

การสร้างแบบจำลองการพยากรณ์แนวโน้มการสมัครเข้าศึกษาต่อในมหาวิทยาลัย
เอกชน: การศึกษาจากพฤติกรรมผู้เข้าร่วมกิจกรรมแนะแนวการศึกษา
Predictive Modeling for Private University Enrollment Trends: A Study Based on
Educational Fair Participant Behavior
กันต์ปานัสม์ อำไพวรรณ^{1*} และ วรภัทร ไพรีเกรง²
Kanpanus Amphaivan¹, Worapat Paireekreng²

เทคโนโลยีสารสนเทศ วิทยาลัยนวัตกรรมการเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์¹ และ
คณะบริหารธุรกิจ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง²
Information Technology, College of Technology and Engineering, Dhurakij Pundit University.
Faculty of Business Administration, King Mongkut's Institute of Technology Ladkrabang.
E-Mail: Kanpanus.prince@gmail.com, Worapat.pa@kmitl.ac.th
(Received : September 22, 2025; Revised : November 29, 2025; Accepted : December 15, 2025)

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการพยากรณ์แนวโน้มการสมัครเข้าศึกษาต่อในมหาวิทยาลัยเอกชน มีวัตถุประสงค์ดังนี้ 1) เพื่อวิเคราะห์ปัจจัยเชิงพฤติกรรม ภายในงาน Open house ที่สัมพันธ์กับการตัดสินใจสมัครเข้าศึกษา 2) เพื่อศึกษาความสัมพันธ์ระหว่างพฤติกรรมการเข้าร่วมกิจกรรมกับการตัดสินใจสมัครเข้าศึกษา และ 3) เพื่อพัฒนาแบบจำลองการพยากรณ์แนวโน้มการสมัครเข้าศึกษาต่อในมหาวิทยาลัยเอกชน โดยวิเคราะห์จากข้อมูลผู้เข้าร่วมกิจกรรมแนะแนวการศึกษา Open House โดยใช้ข้อมูลจากผู้ลงทะเบียนเข้าร่วมกิจกรรมจำนวน 1,321 คน มีผู้ลงทะเบียนเข้าร่วมกิจกรรมจริง 564 คน จำนวนที่ไม่มาร่วมงาน 625 คน และมีนักเรียนที่ลงทะเบียนและสมัครเข้าศึกษาภายในงาน 132 คน การศึกษาใช้แบบจำลอง Naive Bayes Logistic Regression Deep Learning Decision Tree และ Random Forest

ผลการวิจัยที่สำคัญ พบว่า 1) ปัจจัยสำคัญที่สุดที่ส่งผลต่อการเข้าร่วมกิจกรรม คือ ระยะเวลาการเข้าร่วมกิจกรรมและจำนวนกิจกรรมย่อยที่เข้าชม 2) พฤติกรรมการเข้าร่วมกิจกรรมจริง มีความสัมพันธ์อย่างมีนัยสำคัญกับการตัดสินใจสมัครเข้าศึกษา ปัจจัยที่มีอิทธิพลต่อการตัดสินใจสมัครเรียนมากที่สุด ได้แก่ ระยะเวลาการเข้าร่วมกิจกรรม (Importance Score = 0.38) และจำนวนกิจกรรมย่อยที่เข้าชม (Importance Score = 0.25) ซึ่งเป็นตัวบ่งชี้ความมีส่วนร่วมและความตั้งใจสมัครเรียน และ 3) การพัฒนาแบบจำลองการพยากรณ์แนวโน้มการสมัครเข้าศึกษาต่อในมหาวิทยาลัยเอกชน พบว่า แบบจำลอง Deep Learning มีประสิทธิภาพสูงสุด (Accuracy 0.94 และ AUC 0.98) รองลงมาคือ Random Forest

ผลการทดลองพบว่า แบบจำลอง Random Forest (Accuracy 0.93 และ AUC 0.96) มีประสิทธิภาพอยู่ในเกณฑ์ที่สูงมากและใกล้เคียงกับ Deep Learning แบบจำลอง Deep Learning ในการศึกษา สะท้อนให้เห็นถึงขีดความสามารถของโครงข่ายประสาทเทียมในการเรียนรู้รูปแบบความสัมพันธ์ที่ซับซ้อนและไม่เป็นเชิงเส้น (Non-linear Patterns) ของตัวแปรเชิงพฤติกรรม เช่น ความสัมพันธ์ระหว่างระยะเวลาที่ใช้ในงานและจำนวนกิจกรรมย่อยที่เข้าชม ซึ่งส่งผลต่อการตัดสินใจสมัครเรียนได้ดีกว่าแบบจำลองทางสถิติทั่วไป ดังนั้น จึงสรุปได้ว่า Deep Learning เป็นแบบจำลองที่เหมาะสมและดีที่สุด สำหรับการนำมาใช้เป็นเครื่องมือช่วยตัดสินใจเชิงกลยุทธ์ในการพยากรณ์การรับสมัครนักศึกษาของมหาวิทยาลัยเอกชนในการศึกษานี้

การใช้ข้อมูลพฤติกรรมแบบการประมวลผลเพื่อแสดงข้อมูลทันทีร่วมกับเทคนิคการเรียนรู้ของ Machine Learning สามารถเพิ่มประสิทธิภาพการพยากรณ์และสนับสนุนการวางแผนเชิงกลยุทธ์ด้านการรับสมัครนักศึกษาได้อย่างมีประสิทธิภาพสำหรับสถาบันอุดมศึกษาเอกชน

คำสำคัญ : การพยากรณ์แนวโน้ม, เหมืองข้อมูล, การจำแนกข้อมูล

ABSTRACT

This study aimed to develop a predictive model for forecasting the likelihood of student applications to private universities. The specific objectives were 1) to analyze behavioral factors during open house events that are associated with application decisions; 2) to examine the relationship between participation behavior and application decisions; and 3) to develop a predictive model for application likelihood using behavioral data from open house participants. The dataset consisted of 1,321 registered participants, of whom 564 actually attended the event, 625 did not attend, and 132 students both registered and submitted applications during the event. The study employed Naive Bayes, Logistic Regression, Deep Learning, Decision Tree, and Random Forest models.

The key findings were as follows 1) the most important factors influencing event participation were the duration of attendance and the number of sub-activities visited; 2) actual participation behavior at the event was significantly associated with the decision to apply. The most influential factors in application decisions were duration of activity participation (Importance Score = 0.38) and number of sub-activities attended (Importance Score = 0.25), both reflecting levels of engagement and intention to apply; and 3) in developing a predictive model for application likelihood to private universities, the Deep Learning model achieved the highest performance (Accuracy = 0.94, AUC = 0.98), followed by the Random Forest model (Accuracy = 0.93, AUC = 0.96), which also demonstrated very high and comparable performance.

The results highlight the capability of the Deep Learning model, based on artificial neural networks, to learn complex and non-linear behavioral patterns, such as the interaction between time spent at the event and the number of sub-activities attended, thereby outperforming traditional statistical models in predicting application decisions. Therefore, Deep Learning is concluded to be the most appropriate and effective model for use as a strategic decision-support tool in forecasting student admissions to private universities in this study.

The use of behavioral data processed for real-time visualization, in combination with machine learning techniques, can substantially enhance predictive accuracy and effectively support strategic planning for student recruitment in private higher education institutions.

Keywords : Trend Forecasting, Data Mining, Data Classification

บทนำ

การแข่งขันในภาคอุดมศึกษาของประเทศไทย โดยเฉพาะอย่างยิ่งในกลุ่มมหาวิทยาลัยเอกชน มีความแข่งขันที่เพิ่มขึ้นอย่างต่อเนื่อง การพยากรณ์แนวโน้มการสมัครเข้าศึกษาต่อจึงเป็นปัจจัยสำคัญเชิงกลยุทธ์สำหรับการวางแผนทรัพยากรและการตลาด การพยากรณ์ที่แม่นยำจะช่วยให้มหาวิทยาลัยสามารถระบุกลุ่มเป้าหมายที่มีศักยภาพสูงและจัดสรรทรัพยากรได้อย่างมีประสิทธิภาพ

ในอดีตการพยากรณ์แนวโน้มการสมัครมักอาศัยข้อมูลทางประชากรศาสตร์ (Socio-demographic) และข้อมูลผลการเรียน (Academic Performance) เป็นหลักแต่ในปัจจุบัน การวิจัยได้ให้ความสำคัญกับการใช้เทคนิคการเรียนรู้ของ Machine Learning เพื่อสร้างแบบจำลองการพยากรณ์ที่มีความแม่นยำสูงขึ้น [1]-[3] เทคนิคเหล่านี้

ไม่เพียงแต่ใช้ในการพยากรณ์ผลการเรียนหรือการคงอยู่ของนักศึกษาเท่านั้น แต่ยังสามารถขยายขอบเขตมาใช้ในการพยากรณ์การตัดสินใจสมัครเข้าศึกษาได้ด้วย

อย่างไรก็ตาม งานวิจัยส่วนใหญ่ยังคงเน้นไปที่ปัจจัยภายนอกหรือปัจจัยคงที่ (Static Factors) ในขณะที่ปัจจัยเชิงพฤติกรรมหรือ Engagement Metrics ที่ได้จากกิจกรรมจริง ยังไม่ถูกนำมาศึกษาอย่างแพร่หลาย การวิจัยของ Matz et al. [4] ได้แสดงให้เห็นถึงศักยภาพของการใช้ Engagement Metrics ที่รวบรวมผ่านแอปพลิเคชันมาทำนายการคงอยู่ของนักศึกษา ซึ่งเป็นแนวคิดที่สามารถนำมาประยุกต์ใช้ในการทำนายการสมัครเข้าศึกษาต่อได้

การศึกษางานวิจัยฉบับนี้จึงมีเป้าหมายเพื่อเติมเต็มช่องว่างทางวิชาการดังกล่าว โดยมุ่งเน้นการพัฒนาแบบจำลองการพยากรณ์แนวโน้มการสมัครเข้าศึกษาต่อในมหาวิทยาลัยเอกชน โดยเน้นการวิเคราะห์จาก ปัจจัยเชิงพฤติกรรมภายในงาน Open House ซึ่งรวมถึง ระยะเวลาการเข้าร่วมกิจกรรม และ จำนวนกิจกรรมย่อยที่เข้าชม ซึ่งเป็นตัวแปรที่สะท้อนถึงความสนใจและความตั้งใจจริงของผู้เข้าร่วมงาน

1. วัตถุประสงค์การวิจัย

- 1.1 เพื่อวิเคราะห์ปัจจัยเชิงพฤติกรรม ภายในงาน Open House ที่สัมพันธ์กับการตัดสินใจสมัครเข้าศึกษา
- 1.2 เพื่อศึกษาความสัมพันธ์ระหว่างพฤติกรรมเข้าร่วมกิจกรรมกับการตัดสินใจสมัครเข้าศึกษา
- 1.3 เพื่อพัฒนาแบบจำลองการพยากรณ์แนวโน้มการสมัครเข้าศึกษาต่อในมหาวิทยาลัยเอกชน

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

การประยุกต์ใช้การเรียนรู้ของ Machine Learning และการทำเหมืองข้อมูล (Data Mining) ในภาคการศึกษาหรือ Educational Data Mining (EDM) มีความก้าวหน้าอย่างต่อเนื่อง งานวิจัยก่อนหน้านี้ส่วนใหญ่ มุ่งเน้นไปที่การทำนายผลการเรียนของนักศึกษา [5] หรือการทำนายการออกกลางคัน (Student Dropout) [1] เป็นหลัก อย่างไรก็ตาม การนำเทคนิคเหล่านี้มาใช้ในการพยากรณ์พฤติกรรมของนักศึกษาก่อนการสมัครเรียน (Pre-Enrollment Behavior) จากกิจกรรมทางการตลาด เช่น การเข้าร่วมงานแนะแนวหรือ Open House ยังมี การศึกษาค้นคว้าจำกัด

2.1 งานวิจัยที่เกี่ยวข้องกับการพยากรณ์ในบริบทก่อนเข้าศึกษา [2] ใช้แบบจำลอง Logistic Regression ในการทำนายการสมัครเรียนของนักศึกษามหาวิทยาลัยจากข้อมูลประชากรศาสตร์และประวัติการมีปฏิสัมพันธ์กับสถาบัน ผลการศึกษาชี้ให้เห็นว่าการมีปฏิสัมพันธ์โดยตรง (Direct Interaction) เป็นปัจจัยทำนายที่สำคัญ ซึ่งสอดคล้องกับแนวคิดของการเข้าร่วมงาน Open House ในงานวิจัยฉบับปัจจุบัน การศึกษาของ [2] เป็นพื้นฐานสำคัญที่แสดงให้เห็นถึงความสัมพันธ์ระหว่างการมีส่วนร่วมและความตั้งใจที่จะสมัคร

2.2 งานวิจัยด้านการรับสมัครนักศึกษามักเผชิญกับปัญหา ข้อมูลไม่สมดุล (Imbalanced Data) เนื่องจากจำนวนผู้สมัครเรียนจริงน้อยกว่าผู้ที่ไม่สมัคร จากข้อมูลผู้ลงทะเบียนเข้าร่วมกิจกรรม จำนวน 1,321 คน พบว่ามี ผู้สมัครเข้าศึกษาจริงเพียง 132 คน คิดเป็นร้อยละ 10 และส่วนผู้ไม่สมัครคิดเป็นร้อยละ 90 จากข้อมูลความไม่สมดุลนี้ทำให้แบบจำลองมีแนวโน้มที่จะเกิดอคติในการพยากรณ์ [4] จากงานวิจัยที่เกี่ยวข้องนี้แสดงให้เห็นว่าเทคนิค Ensemble Learning อย่าง Random Forest มีประสิทธิภาพสูงในการจัดการกับชุดข้อมูลที่ไม่สมดุลและให้ผลการทำนายที่มั่นคงด้วยเหตุนี้ผู้วิจัยจึงได้เลือกใช้ Random Forest เป็นหนึ่งในแบบจำลอง

2.3 งานวิจัยที่เกี่ยวข้องกับเทคนิค Deep Learning สำหรับเทคนิคขั้นสูง [3] ได้ทบทวนวรรณกรรมอย่างเป็นระบบเกี่ยวกับการใช้ Deep Learning ในการทำนายผลสัมฤทธิ์ทางการเรียนในสภาพแวดล้อมการเรียนรู้เสมือน ผลการทบทวนพบว่าโครงข่ายประสาทเทียม (Neural Networks) มีความสามารถโดดเด่นในการตรวจจับและเรียนรู้รูปแบบความสัมพันธ์ที่ซับซ้อน (Complex Non-linear Patterns) จากข้อมูลเชิงพฤติกรรม ซึ่งแบบจำลองดั้งเดิมอาจไม่สามารถจับได้ งานวิจัยนี้สนับสนุนกรอบความคิดในการนำ Deep Learning มาใช้เพื่อวิเคราะห์พฤติกรรมการมีส่วนร่วมที่ซับซ้อนของผู้เข้าร่วมงาน Open House

2.4 งานวิจัยที่เกี่ยวข้องกับการจัดการข้อมูลไม่สมดุล นอกจากการเลือกอัลกอริทึมแล้ว การเตรียมข้อมูลก็มีความสำคัญอย่างยิ่งต่อการสร้างแบบจำลองที่มีประสิทธิภาพ โดยเฉพาะการจัดการกับคลาสที่ไม่สมดุล [6] ได้เสนอเทคนิค SMOTE (Synthetic Minority Over-Sampling Technique) ซึ่งเป็นวิธีการ Oversampling ในระดับข้อมูล (Data Level) ที่มีประสิทธิภาพสูงโดยการสร้างข้อมูลสังเคราะห์สำหรับคลาส minority แทนการทำซ้ำข้อมูลแบบง่าย ๆ เทคนิคนี้ได้รับการยอมรับและประยุกต์ใช้อย่างกว้างขวางในงานวิจัยด้าน EDM งานศึกษานี้จึงได้นำเทคนิค SMOTE มาใช้เพื่อปรับสมดุลของคลาสระหว่าง ผู้สมัคร (Minority Class) และ ผู้ไม่สมัคร (Majority Class) ก่อนกระบวนการฝึกแบบจำลอง

2.5 งานวิจัยที่เกี่ยวข้องกับอัลกอริทึมพื้นฐาน การศึกษานี้ยังได้นำอัลกอริทึมพื้นฐานอื่นๆ มาทดสอบเพื่อใช้เป็นเกณฑ์เปรียบเทียบประสิทธิภาพ โดยอ้างอิงจากหลักการและงานวิจัยดั้งเดิม เช่น Naive Bayes [7] ซึ่งเป็นแบบจำลองที่คำนวณเร็วและมีประสิทธิภาพดีกับข้อมูลขนาดใหญ่จนเกินไป และ Decision Tree [8] ซึ่งเป็นแบบจำลองที่ตีความได้ง่ายและเป็นพื้นฐานของอัลกอริทึม Ensemble อย่าง Random Forest

จากวรรณกรรมที่เกี่ยวข้องดังกล่าว จะเห็นได้ว่ายังมีช่องว่างในการศึกษาการพยากรณ์การสมัครเรียนจากข้อมูลพฤติกรรมที่มีส่วนร่วมในกิจกรรมทางการตลาดแบบการประมวลผลเพื่อแสดงข้อมูลทันที งานวิจัยฉบับนี้จึงมีวัตถุประสงค์เพื่อเติมเต็มช่องว่างดังกล่าว โดยการพัฒนาและเปรียบเทียบแบบจำลองการพยากรณ์หลายๆ แบบจำลอง ด้วยเทคนิคที่ทันสมัย เช่น Deep Learning และ Random Forest พร้อมทั้งจัดการกับปัญหาข้อมูลไม่สมดุลอย่างเหมาะสม เพื่อสร้างองค์ความรู้และเครื่องมือที่สามารถนำไปประยุกต์ใช้ได้จริงสำหรับสถาบันอุดมศึกษา

วิธีดำเนินการวิจัย

1. เครื่องมือการวิจัย

1.1 ระบบลงทะเบียนเข้าออนไลน์ (Online Registration System) เป็นแบบฟอร์มดิจิทัลที่ออกแบบมาเฉพาะสำหรับการลงทะเบียนเข้าร่วมกิจกรรม Open House ระบบสแกน QR Code หรือการค้นหาชื่อเพื่อยืนยันการเข้าร่วมงานจริง ณ จุดลงทะเบียนในวันงาน

1.2 ซอฟต์แวร์สำหรับวิเคราะห์ข้อมูลและสร้างแบบจำลอง RapidMiner Studio ร่วมกับไลบรารีมาตรฐานด้านวิทยาศาสตร์ข้อมูล

1.3 ตัวชี้วัดการประเมินแบบจำลอง ใช้เมตริกมาตรฐานในการประเมินประสิทธิภาพของแบบจำลอง ได้แก่ ความแม่นยำ (Accuracy) ความเที่ยง (Precision) ความครบถ้วน (Recall) ค่า F1-Score และพื้นที่ Under curve (AUC) ของ ROC Curve

2. กลุ่มเป้าหมาย

2.1 ข้อมูลที่ใช้ในการศึกษาได้มาจากกิจกรรม Open House ของมหาวิทยาลัยเอกชนแห่งหนึ่ง ในปีการศึกษา 2567 ข้อมูลการลงทะเบียน จำนวน 1,321 คน (ชื่อ นามสกุล เบอร์โทรศัพท์ โรงเรียน และสาขาที่สนใจ)

2.2 ข้อมูลการเข้าร่วมกิจกรรมจริง จำนวน 564 คน ผู้ที่สแกน QR Code หรือลงทะเบียน ณ วันงาน

3. ขั้นตอนการดำเนินการวิจัย

การวิจัยนี้ดำเนินการตามขั้นตอนอย่างเป็นระบบ เริ่มจากการวางแผนและจัดกิจกรรม Open House ของมหาวิทยาลัยเอกชนแห่งหนึ่งในช่วงวันที่ 15 ตุลาคม – 12 พฤศจิกายน 2567 เพื่อเปิดโอกาสให้นักเรียนมัธยมศึกษาตอนปลายและกลุ่มนักเรียนสายอาชีพศึกษาได้เข้าชมมหาวิทยาลัยและทดลองทำกิจกรรมต่าง ๆ ภายในงาน โดยมีขั้นตอนเพื่อวิจัยดังนี้

3.1 การรวบรวมข้อมูล

- 1) รวบรวมข้อมูลผู้ลงทะเบียนเข้าร่วมกิจกรรมแนะแนวการศึกษา (Open House) จำนวน 1,321 คน
- 2) บันทึกจำนวนผู้เข้าร่วมกิจกรรมจริง 564 คน
- 3) จำนวนผู้สมัครเข้าศึกษาในวันที่ยื่นงานจริง 132 คน
- 4) ระยะเวลาที่ใช้ในงานและจำนวนกิจกรรมย่อยที่เข้าชม เก็บจากการสแกน QR code บันทึก

ข้อมูลแบบ Real-time เช่น เวลาเข้า-ออกกิจกรรมย่อย

3.2 การเตรียมข้อมูล (Data Preparation)

- 1) ตรวจสอบความครบถ้วนและความถูกต้องของข้อมูล
- 2) ทำความสะอาดข้อมูล (เช่น กำจัดข้อมูลซ้ำหรือผิดพลาด)
- 3) แปลงข้อมูลหมวดหมู่เป็นตัวเลขโดยใช้เทคนิค Encoding
- 4) จัดการกับค่าที่ขาดหายและค่าผิดปกติ (Missing values & Outliers)

3.3 การสร้างและฝึกสอนแบบจำลอง (Model Training)

- 1) สร้างแบบจำลองด้วยเทคนิค Naive Bayes
- 2) สร้างแบบจำลองด้วย Logistic Regression
- 3) สร้างแบบจำลองด้วย Deep Learning (เช่น Neural Networks)
- 4) สร้างแบบจำลองด้วย Decision Tree
- 5) สร้างแบบจำลองด้วย Random Forest

3.4 การประเมินผลแบบจำลอง (Model Evaluation)

ใช้ชุดข้อมูลทดสอบประเมินผลแบบจำลองแต่ละประเภท

3.5 การวิเคราะห์ผลและเปรียบเทียบ

เปรียบเทียบประสิทธิภาพของโมเดลแต่ละแบบเพื่อปรับใช้

4. สถิติที่ใช้ในการวิจัย

ในการวิจัยเรื่องการสร้างแบบจำลองการพยากรณ์แนวโน้มการสมัครเข้าศึกษาต่อในมหาวิทยาลัยเอกชน ดังที่ได้นำเสนอไป นักวิจัยได้ใช้เทคนิคทางสถิติและการเรียนรู้ของ Machine Learning หลายประเภท ซึ่งสามารถแบ่งออกได้เป็น 2 กลุ่มหลัก ดังนี้

4.1 สถิติเชิงพรรณนา (Descriptive Statistics)

ใช้เพื่ออธิบายลักษณะพื้นฐานของกลุ่มตัวอย่างก่อนการสร้างแบบจำลอง

ความถี่ (Frequency) และ ร้อยละ (Percentage) ใช้บรรยายลักษณะของกลุ่มตัวอย่าง เช่น

จำนวนผู้ลงทะเบียน (N=1,321 คน)

จำนวนและร้อยละของผู้เข้าร่วมงานจริง 564 คน หรือ ร้อยละ 42.7

จำนวนไม่มาร่วมงาน 625 คน หรือ ร้อยละ 47.31

จำนวนลงทะเบียนและสมัครเข้าศึกษาภายในงาน 132 คน หรือร้อยละ 10

4.2 สถิติเชิงอนุมาน/การวิเคราะห์เพื่อการพยากรณ์ (Inferential/Predictive Statistics) ใช้เพื่อสร้างแบบจำลองและประเมินประสิทธิภาพของการพยากรณ์

4.2.1 เทคนิคก่อนการประมวลผลข้อมูล (Data Preprocessing Techniques)

SMOTE (Synthetic Minority Over-Sampling Technique) เป็นเทคนิคทางสถิติขั้นสูงที่ใช้ในการจัดการกับปัญหาชุดข้อมูลไม่สมดุล (Imbalanced Dataset) ในงานวิจัยนี้ข้อมูลการลงทะเบียนมีจำนวนทั้งสิ้น 1,321 คน แต่มีการกระจายตัวของคลาสที่ไม่สมดุลอย่างมาก โดยมีกลุ่มผู้สมัครเข้าศึกษา เพียง 132 คน หรือคิดเป็นร้อยละ 10 ของทั้งหมด ในขณะที่กลุ่ม ผู้ไม่สมัครมีจำนวนสูงถึง 1,189 คน คิดเป็นร้อยละ 90 ดังนั้นผู้วิจัยจึง

ใช้ SMOTE เพื่อแก้ไขความไม่สมดุลนี้ โดยการสร้างข้อมูลสังเคราะห์ (Synthetic Data) ของคลาส Minority คือ กลุ่ม ผู้สมัคร แทนการเพิ่มความถี่แบบง่าย ๆ (Duplication) เพื่อให้สัดส่วนของแต่ละคลาสสมดุลขึ้นก่อนการฝึกแบบจำลอง ซึ่งช่วยป้องกันไม่ให้แบบจำลองเกิดอคติ (Bias) ไปพยากรณ์กลุ่มไม่สมัคร เทคนิคการแปลงข้อมูล เช่น ชื่อโรงเรียน จังหวัด ให้เป็นรูปแบบ numerical ตัวเลข 0 และ 1 ที่แบบจำลองแมชชีนเลิร์นนิงสามารถเข้าใจได้

4.3 อัลกอริทึมสำหรับการสร้างแบบจำลองการจัดประเภท (Classification Algorithms)

อัลกอริทึมเหล่านี้เป็นแกนของงานวิจัย เป็นอัลกอริทึมทางสถิติและการเรียนรู้ของ Machine Learning ที่ใช้สำหรับการพยากรณ์ผลลัพธ์ 2 กลุ่ม (Binary Classification)

Naive Bayes เป็นแบบจำลองอิงจากพื้นฐานบนทฤษฎีบทของเบย์ (Bayes' Theorem) โดยสมมติว่าคุณลักษณะต่าง ๆ เป็นอิสระต่อกัน (Feature Independence) เป็นโมเดลที่คำนวณเร็วและใช้เป็นแบบจำลองพื้นฐาน

Logistic Regression เป็นอัลกอริทึมเชิงสถิติคลาสสิกที่ใช้สำหรับการพยากรณ์ความน่าจะเป็นของเหตุการณ์สองสถานะ โดยการนำข้อมูลไปสู่ฟังก์ชัน Logistic (S-Curve) ให้ผลลัพธ์เป็นความน่าจะเป็น ระหว่าง 0 ถึง 1 โดยยังไม่ได้จัดประเภทข้อมูล

Decision Tree เป็นอัลกอริทึมที่ใช้โครงสร้างแผนภาพต้นไม้คล้ายกัน Flowchart โดยแต่ละโหนดแสดงถึงการทดสอบบนคุณลักษณะหนึ่ง แต่ละสาขาแสดงถึงผลลัพธ์ของการทดสอบนั้น ทำให้ตีความและอธิบายได้ง่าย

Random Forest เป็นอัลกอริทึมประเภท Ensemble Learning ที่นำหลักการ Strength in Numbers มาใช้ โดยการสร้างแผนผังการตัดสินใจจำนวนมาก (a Forest of Trees) และให้ Trees แต่ละต้นทำนายผล จากนั้นนำผลลัพธ์ทั้งหมดมาโหวตกัน (Voting) เพื่อได้คำตอบสุดท้าย ช่วยลด Overfitting และส่วนใหญ่ให้ความแม่นยำสูงสุด

Deep Learning เป็นโครงข่ายประสาทเทียมเทียมที่มีการทำงานหลายชั้นโหนด (Neurons) ที่ต่อกันอย่างซับซ้อน สามารถเรียนรู้รูปแบบและ เชิงเส้นที่ซับซ้อนในข้อมูล

4.4 การวัดค่าสำหรับประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation Metrics)

เนื่องจากปัญหาข้อมูลไม่สมดุลการใช้เพียงความแม่นยำอย่างเดียวจะต้องใช้เมตริกที่หลากหลายที่เหมาะสมซึ่งเป็นเมตริกทางสถิติดังต่อไปนี้

Accuracy สัดส่วนของการทำนายที่ถูกต้องทั้งหมดต่อจำนวนข้อมูลทั้งหมด

Precision สัดส่วนของการทำนายสมัคร ที่ถูกต้อง จากทั้งหมดที่แบบจำลองทำนายว่าสมัคร

Recall สัดส่วนของการทำนายการสมัครที่ถูกต้อง จากจำนวนผู้สมัครจริงทั้งหมด เมตริกนี้มีความสำคัญ ซึ่งเป็นแนวโน้มที่สมัครเพื่อศึกษาต่อ

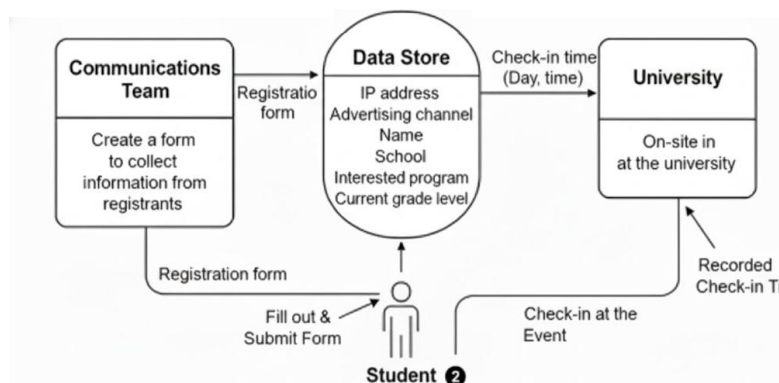
F1-Score เป็นเมตริกที่แสดงถึงความสำคัญของทั้งสองค่า เหมาะสำหรับการประเมินข้อมูล

AUC-ROC (Area Under the Receiver Operating Characteristic Curve) วัดประสิทธิภาพของแบบจำลองในการแยกแยะระหว่างสองคลาส ผู้สมัคร และ ไม่สมัคร โดยค่า AUC ใกล้ค่า 1 แสดงให้เห็นแบบจำลองมีความสามารถในการแยกแยะได้ดีเป็นหนึ่งในมาตรวัดที่ดีที่สุด

ผลการวิจัย

1. ผลการพัฒนา

ผู้วิจัยได้ดำเนินการพัฒนาระบบจากข้อมูลที่ใช้ในการศึกษาได้มาจากกิจกรรม Open House ของมหาวิทยาลัยเอกชนแห่งหนึ่ง ในปีการศึกษา 2567 ข้อมูลการลงทะเบียนเข้าร่วมงาน จำนวน 1,321 คน ตามขั้นตอนการวิจัย โดยนำข้อมูลจากการศึกษา และวิเคราะห์เพื่อจัดทำระบบโมเดลและเครื่องมือของกิจกรรม แสดงดังภาพที่ 1 ตารางที่ 1 และ ตารางที่ 2



ภาพที่ 1 แผนผังการไหลของข้อมูล

จากภาพที่ 1 กระบวนการลงทะเบียนและติดตามผู้สนใจเข้าร่วมกิจกรรมของมหาวิทยาลัย เริ่มต้นขึ้นจากออกแบบและก่อให้เกิด แบบฟอร์มลงทะเบียน (Registration Form) ขึ้นมา แบบฟอร์มนี้ทำหน้าที่ดังประตูบานแรก ที่เชื่อมระหว่างมหาวิทยาลัยกับผู้สนใจที่มาร่วมงาน โดยออกแบบมาเพื่อรวบรวมข้อมูลส่วนบุคคลที่สำคัญ

เมื่อถึงเวลาที่กำหนด นักเรียนซึ่งลงทะเบียนไว้แล้วก็จะเดินทางมาเยือนมหาวิทยาลัยจริง ๆ ณ จุดนี้ มหาวิทยาลัย (University) จะดำเนินการ เช็คอิน (Check in) ให้กับนักเรียน การเช็คอินนี้เปรียบเสมือนการยืนยันตัวตนและการมาร่วมกิจกรรมตรวจสอบผ่านการสแกนคิวอาร์โค้ดหรือการค้นหาชื่อในระบบฐานข้อมูลลงทะเบียน

ขณะที่เกิดการเช็คอิน ระบบจะทำการส่งสิ่งสำคัญอย่างรวดเร็ว นั่นคือการดึงข้อมูลของนักเรียนคนนั้น ๆ จากคลังข้อมูลขึ้นมา พร้อมทั้งบันทึก เวลาในการเช็คอิน (Check-in Time) ทั้งวันและเวลาที่แน่นอนเพิ่มลงไป ในประวัติของนักเรียน ทุกการกระทำนี้ล้วนส่งผ่าน การไหลของข้อมูล (Data Flow) ที่ต่อเนื่องกันโดยมีรายละเอียดตามตารางที่ 1 ดังนี้

ตารางที่ 1 แสดงลักษณะทางสถิติของชุดข้อมูลเบื้องต้น

รายละเอียดของข้อมูล	จำนวน (คน)	ร้อยละ (Percentage)
จำนวนผู้เข้าร่วมงานจริง	564	42.70
จำนวนลงทะเบียนและสมัครเข้าศึกษาภายในงาน	132	10.00
จำนวนที่ไม่มาร่วมงาน	625	47.31
จำนวนผู้ลงทะเบียนทั้งหมด	1,321	100.00

การศึกษานี้ใช้ข้อมูลจากกิจกรรมแนะแนวการศึกษาของมหาวิทยาลัยเอกชนแห่งหนึ่ง โดยมีผู้ลงทะเบียนเข้าร่วมกิจกรรมจำนวนทั้งสิ้น 1,321 คน ในจำนวนนี้มีผู้เข้าร่วมกิจกรรมจริง 564 คน หรือคิดเป็นร้อยละ 42.7 ของผู้ลงทะเบียนทั้งหมด และมีนักเรียนที่ลงทะเบียนและสมัครเข้าศึกษาภายในงาน 132 คน ซึ่งคิดเป็นร้อยละ 10 ของผู้ลงทะเบียนทั้งหมด และเป็นร้อยละ 23.4 ของผู้เข้าร่วมงานจริง จำนวนไม่มาร่วมงาน 625 คน คิดเป็นร้อยละ 43.31 ข้อมูลเหล่านี้ถูกนำมาใช้เพื่อสร้างแบบจำลองการจำแนกประเภทแบบไบนารี โดยคลาสบวก (Positive Class) คือ ผู้สมัครเข้าศึกษา และคลาสลบ (Negative Class) คือ ผู้ไม่สมัครเข้าศึกษา จากข้อมูลดังกล่าวจะเห็นว่ามีการกระจายตัวของคลาสที่ไม่สมดุลอย่างมาก (Class imbalance) ซึ่งเป็นความท้าทายสำคัญที่ส่งผลต่อการประเมินประสิทธิภาพของแบบจำลอง

2. ผลการทดลองใช้ระบบวิเคราะห์แต่ละโมเดล

ผู้วิจัยดำเนินการทดลองใช้ระบบบัจล่องและทดสอบบนชุดข้อมูลทดสอบที่พัฒนาขึ้น จากข้อมูลที่ใช้ในการศึกษาได้มาจากกิจกรรม Open House ของมหาวิทยาลัยเอกชนแห่งหนึ่ง ในปีการศึกษา 2567 โดยมีข้อมูลการลงทะเบียนจำนวน 1,321 คน

ประเมินประสิทธิภาพของแบบจำลองโดยใช้เมตริกหลักได้แก่ ความแม่นยำ (Accuracy) ความเที่ยง (Precision) ความครบถ้วน (Recall) F1-Score และพื้นที่ under curve (AUC) ของ ROC Curve เน้นค่า Recall และ AUC ชุดข้อมูลไม่สมดุล สรุปผลผลดังตารางที่ 2

ตารางที่ 2 ผลการประเมินประสิทธิภาพของแบบจำลองต่าง ๆ

แบบจำลอง	ความแม่นยำ (Accuracy)	ความเที่ยง (Precision)	ความครบถ้วน (Recall)	F1-Score	ROC AUC
Naive Bayes	0.85	0.52	0.79	0.63	0.82
Logistic Regression	0.91	0.74	0.85	0.79	0.92
Decision Tree	0.89	0.69	0.77	0.73	0.88
Random Forest	0.93	0.81	0.89	0.85	0.96
Deep Learning	0.94	0.83	0.92	0.87	0.98

3. การนำเสนอ Feature Importance

เพื่อระบุปัจจัยเชิงพฤติกรรมของผู้เข้าร่วมงาน Open House ที่มีความสำคัญที่สุดในการพยากรณ์การตัดสินใจสมัครเข้าศึกษา ผู้วิจัยได้วิเคราะห์ค่า Feature Importance จากแบบจำลอง Random Forest โดยผลการวิเคราะห์แสดงดังตารางที่ 3

ตารางที่ 3 แสดงค่าความสำคัญของปัจจัย (Feature Importance) จากแบบจำลอง Random Forest

ปัจจัย (Feature)	ค่าความสำคัญ (Importance Score)	ข้อสรุปเชิงพฤติกรรม
ระยะเวลาที่ใช้ในงาน (Total Time Spent in Event)	0.38	นักเรียนใช้เวลาอยู่ในงานนาน มีแนวโน้มที่จะสมัคร
จำนวนกิจกรรมย่อยที่เข้าชม (Number of Sub-Events Attended)	0.25	นักเรียนเข้าร่วมกิจกรรมย่อยมาก แสดงถึงความสนใจจริง
โรงเรียนที่ผู้เข้าร่วมสังกัด (School Attended)	0.15	สถานที่ตั้งหรือความร่วมมือระหว่างโรงเรียน และมหาวิทยาลัยผลต่อการรับรู้ของนักเรียน
สาขาที่ลงทะเบียนสนใจ (Registered Interested Program)	0.10	ความชัดเจนในสาขาที่เลือกตั้งแต่แรกมีผลต่อการตัดสินใจ
ช่องทางการลงทะเบียน (Registration Channel)	0.08	ช่องทาง (เช่น การแนะนำจากอาจารย์, โซเชียลมีเดีย) อาจบ่งชี้ความตั้งใจ
ระยะเวลาในการลงทะเบียนล่วงหน้า (Pre-Registration Days)	0.04	ผู้ที่ลงทะเบียนล่วงหน้าอย่างกระชั้นชิดหรือนานมากอาจมีพฤติกรรมต่างกัน

จากตารางที่ 3 พบว่า ปัจจัยที่มีความสำคัญสูงสุดในการทำนายแนวโน้มการสมัครเข้าศึกษาคือ ระยะเวลาที่ใช้ในงาน (Total Time Spent in Event) และ จำนวนกิจกรรมย่อยที่เข้าชม (Number of Sub-Events Attended) โดยปัจจัยทั้งสองนี้เป็นตัวแปรที่สะท้อนถึง พฤติกรรมการมีส่วนร่วมและความตั้งใจจริงภายในงาน

(In-event Engagement and Intent) ซึ่งให้ค่าความสำคัญรวมกันสูงกว่าปัจจัยด้านประชากรศาสตร์ (เช่น โรงเรียนหรือสาขาที่สนใจ) อย่างมีนัยสำคัญ

ผลการวิเคราะห์นี้ตอกย้ำว่า ความลึกและความหลากหลายของการมีส่วนร่วม (Engagement Depth) มีความสัมพันธ์โดยตรงกับโอกาสในการเปลี่ยนจากผู้สนใจมาเป็นผู้สมัครเรียน ผู้เข้าร่วมที่ใช้เวลาอยู่ในงานนานและเข้าเยี่ยมชมกิจกรรมบ่อยจำนวนมากแสดงให้เห็นถึงความสนใจที่สูงและมีแนวโน้มที่จะตัดสินใจสมัครเรียนในอัตราที่สูงขึ้น

อภิปรายผลการวิจัย

การวิเคราะห์ปัจจัยที่ส่งผลต่อการเข้าร่วมกิจกรรมและแนวทางการศึกษา Open House พบว่า ผลการวิจัยแสดงให้เห็นว่า พฤติกรรมการมีส่วนร่วมของผู้เข้าร่วมกิจกรรมและแนวทางการศึกษา Open House มีอิทธิพลอย่างมีนัยสำคัญต่อการตัดสินใจสมัครเข้าศึกษาต่อในมหาวิทยาลัยเอกชน โดยเฉพาะ ระยะเวลาที่ใช้ภายในงาน และ จำนวนกิจกรรมย่อยที่เข้าชม ซึ่งมีค่า Feature Importance สูงสุด (0.38 และ 0.25 ตามลำดับ) จากแบบจำลอง Random Forest ปัจจัยทั้งสองนี้สะท้อนระดับความมีส่วนร่วมเชิงลึก (Engagement Depth) และความตั้งใจจริงในการสมัครเรียน (Behavioral Intention) ผลลัพธ์ดังกล่าวสอดคล้องกับแนวคิดด้าน Engagement Based Decision Model ที่ชี้ว่า ความถี่และความลึกของการมีส่วนร่วมเป็นตัวทำนายที่มีน้ำหนักต่อการตัดสินใจของผู้บริโภค [4]

นอกจากนี้ การเข้าร่วมกิจกรรมจริงยังมีความสัมพันธ์เชิงบวกกับการตัดสินใจสมัครเรียน โดยกลุ่มผู้เข้าร่วมกิจกรรมจริงมีอัตราการสมัครเรียนสูงถึงร้อยละ 23.4 เทียบกับเพียงร้อยละ 10 ของผู้ลงทะเบียนทั้งหมด ซึ่งสะท้อนถึงอิทธิพลของประสบการณ์ตรง (Experiential Marketing) ที่ช่วยสร้างความเชื่อมั่นและลดความไม่แน่นอนในการตัดสินใจสมัครเรียน การได้สัมผัสสภาพแวดล้อมจริงและปฏิสัมพันธ์กับบุคลากรของมหาวิทยาลัยมีผลต่อความเชื่อมโยงทางอารมณ์ (Emotional Connection) และความตั้งใจในการสมัครเรียนในเชิงพฤติกรรม สอดคล้องกับผลการศึกษาของ [2] ที่พบว่า การมีปฏิสัมพันธ์โดยตรงกับสถาบันส่งผลต่อความเป็นไปได้ในการสมัครเรียนอย่างมีนัยสำคัญ ในด้านการวิเคราะห์เชิงเทคนิค แบบจำลอง Deep Learning ให้ประสิทธิภาพสูงสุด (Accuracy = 0.94, ROC AUC = 0.98) รองลงมาคือ Random Forest (Accuracy = 0.93, ROC AUC = 0.96). ความสามารถของ Deep Learning ในการตรวจจับรูปแบบเชิงซับซ้อนและไม่เป็นเชิงเส้น (Non-linear Relationships) ทำให้สามารถเรียนรู้ความสัมพันธ์ระหว่างคุณลักษณะพฤติกรรม เช่น ระยะเวลาและจำนวนกิจกรรมย่อย ได้อย่างมีประสิทธิภาพสูง ซึ่งแบบจำลองเชิงเส้น เช่น Logistic Regression ไม่สามารถจับได้ครบถ้วน ผลนี้สอดคล้องกับการศึกษาของ [3] ที่ระบุว่า Deep Learning เหมาะอย่างยิ่งสำหรับการทำนายพฤติกรรมของผู้เรียนในสภาพแวดล้อมข้อมูลที่ซับซ้อน การที่แบบจำลอง Random Forest แสดงประสิทธิภาพใกล้เคียงกับ Deep Learning สอดคล้องกับผลการศึกษาของ [4] ที่แสดงให้เห็นว่าเทคนิค Ensemble Learning สามารถจัดการกับข้อมูลไม่สมดุล (Imbalanced Data) ได้อย่างมีประสิทธิภาพและให้ผลการทำนายที่มั่นคง โดยเฉพาะในข้อมูลเชิงพฤติกรรมของผู้เรียน

โดยสรุป การอภิปรายผลนี้ชี้ให้เห็นว่า การบูรณาการเทคโนโลยีการเรียนรู้ของ Machine Learning เข้ากับข้อมูลพฤติกรรมเชิงลึกจากกิจกรรม Open House ไม่เพียงเพิ่มความแม่นยำในการพยากรณ์แนวโน้มการสมัครเรียน แต่ยังสร้างฐานข้อมูลเชิงกลยุทธ์ที่สามารถนำไปใช้ในการออกแบบกิจกรรมและแนว การสื่อสารเชิงกลุ่มเป้าหมาย และการบริหารทรัพยากรด้านการตลาดการศึกษาอย่างมีประสิทธิภาพสำหรับมหาวิทยาลัยเอกชนในอนาคต

ข้อเสนอแนะ

การศึกษานี้สำเร็จในการพัฒนาแบบจำลองการพยากรณ์แนวโน้มการสมัครเรียนโดยอ้างอิงพฤติกรรมผู้เข้าร่วมกิจกรรม Open House โดยแสดงให้เห็นว่าแบบจำลอง Deep Learning และ Random Forest มีศักยภาพสำหรับงานนี้ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

- 1) เพิ่มจำนวนคุณลักษณะ โดยเฉพาะข้อมูลหลังงาน เช่น การติดตามผลผ่านโซเชียลมีเดีย หรืออีเมล
- 2) ทดสอบกับชุดข้อมูลจากหลายมหาวิทยาลัยเพื่อเพิ่ม ความ generalize ของแบบจำลอง
- 3) พัฒนาระบบแนะนำอย่าง Real-time integrated with mobile application during the event.
- 4) สำรวจเทคนิคอื่น ๆ เช่น Gradient Boosting (XGBoost, LightGBM) หรือการทำ Natural Language Processing (NLP) บนข้อมูล feedback จากผู้เข้าร่วมงาน

เอกสารอ้างอิง

- [1] E. Agyemang et al., "Predicting Students' Academic Performance Via Machine Learning Algorithms: An Empirical Review and Practical Application," *Computer Engineering and Intelligent Systems*, vol. 15, pp. 86–102, Oct. 2024, doi: 10.7176/CEIS/15-1-09.
- [2] O. Lei et al., "A Logistic Regression Model to Predict Graduate Student Matriculation," *Journal of International Education and Practice*, vol. 4, pp. 24–34, Oct. 2021, doi: 10.30564/jiep.v4i1and2.2628.
- [3] B. Alnasyan, M. Basher, and M. Alassafi, "The Power of Deep Learning Techniques for Predicting Student Performance in Virtual Learning Environments: A Systematic Literature Review," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100231, 2024, doi: 10.1016/j.caeai.2024.100231.
- [4] S. C. Matz et al., "Using Machine Learning to Predict Student Retention from Socio-Demographic Characteristics and App-Based Engagement Metrics," *Scientific Reports*, vol. 13, no. 1, p. 5705, 2023, doi: 10.1038/s41598-023-32484-w.
- [5] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, Nov. 2010.
- [6] N. V. Chawla et al., "SMOTE: Synthetic Minority Over-Sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [7] I. Rish, "An Empirical Study of the Naive Bayes Classifier," in *Proc. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, Seattle, WA, USA, vol. 3, no. 22, pp. 41–46, 2001.
- [8] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.