.17% กลุ่มวิชาเทคโนโลยีคอมพิวเตอร์มีค่าความถูกต้อง ร้อยละ 88.72% และกลุ่มวิชาการจัดการคอมพิวเตอร์เพื่อการศึกษา มีค่าความถูกต้อง ร้อยละ 92.87% ซึ่งสามารถสรุปได้ว่าตัวแบบจำลองต้นไม้การตัดสินใจนั้นมีประสิทธิภาพดีที่สุด

ตารางที่ 1 เปรียบเทียบงานวิจัยที่เกี่ยวข้อง

งานวิจัย	เครื่องมือการทำเหมืองข้อมูล									ค่าประสิทธิภาพที่ใช้เปรียบเทียบ			
	เทคนิคต้นไม้การตัดสินใจ	เทคนิคการจัดลำดับ	เทคนิคกฎการจำแนก	เทคนิคสุ่มป่าไม้	เทคนิคแบ็กกิง	เทคนิคโครงข่ายประสาทเทียม	เทคนิคเบย์เขียนแบบง่าย	เทคนิควิเคราะห์การถดถอยโลจิสติกส์พหุกลุ่ม	เทคนิคเพื่อนบ้านใกล้สุด	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าความระลึก	ค่าความถ่วงดุล
สุกลักษณ์ (2559)	X	X	X								X	X	
รัชฎา และจรัญ (2562)	X									X			
สำราญ และอริช (2561)	X			X	X					X	X	X	
อรไท และชำนาญ (2557)	X							X		X	X	X	
กมลรัตน์ (2563)	X					X	X			X	X	X	X
อนันต์ (2560)	X		X				X			X	X	X	
งานวิจัยนี้	X			X					X	X	X	X	X

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่าแต่ละงานวิจัยที่เกี่ยวข้องได้นำเทคนิคต่าง ๆ มาใช้ในการศึกษาปัจจัยที่มีผลต่อการเลือกศึกษา รวมไปถึงการเลือกอาชีพ และเลือกตำแหน่งงานในอนาคต นำมาใช้พัฒนาตัวแบบจำลองการตัดสินใจ และใช้ในการเปรียบเทียบค่าประสิทธิภาพต่าง ๆ ของเทคนิคเหมือนข้อมูลที่นำมาใช้ในการทำวิจัย ซึ่งจากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่าเทคนิคที่ได้รับความนิยมมากที่สุด คือ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยเหตุนี้เอง ผู้วิจัยจึงสนใจนำเทคนิคดังกล่าวมาปรับใช้ รวมถึงเทคนิคการทำเหมือนข้อมูลอื่น ๆ เพื่อศึกษาเปรียบเทียบกับเทคนิคต่าง ๆ และเลือกใช้เทคนิคที่เหมาะสมกับข้อมูลทดสอบที่ต้องการนำวิเคราะห์ต่อไป

### วัสดุ อุปกรณ์ และวิธีเนิการวิจัย



ภาพที่ 6 กระบวนการวิเคราะห์ข้อมูลด้วยคริปส์ ดีเอ็ม

จากภาพที่ 6 วิธีการดำเนินงานวิจัยในการสร้างตัวแบบจำลองทำนายปัจจัยที่มีผลต่อการเลือกอาชีพในอนาคตของนักศึกษาสำนักวิทยาศาสตร์และศิลปดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี โดยนำกระบวนการการวิเคราะห์ข้อมูลด้วยคริปส์ ดีเอ็ม มาปรับใช้ให้เหมาะสมกับงานวิจัย โดยมีขั้นตอนและรายละเอียด ดังนี้

**ขั้นตอนที่ 1 การทำความเข้าใจปัญหา (Business Understanding)** ทำความเข้าใจจากข้อมูลที่ได้มา ซึ่งเป็นข้อมูลเกี่ยวกับปัจจัยในการเลือกอาชีพของนักศึกษา สำนักวิทยาศาสตร์และศิลปดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี ในแต่ละชั้นปีการศึกษา โดยในงานวิจัยนี้ผู้วิจัยต้องการศึกษาเรื่องการตัดสินใจปัจจัยที่เกี่ยวข้องกับการเลือกอาชีพของนักศึกษา โดยดูจากแนวโน้มการเลือกตอบคำถามปัจจัยด้านอาชีพ เพื่อให้ได้ข้อมูลที่ตรงกับวัตถุประสงค์ในการเลือกปัจจัยที่ใช้ตัดสินใจในการประกอบอาชีพ

**ขั้นตอนที่ 2 การทำความเข้าใจข้อมูล (Data Understanding)** ข้อมูลนี้ได้มีการรวบรวมข้อมูลนักศึกษาไว้ โดยมีทั้งหมด 310 รายการ และมีจำนวนแอตทริบิวต์ 262 แอตทริบิวต์ ซึ่งในงานวิจัยนี้จะใช้รายละเอียดข้อมูล ในการนำมาวิเคราะห์ จำนวนทั้งสิ้น 16 แอตทริบิวต์ โดยมีกลุ่มเป้าหมายเป็นกลุ่มอาชีพรายละเอียดของข้อมูลในแต่ละแอตทริบิวต์ ที่จะนำข้อมูลไปวิเคราะห์ (Data Directory) ตามตารางที่ 2

## ตารางที่ 2 รายละเอียดข้อมูลนำเข้า

ลำดับ	รายการข้อมูล	คำอธิบายข้อมูล	ประเภทข้อมูล	รายการข้อมูล	หมายเหตุ
1	ชั้นปี	ชั้นปีที่กำลังศึกษา	ข้อความอักษร	ชั้นปีที่ 1 (B65) ชั้นปีที่ 2 (B64) ชั้นปีที่ 3 (B63)	“Year1” “Year2” “Year3”
2	เพศ	เพศ	ข้อความอักษร	เพศชาย เพศหญิง ผู้มีความหลากหลายทางเพศ (LGBTQ+)	“M” “F” “L”
3	เกรดเฉลี่ย	เกรดเฉลี่ยสะสม (GPAX) ในเทอมล่าสุด	ข้อความอักษร	0.00 – 2.00 2.01 – 2.75 2.76 - 3.25 3.26 – 3.75 3.76 – 4.00	“Poor” “Fair” “Good” “Very Good” “Excellent”
4	กลุ่มวิชาชีพ	กลุ่มวิชาชีพที่เลือก	ข้อความอักษร	DT - กลุ่มวิชาชีพด้านการจัดการข้อมูล (Data Handling Professionals) DT - กลุ่มวิชาชีพด้านซอฟต์แวร์วิสาหกิจ (Enterprise Software Professionals) DT - กลุ่มวิชาชีพด้านเทคโนโลยีคลาวด์ (Cloud Technology Professionals) DT - กลุ่มวิชาชีพด้านโปรแกรมประยุกต์บนมือถือ (Mobile Application Professionals) DT - กลุ่มวิชาชีพด้านโปรแกรมประยุกต์บนเว็บ (Web Application Professionals) DT - กลุ่มวิชาชีพด้านวิทยาศาสตร์ข้อมูล (Data Science Professionals) DC - กลุ่มวิชาชีพด้านการรายงานข่าวแบบเรียลไทม์ (Real-time Reporting Professionals) DC - กลุ่มวิชาชีพด้านการวิจัยสื่อดิจิทัล (Digital Media Research Professionals) DC - กลุ่มวิชาชีพด้านการออกแบบบิววล (Visual Design Professionals) DC - กลุ่มวิชาชีพด้านเกมคอมพิวเตอร์ (Computer Game Professionals) DC - กลุ่มวิชาชีพด้านเนื้อหาดิจิทัล (Digital Content Professionals) DC - กลุ่มวิชาชีพด้านโปรแกรมประยุกต์บนมือถือ (Mobile Application Professionals) DC - กลุ่มวิชาชีพด้านวีดิทัศน์ดิจิทัล (Digital Video Professionals)	“DT1” “DT2” “DT3” “DT4” “DT5” “DT6” “DC1” “DC2” “DC3” “DC4” “DC5” “DC6” “DC7”

ลำดับ	รายการข้อมูล	คำอธิบายข้อมูล	ประเภทข้อมูล	รายการข้อมูล	หมายเหตุ
				DC - กลุ่มวิชาชีพด้านแอนิเมชัน (Animation Professionals) DC - กลุ่มอาชีพด้านการวางแผนสื่อดิจิทัล (Digital Media Planning Professionals)	“DC8”  “DC9”
5	ปัจจัยด้านการประกอบอาชีพ (career1)	หางานทำได้ง่าย เมื่อสำเร็จการศึกษาในกลุ่มวิชาชีพนั้น	ข้อความอักษร	มากที่สุด มาก ปานกลาง น้อย น้อยที่สุด	“EXCELLENT” “GOOD” “OK” “NOT” “NOT GOOD”
6	ปัจจัยด้านการประกอบอาชีพ (career2)	มีโอกาสดำรงตำแหน่งสูง หากทำงานในกลุ่มวิชาชีพนั้น	ข้อความอักษร	มากที่สุด มาก ปานกลาง น้อย	“EXCELLENT” “GOOD” “OK” “NOT”
7	ปัจจัยด้านการประกอบอาชีพ (career3)	เป็นกลุ่มวิชาชีพที่สามารถหารายได้พิเศษได้หลายทาง	ข้อความอักษร	มากที่สุด มาก ปานกลาง น้อย	“EXCELLENT” “GOOD” “OK” “NOT”
8	ปัจจัยด้านการประกอบอาชีพ (career4)	มีแนวโน้มความต้องการในตลาดแรงงานสูงขึ้นในกลุ่มวิชาชีพนั้น	ข้อความอักษร	มากที่สุด มาก ปานกลาง น้อย น้อยที่สุด	“EXCELLENT” “GOOD” “OK” “NOT” “NOT GOOD”
9	ปัจจัยด้านการประกอบอาชีพ (career5)	เป็นกลุ่มวิชาชีพที่องค์กรให้ความสำคัญเป็นอย่าง	ข้อความอักษร	มากที่สุด มาก ปานกลาง น้อย	“EXCELLENT” “GOOD” “OK” “NOT”
10	ปัจจัยด้านการประกอบอาชีพ (career6)	มีตำแหน่งงานรองรับทั้งในภาครัฐบาลและเอกชน	ข้อความอักษร	มากที่สุด มาก ปานกลาง น้อย น้อยที่สุด	“EXCELLENT” “GOOD” “OK” “NOT” “NOT GOOD”
11	ปัจจัยด้านการประกอบอาชีพ (career7)	เป็นกลุ่มวิชาชีพที่สอดคล้องกับสถานะเศรษฐกิจปัจจุบันและอนาคต	ข้อความอักษร	มากที่สุด มาก ปานกลาง น้อย น้อยที่สุด	“EXCELLENT” “GOOD” “OK” “NOT” “NOT GOOD”
12	ปัจจัยด้านการประกอบอาชีพ (career8)	ความรู้ในกลุ่มวิชาชีพนั้นสามารถ	ข้อความอักษร	มากที่สุด มาก ปานกลาง	“EXCELLENT” “GOOD” “OK”

ลำดับ	รายการข้อมูล	คำอธิบายข้อมูล	ประเภทข้อมูล	รายการข้อมูล	หมายเหตุ
		นำไปประยุกต์ใช้งาน ได้เกือบทุกอาชีพ		น้อย น้อยที่สุด	“NOT” “NOT GOOD”
13	ปัจจัยด้านการ ประกอบอาชีพ (career9)	เป็นกลุ่มวิชาชีพที่ น่าจะสามารถสร้าง ความเจริญก้าวหน้า ได้อย่างรวดเร็ว	ข้อความอักษร	มากที่สุด มาก ปานกลาง น้อย น้อยที่สุด	“EXCELLENT” “GOOD” “OK” “NOT” “NOT GOOD”
14	ปัจจัยด้านการ ประกอบอาชีพ (career10)	เป็นกลุ่มวิชาชีพที่ สามารถสร้างความ มั่นคงในชีวิตได้	ข้อความอักษร	มากที่สุด มาก ปานกลาง น้อย น้อยที่สุด	“EXCELLENT” “GOOD” “OK” “NOT” “NOT GOOD”
15	ปัจจัยด้านการ ประกอบอาชีพ (career11)	เป็นกลุ่มวิชาชีพที่มี บรรยากาศและ สิ่งแวดล้อมในการ ทำงานอยู่ใน สำนักงาน	ข้อความอักษร	มากที่สุด มาก ปานกลาง น้อย น้อยที่สุด	“EXCELLENT” “GOOD” “OK” “NOT” “NOT GOOD”
16	ปัจจัยด้านการ ประกอบอาชีพ (career12)	เป็นกลุ่มวิชาชีพที่ เปิดโอกาสให้ได้ใช้ สติปัญญาและ ความคิดอยู่ ตลอดเวลา	ข้อความอักษร	มากที่สุด มาก ปานกลาง	“EXCELLENT” “GOOD” “OK”

**ขั้นตอนที่ 3 การเตรียมข้อมูลก่อนการวิเคราะห์ (Data Preparation)** โดยขั้นตอนนี้จะเริ่มจากการสำรวจ วิเคราะห์ข้อมูล (Exploratory Data Analysis) และเตรียมข้อมูลให้เหมาะสม โดยมีขั้นตอนในการจัดเตรียมข้อมูลก่อนนำไปวิเคราะห์ โดยมี 3 ขั้นตอนย่อย ดังนี้

ขั้นตอนที่ 1 การคัดเลือกข้อมูล (Data Selection) เลือกข้อมูลที่เกี่ยวข้อง โดยงานวิจัยนี้วิเคราะห์การเลือกอาชีพ จากปัจจัยการตอบแบบสอบถามด้านอาชีพ ว่าปัจจัยใดบ้างที่ส่งผลต่อการเลือกอาชีพด้านนั้น จึงได้ข้อมูลที่เป็นกลุ่มเป้าหมายที่ใช้วิเคราะห์ทั้งหมด 16 แอตทริบิวต์ จาก 262 แอตทริบิวต์ ดังตารางที่ 2

ขั้นที่ 2 การทำความสะอาดข้อมูล (Data Cleaning) เมื่อได้ข้อมูลที่เกี่ยวข้อง แล้วการทำความสะอาดข้อมูล พบว่าปัญหาการจำแนกข้อมูลที่ไม่สมดุล จึงใช้เทคนิคสโมต (Synthetic Minority Oversampling Technique: SMOTE) มาสังเคราะห์ข้อมูลเพิ่มในแอตทริบิวต์ กลุ่มวิชาชีพ และทำการลบรายการข้อมูลที่มีค่าน้อยและไม่ส่งผลต่อการนำไปวิเคราะห์ข้อมูลออกไป ส่งผลให้จากเดิมมีข้อมูลจำนวน 310 รายการ ซึ่งเมื่อปรับข้อมูลแล้วมีจำนวนเพิ่มขึ้นเป็นจำนวน 594 รายการ

ขั้นที่ 3 การแปลงรูปแบบของข้อมูล (Data Transformation) เมื่อทำการคัดเลือกข้อมูล และทำความสะอาดเรียบร้อยแล้วส่วนที่เกี่ยวข้องได้แล้ว ก่อนที่จะนำไปวิเคราะห์ได้นั้น ต้องทำให้ไฟล์ข้อมูลอยู่ในรูปแบบเดียวกัน โดยไฟล์ข้อมูลที่ได้รวบรวมมาอยู่ในรูปแบบไฟล์ (.xlsx) และเป็นข้อความภาษาไทยทั้งหมด ทำให้ไม่สามารถอ่านไฟล์ข้อมูลในโปรแกรมเวก้า (Weka) ได้ เมื่อพบปัญหาเหล่านี้แล้ว จึงทำการแปลงข้อมูลเป็นไฟล์ (.csv) และแทนค่าข้อมูล (Replace) จากภาษาไทยเป็นภาษาอังกฤษ โดยในงานวิจัยนี้มีการกำหนดป้ายกำกับให้กับคุณลักษณะที่คัดเลือกเพื่อการวิเคราะห์ ตามหมายเหตุในตารางที่ 2 ซึ่งเป็นการแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์ หรือทำเหมืองข้อมูล

**ขั้นตอนที่ 4 การสร้างตัวแบบจำลอง (Modeling)** เมื่อดำเนินการขั้นตอนการเข้าใจปัญหา การทำความเข้าใจข้อมูล การเตรียมข้อมูลเรียบร้อยแล้ว ผู้วิจัยจะทำการสร้างตัวแบบจำลองการจำแนกประเภทข้อมูล โดยเลือกใช้เทคนิคในการสร้างตัวแบบจำลองทั้งหมด 2 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ ด้วยชุดคำสั่งเจ48 และชุดคำสั่งการสุ่มป่าไม้ เทคนิคเพื่อนบ้านใกล้สุด ใช้ชุดคำสั่งไอบีเค ด้วยโปรแกรมเวก้า (Weka) และแบ่งข้อมูลออกเป็นส่วน ๆ (Cross-Validation) มีค่าเท่ากับ 20 โดยกำหนดค่าพารามิเตอร์ในแต่ละเทคนิค ดังนี้

เทคนิคต้นไม้ตัดสินใจ ด้วยชุดคำสั่ง เจ48 (J48) กำหนดค่าพารามิเตอร์ ดังนี้ ขนาดแบทช์ (batchSize) มีค่าเท่ากับ 100 แยกไบนารี (binarySplits) คือ True การล้มของต้นไม้ (collapseTree) คือ False ค่าตัวแปรความมั่นใจ (confidenceFactor) มีค่าเท่ากับ 0.25 เลขการแบ่งจำนวน(numFolds) มีค่าเท่ากับ 3

เทคนิคต้นไม้ตัดสินใจ ด้วยชุดคำสั่ง การสุ่มป่าไม้ (Random Forest) กำหนดค่าพารามิเตอร์ ดังนี้ ขนาดกระเป๋เปอร์เซ็นต์ (bagSizePercent) มีค่าเท่ากับ 100 ขนาดแบทช์ (batchSize) มีค่าเท่ากับ 100 ทำลายความสัมพันธ์แบบสุ่ม (breakTiesRandomly) เป็น False หน่วยลึกที่สุด (maxDepth) เป็น 0

เทคนิคเพื่อนบ้านใกล้สุด ใช้ชุดคำสั่ง ไอบีเค (IBK) กำหนดค่าพารามิเตอร์ การแบ่งข้อมูล(KNN) มีค่าเท่ากับ 1 ขนาดแบทช์ (batchSize) มีค่าเท่ากับ 100 การไขว้ชุดข้อมูล (crossValidate) เป็น False ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (meanSquared) เป็น False

**ขั้นตอนที่ 5 การวัดประสิทธิภาพของแบบจำลอง (Evaluation)** ทำการวัดประสิทธิภาพจากการประเมินแบบจำลอง โดยพิจารณาผลลัพธ์จากโปรแกรมเวก้า ดังนี้

- 1) ค่าความถูกต้อง (Accuracy) วัดค่าความถูกต้องโดยรวมของโมเดล
- 2) ค่าความแม่นยำ (Precision) วัดสัดส่วนของข้อมูลที่ถูกต้อง และตรงตามความต้องการ



- 3) ค่าความระลึก (Recall) วัดค่าสัดส่วนของข้อมูลที่ตรงตามความต้องการที่ถูกค้นคืนทั้งหมด
- 4) ค่าความถ่วงดุล (F-measure) วัดค่าเฉลี่ยค่าความแม่นยำในการตรวจพบและค่าความครบถ้วนในการตรวจพบเข้าด้วยกัน

**ขั้นตอนที่ 6 การนำผลลัพธ์ไปใช้งาน (Deployment)** ผลที่ได้จากการสร้างตัวแบบจำลองและเปรียบเทียบประสิทธิภาพของตัวแบบจำลองเรียบร้อยแล้ว สามารถนำไปพัฒนาเป็นระบบวิเคราะห์และเผยแพร่ข้อมูลปัจจัยหรือพฤติกรรมในการเลือกอาชีพของนักศึกษา เพื่อประโยชน์ในการพัฒนาหลักสูตรให้ตรงกับความต้องการของนักศึกษาในยุคปัจจุบัน

### ผลการดำเนินงาน

งานวิจัยนี้ผู้วิจัยต้องการศึกษาเรื่องการตัดสินใจปัจจัยที่เกี่ยวข้องกับการเลือกอาชีพของนักศึกษา โดยดูจากแนวโน้มการเลือกตอบแบบสอบถามปัจจัยด้านอาชีพ เพื่อเป็นแนวทางในการประกอบการตัดสินใจของนักศึกษาเอง จากผลการดำเนินงาน ผู้วิจัยได้นำข้อมูลมาวิเคราะห์ จากข้อมูลการรวบรวมข้อมูลนักศึกษาไว้ โดยมีจำนวน 311 รายการ แต่ทำการปรับค่าที่ไม่สมดุล ด้วยวิธีสโมตเพื่อสังเคราะห์ข้อมูลเพิ่ม ทำให้มีข้อมูลนำมาใช้สร้างตัวแบบจำลองทั้งหมด 594 รายการ และนำแอตทริบิวต์มาวิเคราะห์จำนวนทั้งสิ้น 16 แอตทริบิวต์ ตามตารางนำเข้าข้อมูลตารางที่ 2 มาสร้างตัวแบบจำลองทั้งหมด 2 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ ด้วยชุดคำสั่งเจ48 และชุดคำสั่งการสุ่มป่าไม้ เทคนิคเพื่อนบ้านใกล้สุด ใช้ชุดคำสั่งไอบีเค โดยใช้ตัววัดประสิทธิภาพตัวแบบจำลองจากการประเมินผลลัพธ์การทำนายเมทริกซ์ข้อผิดพลาด (Confusion matrix) จากค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล ได้ผลการทำนาย ดังตารางที่ 3

**ตารางที่ 3** ตารางเปรียบเทียบผลการวิเคราะห์ค่าความถูกต้องของตัวแบบจำลอง

ลำดับ	ตัวแบบจำลอง	ค่าความถูกต้อง(%)	ค่าความแม่นยำ(%)	ค่าความระลึก(%)	ค่าความถ่วงดุล(%)
1	การสุ่มป่าไม้	58.08	56.30	58.10	56.70
2	ไอบีเค	56.90	56.60	56.90	56.30
3	เจ48	55.39	54.60	55.40	54.80
	ผลรวมเฉลี่ย	56.79	55.83	56.80	55.93

จากตารางที่ 3 พบว่า ตัวแบบจำลองที่สร้างจากเทคนิคต้นไม้ตัดสินใจ ด้วยชุดคำสั่งการสุ่มป่าไม้ ให้ค่าความถูกต้องสูงที่สุด ซึ่งมีค่า 58.08% ค่าความแม่นยำ 56.30% ค่าความระลึก 58.10% ค่าความถ่วงดุล 56.70% ตัวแบบจำลองที่สร้างจากเทคนิคเพื่อนบ้านใกล้สุด ใช้ชุดคำสั่ง ไอบีเค ให้ค่าความถูกต้อง ซึ่งมีค่า 56.90% ค่าความแม่นยำ 56.60% ค่าความระลึก 56.90% ค่าความถ่วงดุล 56.30% และตัวแบบจำลองที่สร้างจากเทคนิคต้นไม้

ตัดสินใจ ด้วยชุดคำสั่งเจ48 ให้ค่าความถูกต้องน้อยที่สุด ซึ่งมีค่า 55.39% ค่าความแม่นยำ 54.60% ค่าความ  
 ระลึก 55.40% ค่าความถ่วงดุล 54.80% และแสดงตารางเมทริกซ์ข้อผิดพลาดในชุดคำสั่งต่าง ๆ ดังนี้

ตารางที่ 4 ตารางเมทริกซ์ข้อผิดพลาด (Confusion Matrix) ของเจ48

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	<-- classified as
17	6	3	2	4	1	1	5	7	1	2	2	a = DT2
7	11	5	3	5	6	1	6	1	1	5	1	b = DT5
2	2	22	2	4	4	5	4	0	4	2	0	c = DT4
0	2	1	41	1	1	0	2	0	1	0	0	d = DC4
5	6	7	1	14	3	1	2	3	0	2	4	e = DT1
2	8	2	1	1	27	1	0	1	0	3	2	f = DT6
1	2	2	0	2	2	32	2	4	0	0	1	g = DT3
1	6	4	5	1	1	0	23	1	1	4	2	h = DC7
2	1	2	2	2	0	0	2	34	1	2	2	i = DC5
2	1	2	0	1	1	0	0	0	39	0	2	j = DC3
2	5	3	1	0	0	1	3	2	1	30	2	k = DC8
0	0	1	1	1	0	1	4	0	0	3	39	l = DC9

ตารางที่ 5 ตารางเมทริกซ์ข้อผิดพลาด (Confusion Matrix) ของไอบีเค

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	<-- classified as
19	5	5	0	4	0	2	4	6	0	3	3	a = DT2
6	13	6	4	3	3	4	6	1	4	1	1	b = DT5
8	2	20	2	5	1	4	3	1	2	1	2	c = DT4
0	2	1	42	0	0	0	1	1	2	0	0	d = DC4
5	4	1	0	20	2	2	3	3	1	1	6	e = DT1
6	4	0	1	3	21	3	1	4	0	2	3	f = DT6
1	2	0	0	2	1	35	0	4	0	2	1	g = DT3
6	6	3	4	5	1	0	19	2	1	1	1	h = DC7
5	2	2	0	4	0	0	1	34	1	0	1	i = DC5
1	3	0	1	0	0	0	0	1	41	0	1	j = DC3
2	5	1	0	2	0	1	0	1	2	35	1	k = DC8
1	0	3	0	0	1	0	3	0	1	2	39	l = DC9

ตารางที่ 6 ตารางเมทริกซ์ข้อผิดพลาด (Confusion Matrix) ของการสุ่มป่าไม้

=== Confusion Matrix ===

	a	b	c	d	e	f	g	h	i	j	k	l		<-- classified as
18	5	6	0	3	3	2	2	6	0	4	2			a = DT2
4	11	5	4	3	7	3	8	1	3	2	1			b = DT5
3	2	23	3	4	3	3	3	2	2	1	2			c = DT4
0	1	2	42	1	0	0	2	0	1	0	0			d = DC4
3	3	0	3	19	5	1	4	2	1	3	4			e = DT1
3	6	0	1	3	18	3	2	5	0	3	4			f = DT6
1	2	1	0	0	2	35	1	3	0	3	0			g = DT3
5	1	3	4	3	1	0	26	2	1	2	1			h = DC7
2	2	2	2	3	1	0	0	34	1	2	1			i = DC5
0	1	0	0	1	0	1	2	0	40	2	1			j = DC3
1	2	0	1	2	0	0	1	1	3	39	0			k = DC8
1	0	2	0	0	0	0	4	0	1	2	40			l = DC9

## สรุปและอภิปรายผล

งานวิจัยนี้ผู้วิจัยต้องการศึกษาเรื่องการตัดสินใจปัจจัยที่เกี่ยวข้องกับการเลือกอาชีพของนักศึกษา โดยดูจากแนวโน้มการเลือกตอบคำถามปัจจัยด้านอาชีพ เลือกใช้เทคนิคในการสร้างตัวแบบจำลองทั้งหมด 2 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ ด้วยชุดคำสั่งเจ48 และชุดคำสั่งการสุ่มป่าไม้ เทคนิคเพื่อนบ้านใกล้สุด ใช้ชุดคำสั่งไอบีเค ซึ่งเป็นตัวแบบจำลองที่สอดคล้องกับข้อมูล และใช้ตัววัดประสิทธิภาพมาเปรียบเทียบตัวแบบจำลอง เพื่อหาปัจจัยที่ส่งผลต่อการเลือกอาชีพของนักศึกษา ในการทำเหมืองข้อมูลครั้งนี้ผู้วิจัยเลือกใช้โปรแกรมเวก้า ซึ่งเป็นโปรแกรมที่ใช้งานง่าย และมีหลายเทคนิคให้เลือกใช้ โดยในงานวิจัยนี้เลือกใช้เทคนิคต้นไม้การตัดสินใจ เพราะมีค่าความถูกต้องที่สูงที่สุด ซึ่งมีความสอดคล้องกับงานวิจัยของสำราญ วานนท์ และ ธรัช แสนราช (2561) ที่มีการศึกษาในเรื่องของการคาดการณ์อาชีพอนาคตของบัณฑิต ซึ่งใช้เทคนิคต้นไม้ตัดสินใจ และได้ค่าความถูกต้องสูงที่สุด นอกจากนี้งานวิจัยของรัชฎา เทพประสิทธิ์ และ จริญญา แสนราช (2562) ได้มีการวิเคราะห์ปัจจัยที่มีผลต่อการเลือกเรียนสาขาวิชาในคณะครุศาสตร์ โดยใช้เทคนิคต้นไม้การตัดสินใจ ซึ่งได้ค่าความถูกต้องสูงที่สุดเช่นเดียวกัน จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่า การจำแนกประเภทข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ นิยมใช้มากที่สุด โดยงานวิจัยการทำเหมืองข้อมูลนี้สามารถนำมาเป็นแนวทางในการเลือกอาชีพของนักศึกษา และสามารถนำผลมาประยุกต์ใช้ในการสร้างแผนการเรียนการสอนให้เหมาะสมกับความต้องการของนักศึกษาต่อไป

## ข้อเสนอแนะ/งานวิจัยในอนาคต

1. ข้อมูลวิจัยที่เก็บรวบรวมมาใช้ในการวิเคราะห์มีจำนวนน้อยเกินไป ดังนั้น ในอนาคตควรมีการรวบรวมข้อมูลให้มีจำนวนที่เพิ่มมากขึ้น เพื่อให้ได้ผลการวิเคราะห์ที่มีความสมบูรณ์มากยิ่งขึ้น
2. งานวิจัยนี้เก็บรวบรวมข้อมูลของผู้ตอบแบบสอบถามในแต่ละชั้นปีได้น้อย จึงทำให้มีข้อจำกัดในการจำแนกประเภทข้อมูลของนักศึกษาในแต่ละชั้นปี โดยนักศึกษาปีที่ 1 อาจจะยังไม่พบความต้องการที่แท้จริงในการตัดสินใจ ในอนาคตอาจจะเลือกผู้ตอบแบบสอบถามชั้นปีที่ 3 และ 4 แทน
3. งานวิจัยนี้ใช้เทคนิคสโมตในการปรับข้อมูลให้เท่ากันทุกๆคลาสของกลุ่มอาชีพเพื่อให้ได้ค่าความถูกต้องที่มากขึ้น หากนักวิจัยท่านใดพบเจอปัญหาสามารถนำเทคนิคนี้ไปประยุกต์ใช้ได้
4. คณบดีสำนักวิชาศาสตร์และศิลป์ดิจิทัลสามารถนำผลวิจัยนี้ไปเป็นแนวทาง ในการทำแบบสอบถามเพื่อประเมินตนเองสำหรับนักศึกษาที่ไม่รู้ว่าจะเลือกอาชีพไหน มาทำแบบสอบถามเพื่อดูว่าตนเองเหมาะกับอาชีพไหนเพื่อช่วยในการตัดสินใจ
5. นักวิจัยหรือองค์กรทางด้านการศึกษา สามารถนำกระบวนการในการสร้างแบบจำลองการตัดสินใจเลือกกลุ่มวิชา ไปใช้เป็นแนวทางในการสร้างแบบจำลองตัดสินใจอื่น ๆ ได้

## เอกสารอ้างอิง

ALEXANDRA, T. (2022). *What Is Data Mining? How It Works, Benefits, Techniques, and Examples*.

Retrieved from <https://www.investopedia.com/terms/d/datamining.asp>

Alfred L., (1932). *"Quantitative Expression of Cultural Relationships"*. เข้าถึงได้จาก

[www.pratya.nuankaew.com/wp-content/uploads/2019/01/2561\\_2\\_221203\\_02.pdf](http://www.pratya.nuankaew.com/wp-content/uploads/2019/01/2561_2_221203_02.pdf)

Author: tmtyai. (ม.ป.ป.). *ความน่าจะเป็น กฎทฤษฎีเบย์ (Bayes' Theorem)*. เข้าถึงได้จาก

<https://tuemaster.com/blog/%E0%B8%84%E0%B8%A7%E0%B8%B2%E0%B8%A1%E0%B8%99%E0%B9%88%E0%B8%B2%E0%B8%88%E0%B8%B0%E0%B9%80%E0%B8%9B%E0%B9%87%E0%B8%99%E2%80%8E-%E0%B8%81%E0%B8%8E%E0%B8%97%E0%B8%A4%E0%B8%A9%E0%B8%8E%E0%B8%B5%E0%B9%80/>

chengz. (2019, Oct3). *Confusion*. เข้าถึงได้จาก

<https://medium.com/@cheng3374/%E0%B8%A7%E0%B8%B1%E0%B8%94%E0%B8%9B%E0%B8%A3%E0%B8%B0%E0%B8%AA%E0%B8%B4%E0%B8%97%E0%B8%98%E0%B8%B4%E0%B8%A0%E0%B8%B2%E0%B8%9E-model-%E0%B8%88%E0%B8%B2%E0%B8%81-confusion-matrix-69d391bcd48#:~:text=Confusion%20Matrix%20%E0%B8>

Kong Ruksiam. (2020). *การตัดสินใจเลือกอาชีพด้วยโมเดลต้นไม้ตัดสินใจ*. เข้าถึงได้จาก

<https://lb.mol.go.th/%E0%B8%84%E0%B8%99%E0%B8%AB%E0%B8%B2%E0%B8%87%E0%B8%B2%E0%B8%99-%E0%B8%84%E0%B8%99%E0%B8%A7%E0%B9%88%E0%B8%B2%E0%B8%87%E0%B8%87%E0%B8%B2%E0%B8%99/%E0%B9%81%E0%B8%99%E0%B8%B0%E0%B9%81%E0%B8%99%E0%B8%A7%E0%B8%AD%E0%B8%B2%E0%B8%8A%E0%B8%B5%E0%B8%9E/%E0%B8%81%E0%B8%B2%E0%B8%A3%E0%B8%95%E0%B8%B1%E0%B8%94%E0%B8%AA%E0%B8%B4%E0%B8%99%E0%B9%83%E0%B8%88%E0%B9%80%E0%B8%A5%E0%B8%B7%E0%B8%AD%E0%B8%81%E0%B8%AD%E0%B8%B2%E0%B8%8A%E0%B8%B5%E0%B8%9E?fbclid=IwAR3Nb7qFCZMvbirEINVuTHNpXRlNj8jSPzfjvhpBT-8Za2hCPz-VHind4wg>

Mr. Automated. (2014). *เหมืองข้อมูล (data mining)*. เข้าถึงได้จาก [https://www.autosoft.in.th/data-](https://www.autosoft.in.th/data-platform/%E0%B9%80%E0%B8%AB%E0%B8%A1%E0%B8%B7%E0%B8%AD%E0%B8%87%E0%B8%82%E0%B9%89%E0%B8%AD%E0%B8%A1%E0%B8%B9%E0%B8%A5-data-mining/)

[platform/%E0%B9%80%E0%B8%AB%E0%B8%A1%E0%B8%B7%E0%B8%AD%E0%B8%87%E0%B8%82%E0%B9%89%E0%B8%AD%E0%B8%A1%E0%B8%B9%E0%B8%A5-data-mining/](https://www.autosoft.in.th/data-platform/%E0%B9%80%E0%B8%AB%E0%B8%A1%E0%B8%B7%E0%B8%AD%E0%B8%87%E0%B8%82%E0%B9%89%E0%B8%AD%E0%B8%A1%E0%B8%B9%E0%B8%A5-data-mining/)

NICK HOTZ. (2023). *กระบวนการวิเคราะห์ข้อมูลด้วยคริสปี้. ดีเอ็ม (Cross Industry Standard Process*

- for Data Mining: CRISP DM). เข้าถึงได้จาก  
<https://datacube.ai/crisp-dm>
- Ninenox. (2563). *ค่าความแม่นยำ*. เข้าถึงได้จาก  
<http://www.ninenox.com/2020/09/24/%E0%B8%97%E0%B8%B3%E0%B8%84%E0%B8%A7%E0%B8%B2%E0%B8%A1%E0%B9%80%E0%B8%82%E0%B9%89%E0%B8%B2%E0%B9%83%E0%B8%88-accuracyprecisionrecallf1-score/>
- Pradyasin. (2019). *Random Forest คืออะไร*. เข้าถึงได้จาก  
<https://medium.com/@pradyasin/random-forest-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-74d2a0af3d7>
- Thaiall.com. (2558). *การวัดประสิทธิภาพ*. เข้าถึงได้จาก  
[http://www.thaiall.com/research/efficiency.htm#:~:text=%E0%B8%84%E0%B9%88%E0%B8%B2%E0%B8%84%E0%B8%A7%E0%B8%B2%E0%B8%A1%E0%B9%81%E0%B8%A1%E0%B9%88%E0%B8%99%E0%B8%A2%E0%B8%B3%20\(recall\)%20%E0%B8%84%E0%B8%B7%E0%B8%AD,Recall%20%2F%20\(Precision%20%2B%20Recall\)](http://www.thaiall.com/research/efficiency.htm#:~:text=%E0%B8%84%E0%B9%88%E0%B8%B2%E0%B8%84%E0%B8%A7%E0%B8%B2%E0%B8%A1%E0%B9%81%E0%B8%A1%E0%B9%88%E0%B8%99%E0%B8%A2%E0%B8%B3%20(recall)%20%E0%B8%84%E0%B8%B7%E0%B8%AD,Recall%20%2F%20(Precision%20%2B%20Recall))
- กมลรัตน์ สมใจ. (2563). *การเปรียบเทียบประสิทธิภาพโมเดลการตัดสินใจเลือกกลุ่มวิชาของนักศึกษา สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์. มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย, จังหวัดนครศรีธรรมราช*. เข้าถึงได้จาก  
<https://dspace.bru.ac.th/xmlui/bitstream/handle/123456789/6006/Template%20NSCIC%202020-%E0%B8%81%E0%B8%A1%E0%B8%A5%E0%B8%A3%E0%B8%B1%E0%B8%95%E0%B8%99%E0%B9%8C-%E0%B9%81%E0%B8%81%E0%B9%89%E0%B9%84%E0%B8%82.pdf?sequence=1&isAllowed=y>
- กรมการจัดหางาน. (2561). *การเปลี่ยนแปลงของตลาดแรงงานในยุคดิจิทัล. กองบริหารข้อมูลตลาดแรงงาน*. เข้าถึงได้จาก  
[https://www.doe.go.th/prd/assets/upload/files/lmia\\_th/c96a3a1c552a7148c0aba5f5959214e9.pdf](https://www.doe.go.th/prd/assets/upload/files/lmia_th/c96a3a1c552a7148c0aba5f5959214e9.pdf)
- กรมจัดหางาน. (2561). *คลังความรู้ด้านอาชีพ*. เข้าถึงได้จาก  
<https://www.doe.go.th/prd/main/knowledge/param/site/1/cat/94/sub/0/pull/category/view/cover-view>
- กรมแรงงาน. (2555). *บทความการสมัครอาชีพ*. เข้าถึงได้จาก บทความวิชาการ (labour.go.th)

กรุงเทพธุรกิจ. (2565). *ส่องทิศทางตลาดงานปี65 กับ10 อันดับ "สายงาน" ที่เป็นที่ต้องการของตลาด*. เข้าถึงได้จาก [https://www.bangkokbiznews.com/social/983267?fbclid=IwAR0sh92ZTHPpTGQyCi-zNvRwLniWod1pniqfC8-F5MQ8BB\\_\\_f6it-o3XxiE](https://www.bangkokbiznews.com/social/983267?fbclid=IwAR0sh92ZTHPpTGQyCi-zNvRwLniWod1pniqfC8-F5MQ8BB__f6it-o3XxiE)

ทวีผล.(2555). accuracy ค่าความถูกต้อง ความแม่นยำ. เข้าถึงได้จาก <https://www.foodnetworksolution.com/wiki/word/4289/accuracy-%E0%B8%84%E0%B8%A7%E0%B8%B2%E0%B8%A1%E0%B8%96%E0%B8%B9%E0%B8%81%E0%B8%95%E0%B9%89%E0%B8%AD%E0%B8%87-%E0%B8%84%E0%B8%A7%E0%B8%B2%E0%B8%A1%E0%B9%81%E0%B8%A1%E0%B9%88%E0%B8%99%E0%B8%A2%E0%B8%B3>

พุทธิพร ธนธรรมเมธี และ ยาวเรศ ศิริสฤตย์กุล. (2562). *เทคนิคการจำแนกข้อมูลทีพัฒนาสำหรับชุดข้อมูลที่ไม่สมดุลของภาวะข้อเข่าเสื่อมในผู้สูงอายุ*. หลักสูตรวิศวกรรมซอฟต์แวร์ สำนักวิชาสารสนเทศศาสตร์ มหาวิทยาลัยวลัยลักษณ์. ตำบลไทยบุรี อำเภอท่าศาลา จังหวัดนครศรีธรรมราช 80160 : วารสารวิทยาศาสตร์และเทคโนโลยี. เข้าถึงได้จาก <https://li01.tci-thaijo.org/index.php/tstj/article/download/205346/143001/635829>

(ม.ป.ป.). *การจำแนก (Classification)*. (2561)เข้าถึงได้จาก <https://csit.nu.ac.th/kraisak/ds/ds/chapter05/Chapter05.pdf>

(ม.ป.ป.). *ทฤษฎี แนวคิดและงานวิจัยที่เกี่ยวข้อง*. มหาวิทยาลัยเชียงใหม่. เข้าถึงได้จาก [https://archive.lib.cmu.ac.th/full/T/2555/econ30455tr\\_ch2.pdf](https://archive.lib.cmu.ac.th/full/T/2555/econ30455tr_ch2.pdf)

มหาวิทยาลัยเทคโนโลยีสุรนารี. (2563). *อาชีพแนะนำของคณะศาสตร์และศิลปบัณฑิต* เข้าถึงได้จาก <https://digitech.sut.ac.th>

รัชฎา เทพประสิทธิ์ และ จริญญา แสนราช. (2562). *การวิเคราะห์ปัจจัยที่มีผลต่อการเลือกสาขาวิชาของนักศึกษา ระดับปริญญาตรี คณะครุศาสตร์*. สาขาวิชาคอมพิวเตอร์และเทคโนโลยีการศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏเชียงราย. เข้าถึงได้จาก <https://so02.tci-thaijo.org/index.php/JournalGradVRU/article/view/148746/163892>

สำราญ วานนท์ และ ธรัช อารีราษฎร์, จริญญา แสนราช. (2561). *การศึกษาเทคนิคพยากรณ์อาชีพสำหรับ นักศึกษาระดับปริญญาตรีสาขาคอมพิวเตอร์*. สาขาวิชาการจัดการเทคโนโลยี คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. เข้าถึงได้จาก <https://ph02.tci-thaijo.org/index.php/itm-journal/article/download/140249/104035/>

สำนักงานปลัดกระทรวงแรงงาน. (2552). *การตัดสินใจเลือกอาชีพ*. ถนนมิตรไมตรี แขวงดินแดง เขตดินแดง กรุงเทพมหานคร 10400: e-Labour บริการประชาชน. เข้าถึงได้จาก

โสธยา สุภาพล,สุวรรณณี หงษ์วิจิตร, สำราญ ถัดดาวัลย์, ชุติพร ลักษณะพิพัฒน์, และ เพ็ญพิชชา โพธากุล. (2563). ปัจจัยสภาพแวดล้อมในการทำงานและปัจจัยการทำงานเป็นทีมที่ส่งผลต่อประสิทธิภาพในการทำงานของพนักงานโรงพยาบาลแห่งหนึ่งในจังหวัดสุพรรณบุรี. มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ. วารสาร มจร สังคมศาสตร์ปริทรรศน์ Journal of MCU Social Science Review. เข้าถึงได้จาก jomcusocuserGroup018-211-225.pdf

สัญญา พันธุ์แพง. (2562). การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลสำหรับการพยากรณ์การศึกษาต่อนักศึกษาใหม่ระดับปริญญาตรี ในมหาวิทยาลัยราชภัฏเชียงใหม่. เข้าถึงได้จาก <http://backoffice.thaiedresearch.org/uploads/paper/0a43e6343fdb09aeadb31d9dad0718f6.pdf>

สถาบันนวัตกรรมและธรรมาภิบาลข้อมูล. (2565). เข้าใจใน 5 นาที! Classification Model คืออะไร. วิทยาศาสตร์ เทคโนโลยีดิจิทัลและนวัตกรรม. เข้าถึงได้จาก <https://digi.data.go.th/blog/what-is-classification-model/>

สุกฤษณ์ สีสุทอง. (2559). การพัฒนาแบบจำลองการตัดสินใจในการเลือกกลุ่มวิชาสำหรับนักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ. มหาวิทยาลัยเทคโนโลยีสุรนารี. เข้าถึงได้จาก <http://sutir.sut.ac.th:8080/jspui/bitstream/123456789/7074/1/Fulltext-watermark.pdf>

สมฤทัย กลัดแก้ว, อรไท ชั่วเจริญ และ ชำนาญ เจริญรุ่งเรือง. (2557). ระบบสนับสนุนการตัดสินใจการเลือกตำแหน่งงานให้สอดคล้องกับความสามารถของบัณฑิต. เข้าถึงได้จาก 2-1435149622\_AbstractBeer.pdf (ru.ac.th)

สุรพงศ์ เอื้อวัฒนามงคล. (2557). การทำเหมืองข้อมูล (Data Mining). เข้าถึงได้จาก <http://backoffice.thaiedresearch.org/uploads/paper/0a43e6343fdb09aeadb31d9dad0718f6.pdf>

หนึ่งทัย ชัยอากร. (2559). การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมนิง. เข้าถึงได้จาก <https://erp.mju.ac.th/acticleDetail.aspx?qid=551>

เอกสิทธิ์ พัทธวงศ์ศักดา. (2557). กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM และตัวอย่างการประยุกต์ใช้ทาง



ทางด้านการศึกษา. เข้าถึงได้จาก

<http://backoffice.thaiedresearch.org/uploads/paper/0a43e6343fdb09aeadb31d9dad0718f6.pdf>อาจารย์ปรัชญา นวนแก้ว. (2558). เทคโนโลยีเพื่องานประยุกต์ทางธุรกิจ. เข้าถึงได้จาก

[http://www.pratya.nuankaew.com/wp-content/uploads/2019/01/2561\\_2\\_221203\\_02.pdf](http://www.pratya.nuankaew.com/wp-content/uploads/2019/01/2561_2_221203_02.pdf)

อนุพงศ์ สุขประเสริฐ. (2561). *Classification Performance*. คณะบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม. เข้าถึงได้จาก

<https://slideplayer.in.th/slide/14964390/>

อรนุช พันโท, และ มนต์ชัย เทียนทอง. (2556). *การเปรียบเทียบประสิทธิภาพการจำแนกรูปแบบการเรียนรู้ VARK*. คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. 1518 ถนนประชากรราษฎร์ 1 แขวงวงศ์สว่าง เขตบางซื่อ กรุงเทพฯ 10800: วารสาร เทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏอุบลราชธานี. เข้าถึงได้จาก

<http://www.cmruir.cmru.ac.th/bitstream/123456789/423/1/58300-136165-1-SM.pdf>