

การจำแนกผู้สมัครที่ลังเลในการตัดสินใจศึกษาต่อเพื่อสร้างกลยุทธ์สนับสนุน  
การตัดสินใจให้เข้าศึกษาต่อ กรณีศึกษามหาวิทยาลัยราชภัฏจันทรเกษม  
Identification of Applicants Who are Hesitant to Decide to Study  
in Order to Create Strategies to Support Decision-making for  
Admission Case Study of Chandrakasem Rajabhat University

ชัยศิริ สนิทพลกลาง<sup>1\*</sup> และ สัตว์วิภูณ์ คำศิริ<sup>1</sup>

Chaisiri Sanitphonklang<sup>1</sup> and Sattawat Khamisiri<sup>1</sup>

<sup>1</sup>สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏจันทรเกษม

<sup>1</sup>Computer Science Program, Faculty of Science, Chandrakasem Rajabhat University

วันที่ส่งบทความ : 26 มกราคม 2564 วันที่แก้ไขบทความ : 12 เมษายน 2564 วันที่ตอบรับบทความ : 23 กรกฎาคม 2564

Received: 26 January 2021, Revised: 12 April 2021, Accepted: 23 July 2021

### บทคัดย่อ

ปัจจุบันนักเรียนเข้าศึกษาต่อระดับอุดมศึกษามีอัตราการลดลงอย่างต่อเนื่องทุกปี ในทางตรงข้ามจำนวนที่มหาวิทยาลัยประกาศรับนักศึกษาใหม่ยังคงเดิมซึ่งมีแนวโน้มสูงกว่าจำนวนผู้สมัคร ก่อให้เกิดการแข่งขันของมหาวิทยาลัยทั้งภาครัฐและเอกชนด้วยกลยุทธ์ที่แตกต่างกันในแต่ละสถาบัน การวิจัยที่นำเสนอในครั้งนี้เป็นอีกหนึ่งกลยุทธ์ โดยการค้นหาผู้ลังเลในการตัดสินใจศึกษาต่อเพื่อนำเข้าสู่กลยุทธ์โน้มน้าวใจ ด้วยเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) จากอัลกอริทึมต้นไม้ตัดสินใจเพื่อสกัดหาผู้ลังเลในการตัดสินใจศึกษาต่อ การเรียนรู้ของเครื่องงานวิจัยนี้ใช้ข้อมูลผู้สมัครในอดีตของมหาวิทยาลัยราชภัฏจันทรเกษมจำนวน 6,000 ระเบียบ จาก พ.ศ. 2560 ถึง พ.ศ. 2562 ได้ค้นพบว่าแอตทริบิวต์ผู้สมัครที่สามารถจำแนกกลุ่มผู้ลังเลในการตัดสินใจศึกษาต่อโดยเรียงลำดับตามความสำคัญตามกฎของต้นไม้ดังนี้ สาขาวิชา เกรดเฉลี่ย รายได้ครอบครัว ภูมิลำเนา ประเภทโรงเรียน และเพศ ซึ่งมีประสิทธิภาพในการจำแนกถูกต้องสูงถึง 95.89% เมื่อเทียบกับ อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) อัลกอริทึมเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors: KNN) และอัลกอริทึมเทคนิคนาอิวเบย์ (Naive Bayes) ที่มีประสิทธิภาพการจำแนกเป็น 87.67%, 85.50% และ 82.63% ตามลำดับ

**คำสำคัญ :** การเรียนรู้ของเครื่อง ต้นไม้ตัดสินใจ การจำแนก ผู้สมัครที่ลังเลใจ

---

\*ที่อยู่ติดต่อ E-mail address: chaisiri.s@chandra.ac.th

## Abstract

Nowadays, The rate of admission to higher education was reduced but contrary to the number that the university has announced. Therefore, there is a competition for recruiting students to study in universities, both public and private sectors, with different strategies for each institution. The research was presented this time is another strategy. Finding those who are reluctant to choose to study with a university as a way of persuasion By using machine learning techniques with algorithms, the decision tree was extracted from the applicants to the university. This machine learning uses 6,000 records of past applicants from Chandrakasem Rajabhat University from 2017 to 2019. It was found that the qualifications of applicants that could be classified into a reluctant group were application field, GPA, family income, domicile, school and gender. The classification efficiency was 95.89% compared to SVM, KNN, Naive Bayes with classification efficiency of 87.67%, 85.50%, and 82.63%, respectively.

**Keywords:** Machine learning, Decision tree, Classification, Hesitant candidate

### 1. บทนำ

อัตราการเกิดประชากรไทยเฉลี่ยอยู่ที่ 1.5 คนต่อผู้หญิงที่ให้กำเนิดหนึ่งคนในช่วงสิบปีหลังที่ผ่านมา เมื่อเทียบกับยี่สิบปีก่อนหน้าอัตราการเกิดเฉลี่ยอยู่ที่ 3.39 คนต่อผู้หญิงที่ให้กำเนิดหนึ่งคน [1] กลุ่มประชากรเหล่านี้ในอนาคตต้องเข้าศึกษาต่อในระดับอุดมศึกษาของประเทศไทยเพื่อศึกษาหาความรู้ไปสู่การพัฒนาประเทศชาติ และสถาบันระดับอุดมศึกษาของประเทศไทยถือกำเนิดขึ้นมาได้ระยะเวลาเร็วกว่าปี โดยมีการเจริญเติบโตตามยุคตามสมัยเรื่อยมาจนถึงปัจจุบัน มหาวิทยาลัยในประเทศไทยมีทั้งสิ้น 175 สถาบัน [2] ประกอบไปด้วยกลุ่มดังต่อไปนี้ 1) สถาบันอุดมศึกษาในกำกับของรัฐ 26 สถาบัน 2) สถาบันอุดมศึกษาของรัฐ 11 สถาบัน 3) มหาวิทยาลัยราชชมงคล 9 สถาบัน 4) มหาวิทยาลัยราชภัฏ 38 สถาบัน 5) วิทยาลัยชุมชน 20 สถาบัน 6) มหาวิทยาลัยเอกชน 42 สถาบัน 7) วิทยาลัยเอกชน 18 สถาบัน และ 8) สถาบันเอกชน 11 สถาบัน กระจายอยู่ทุกภูมิภาคของประเทศไทย ในแต่ละปีการศึกษาสามารถรับนักศึกษาได้ประมาณ 4 แสนคนต่อปีการศึกษา แต่มีผู้สมัครเข้าศึกษาต่อประมาณ 3 แสนคนต่อปีการศึกษา ถือว่าน้อยกว่าความต้องการของสถาบันการศึกษาที่ประกาศรับ สรุปได้ว่าจำนวนนักศึกษาในหลายสถาบันมีจำนวนผู้สมัครลดลงส่งผลให้หลายหลักสูตรมีจำนวนนักศึกษาไม่เป็นไปตามเป้าที่วางไว้ ซึ่งมีผลกระทบหลายด้านต่อสถาบันการศึกษา อาทิเช่น ด้านการประเมินคุณภาพการศึกษาในเรื่องอัตราส่วนบุคลากรต่อจำนวนนักศึกษา การเปิดและปิดหลักสูตรที่มีจำนวนนักศึกษาน้อยเพราะส่งผลกระทบต่อต้นทุนการผลิตบัณฑิต หรือการปรับโครงสร้างหลักสูตรให้ตอบสนองความต้องการของผู้เรียนมากขึ้น ขึ้นร้ายแรงอาจมีการปิดตัวของมหาวิทยาลัยอันเกิดจากจำนวนนักศึกษาที่ลดลงทำให้รายได้ของมหาวิทยาลัยไม่คล่องตัวในการบริหารการเรียนการสอน จากผลกระทบดังกล่าวหลายสถาบันเล็งเห็นถึงความสำคัญของปัญหาจึงหา กล

ยุทธ์ต่าง ๆ เข้ามาช่วยแก้ปัญหา เช่น การออกประชาสัมพันธ์ตามโรงเรียนโดยใช้ชั่วโมงแนะแนวของทางโรงเรียน การออกบูธประชาสัมพันธ์ตามงานมหกรรมตลาดวิชาการในแต่ละภูมิภาค การประชาสัมพันธ์ผ่านช่องทางสังคมออนไลน์ต่าง ๆ เช่น เฟซบุ๊ก ไลน์แอปพลิเคชัน ฯลฯ การส่งประกาศพร้อมใบสมัครไปตามโรงเรียนทั่วประเทศ กลยุทธ์เหล่านี้เป็นการประชาสัมพันธ์ในลักษณะแบบวงกว้างไม่เจาะจงเป็นรายบุคคล ซึ่งผลลัพธ์ที่ได้เมื่อเทียบกับต้นทุนในการประชาสัมพันธ์ถือว่าไม่สมดุล

การวิจัยครั้งนี้ได้คิดค้นเครื่องมือเพื่อเข้ามาช่วยค้นหาและเจาะจงเข้าไปยังกลุ่มเป้าหมายแบบรายบุคคล [3] เพื่อการประชาสัมพันธ์แบบโน้มน้าวจิตใจเป็นรายบุคคลและตรงจุดมากขึ้น ให้ผลลัพธ์ที่ดีขึ้นกว่ากลยุทธ์อื่นที่กล่าวมาแล้วข้างต้น ด้วยการใช้เทคโนโลยีที่เรียกว่าปัญญาประดิษฐ์ หรือ เอไอ (Artificial intelligence: AI) สำหรับสอนเครื่องให้เรียนรู้ข้อมูลผู้สมัครเข้าศึกษาต่อในมหาวิทยาลัยราชภัฏจันทรเกษม ด้วยอัลกอริทึมต้นไม้ตัดสินใจ การเรียนรู้ครั้งนี้ใช้ข้อมูลในการเรียนรู้ทั้งสิ้น 6,000 รายการ เป็นข้อมูลผู้สมัครเข้าศึกษาาระหว่าง พ.ศ. 2560 ถึง พ.ศ. 2562 ที่ประกอบไปด้วยแอตทริบิวต์ของผู้สมัครดังนี้ ชื่อ-สกุล คณะ สาขา ที่อยู่ เพศ ประเภทโรงเรียน เกรดเฉลี่ย อาชีพบิดา อาชีพมารดา รายได้ครอบครัว ภูมิลำเนา เบอร์ติดต่อ ชื่อเฟซบุ๊ก และไอทีไลน์ โดยผ่านขั้นตอนการสร้างการเรียนรู้ของเครื่อง 5 ขั้นตอน [4] ดังต่อไปนี้ ทำความเข้าใจข้อมูล จัดเตรียมข้อมูล แปลงข้อมูล สร้างโมเดล และวัดประสิทธิภาพ ระหว่างการทดลองได้ทำการลดมิติของข้อมูลลงจากแอตทริบิวต์ 14 แอตทริบิวต์ ปรับลดให้เหลือเพียง 6 แอตทริบิวต์ แต่ประสิทธิภาพในการจำแนกมีความถูกต้องสูงถึง 95.89% หลังจากนั้นนำข้อมูลเดียวกันที่ผ่านการลดขนาดแอตทริบิวต์ทำการทดลองเปรียบเทียบกับอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน อัลกอริทึมเพื่อนบ้านใกล้ที่สุด และอัลกอริทึมเทคนิคนาอิวเบย์ ปรากฏว่าประสิทธิภาพการจำแนกได้ถูกต้องสูงสุดคือ อัลกอริทึมต้นไม้ตัดสินใจ และนำไปใช้ในการจำแนกผู้สมัครที่ลงทะเบียนในการตัดสินใจเข้าศึกษาต่อในระบบรับสมัครผู้เข้าศึกษาต่อ โดยสร้างเป็นเครื่องมือจำแนกในรูปแบบของเว็บไซต์

ในงานวิจัยนี้แบ่งเนื้อหาออกเป็น 5 ส่วน ประกอบไปด้วยส่วนแรกบทนำ ส่วนที่สองทบทวนวรรณกรรมที่เกี่ยวข้อง ส่วนที่สามขั้นตอนดำเนินงาน ส่วนที่สี่ผลการดำเนินงานและขั้นตอนสุดท้ายสรุปผล โดยสี่ส่วนหลังนี้ถูกนำเสนอในหัวข้อถัดไป

## 2. ทบทวนวรรณกรรมที่เกี่ยวข้อง

การจำแนกผู้ลงทะเบียนในการตัดสินใจเข้าศึกษาต่อเพื่อสร้างกลยุทธ์สนับสนุนการตัดสินใจเข้าศึกษาต่อกรณีศึกษามหาวิทยาลัยราชภัฏจันทรเกษม ผู้วิจัยทำการศึกษาและค้นคว้าข้อมูลต่าง ๆ ที่เกี่ยวข้องเพื่อใช้ประกอบในการจัดทำงานวิจัยนี้ โดยมีรายละเอียดดังนี้

### 2.1 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่องเป็นการศึกษาอัลกอริทึมคอมพิวเตอร์ที่ปรับปรุงโดยอัตโนมัติผ่านประสบการณ์ [5]-[6] การใช้งานเริ่มจากกระบวนการค้นพบความสัมพันธ์ของข้อมูลจากชุดข้อมูลขนาดใหญ่ไปจนถึงระบบกรองข้อมูลที่สนใจสำหรับการเรียนรู้แบบอัตโนมัติ การเรียนรู้ของเครื่องเป็นส่วนหนึ่งของปัญญาประดิษฐ์ โดยอาศัยอัลกอริทึมสร้างแบบจำลองทางสถิติและคณิตศาสตร์จากข้อมูลตัวอย่างหรือที่เรียกว่าข้อมูลสอน (Data Training Set) เพื่อให้เครื่องเรียนรู้แล้วนำไปคาดการณ์หรือตัดสินใจได้อย่าง

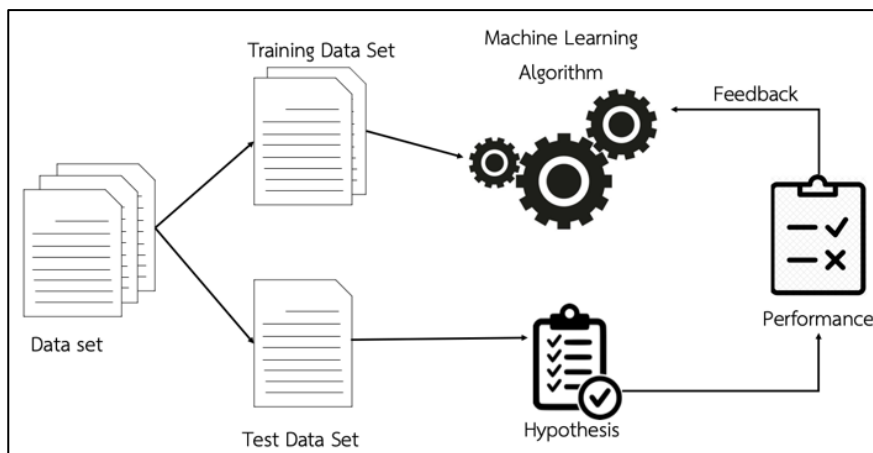
แม่นยำ หัวใจสำคัญของการเรียนรู้เครื่องคือการสร้างโมเดลเรียนรู้กับข้อมูลให้ครอบคลุมกรณีหรือเหตุการณ์ต่าง ๆ ให้มากที่สุดในเรื่องที่กำลังเรียนรู้เพื่อส่งผลให้เครื่องสามารถพยากรณ์หรือทำงานกับตัวอย่างข้อมูลที่ไม่เคยเห็นมาก่อนได้อย่างแม่นยำมากขึ้น การเรียนรู้ของเครื่องสามารถแบ่งออกเป็นประเภทได้ดังนี้

1) การเรียนรู้แบบมีผู้สอน (Supervised Learning) คือ การสอนเครื่องด้วยการนำเข้าข้อมูลที่มีป้ายกำกับ (Label) และไว้กับชุดข้อมูลสำหรับเรียนรู้เพื่อหาความเชื่อมโยงข้อมูลขาเข้าและขาออก

2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) คือ การเรียนรู้ของเครื่องจากข้อมูลที่ไม่มีการป้ายกำกับ (Non Label) โดยให้เครื่องหาโครงสร้างของข้อมูลขาเข้าเองโดยอาศัยความคล้ายคลึงกันของข้อมูล

3) การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) คือ เครื่องมีปฏิสัมพันธ์กับสิ่งแวดล้อมที่เปลี่ยนแปลงตลอดเวลา ซึ่งเครื่องทำงานบางอย่างที่ปราศจากผู้สอนที่คอยบอกหรือที่เรียกว่าการสังเกตและเลียนแบบ

การเรียนรู้ของเครื่องสัมพันธ์กับการสร้างแบบจำลอง ซึ่งเกิดจากการสอนข้อมูลช่วยสอนและสามารถพยากรณ์ข้อมูลที่เข้ามาใหม่ได้ มีด้วยกันหลายรูปแบบที่ถูกนำมาประยุกต์ใช้ในปัจจุบัน อาทิเช่น ระบบทำนายระดับความเครียดด้วยเทคนิคต้นไม้ตัดสินใจที่มีประสิทธิภาพในการทำนายระดับความเครียดได้แม่นยำสูงถึง 91.10% [7] แบบจำลองแบบผสมเพื่อสร้างกฎสำหรับคำถาม-คำตอบ โดยใช้เทคนิคการทำเหมืองข้อมูลที่ใช้ตัวแบบวิธีการเพื่อนบ้านที่ใกล้ที่สุดร่วมกับอัลกอริทึมแอฟริอริเพื่อหาความสัมพันธ์ระหว่างคำถามที่ถูกจำแนกกลุ่มหาคำตอบที่เหมาะสมกับคำถามโดยมีประสิทธิภาพในการจำแนกถูกต้องสูงถึง 91.43% [8] เป็นต้น กระบวนการในการเรียนรู้ของเครื่องได้แสดงไว้ดังรูปที่ 1



รูปที่ 1. กระบวนการเรียนรู้ของเครื่อง

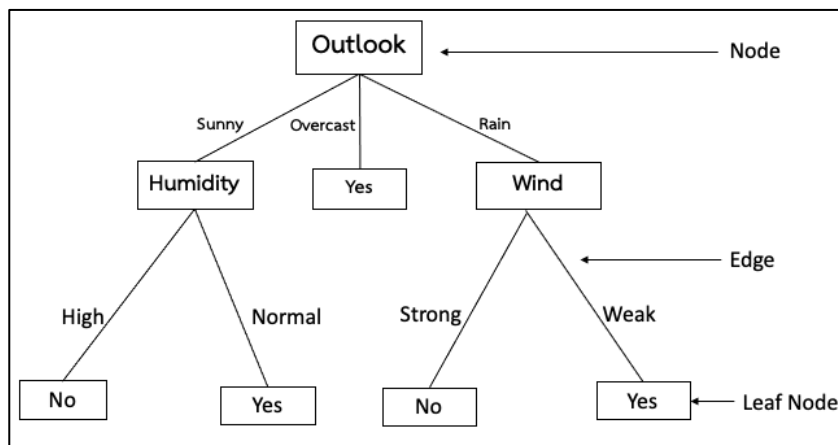
## 2.2 การจำแนกประเภทข้อมูล (Classification)

การจำแนกประเภทข้อมูลเป็นเทคนิคหนึ่งของเหมืองข้อมูล (Data Mining) โดยข้อมูลตัวอย่างถูกแบ่งออกจากข้อมูลทั้งหมดในอัตราส่วน 80% [9] ข้อมูลชุดนี้เรียกว่า ข้อมูลสำหรับสอน (Training Data Set) โดยแต่ละระเบียนของข้อมูลต้องมีป้ายกำกับ (Label) ซึ่งป้ายกำกับเป็นหัวใจของการจำแนกประเภท

ข้อมูล ป้ายกำกับถูกใช้สอนเครื่องทำการเรียนรู้เพื่อนำไปใช้กับข้อมูลใหม่ที่เครื่องยังไม่เคยรู้จักมาก่อนว่าต้องจำแนกให้อยู่ในกลุ่มใด การจำแนกข้อมูลใหม่อาศัยหลักการวัดระยะห่างบนระนาบสองมิติว่าใกล้เคียงกับพิกัดใดที่สุดแล้วจัดเข้ากลุ่มข้อมูลชุดนั้น สำหรับการจำแนกข้อมูลในส่วนข้อมูลให้ทำการพิจารณาและใส่ใจเป็นอย่างมาก ถ้าข้อมูลที่นำมาใช้ไม่มีความยุติธรรมจะส่งผลให้เกิดปรากฏการณ์จำแนกข้อมูลที่มีความพอดีเกินไป (Over Fitting) จากข้อมูลสำหรับสอน ซึ่งทำให้เครื่องจำแนกถูกต้องกับข้อมูลสำหรับสอนเท่านั้น แต่ใช้ไม่ได้กับข้อมูลทั่วไป การจำแนกประเภทข้อมูลได้ถูกนำมาใช้กับการสร้างโมเดลสำหรับการเรียนรู้ของเครื่องอย่างแพร่หลายดังเห็นได้จากงานวิจัยการเปรียบเทียบประสิทธิภาพการวิเคราะห์ความพึงพอใจเกี่ยวกับการจัดการเรียนการสอนด้วยกระบวนการวิเคราะห์ความรู้สึกโดยใช้เทคนิคการรวมกลุ่มเพื่อจำแนกข้อมูล [10] ของ พิชญะ พรหมลา และจรัญ แสนราช โดยผลการวิจัยพบว่าการจำแนกกลุ่มแบบ Vote มีประสิทธิภาพมากที่สุด และงานวิจัย การเปรียบเทียบเทคนิคการสุ่มตัวอย่างเพื่อการจำแนกข้อมูล ที่ไม่สมดุลของ กาญจน์ ฒ ศรีระ และคณะ [11] โดยมีผลสรุปของงานวิจัยว่าการสุ่มตัวอย่างซ้ำสามารถปรับปรุงข้อมูลที่ไม่สมดุลได้ดีกว่าเทคนิคการเพิ่มตัวอย่างส่วนน้อย

### 2.3 ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจคือเครื่องมือใช้สำหรับช่วยตัดสินใจ [12] โดยใช้โครงสร้างข้อมูลแบบต้นไม้ที่ประกอบไปด้วยโหนดรากที่แตกกิ่งก้านออกเป็นโหนดย่อยลงเป็นลำดับชั้น โดยชั้นล่างสุดหรือโหนดล่างสุดจะเป็นใบของต้นไม้ซึ่งถูกใช้เก็บข้อมูลป้ายกำกับ และต้นไม้ตัดสินใจสามารถเป็นได้ทั้งแบบสองคลาส (Binary Class) หรือแบบหลายคลาส (Multiple Class) โครงสร้างแบบต้นไม้สร้างความเข้าใจและแปลความได้ง่าย รวมถึงแสดงโครงสร้างในฐานข้อมูลเพื่อความรวดเร็วในการค้นหา โครงสร้างต้นไม้ได้แสดงให้เห็นดังรูปที่ 2 โดยมี 3 องค์ประกอบ ได้แก่ 1) โหนด (Node) ใช้สำหรับระบุชื่อของแอตทริบิวต์ 2) เส้นเชื่อมโหนด (Edge) ใช้เชื่อมกับโหนดชั้นถัดไปจากโหนดรากซึ่งเป็นโหนดที่ใช้ระบุค่าของคุณลักษณะของโหนดลำดับชั้นก่อนหน้า 3) โหนดใบ (Leaf Node) ใช้สำหรับระบุค่าป้ายกำกับ (Label) หรือคลาส (Class)



รูปที่ 2. โครงสร้างต้นไม้ตัดสินใจ (Decision Tree) [12]

การสร้างต้นไม้ตัดสินใจมีความสำคัญคือ การเลือกแอตทริบิวต์สร้างเป็นโหนดรากโดยเลือกแอตทริบิวต์ที่สามารถแยกข้อมูลได้อย่างสมบูรณ์ก่อนเสมอ โดยอาศัยทฤษฎีข่าวสาร (Information Theory) [13] ซึ่งประกอบไปด้วยค่าเอนโทรปี (Entropy) และค่า Information Gain (IG) โดยค่า IG นำมาใช้สำหรับเลือกคุณลักษณะข้อมูลที่ดีที่สุดแต่ต้องอาศัยค่าเอนโทรปีเข้ามาช่วยคำนวณ ดังแสดงไว้ในสมการ (1) และ (2)

Entropy

$$H(x) = - \sum_{i=1}^n P(x_i) \log_b P(x_i) \quad (1)$$

Information Gain

$$IG(x, a) = H(x) - \sum_{v=values(a)} \frac{|x_v|}{|x|} H(x_v) \quad (2)$$

โดย

Entropy คือ ค่าที่ใช้วัดความแตกต่าง การกระจายของข้อมูล

Information Gain คือ ค่าที่สามารถแบ่งแยกข้อมูล

H คือ เอนโทรปีของตัวอย่าง

X คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลาย ๆ กรณี

P (X<sub>i</sub>) คือ อัตราส่วนของกรณีใน X ที่ตัวแปรตามหรือผลลัพธ์มีค่า i

a คือ ตัวแปรต้นที่พิจารณา

Value (a) คือ เซตของค่าของ a ที่เป็นไปได้

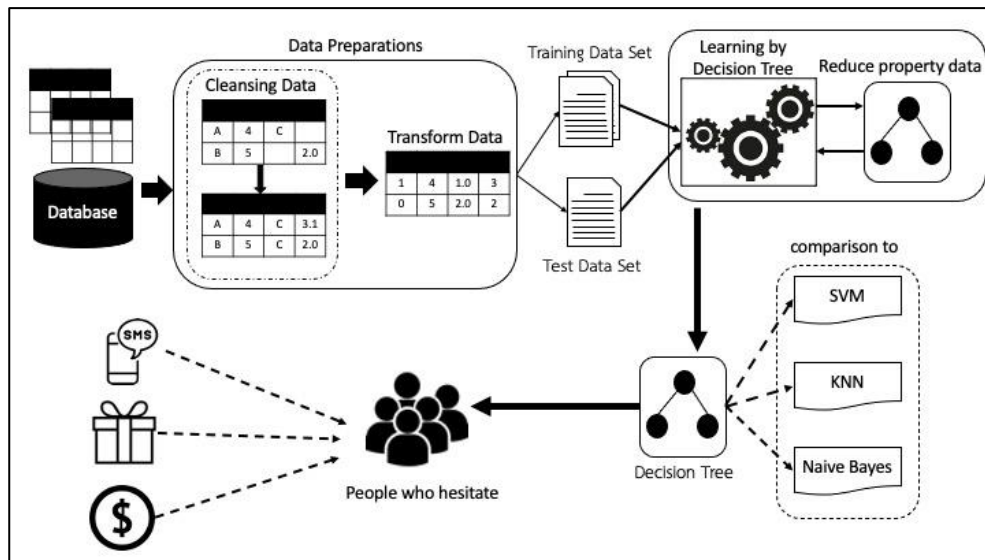
X<sub>v</sub> คือ ตัวอย่างที่ a มีค่า v ทั้งหมด

#### 2.4 ระบบการรับสมัครนักศึกษา

มหาวิทยาลัยราชภัฏจันทรเกษมรับสมัครนักศึกษาใหม่ผ่านระบบ TCAS [14] ทั้งหมด 5 รอบ ประกอบไปด้วย 1) รอบแฟ้มสะสมผลงาน (Portfolio) 2) รอบโควตา (Quota) 3) รอบแอดมิชชั่น (Admission) 4) รอบรับตรงร่วม และ 5) รอบรับตรงอิสระ โดยทั้ง 5 รอบผู้สมัครทำการส่งข้อมูลการสมัครที่ประกอบไปด้วยประวัติส่วนตัว คณะและสาขาที่สนใจเข้าศึกษาต่อในรูปแบบออนไลน์ ด้วยวิธีป้อนผ่านระบบออนไลน์หรือสมัครโดยตรงผ่านศูนย์รับสมัครกับทางมหาวิทยาลัยด้วยตนเอง เมื่อทำการสมัครเสร็จสิ้น ขั้นตอนถัดไปเป็นการสอบในรูปแบบข้อเขียน หรือสอบสัมภาษณ์ จากนั้นเป็นขั้นตอนการประกาศผล ซึ่งมีทั้งผู้ผ่านและไม่ผ่านเข้าศึกษาต่อ โดยผู้ที่ผ่านการคัดเลือกมีบางส่วนไม่ตอบรับและไม่ยืนยันการเข้าศึกษาต่อกับทางมหาวิทยาลัย บุคคลกลุ่มนี้ถือว่าเป็นบุคคลที่น่าสนใจโดยทางมหาวิทยาลัยต้องดำเนินการโน้มน้าวจิตใจให้ผู้สมัครตอบรับและยืนยันเข้าศึกษาต่อ เนื่องจากระบบสอบ TCAS ให้ผู้สมัครสามารถสมัครได้หลายสถาบันและหลายรอบทำให้เกิดการลังเลของผู้สมัครในการเลือกเข้าศึกษาต่อในสถาบันและผู้สมัครกลุ่มที่เลือกสมัครกับทางมหาวิทยาลัยราชภัฏจันทรเกษมมีแนวโน้มที่สนใจเข้าศึกษาต่อกับทางมหาวิทยาลัยแต่อยู่ในช่วงลังเลในการตัดสินใจเข้าศึกษาต่อ ในช่วงนี้ทางสถาบันควรดำเนินการโน้มน้าวจิตใจด้วยกลยุทธ์ที่เข้าตรงถึงผู้สมัครเพราะส่งผลให้ผู้สมัครตัดสินใจเลือกเข้าศึกษากับทางสถาบันได้ ด้วยเหตุนี้จึงเป็นแรงบันดาลใจให้เกิดงานวิจัยนี้ขึ้น

### 3. ระเบียบวิธีดำเนินงาน

ระเบียบวิธีดำเนินงานได้แสดงไว้ดังรูปที่ 3 โดยใช้ชุดข้อมูลจากฐานข้อมูลประวัติผู้สมัครเข้าศึกษาต่อจากสำนักส่งเสริมวิชาการและงานทะเบียน ระเบียบวิธีดำเนินงานเริ่มต้นด้วยการจัดเตรียมข้อมูลเพื่อให้ข้อมูลมีความสะอาดสมบูรณ์ การเพิ่มข้อมูลที่สูญหายด้วยหลักการทางสถิติ การตัดค่าหรือข้อมูลที่ไม่ได้ใช้ประโยชน์ และทำการแปลงข้อมูลให้อยู่ในรูปแบบที่นำเข้าสู่โมเดลต้นไม้ตัดสินใจ ก่อนนำเข้าสู่โมเดลแบ่งข้อมูลออกเป็น 2 ส่วน ส่วนแรกใช้เป็นข้อมูลสำหรับสอน (Training Set) คิดเป็น 80% ของข้อมูลทั้งหมดใช้สอนเครื่องเพื่อให้ได้ต้นไม้ตัดสินใจที่เหมาะสม ส่วนที่สอง 20% ของข้อมูลทั้งหมดนำมาเป็นชุดข้อมูลทดสอบ (Test Set) ต้นไม้ตัดสินใจและวัดประสิทธิภาพการทำงานของโมเดลเป็นลำดับถัดไป



รูปที่ 3. ระเบียบวิธีดำเนินงาน

เพื่อเพิ่มประสิทธิภาพความเร็วในการทำงานของต้นไม้ตัดสินใจมีการลดจำนวนแอตทริบิวต์ข้อมูลสำหรับสอนด้วยวิธีลดที่ละแอตทริบิวต์ที่มีความสำคัญน้อยออกแล้วสอนเครื่องเพื่อหาความแม่นยำในการจำแนก หลังสรุปแอตทริบิวต์ที่สามารถจำแนกข้อมูลได้อย่างสมบูรณ์ได้นำชุดข้อมูลไปใช้กับโมเดลซัพพอร์ตเวกเตอร์แมชชีน วิธีการเพื่อนบ้านใกล้ที่สุด และเทคนิคคานาอีฟเบย์ เพื่อเปรียบเทียบความแม่นยำในการจำแนกข้อมูล โดยผลการเปรียบเทียบได้นำเสนอไว้ในหัวข้อ 3.5 ในการทดลองครั้งนี้ปรากฏว่าต้นไม้ตัดสินใจมีความแม่นยำในการจำแนกสูงสุด และถูกนำมาใช้สำหรับสร้างซอฟต์แวร์ในการจำแนกผู้สมัครที่ลังเลในการตัดสินใจเข้าศึกษาต่อกับมหาวิทยาลัยราชภัฏจันทรเกษมเพื่อนำกลุ่มคนเหล่านี้เข้าสู่กระบวนการโน้มน้าวใจเพื่อให้ตัดสินใจเลือกศึกษาต่อกับทางมหาวิทยาลัยซึ่งเป็นขั้นตอนสุดท้ายของการดำเนินการวิจัยในครั้งนี้

### 3.1 ชุดข้อมูล (Data Set)

ข้อมูลที่น่ามาใช้ในการวิจัยครั้งนี้คือข้อมูลประวัติผู้สมัคร คณะและสาขาที่ผู้สมัครเลือก พร้อมทั้งป้ายกำกับแต่ละระเบียบในรูปแบบตารางข้อมูลผู้สมัครเข้าศึกษาต่อจากสำนักส่งเสริมวิชาการและงานทะเบียน จำนวน 6,000 รายการ ระหว่าง พ.ศ. 2560 ถึง พ.ศ. 2562 ประกอบไปด้วยแอตทริบิวต์ ชื่อ-สกุล คณะที่สมัคร สาขาที่สมัคร ที่อยู่ เพศ โรงเรียน เกรดเฉลี่ย อาชีพบิดา อาชีพมารดา รายได้ครอบครัว ภูมิลำเนา เบอร์ติดต่อ ชื่อเฟสบุ๊ค และไอดีไลน์ มีลักษณะดังแสดงไว้ในตารางที่ 1

ตารางที่ 1. ข้อมูลผู้สมัครเข้าศึกษาต่อ

ชื่อ-สกุล	คณะ	สาขา	ที่อยู่	.....	คลาส
ผู้สมัคร 1	วิทยาศาสตร์	วิทยาการคอมพิวเตอร์	45/6 กทม		1
ผู้สมัคร 2	วิทยาการจัดการ	บัญชี	3/23 ขอนแก่น		0
ผู้สมัคร 4	ศึกษาศาสตร์	คณิตศาสตร์	5/4 นนทบุรี		1

### 3.2 เตรียมข้อมูล (Data Preparation)

1) การเลือกแอตทริบิวต์ (Feature Selection) วิเคราะห์แอตทริบิวต์ที่สามารถนำมาใช้งานหรือมีประโยชน์ต่อการจำแนกข้อมูลจากแอตทริบิวต์ทั้งหมด 14 แอตทริบิวต์ ประกอบไปด้วย ชื่อ-สกุล คณะ สาขา ที่อยู่ เพศ ประเภทโรงเรียน เกรดเฉลี่ย อาชีพบิดา อาชีพมารดา รายได้ครอบครัว ภูมิลำเนา เบอร์ติดต่อ ชื่อเฟสบุ๊ค และไอดีไลน์ ปรากฏว่ามี 5 แอตทริบิวต์ ไม่มีคุณสมบัติใช้จำแนกผู้สมัครที่ลงใจได้ ซึ่งได้แก่ ชื่อ-สกุล ที่อยู่ เบอร์ติดต่อ ชื่อเฟสบุ๊ค และไอดีไลน์ เพราะข้อมูลคุณสมบัติเหล่านี้ไม่มีความจำเพาะเจาะจงในการจำแนกได้อย่างชัดเจน ซึ่งได้คัดออกจากชุดข้อมูลสำหรับสอน แอตทริบิวต์ที่เหลือจำนวน 9 แอตทริบิวต์สามารถนำไปใช้งานได้ ประกอบด้วย คณะ สาขา เพศ ประเภทโรงเรียน เกรดเฉลี่ย อาชีพบิดา อาชีพมารดา รายได้ครอบครัว และภูมิลำเนา

2) การแทนค่าข้อมูลที่สูญหาย จากชุดข้อมูลผู้สมัครมีบางรายการที่บางแอตทริบิวต์สูญหาย ดังแสดงไว้ในตารางที่ 2 เช่น เกรดเฉลี่ย รายได้ผู้ปกครอง ซึ่งเป็นข้อมูลในรูปแบบของตัวเลข การแทนค่าข้อมูลที่สูญหายได้ใช้ค่าเฉลี่ยของแอตทริบิวต์ไปทดแทน ส่วนข้อมูลที่อยู่ในรูปแบบหมวดหมู่ เช่น อาชีพผู้ปกครอง ภูมิลำเนา ใช้หลักการจำนวนรายสูงสุดของแอตทริบิวต์นั้นทดแทนข้อมูลที่สูญหาย ดังแสดงไว้ในตารางที่ 3

ตารางที่ 2. ตัวอย่างข้อมูลที่สูญหาย

คณะ	สาขา	เกรด	ประเภท ร.ร.	.....	คลาส
วิทยาศาสตร์	วิทยาการคอมพิวเตอร์	3.4	-		1
วิทยาการจัดการ	บัญชี	-	ขนาดกลาง		0
ศึกษาศาสตร์	คณิตศาสตร์	2.7	ขนาดเล็ก		1



ตารางที่ 3. การแทนค่าให้กับข้อมูลที่สูญหาย

คณะ	สาขา	เกรด	ประเภท ร.ร.	.....	คลาส
วิทยาศาสตร์	วิทยาการคอมพิวเตอร์	3.4	ขนาดกลาง		1
วิทยาการจัดการ	บัญชี	3.1	ขนาดกลาง		0
ศึกษาศาสตร์	คณิตศาสตร์	2.7	ขนาดเล็ก		1

3) การแปลงรูปข้อมูล (Data Transform) หลังทำความสะอาดข้อมูล (Data Cleansing) ขั้นตอนถัดมาแปลงข้อมูลที่อยู่ในรูปแบบหมวดหมู่ให้กลายเป็นตัวเลข ดังแสดงไว้ในตารางที่ 4

ตารางที่ 4. แปลงข้อมูลคุณสมบัติให้กลายเป็นข้อมูลตัวเลข

ข้อมูลคุณสมบัติ	ข้อมูลแปลง
คณะ	1 = วิทยาศาสตร์ 2 = วิทยาการจัดการ 3 = มนุษยศาสตร์และสังคมศาสตร์ 4 = ศึกษาศาสตร์ 5 = เกษตรและเทคโนโลยีชีวภาพ 6 = วิทยาลัยแพทย์ทางเลือก
สาขา	1 = วิทยาการคอมพิวเตอร์ 2 = เทคโนโลยีสารสนเทศ 3 = เทคโนโลยีมีเดีย 4 = วิทยาศาสตร์สิ่งแวดล้อม
เพศ	1 = ชาย 2 = หญิง
ประเภทโรงเรียน	1 = ขนาดเล็ก 2 = ขนาดกลาง 3 = ขนาดใหญ่ 4 = ขนาดใหญ่พิเศษ
อาชีพบิดาและมารดา	1 = ราชการ/รัฐวิสาหกิจ 2 = ธุรกิจ 3 = ค้าขาย 4 = เกษตรกร 5 = รับจ้างทั่วไป 6 = อื่น ๆ
รายได้ผู้ปกครอง	1 = ต่ำกว่า 10,000 2 = 10,001 – 20,000 3 = 20,001 – 30,000 4 = 30,001 – 40,000 5 = 40,001 – 50,000 6 = 50,001 ขึ้นไป

ตารางที่ 4. (ต่อ) แปลงข้อมูลคุณสมบัติให้กลายเป็นข้อมูลตัวเลข

ข้อมูลคุณสมบัติ	ข้อมูลแปลง
ภูมิลำเนา	5 = กรุงเทพมหานคร
การตัดสินใจ	0 = ไม่เข้าศึกษาต่อ 1 = เข้าศึกษาต่อ

### 3.3 การแบ่งข้อมูลสำหรับสอนและทดสอบ

จากข้อมูลทั้งหมด 6,000 รายการ แบ่งออกเป็นสองส่วน ส่วนแรกสำหรับสอนเครื่องจำนวน 4,800 รายการ คิดเป็น 80% ส่วนที่สองไว้สำหรับทดสอบการเรียนรู้ของเครื่องจำนวน 1,200 รายการ คิดเป็นร้อยละ 20 ในข้อมูลสำหรับสอนทำการเลือกรายการที่มีป้ายกำกับที่เป็น 1 (เข้าศึกษาต่อ) และ 0 (ไม่เข้าศึกษาต่อ) ในปริมาณที่ใกล้เคียงกันเป็น 2,365 และ 2,435 รายการตามลำดับ เพื่อไม่ให้เกิด Overfit ของเครื่องสำหรับป้องกันการไม่เข้าใจข้อมูลทั่วไปของเครื่อง

### 3.4 การเรียนรู้ของเครื่องด้วยต้นไม้ตัดสินใจ

ขั้นตอนนี้ผู้วิจัยนำข้อมูลสำหรับสอนมาทำการเรียนรู้ด้วยอัลกอริทึมต้นไม้ตัดสินใจ แล้วใช้ชุดข้อมูลทดสอบทำการทดสอบการจำแนกข้อมูลและวัดประสิทธิภาพการจำแนก ทำการลดจำนวนข้อมูลแอตทริบิวต์ ดังแสดงไว้ในตารางที่ 5 ผลปรากฏว่าจำนวนข้อมูลแอตทริบิวต์ที่สามารถจำแนกได้แม่นยำและมีค่าความถูกต้องสูงสุดมีด้วยกัน 2 แบบ แบบที่ 1 ข้อมูลแอตทริบิวต์ 7 แอตทริบิวต์ และข้อมูลแอตทริบิวต์ 6 แอตทริบิวต์ ซึ่งมีค่าความถูกต้องอยู่ที่ 0.958 คิดเป็น 95.80% ในการวิจัยครั้งนี้เลือกแบบที่ 2 เป็นโมเดลสำหรับจำแนกผู้สมัครที่ลงทะเบียนเข้าศึกษาต่อ โดยข้อมูลแอตทริบิวต์ประกอบไปด้วย เพศ ประเภทโรงเรียนเกรดเฉลี่ย รายได้ผู้ปกครอง ภูมิลำเนา และสาขา

ตารางที่ 5. ค่าความถูกต้องในการจำแนกเมื่อทำการลดจำนวนคุณสมบัติลงทีละคุณสมบัติ

จำนวนแอตทริบิวต์	ค่าความถูกต้อง
9	0.902
8	0.917
7	0.958
6	0.958
5	0.903

### 3.5 เปรียบเทียบประสิทธิภาพโมเดลสำหรับจำแนก

การวิจัยครั้งนี้ผู้วิจัยเพิ่มความเชื่อมั่นในตัวโมเดลต้นไม้ตัดสินใจถึงประสิทธิภาพการทำงานที่มีความน่าเชื่อถือด้วยการนำข้อมูลชุดทดสอบเดียวกันนี้ใช้ทดสอบกับสามอัลกอริทึมได้แก่ 1) อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน 2) อัลกอริทึมเพื่อนบ้านใกล้ที่สุด และ 3) อัลกอริทึมเทคนิคนาอีฟเบย์ ผลสรุปได้ว่าอัลกอริทึมต้นไม้ตัดสินใจมีความถูกต้องในการจำแนกสูงที่สุดถึง 95.89% โดยผลการทดลองแสดงรายละเอียดในหัวข้อที่ 4



$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F - Measure = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (6)$$

โดย

Accuracy คือ ค่าความถูกต้อง

Precision คือ ค่าความแม่นยำ

Recall คือ ค่าความระลึก

F-Measure คือ ค่าเฉลี่ยฮาร์โมนิกของความแม่นยำ

True Positives (TP) คือ ข้อมูลที่เป็น Positives และจำแนกว่าเป็น Positive ได้ถูกต้อง

True Negatives (TN) คือ ข้อมูลที่เป็น Negatives และจำแนกว่าเป็น Negatives ได้ถูกต้อง

False Positives (FP) คือ ข้อมูลที่เป็น Negatives และจำแนกว่าเป็นผิดว่าเป็น Positives

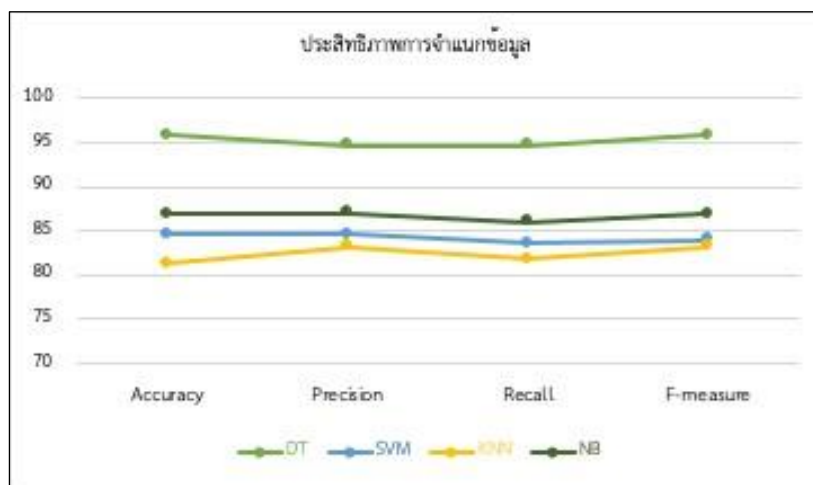
False Negatives (FN) คือ ข้อมูลที่เป็น Positives และจำแนกว่าเป็นผิดว่าเป็น Negatives

การทดสอบประสิทธิภาพการจำแนกผู้สมัครที่ลงทะเบียนในการตัดสินใจเข้าศึกษาต่อโดยใช้อัลกอริทึมต้นไม้ตัดสินใจ อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน อัลกอริทึมเพื่อนบ้านใกล้ที่สุด และอัลกอริทึมเทคนิคนาอิวเบย์ ได้ผลดังนี้

ตารางที่ 6. ผลการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูล

Method / Algorithm	DT	SVM	KNN	NB
Accuracy (%)	95.89	84.64	81.23	86.89
Precision (%)	94.64	84.60	83.19	86.99
Recall (%)	94.75	83.55	81.85	85.97
F-measure (%)	95.87	83.97	83.21	86.94

จากการแสดงไว้ในตารางที่ 6 พบว่าตัวแบบจากอัลกอริทึมต้นไม้ตัดสินใจ (DT) มีค่าความถูกต้อง 95.89% ค่าความแม่นยำ 94.64% ค่าความระลึก 94.75% ค่า F-measure 95.87% ตัวแบบอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน มีค่าความถูกต้อง 84.64% ค่าความแม่นยำ 84.60% ค่าความระลึก 83.55% และค่า F-measure 83.97%



รูปที่ 5. กราฟเปรียบเทียบประสิทธิภาพการจำแนกข้อมูล

ตัวแบบที่สร้างด้วยอัลกอริทึมเพื่อนบ้านใกล้ที่สุด มีค่าความถูกต้อง 81.23% ค่าความแม่นยำ 83.19% ค่าความระลึกลับ 81.85% ค่า F-measure 83.21% และตัวแบบอัลกอริทึมเทคนิคนาอิวเบย์ ค่าความถูกต้อง 86.89% ค่าความแม่นยำ 86.99% ค่าความระลึกลับ 85.97% ค่า F-measure 86.94% เมื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลแล้วได้แสดงไว้ในกราฟดังรูปที่ 5 เห็นได้ว่าตัวแบบจากอัลกอริทึมต้นไม้ตัดสินใจมีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึกลับ และค่า F-measure สูงกว่าทั้งสามอัลกอริทึมทุกมิติ

## 5. ผลการดำเนินการ

จากผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึมต้นไม้ตัดสินใจกับอีก 3 อัลกอริทึม ได้แก่ 1) อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (SVM) 2) อัลกอริทึมเพื่อนบ้านใกล้ที่สุด (KNN) และ 3) อัลกอริทึมเทคนิคนาอิวเบย์ ผลสรุปว่าตัวแบบอัลกอริทึมต้นไม้ตัดสินใจมีประสิทธิภาพการจำแนกข้อมูลได้ดีที่สุดในทุกมิติ ดังต่อไปนี้ ค่าความถูกต้อง 95.89% ค่าความแม่นยำ 94.64% ค่าความระลึกลับ 94.75% และค่า F-measure 95.87% ดังแสดงไว้ในตารางที่ 7

ตารางที่ 7. ประสิทธิภาพการจำแนกของตัวแบบอัลกอริทึมต้นไม้ตัดสินใจ

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
Decision Tree	95.89	94.64	94.75	95.87

สำหรับแอตทริบิวต์ที่สามารถใช้ในการจำแนกผู้สมัครที่ลงทะเบียนในการตัดสินใจเข้าศึกษาต่อสำหรับใช้กับตัวแบบอัลกอริทึมต้นไม้ตัดสินใจ ประกอบไปด้วย สาขาวิชา เกรดเฉลี่ย รายได้ครอบครัว ภูมิลำเนา ประเภทโรงเรียน และเพศ ตามกฎของต้นไม้ตัดสินใจได้แสดงไว้ดังรูปที่ 4

อีกหนึ่งผลลัพธ์จากการวิจัยคือ ระบบจำแนกผู้สมัครที่ลงทะเบียนในการตัดสินใจศึกษาต่อเพื่อสร้างกลยุทธ์สนับสนุนการตัดสินใจให้เข้าศึกษาต่อ ในรูปแบบเว็บแอปพลิเคชันที่พัฒนาจากภาษา C# เชื่อมต่อฐานข้อมูล SQL Server ประกอบไปด้วยส่วนติดต่อกับผู้สมัคร และส่วนรายงานจำนวนผู้สมัครที่ลงทะเบียนให้ผู้บริหารและผู้เกี่ยวข้องใช้สำหรับดำเนินการโน้มน้าวใจได้แสดงไว้ดังรูปที่ 6 และรูปที่ 7

รูปที่ 6. หน้าจอส่วนการรับสมัครนักศึกษาใหม่

ชื่อ-สกุล	สาขา	ผลการนิสถานะ	ข้อมูลติดต่อ
นายนครินทร์ สมุทรไทย	วิทยาการคอมพิวเตอร์	เรียน	🔍
นายอรรถ สิริธรรสยา	วิทยาการคอมพิวเตอร์	เรียน	🔍
นางสาวอชิราพร กุลขนิษฐา	วิทยาการคอมพิวเตอร์	ไม่เรียน	🔍

รูปที่ 7. หน้าจอสำหรับรายงานการจำแนกผู้สมัคร

## 6. สรุปผล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อสร้างเครื่องเรียนรู้สำหรับการจำแนกผู้สมัครที่ลงทะเบียนในการตัดสินใจศึกษาต่อเพื่อสร้างกลยุทธ์สนับสนุนการตัดสินใจให้เข้าศึกษาต่อ ด้วยอัลกอริทึมต้นไม้ตัดสินใจในการจำแนกผู้สมัครที่ลงทะเบียนจากข้อมูลผู้สนใจเข้าสมัครศึกษาของราชภัฏจันทรเกษมจำนวน 6,000 ระเบียบ ใช้ข้อมูล 80% จากข้อมูลทั้งหมดสำหรับใช้สอนเครื่อง และอีก 20% ใช้สำหรับทดสอบเครื่อง ซึ่งประสิทธิภาพในการจำแนกของอัลกอริทึมต้นไม้ตัดสินใจมีความแม่นยำสูงถึง 95.89% จากแอดทริบิวต์ผู้สมัครซึ่งประกอบด้วยสาขาวิชา เกรดเฉลี่ย รายได้ครอบครัว ภูมิลำเนา ประเภทโรงเรียน และเพศ และพัฒนาระบบการจำแนกผู้สมัครลงทะเบียนในรูปแบบของเว็บแอปพลิเคชันเพื่อใช้สำหรับค้นหาผู้ลงทะเบียนเพื่อใช้เป็นเครื่องมือจำแนกผู้สมัครที่ลงทะเบียนในการตัดสินใจเข้าศึกษาต่อ

## เอกสารอ้างอิง (References)

- [1] สถาบันวิจัยประชากรและสังคม มหาวิทยาลัยมหิดล. 2563. สารประชากร มหาวิทยาลัยมหิดล. แหล่งข้อมูล : <http://pathumthani.dop.go.th/th/gallery/4/1328>. ค้นเมื่อวันที่ 22 กรกฎาคม 2563.
- [2] สำนักงานคณะกรรมการอุดมศึกษา. 2561. สถิติอุดมศึกษา 2558-2560. แหล่งข้อมูล : <http://www.info.mua.go.th/info/>. ค้นเมื่อวันที่ 22 กรกฎาคม 2563.
- [3] วาสิตา บุรณศิริปิน. 2552. การใช้กลยุทธ์การบริหารความสัมพันธ์กับลูกค้าขององค์กรธุรกิจไทย. นิเทศศาสตรมหาบัณฑิต. สาขาการประชาสัมพันธ์. จุฬาลงกรณ์มหาวิทยาลัย. [Wasita Euranasilpin. 2009. The Use of Customer Relationship Management in Thai Business Organization. M.A. Communication Arts. Public Relations Program. Chulalongkorn University (in Thai)]
- [4] Mitchell, T. 1997. Machine Learning. 2<sup>nd</sup> ed, McGraw Hill.
- [5] Michie, D. 1968. "Memo" Functions and Machine Learning. *Reprinted from Nature*. 218(5136), 19-22.
- [6] Mohri, M., Rostamizadeh, A. and Talwalkar, A. 2018. Foundations of Machine Learning. 2<sup>nd</sup> ed. The MIT Press.
- [7] อัชฌาพร กว่างสวาสดี, เพียงฤทัย หนูสวัสดิ์, วราลี คงเหมาะ, ปวีณา ทิพย์ากุลรักษ์ และบุษกร สังขนันท์. 2562. ระบบทำนายระดับความเครียด ด้วยเทคนิคต้นไม้การตัดสินใจ. *วารสารวิทยาศาสตร์และเทคโนโลยีแห่งรัตนโกสินทร์*, 1(2), 13-26. [Achaporon Kwangsawad, Painingruthai Nusawat, Walari Kongmo, Paveena Thippayakunrak and Busakorn Sangkhanan. 2019. The System of Stress level prediction using Decision tree. *Rattanakosin Journal of Science and Technology*, 1(2), 13-26. (in Thai)]
- [8] ธัญชัย เพ็งพรหม และวารรัตน์ สงฆ์แป้น. 2562. แบบจำลองแบบผสมเพื่อสร้างกฎสำหรับคำถามคำตอบ โดยใช้เทคนิคการทำเหมืองข้อความ. *วลัยลักษณ์วิจัย*, 2019(7), NCIs20 (1-12). [Thanachai Penprom and Wararat Songpan. 2019. A Hybrid Model for Rule-Based

- Generating of Question-Answering Using Text Mining. *Walailak Procedia*, 2019(7), NCI20 (1-12). (in Thai)]
- [9] อนัตต์ชัย ชูติภาสเจริญ และจรัญ แสนราช. 2561. การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไป. *วารสารวิจัย มหาวิทยาลัยขอนแก่น (ฉบับบัณฑิตศึกษา) สาขามนุษยศาสตร์และสังคมศาสตร์*, 6(3), 105-113. [Anutchai Chutipascharoen and Charun Sanrach. 2018. A Comparison of the Efficiency of Algorithms and Feature Selection Methods for Predicting the Success of Personal Overseas Money Transfer. *KKU Research Journal of Humanities and Social Sciences (Graduate Studies)*, 6(3), 105-113. (in Thai)]
- [10] พิชญะ พรหมลา และจรัญ แสนราช. 2563. การเปรียบเทียบประสิทธิภาพการวิเคราะห์ความพึงพอใจเกี่ยวกับการจัดการเรียนการสอนด้วยกระบวนการวิเคราะห์ความรู้สึกโดยใช้เทคนิคการรวมกลุ่มเพื่อจำแนกข้อมูล. *วารสารวิจัย มข. (ฉบับบัณฑิตศึกษา)*, 20(4), 140-149. [Pichaya Promla and Charun Sanrach. 2020. The Comparison of Efficiency on The Analysis of Satisfaction on Teaching Performance using Sentiment Analysis by Ensemble Technique. *KKU Research Journal (Graduate Studies)*, 20(4), 140-149. (in Thai)]
- [11] กาญจน์ ณ ศรีระ, กิตติศักดิ์ เกิดประสพ และนิตยา เกิดประสพ. 2560. การเปรียบเทียบเทคนิคการสุ่มตัวอย่างเพื่อการจำแนกข้อมูลที่ไม่สมดุล. *วารสารวิทยาการสารสนเทศและเทคโนโลยีประยุกต์*, 1(1), 20-37. [Karn Nasritha, Kittisak Kerdprasop and Nittaya Kerdprasop. 2017. Comparison of sampling techniques for imbalanced data classification. *Journal of Applied Informatics and Technology*, 1(1), 20-37. (in Thai)]
- [12] ปริญญา สงวนสัตย์. 2562. Artificial Intelligence with Machine Learning AI สร้างได้ด้วยแมชชีนเลิร์นนิ่ง. พิมพ์ครั้งที่ 1, ไอทีซี พรีเมียร์, นนทบุรี. [Parinya Sanguansat. 2019. Artificial Intelligence with Machine Learning AI. 1<sup>st</sup> ed, IDC Premier, Nonthaburi. (in Thai)]
- [13] Quinlan, J.R. 1986. Induction of decision trees. *Machine learning*, 1(1), 81-106.
- [14] ที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ). 2563. คู่มือระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษา. แหล่งข้อมูล : <https://www.mytcas.com/>. ค้นเมื่อวันที่ 16 กรกฎาคม 2563.