JSSE
JOURNAL OF SCIENCE & SCIENCE EDUCATION
Faculty of Science, Ubon Ratchathani University
Vol. 7 No. 1 (Jan – Jun 2024): 10 – 20
http://doi.org/10.14456/jsse.2024.2

Research Article

# Enrollment confirmation prediction for prospective university students by applying deep learning techniques

**Pattadon Jaisin[1], Supawadee Hiranpongsin[2] and Phaichayon Kongchai[2,*]**

*[1]M.Sc. Student, Major of Information Technology, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani*

*[2]Department of Mathematics, Statistics and Computer, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani*

**Email: phaichayon.k@ubu.ac.th*

Received <1 January 2023 >; Revised < 20 January 2024 >; Accepted < 30 January 2024>

## Abstract

The objectives of this research are 1) to study the models of enrollment confirmation prediction for prospective university students, 2) to study and find the appropriate parameter values to enhance the efficiency of the models, 3) to compare the model performance of deep learning techniques and traditional machine learning techniques for predicting the university students' admission confirmation. The applied models include Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Random Forest, Naive Bayes, and Support Vector Machine (SVM). The samples are 4,479 records of students who applied for admission to universities during the academic year 2018-2021. The modeling of traditional machine learning and deep learning uses 11 features consisting of prefixes, schools, applicant types, and 8-subjects grades. The performance comparison between the two models revealed that the deep learning model outperformed the traditional machine learning model. Therefore, a study was conducted to find suitable parameters to enhance the efficiency of the deep learning model. The results of the study found that the CNN model had the highest accuracy at 64.68 percent, and the LSTM needed to be more suitable due to overfitting. In conclusion, the CNN was suitable for applying to the system of enrollment confirmation prediction for prospective university students.

**Keywords**: Deep Learning, Convolutional Neural Network, Long Short-Term Memory, Enrollment Confirmation, University

บทความวิจัย

# วิธีทำนายการยืนยันสิทธิเข้าศึกษาต่อของนักศึกษามหาวิทยาลัยด้วยการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึก

**พัทธดนย์ จัยสิน[1] สุภาวดี หิรัญพงศ์สิน[2] และไพชยนต์ คงไชย[2,*]**

*[1]นักศึกษาหลักสูตรวิทยาศาสตร์มหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี จังหวัดอุบลราชธานี*

*[2]ภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี จังหวัดอุบลราชธานี*

**Email: phaichayon.k@ubu.ac.th*

รับบทความ: 1 ธันวาคม 2566 แก้ไขบทความ: 20 มกราคม 2567 ยอมรับตีพิมพ์: 30 มกราคม 2567

## บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ 1) เพื่อศึกษาโมเดลทำนายการยืนยันสิทธิเข้าศึกษาต่อของนักศึกษามหาวิทยาลัย 2) เพื่อศึกษาหาค่าพารามิเตอร์ที่เหมาะสมกับการนำไปเพิ่มประสิทธิภาพของโมเดล 3) เพื่อเปรียบเทียบประสิทธิภาพระหว่างเทคนิคการเรียนรู้เชิงลึกและเทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิมในการทำนายการยืนยันสิทธิเข้าศึกษาต่อของนักศึกษามหาวิทยาลัย ซึ่งประกอบด้วย โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network) หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) แรนดอมฟอเรสต์ (Random Forest) นาอีฟเบย์ (Naive Bayes) และซัพพอร์ทเวกเตอร์แมทชีน (Support Vector Machine) กลุ่มตัวอย่างทดสอบ คือ ข้อมูลนักเรียนที่สมัครเข้าเรียนมหาวิทยาลัยในปีการศึกษา 2561-2564 จำนวน 4,479 คน การสร้างโมเดลการเรียนรู้ของเครื่องแบบดั้งเดิมและการเรียนรู้เชิงลึกใช้ 11 คุณลักษณะ ได้แก่ คำนำหน้าชื่อ โรงเรียน ประเภทผู้สมัคร และเกรดรายวิชา 8 รายวิชา ผลการเปรียบเทียบประสิทธิภาพระหว่างโมเดลทั้งสอง พบว่า โมเดลการเรียนรู้เชิงลึกมีประสิทธิภาพดีกว่าโมเดลการเรียนรู้ของเครื่องแบบดั้งเดิม จึงทำการศึกษาหาค่าพารามิเตอร์ที่เหมาะสมเพื่อเพิ่มประสิทธิภาพโมเดลการเรียนรู้เชิงลึก ผลการศึกษา พบว่า โมเดลที่ดีที่สุดคือโครงข่ายประสาทแบบคอนโวลูชันโดยมีค่าความถูกต้องสูงถึงร้อยละ 64.68 และหน่วยความจำระยะยาว-ระยะสั้นมีสภาพเป็น overfitting ซึ่งไม่เหมาะสำหรับการนำไปใช้งาน สรุปได้ว่า โมเดลโครงข่ายประสาทแบบคอนโวลูชันเหมาะสมที่จะนำไปใช้กับระบบทำนายการยืนยันสิทธิเข้าศึกษาต่อของนักศึกษามหาวิทยาลัย

**คำสำคัญ**: เทคนิคการเรียนรู้เชิงลึก โครงข่ายประสาทแบบคอนโวลูชัน หน่วยความจำระยะยาว-ระยะสั้น การยืนยันสิทธิเข้าศึกษาต่อ มหาวิทยาลัย

## บทนำ

ปัจจุบันประเทศไทยมีอัตราการเกิดใหม่ของประชากรที่ลดน้อยลง ส่งผลให้นักเรียนในระบบการศึกษามีจำนวดลดลง มหาวิทยาลัยหลายแห่งจึงมีการปรับแผนการรับสมัครและการประชาสัมพันธ์นักศึกษาใหม่ในรูปแบบเชิงรุกมากขึ้น แต่เนื่องจากหลักสูตรภายในมหาวิทยาลัยไม่มีข้อมูลปัจจัยหรือเหตุผลของนักเรียนที่จะตัดสินใจเลือกเข้าศึกษาต่อ ทำให้การปรับแผนการรับสมัครและการประชาสัมพันธ์เป็นเรื่องยาก อีกทั้งยังส่งผลให้มหาวิทยาลัยสูญเสียทรัพยากรและรายได้ เพื่อให้มีการปรับแผนการรับสมัครให้มีความเหมาะสมและสอดคล้องกับสถานการณ์ จึงจำเป็นที่จะต้องมีเครื่องมือมาช่วยในการสนับสนุนการตัดสินใจเพื่อให้เกิดประโยชน์สูงสุดในด้านการบริหารจัดการทรัพยากร

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่า งานวิจัยเพื่อหาปัจจัยที่เกี่ยวข้องในการศึกษาต่อระดับปริญญาตรี มี 2 ประเภท คืองานวิจัยที่ใช้สถิติในการวิจัย ซึ่งมักจะใช้แบบสอบถามในการเก็บรวบรวมข้อมูล โดยจะสุ่มกลุ่มประชากรตัวอย่างและเน้นการหาความสัมพันธ์ของปัจจัยที่เกี่ยวข้อง และงานวิจัยที่ใช้เทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิมในการวิจัย ซึ่งจะใช้ข้อมูลที่มีอยู่ในฐานข้อมูล โดยจะมีขั้นตอนในการหาปัจจัยที่เกี่ยวข้องจากประชากรทั้งหมดแล้วนำมาสร้างโมเดล จากนั้นจึงประเมินประสิทธิภาพซึ่งผลลัพธ์จะต้องอยู่ในระดับที่ยอมรับได้ของงานนั้น ๆ ทำให้งานวิจัยประเภทนี้ได้รับความนิยมและสามารถนำไปประยุกต์ใช้งานต่าง ๆ ตัวอย่างเช่น งานวิจัยของ Tongpuang and Sanrach (2021) ได้เสนอวิธีการเปรียบเทียบการจำแนกข้อมูลสำหรับทำนาย การได้รับทุนการศึกษาของนักศึกษาระดับปริญญาตรีโดยใช้วิธี Information Gain ในการหาปัจจัยที่ส่งผลต่อการสร้างโมเดลที่มีประสิทธิภาพมากที่สุด และทำการเปรียบเทียบประสิทธิภาพของ 5 อัลกอริทึม ได้แก่ แรนดอมฟอเรสต์ (Random Forest) ต้นไม้ตัดสินใจ (Decision Tree) นาอีฟเบย์ (Naïve Bayes) เคเนียร์เรสเนเบอร์ (K-Nearest Neighbors) และการเรียนรู้เชิงลึก (Deep Learning) จากนั้นจึงประเมินประสิทธิภาพด้วยค่าความถูกต้อง (Accuracy) พบว่าแรนดอมฟอเรสต์ มีค่าความถูกต้องสูงถึงร้อยละ 94.28 งานวิจัยของ Suepitak, Nissaidee and Buathong (2021) ได้เสนอวิธีการเปรียบเทียบการจำแนกข้อมูลสำหรับทำนายแนวโน้มการสำเร็จการศึกษาของนักเรียน ด้วยวิธีการประเมินประสิทธิภาพของโมเดล 3 แบบ คือ การสุ่มด้วยการแบ่งเป็นร้อยละ (Percentage Split) การตรวจสอบแบบข้ามจำนวน (Cross Validation) 5 กลุ่ม และ 10 กลุ่ม ที่ร้อยละ 66 จาก 3 อัลกอริทึม ได้แก่ ต้นไม้ตัดสินใจ นาอีฟเบย์ และ เคเนียร์เรสเนเบอร์ ผลการทดลองพบว่า อัลกอริทึมต้นไม้ตัดสินใจมีประสิทธิภาพดีที่สุด ซึ่งผลการประเมินโมเดลด้วยวิธีการตรวจสอบแบบข้ามจำนวน 5 กลุ่ม และ 10 กลุ่ม มีค่าความถูกต้องเท่ากัน คือ ร้อยละ 87.40

งานวิจัยของ Nandal (2020) ได้เสนอวิธีการเรียนรู้เชิงลึกสำหรับทำนายการรับเข้าเรียนของนักเรียน โดยใช้คะแนนสอบในแต่ละด้านของนักศึกษามาใช้ในการสร้างโมเดล แล้วทำการเปรียบเทียบประสิทธิภาพอัลกอริทึมทั้ง 3 อัลกอริทึม ได้แก่ อัลกอริทึมเชิงเส้น (Linear Algorithms) อัลกอริทึมจำแนกข้อมูล (Classification Algorithms) และ โครงข่ายประสาทเชิงลึก (Deep Neural Networks) ผลการทดลองพบว่า โครงข่ายประสาทเชิงลึก มีค่าความถูกต้องเฉลี่ยที่ร้อยละ 95.10 ซึ่งสูงกว่าอัลกอริทึมจำแนกข้อมูล ที่มีค่าความถูกต้องร้อยละ 92.10 จากเหตุผลดังกล่าวข้างต้น ผู้วิจัยจึงได้เสนอการพัฒนาระบบทำนายการยืนยันสิทธิเข้าศึกษาต่อในระดับปริญญาตรีของนักศึกษาใหม่ โดยใช้ข้อมูลนักเรียนที่มาสมัครเข้าเรียนในคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี ด้วยเทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิมและการเรียนรู้เชิงลึก เพื่อสร้างแบบจำลองสำหรับทำนายการยืนยันสิทธิและไม่ยืนยันสิทธิการเข้าเรียนของนักศึกษาใหม่ โดยใช้อัลกอริทึม แรนดอมฟอเรสต์ นาอีฟเบย์ และซัพพอร์ทเวกเตอร์แมทชีน (Support Vector Machine) โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network) หน่วยความจำระยะยาว-ระยะสั้น (Long Short-Term Memory) นอกจากนี้ผู้วิจัยได้ศึกษาปรับหาค่าพารามิเตอร์ที่เหมาะสมและทำการเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้เชิงลึก เพื่อให้ได้โมเดลที่มีประสิทธิภาพที่ดีที่สุดในการใช้งาน ดังนั้นวิธีการที่นำเสนอนี้สามารถนำไปทำงานร่วมกับระบบรับเข้า เพื่อสนับสนุนการตัดสินใจการบริหารจัดการทรัพยากรของการปรับแผนการรับสมัครและการประชาสัมพันธ์นักศึกษาใหม่ในรูปแบบเชิงรุกให้มีความเหมาะสมและสอดคล้องกับสถานการณ์ในปัจจุบัน

## วัตถุประสงค์การวิจัย

1. เพื่อศึกษาโมเดลทำนายการยืนยันสิทธิเข้าศึกษาต่อของนักศึกษามหาวิทยาลัย
2. เพื่อศึกษาหาค่าพารามิเตอร์ที่เหมาะสมกับการนำไปเพิ่มประสิทธิภาพของโมเดล
3. เพื่อเปรียบเทียบประสิทธิภาพระหว่างเทคนิคการเรียนรู้เชิงลึกและเทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิม

## วิธีดำเนินการวิจัย

งานวิจัยนี้ได้แบ่งขั้นตอนการดำเนินการวิจัยออกเป็น 4 ขั้นตอนดังนี้

**การทำความเข้าใจข้อมูล (Data Understanding)**

การสร้างโมเดลการเรียนรู้ของเครื่องแบบดั้งเดิมและโมเดลการเรียนรู้เชิงลึกของงานวิจัยนี้ ผู้วิจัยใช้ข้อมูลการรับเข้าของนักเรียนที่จะสมัครเข้าเรียนคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี ในปีการศึกษา 2561 – 2564 จำนวน 4,479 คน แล้วทำการศึกษาข้อมูล ของนักเรียนเพื่อให้เข้าใจถึงความหมายของข้อมูล และเป็นการกรองข้อมูลก่อนนำไปทำการแปลงข้อมูล ตัวอย่างแอททริบิวต์ข้อมูลรับเข้า แสดงดังตารางที่ 1

**ตารางที่ 1** ตัวอย่างแอททริบิวต์ข้อมูลรับเข้า

| ลำดับ | แอททริบิวต์ | คำอธิบาย | ตัวอย่างข้อมูล |
|---|---|---|---|
| 1 | APPLICANTID | หมายเลขผู้สมัคร | 45987, 45893, 45280, 45282, 45999,.... |
| 2 | PREFIXNAME | คำนำหน้าชื่อ | นาย, นาง, นางสาว |
| 3 | APPLICANTSTATUS | สถาณะการสมัคร | ผ่านคัดเลือกและชำระเงินยืนยันสิทธิ์เข้าศึกษา, ....(11), คนสมัครที่มีสิทธิ์สัมภาษณ์ |
| 4 | ACADYEAR | ปีการศึกษาที่สมัคร | 2561, 2562, 2563, 2564 |
| 5 | APPLICANTTYPENAME | ประเภทรอบการรับสมัคร | Quota1, Quota2, Protfolio1, ...(6), Protfolio2 |
| 6 | รอบ TCAS | รอบ TCAS ที่สมัคร | 1, 2, 3, 4, 5 |
| 7 | SCHOOLNAME | ชื่อโรงเรียน | นารีนุกูล, เขมราฐพิทยาคม, ...(721), อำนาจเจริญ |
| 8 | PROVINCENAME | จังหวัด | มหาสารคาม, อุบลราชธานี, ...(67), กาญจนบุรี |
| 9 | QUOTANAME | โควตาที่สมัคร | คณิตศาสตร์, จุลชีววิทยา, ...(7), เคมี |
| 10 | ภาษาไทย | เกรดวิชาภาษาไทย | 2.56, 3.81, 3.34, 2.79, 3.53, .... |
| 11 | คณิตศาสตร์ | เกรดวิชาคณิตศาสตร์ | |
| 12 | วิทยาศาสตร์ | เกรดวิชาวิทยาศาสตร์ | |
| 13 | สังคมศึกษา | เกรดวิชาสังคมศึกษา | |
| 14 | สุขศึกษา | เกรดวิชาสุขศึกษา | |
| 15 | ศิลปะ | เกรดวิชาศิลปะ | |
| 16 | การงานอาชีพ | เกรดวิชาการงานอาชีพ | |
| 17 | ภาษาต่างประเทศ | เกรดภาษาต่างประเทศ | |
| 18 | ENTRYGPAX | เกรดรวม | |
| 19 | สถานะผู้สมัคร | สถานะการยืนยันสิทธิ์ของผู้สมัคร | ไม่เลือกเข้าเรียน, เลือกเข้าเรียน |

**การแปลงข้อมูล (Data Transformation)**

ในขั้นตอนนี้ ผู้วิจัยจะทำการแปลงข้อมูลในแต่ละคอลัมน์ให้อยู่ในรูปแบบตัวแทนของข้อมูลได้แก่คอลัมน์ SCHOOLNAME, PROVINCENAME และตัดคอลัมน์ที่ไม่ได้ใช้งานออกได้แก่ APPLICANTID, APPLICANTSTATUS, ACADYEAR, QUOTANAME เพื่อให้เหมาะสมกับการนำไปสร้างโมเดลการเรียนรู้ของเครื่องแบบดั้งเดิมและโมเดลการเรียนรู้เชิงลึก ตัวอย่างข้อมูลที่ทำการแปลงแล้ว แสดงดังตารางที่ 2

**การสร้างโมเดล (Modeling)**

1. การสร้างโมเดลด้วยวิธีการทำการเรียนรู้ของเครื่องแบบดั้งเดิม ในขั้นตอนนี้ผู้วิจัยได้ทำการสร้างโมเดลการเรียนรู้ของเครื่องแบบดั้งเดิมด้วย เทคนิคแรนดอมฟอเรสต์ นาอีฟเบย์ และ ซัพพอร์ทเวกเตอร์แมทชีน

1.1 แรนดอมฟอเรสต์ (Palwisut, 2016) เป็นเทคนิคที่ใช้ในการสร้างต้นไม้ตัดสินใจ (Decision Tree) แบบอิสระจากกัน โดยการสุ่มเลือกข้อมูลและคุณลักษณะจากชุดข้อมูลแล้วสร้างต้นไม้ตัดสินใจหลายต้น โดยในกระบวนการสร้างแต่ละต้นมีการนำข้อมูลไปสุ่มเลือกตัวอย่างแบบเลือกแล้วใส่กลับและมีบางส่วนของข้อมูลที่ไม่ถูกเลือก ส่วนข้อมูลที่ไม่ถูกเลือกนี้เรียกว่า Out-of-Bag (OOB)

เทคนิคแรนดอมฟอเรสต์ จะนำผลลัพธ์ที่ได้จากแต่ละต้นไม้ตัดสินใจมาคิดเป็นผลโหวต (Voting) และเลือกผลโหวตที่มีความถี่มากที่สุดเพื่อระบุสถานะหรือคลาสของข้อมูล ความเป็นอิสระของต้นไม้ตัดสินใจแต่ละต้นที่สร้างขึ้นทำให้

เทคนิคแรนดอมฟอเรสต์ มีความสามารถในการประมาณความผิดพลาดโดยไม่จำเป็นต้องใช้ชุดข้อมูลทดสอบเพิ่มเติม เนื่องจากข้อมูล OOB ถูกนำมาใช้ในกระบวนการทดสอบต้นไม้ตัดสินใจแต่ละต้นแทน ดังสมการที่ 1

$$Gini\ Impurity\ (node) = 1 - \sum(p(C)\text{^}2) \quad (1)$$

โดยที่

p(C) คือ ความถี่ของคลาส C ในโหนดนั้น

และคำสั่งในการสร้างโมเดลด้วยเทคนิคแรนดอมฟอเรสต์ สามารถอธิบายได้ดังนี้ n_estimators=100 คือการสร้างโมเดล Random Forest Classifier ที่มี 100 ต้นไม้ และ random_state=seed คือการกำหนด seed สำหรับการสุ่ม ดังแสดงในภาพที่ 1

```
ramdom_forest = RandomForestClassifire(n_estimitor=100, randim_state=seed)
```

**ภาพที่ 1** การสร้างโมเดลด้วยการเรียนรู้ของเครื่องแบบดั้งเดิมด้วยเทคนิค แรนดอมฟอเรสต์

**ตารางที่ 2** ตัวอย่างข้อมูลที่ทำการแปลง

| ลำดับ | แอททริบิวต์ | คำอธิบาย | ตัวอย่างข้อมูลที่ทำการแปลง |
|---|---|---|---|
| 1 | PREFIXNAME | คำนำหน้าชื่อ | 0 คือ นาย, 1 คือ นาง, 2 คือ นางสาว |
| 2 | APPLICANTTYPENAME | ประเภทรอบการรับสมัคร | 1 คือ Quota1, 2 คือ Quota2, 3 คือ Protfolio1, 4 คือ Protfolio2, 5 คือ Protfolio3, .... |
| 3 | รอบ TCAS | รอบ TCAS ที่สมัคร | 1, 2, 3, 4, 5 |
| 4 | SCHOOLNAME | ชื่อโรงเรียน | 677 คือ นารีนุกูล, 162 คือ เขมราฐพิทยาคม, .... |
| 5 | PROVINCENAME | จังหวัด | 33 คือ มหาสารคาม, 58 คือ อุบลราชธานี, .... |
| 6 | Thai_Grade | เกรดวิชาภาษาไทย | 0 คือ Grade = 4.00,<br>1 คือ 3.50 <= Grade < 4.00,<br>2 คือ 3.00 <= Grade < 3.49,<br>3 คือ 2.50 <= Grade < 2.99,<br>4 คือ 2.00 <= Grade < 2.49,<br>5 คือ 1.50 <= Grade < 1.99,<br>6 คือ 1 <= Grade < 1.49 |
| 7 | Math_Grade | เกรดวิชาคณิตศาสตร์ | |
| 8 | Science_Grade | เกรดวิชาวิทยาศาสตร์ | |
| 9 | Social_studies_Grade | เกรดวิชาสังคมศึกษา | |
| 10 | Health_education_Grade | เกรดวิชาสุขศึกษา | |
| 11 | Art_Grade | เกรดวิชาศิลปะ | |
| 12 | Home_economics_Grade | เกรดวิชาการงานอาชีพ | |
| 13 | Foreign_language_Grade | เกรดภาษาต่างประเทศ | |
| 14 | GPAX | เกรดรวม | |
| 15 | REGISTERSTATUS | สถานะการยืนยันสิทธิ์ของผู้สมัคร | 0 คือ ไม่เลือกเข้าเรียน, 1 คือ เลือกเข้าเรียน |

1.2 เทคนิคนาอีฟเบย์ (Paphat, Sriurai and Ditcharoen, 2022) เป็นเทคนิคที่อาศัยความทฤษฎีความน่าจะเป็นเข้าใจง่ายไม่มีความซับซ้อน โดยมีพื้นฐานมาจากนาอีฟเบย์ที่จะวิเคราะห์ความสัมพันธ์ของตัวแปรตามและตัวแปรอิสระ ซึ่งคำนวณได้จาก (2)

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

โดยที่

C คือ คลาส (Class)

X คือ คุณลักษณะ (Attributes)

P คือ ความน่าจะเป็น (Probability)

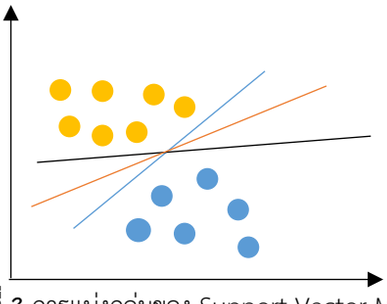
P(C|X) คือ ความน่าจะเป็นที่ข้อมูลที่มีคุณลักษณะ เป็น X จะมีคลาส C
P(X|C) คือ ความน่าจะเป็นข้อมูลที่มีคลาส C และมีคุณลักษณะเป็น X
P(C) คือ ความน่าจะเป็นของคลาส C
P(X) คือ จำนวนคุณลักษณะทั้งหมด

และคำสั่งในการสร้างโมเดลด้วยเทคนิคนาอีฟเบย์ สามารถอธิบายได้ดังนี้ GaussianNB() คือการสร้างโมเดลนาอีฟเบย์ ดังแสดงในภาพที่ 2

```
navie_bayes = GaussianNB()
```

**ภาพที่ 2** การสร้างโมเดลด้วยการเรียนรู้ของเครื่องแบบดั้งเดิมด้วยเทคนิค นาอีฟเบย์

1.3 ซัพพอร์ทเวกเตอร์แมทชีน (Tosasukul and Kuttipon, 2022) เป็นเทคนิคที่มีประสิทธิภาพในการจัดกลุ่มข้อมูลและจำแนกข้อมูล โดยใช้หลักการของการหาเส้นแบ่งแยกที่ดีที่สุดสำหรับข้อมูลที่กำหนดเข้าสู่กระบวนการเรียนรู้เพื่อสร้างระบบจำแนกข้อมูลที่มีประสิทธิภาพมากขึ้น แนวคิดหลักของซัพพอร์ทเวกเตอร์แมทชีนมาจากการนำคุณลักษณะของข้อมูลเข้าสู่เชิงเส้นของพื้นที่ (Feature Space) แล้วหาเส้นแบ่ง (Hyperplane) ที่เหมาะสมที่สุดในการแบ่งกลุ่มข้อมูลออกจากกัน ดังนั้น ซัพพอร์ทเวกเตอร์แมทชีน มีความเร็วและมีประสิทธิภาพในการจัดการกับปัญหาการจำแนกข้อมูลที่ซับซ้อนดังแสดงในภาพที่ 3



**ภาพที่ 3** การแบ่งกลุ่มของ Support Vector Machine

และคำสั่งในการสร้างโมเดลซัพพอร์ทเวกเตอร์แมทชีน สามารถอธิบายได้ดังนี้ kernel='linear' คือการสร้างโมเดลซัพพอร์ทเวกเตอร์แมทชีน โดยใช้ kernel เป็น 'linear', C=1.0 คือ การกำหนดค่าความผิดพลาดในการจำแนกคลาส และ random state=seed คือการกำหนด seed สำหรับการสุ่ม ดังแสดงในภาพที่ 4

```
svm = SVC(kernel= 'liner', c=1.0, radom_state=seed)
```

**ภาพที่ 4** การสร้างโมเดลด้วยการเรียนรู้ของเครื่องแบบดั้งเดิมด้วยเทคนิค ซัพพอร์ทเวกเตอร์แมทชีน

2. การสร้างโมเดลด้วยวิธีการเรียนรู้เชิงลึก ในขั้นตอนนี้ผู้วิจัยได้เลือกใช้อัลกอริทึม 2 อัลกอริทึม คือ โครงข่ายประสาทเทียมแบบคอนโวลูชัน และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะยาว-ระยะสั้น โดยใช้ชุดข้อมูลและมีการกำหนดคลาสเป้าหมายเดียวกัน

2.1 การเรียนรู้เชิงลึกโครงข่ายประสาทแบบคอนโวลูชัน คือ การเรียนรู้เชิงลึกรูปแบบหนึ่งที่จำลองมาจากการมองเห็นของมนุษย์ ถูกออกแบบมาเพื่อประมวลผลข้อมูลที่มีลักษณะพิเศษ เช่น รูปภาพ หรือข้อมูลที่เรียงต่อกันเป็นลำดับ (Cheewaprakobkit, 2022) ในการสร้างโมเดลการเรียนรู้เชิงลึกโครงข่ายประสาทแบบคอนโวลูชัน ผู้วิจัยได้มีการกำหนดจำนวนชั้นและค่าพารามิเตอร์ ดังแสดงในภาพที่ 5 สามารถอธิบายได้ดังนี้ ชั้น Convolutional (Conv2D) คือชั้นแรกในโมเดล CNN และกำหนด filter = 32 โดยมี kernel = 1x3 ใช้สำหรับการสกัดคุณลักษณะจากข้อมูลและActivation Function ที่ใช้คือ ReLU ชั้น MaxPooling2D คือ ชั้นที่ใช้สำหรับการลดขนาดข้อมูลทำให้ข้อมูลที่ผ่านมาจาก Conv2D มีขนาดเล็กลง ชั้น Flatten คือชั้นที่ใช้เปลี่ยนข้อมูลจากรูปแบบที่มีมิติเป็นรูปแบบแบน (flat) เพื่อเตรียมข้อมูลให้สามารถใช้กับชั้น Dense ได้ ชั้น Dense คือชั้นที่ใช้ในการประมวลผลข้อมูลที่ถูกแปลงจาก Flatten ใน ชั้น Dense สุดท้ายมีการใช้ ฟังก์ชัน Softmax เป็น Activation Function

```
model = Sequential()
model.add(Conv2D(filters=32, kernel_size=(1, 3), activation='relu', input_shape=input_shape))
model.add(MaxPooling2D(pool_size=(1, 2)))
model.add(Conv2D(filters=64, kernel_size=(1, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(1, 2)))
model.add(Conv2D(filters=128, kernel_size=(1, 3), activation='relu', padding='same'))
model.add(MaxPooling2D(pool_size=(1, 2)))
model.add(Flatten())
model.add(Dense(256, activation='relu'))
model.add(Dense(128, activation='relu'))
model.add(Dense(2, activation='softmax'))
```

**ภาพที่ 5** การสร้างโมเดลการเรียนรู้เชิงลึกโครงข่ายประสาทแบบคอนโวลูชัน

2.2 การเรียนรู้เชิงลึกหน่วยความจำระยะยาว-ระยะสั้น เป็นโมเดลที่ถูกออกแบบมาเพื่อประมวลผลข้อมูลที่มีลำดับเวลา เช่น ข้อมูลทางการภาษา เสียง ถูกพัฒนาขึ้นเพื่อแก้ไขปัญหาโครงข่ายประสาทวนซ้ำ ทำให้สามารถเก็บสถานะของการคำนวณได้ (Jitboonyapinit, Maneerat and Chirawichitchai, 2022) ในการสร้างโมเดลการเรียนรู้เชิงลึกหน่วยความจำระยะยาว-ระยะสั้น แสดงในภาพที่ 6 สามารถอธิบายได้ดังนี้ ชั้น LSTM Layer แรกใช้หน่วย LSTM จำนวน 64 หน่วย และรับข้อมูลนำเข้า โดยจะแปลงลำดับข้อมูลนำเข้าให้อยู่ในรูปแบบที่เหมาะสำหรับการประมวลผลและเรียนรู้ลำดับข้อมูล ชั้น LSTM Layer ที่สองมีหน่วย LSTM จำนวน 128 หน่วย และกำหนดให้ return_sequences เป็น True ชั้น LSTM Layer ที่สามมีหน่วย LSTM จำนวน 128 หน่วย และไม่มีการส่งออกข้อมูล ชั้น Dense มีทั้งหมด 3 ชั้น Activation Function ที่ใช้คือ ReLU เพื่อประมวลผลข้อมูล Dense ชั้นสุดท้ายมีการใช้ ฟังก์ชัน Softmax เป็น Activation Function

```
model = Sequential()
model.add(LSTM(units=64, input_shape=(x_train.shape[1], x_train.shape[2]), return_sequences=True))
model.add(LSTM(units=128, return_sequences=True))
model.add(LSTM(units=128))
model.add(Dense(units=64, activation='relu'))
model.add(Dense(units=2, activation='softmax'))
```

**ภาพที่ 6** การสร้างโมเดลการเรียนรู้เชิงลึกหน่วยความจำระยะยาว-ระยะสั้น

**การวัดประสิทธิภาพโมเดล**

ในการวัดและหาประสิทธิภาพโมเดลผู้วิจัยได้เลือกใช้ความสูญเสียคอร์สเอนโทรปี (Cross Entropy Loss) เป็นหนึ่งในวิธีการคำนวณความสูญเสียในการฝึกโมเดลแบบ คลาสแบบแบ่งแยก (Classification) โดยเฉพาะการจำแนกคลาสที่มากกว่า 2 คลาส (Multiclass Classification) ซึ่งส่วนใหญ่ถูกนำมาใช้ในเชิงการเรียนรู้เชิงลึกโดยคำนวณได้จากสมการที่ 3

$$H(y,p) = -\sum(y\_i * log(p\_i)) \quad (3)$$

โดยที่

H (y, p) คือความสูญเสียคอร์สเอนโทรปี ระหว่างเวกเตอร์คำตอบจริง (y) และคำตอบที่โมเดลคาดเดา (p)
y_i คือค่าเป้าหมาย (Target) สำหรับคลาส i
p_i คือความน่าจะเป็น (Probability) ที่โมเดลคาดเดาคลาส i ในการจำแนก

และผู้วิจัยยังได้ใช้ ความสูญเสีย cross-entropy ร่วมกับการวัดค่าความแม่นยำ (Accuracy) โดยคำนวณได้จากสมการที่ 4

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

โดยที่

True Positive คือ จำนวนข้อมูลที่ถูกจำแนกถูกต้องว่าเป็นคลาสเป้าหมาย

True Negative คือ จำนวนข้อมูลที่ถูกจำแนกถูกต้องว่าไม่เป็นคลาสเป้าหมาย
False Positive คือ จำนวนข้อมูลที่ถูกจำแนกว่าไม่เป็นคลาสเป้าหมาย
False Negative คือ จำนวนข้อมูลที่ถูกจำแนกว่าเป็นคลาสเป้าหมาย

## ผลการวิจัยและอภิปรายผล

จากการสร้างโมเดลด้วยวิธีการเรียนรู้ของเครื่องแบบดั้งเดิมและการเรียนรู้เชิงลึกพบว่า โมเดลที่สร้างด้วยเทคนิคการเรียนรู้เชิงลึกมีประสิทธิภาพที่ดีกว่า โดยโมเดลการเรียนรู้เชิงลึกแบบโครงข่ายประสาทแบบคอนโวลูชัน มีค่าความถูกต้องร้อยละ 63.27 และหน่วยความจำระยะยาว-ระยะสั้น มีค่าความถูกต้องร้อยละ 63.08 ดังแสดงในตารางที่ 3

**ตารางที่ 3** ตารางแสดงค่า Accuracy ของอัลกอริทึมการเรียนรู้ของเครื่องแบบดั้งเดิมและอัลกอริทึมการเรียนรู้เชิงลึก

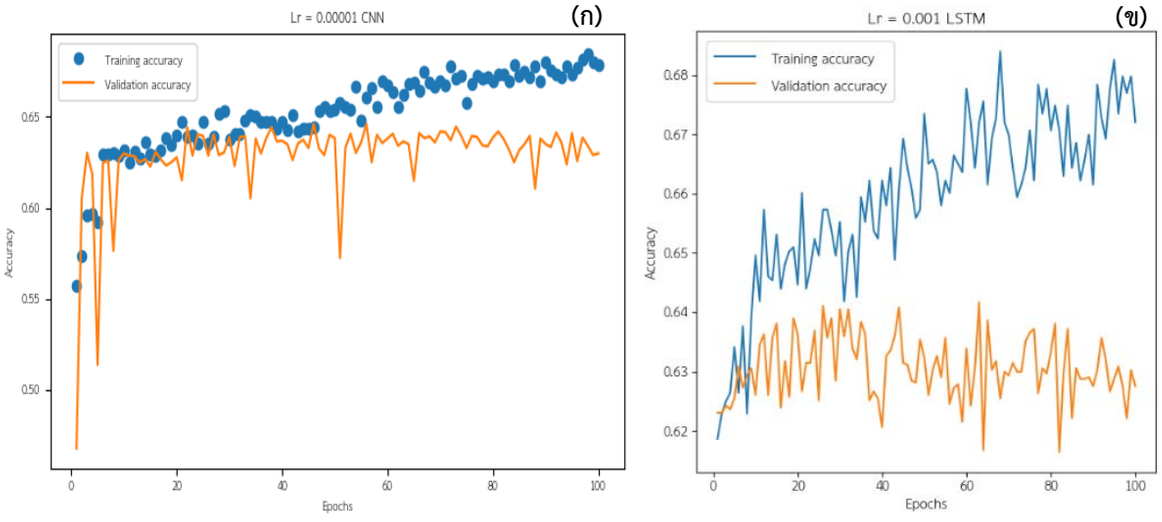
| Algorithms | Accuracy | Precision | |
|---|---|---|---|
| | | **เลือกเข้าเรียน** | **ไม่เลือกเข้าเรียน** |
| Random Forest | 58.86 | 33.27 | 77.14 |
| Naive Bayes | 61.92 | 22.36 | 84.74 |
| Support Vector Machine | 62.37 | 16.96 | 89.43 |
| Convolutional Neural Network | 63.27 | 66.22 | 52.27 |
| Long Short-Term Memory | 63.08 | 64.66 | 48.26 |

จากนั้นผู้วิจัยได้ทำการศึกษาค่าพารามิเตอร์ที่ใช้ในการสร้างโมเดล ที่สร้างด้วยเทคนิคการเรียนรู้เชิงลึกทั้ง 2 ชนิด ได้แก่ โครงข่ายประสาทแบบคอนโวลูชัน และ หน่วยความจำระยะยาว-ระยะสั้น จากข้อมูลการรับเข้าของนักศึกษาคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี ในปีการศึกษา 2561 – 2564 โดยแบ่งชุดข้อมูลที่ใช้ฝึกฝนจำนวนร้อยละ 70 และใช้สำหรับทดสอบร้อยละ 30 โดยมีการศึกษาและการกำหนดค่าพารามิเตอร์ ดังต่อไปนี้

1. Learning Rate เป็นพารามิเตอร์ที่มีส่วนสำคัญในการสร้างโมเดลเนื่องจาก Learning Rate ที่มีค่าสูงอาจทำให้การเรียนรู้ของโมเดลเกิดขึ้นเร็ว แต่มีโอกาสทำให้โมเดลข้ามจุดที่มีค่าสูญเสีย (Loss) ต่ำสุดในขณะที่ Learning Rate มีค่าต่ำเกินไปอาจทำให้โมเดลเรียนรู้ช้าและติดค้างที่ Local Minimum ดังนั้นผู้วิจัย ได้กำหนดค่า Learning Rate ที่แตกต่างกันเพื่อศึกษาพารามิเตอร์ที่ส่งผลต่อประสิทธิภาพของโมเดลที่ดีที่สุด โดยกำหนดค่าพารามิเตอร์อื่น ๆ ดังนี้ Epochs = 100 และ Optimizer = RMSprop สำหรับอัลกอริทึมการเรียนรู้เชิงลึกทั้ง 2 อัลกอริทึม ผู้วิจัยได้แบ่ง Learning Rate ออกเป็น 6 ช่วงดังแสดงในตารางที่ 4 และพบว่าเมื่อพิจารณา Validation Loss และ Validation Accuracy ร่วมกันพบว่า โครงข่ายประสาทแบบคอนโวลูชันที่มีการกำหนดค่า Learning Rate ที่ 0.00001 ให้ผลลัพธ์ที่ดีที่สุดโดย มีค่าความถูกต้องที่ร้อยละ 64.62 และ Learning Rate ที่ 0.001 ของหน่วยความจำระยะยาว-ระยะสั้น มีค่าความถูกต้องที่ร้อยละ 64.25 ดังแสดงในภาพที่ 7

**ตารางที่ 4** การเปรียบเทียบประสิทธิภาพของ CNN และ LSTM ด้วยการกำหนดค่า Learning Rate

| Learning Rate | Validation Accuracy | | Validation Loss | |
|---|---|---|---|---|
| | CNN | LSTM | CNN | LSTM |
| 1 | 62.30 | 62.30 | 66.25 | 66.25 |
| 0.1 | 62.30 | 62.30 | 66.25 | 66.25 |
| 0.01 | 63.89 | 62.30 | 63.75 | 65.60 |
| 0.001 | 63.89 | 64.25* | 63.93 | 63.91* |
| 0.0001 | 63.89 | 64.10 | 63.93 | 63.91 |
| 0.00001 | 64.62* | 62.30 | 63.69* | 64.98 |



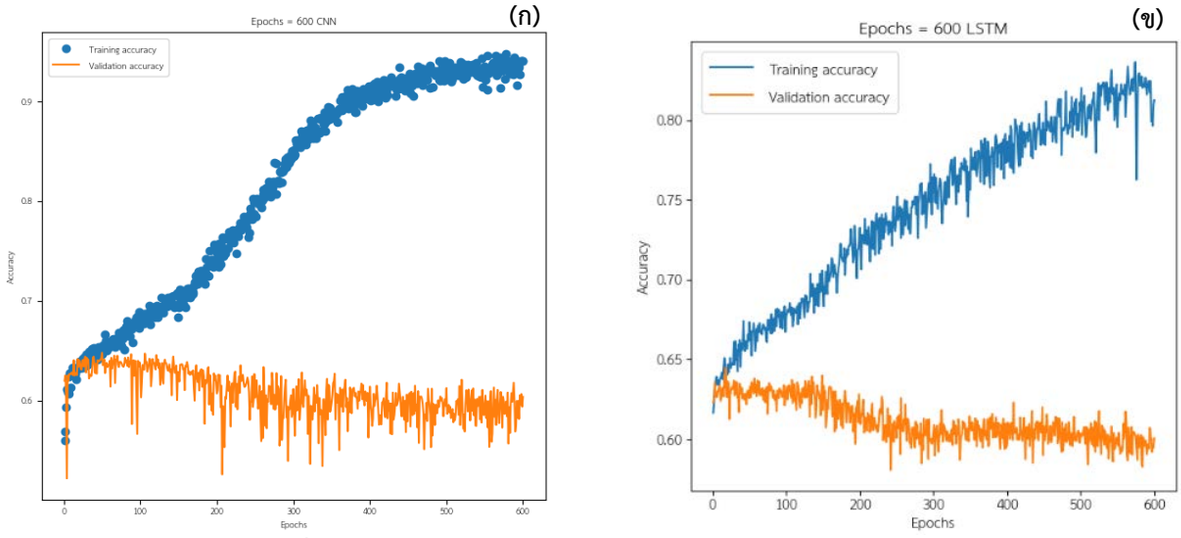

**ภาพที่ 7** Learning Rate ของ CNN = 0.00001 (ก) Learning Rate ของ LSTM = 0.001 (ข)

2. Epochs เป็นพารามิเตอร์ที่กำหนดรอบในการฝึกฝนการสร้างโมเดล การกำหนดค่า Epochs น้อยเกินไปอาจจะส่งผลให้โมเดลมีการเรียนรู้น้อย ทำให้ประสิทธิภาพในการทำนายไม่ดี แต่ถ้ากำหนดค่า Epochs มากเกินไปอาจส่งผลให้โมเดลมีการเรียนรู้มาก ทำให้โมเดลไม่สามารถทำนายข้อมูลใหม่ที่ไม่เคยเห็นมาก่อนได้ และส่งผลต่อกระทบต่อเวลาที่ใช้ในการสร้างโมเดลนานขึ้น ดังนั้นผู้วิจัยได้กำหนดค่า Epochs ที่แตกต่างกันดังแสดงในตารางที่ 5 จำนวน 4 ค่า เพื่อให้ได้ Epochs ที่มีค่า Validation Accuracy สูงที่สุดและค่าสูญเสียน้อยที่สุด โดยกำหนดค่าพารามิเตอร์อื่น ๆ ดังนี้ Learning Rate = 0.00001 Optimizer= RMSprop ของอัลกอริทึมการเรียนรู้เชิงลึกโครงข่ายประสาทแบบคอนโวลูชัน และ Learning Rate = 0.001 Optimizer= RMSprop ของอัลกอริทึมการเรียนรู้เชิงลึกหน่วยความจำระยะยาว-ระยะสั้น พบว่าการปรับค่าพารามิเตอร์ Epochs ของอัลกอริทึมทั้ง 2 อัลกอริทึม มีค่า Training Accuracy เพิ่มสูงขึ้น ในขณะที่ค่า Validation Accuracy ลดลง โดยอัลกอริทึมการเรียนรู้เชิงลึกโครงข่ายประสาทแบบคอนโวลูชัน มีค่า Training Accuracy = 94.80 และ Validation Accuracy = 64.62 ก่อนจะเกิดการ Overfit เมื่อ Epochs = 100 ดังแสดงในรูปที่ 8(ก) และอัลกอริทึมการเรียนรู้เชิงลึกหน่วยความจำระยะยาว-ระยะสั้น มีค่า Training Accuracy = 83.63 และ Validation Accuracy = 64.25 ก่อนจะเกิดการ Overfit เมื่อ Epochs = 100 ดังแสดงในรูปที่ 8(ข)

**ตารางที่ 5** การเปรียบเทียบประสิทธิภาพของ CNN และ LSTM ด้วยการกำหนดค่า Epochs

| epoch | Validation Accuracy | | Validation Loss | |
|---|---|---|---|---|
| | CNN | LSTM | CNN | LSTM |
| 20 | 63.26 | 63.98 | 64.25 | 64.14 |
| 30 | 63.86 | 64.32 | 63.95 | 64.02 |
| 100 | 64.62 | 64.25 | 63.69 | 63.91* |
| 600 | 64.68* | 64.47* | 63.66* | 63.97 |

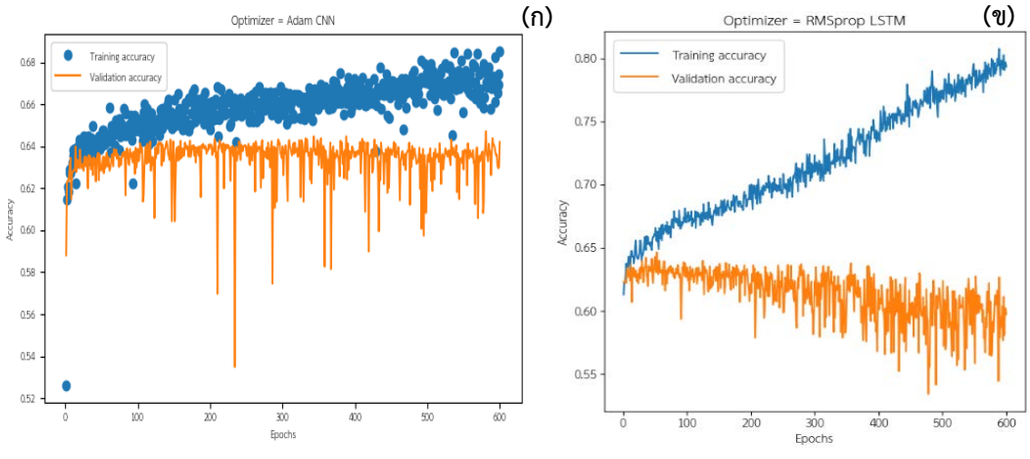
3. Optimizer ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพด้วย Optimizer จำนวน 4 แบบได้แก่ Stochastic Gradient Descent (SGD), Adadelta, Root Mean Squared Propagation (RMSProp) และ Adaptive Moment (Adam) โดยมีผลเปรียบเทียบประสิทธิภาพดังแสดงในตารางที่ 6 พบว่า Optimizer แบบ Adam ของโครงข่ายประสาทแบบคอนโวลูชัน มีค่า Validation Accuracy = 64.68 อีกทั้งแบบจำลองไม่เกิดการ Overfit ตลอดระยะการฝึกแบบจำลอง และ Optimizer ดังแสดงในภาพที่ 9 (ก) และ Optimizer แบบ RMSprop ของหน่วยความจำระยะยาว-ระยะสั้น มีค่า Validation Accuracy = 64.65 และแบบจำลองยังเกิดการ Overfit ดังแสดงในภาพที่ 9(ข)




**ภาพที่ 8** Epochs ของ CNN = 600 (ก) Epochs ของ LSTM = 600 (ข)

**ตารางที่ 6** เปรียบเทียบประสิทธิภาพของ CNN และ LSTM ด้วยการกำหนด Optimizer

| Optimizer | Validation Accuracy | | Validation Loss | |
|---|---|---|---|---|
| | CNN | LSTM | CNN | LSTM |
| SGD | 63.11 | 62.30 | 64.52 | 65.83 |
| RMSprop (Default) | 64.68 | 64.65* | 63.93 | 63.78* |
| Adadelta | 62.48 | 62.39 | 69.54 | 66.42 |
| Adam | 64.71* | 64.47 | 63.66* | 63.97 |




**ภาพที่ 9** Optimizer ของ CNN คือ Adam (ก) Optimizer ของ LSTM คือ RMSprop (ข)

อภิปรายผลจากการศึกษาค่าพารามิเตอร์ทั้ง 3 ประกอบด้วย Learning Rate Epochs และ Optimizer พบว่าค่าพารามิเตอร์ที่เหมาะสมกับการนำไปสร้างโมเดลของโครงข่ายประสาทแบบคอนโวลูชัน คือ Learning Rate = 0.00001 Epochs = 600 Optimizer = Adam ที่มีค่า Validation Accuracy เป็น 64.71 และค่าพารามิเตอร์ที่เหมาะสมกับการสร้างโมเดลของหน่วยความจำระยะยาว-ระยะสั้น คือ Learning Rate = 0.001 Epochs = 600 Optimizer = RMSprop ที่มีค่า Validation Accuracy เป็น 64.65 และพบว่าหน่วยความจำระยะยาว-ระยะสั้นไม่เหมาะสมที่จะนำไปใช้งานเนื่องจากค่า Training Accuracy มีค่าเพิ่มสูงขึ้น ในขณะที่ ค่า Validation Accuracy มีค่าลดลง เมื่อค่า Epochs เพิ่มขึ้น จากเหตุดังกล่าวทำให้โมเดลที่สร้างด้วยอัลกอริทึมของหน่วยความจำระยะยาว-ระยะสั้นเกิดการ Overfit เมื่อเปรียบเทียบประสิทธิภาพของ

โมเดลสรุปได้ว่าโมเดลโครงข่ายประสาทแบบคอนโวลูชันเหมาะสมที่จะนำไปพัฒนาระบบทำนายการยืนยันสิทธิเข้าศึกษาต่อของนักศึกษามหาวิทยาลัย

## สรุปผลการวิจัยและข้อเสนอแนะจากการวิจัย

งานวิจัยนี้ศึกษาวิธีทำนายการยืนยันสิทธิเข้าศึกษาต่อของนักศึกษามหาวิทยาลัยด้วยการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึก จากข้อมูลการรับเข้าของนักเรียนที่จะสมัครเข้าเรียนคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี ในปีการศึกษา 2561 – 2564 จำนวน 4,479 คน ผู้วิจัยได้ทำการแปลงข้อมูลให้อยู่ในรูปแบบของตัวแทนข้อมูล เพื่อให้เหมาะสมกับการนำไปสร้างโมเดลการเรียนรู้ของเครื่องแบบดั้งเดิมและการเรียนรู้เชิงลึก พบว่าโมเดลการเรียนรู้เชิงลึกมีประสิทธิภาพที่ดีกว่า จึงทำการศึกษาหาค่าพารามิเตอร์ที่เหมาะสมกับการสร้างโมเดล ทั้ง 3 ประกอบด้วย Learning Rate, Epochs และ Optimizer พบว่าค่าพารามิเตอร์ที่เหมาะสมกับการนำไปสร้างโมเดลของโครงข่ายประสาทแบบคอนโวลูชัน คือ Learning Rate = 0.00001 Epochs = 600 Optimizer = Adam ที่มีค่า Validation Accuracy เป็น 64.71 และค่าพารามิเตอร์ที่เหมาะสมกับการสร้างโมเดลของหน่วยความจำระยะยาว-ระยะสั้น คือ Learning Rate = 0.001 Epochs = 600 Optimizer = RMSprop ที่มีค่า Validation Accuracy เป็น 64.65 โดยการเลือกโมเดลสำหรับการนำไปใช้จริงควรพิจารณาจากหลายปัจจัย ไม่เพียงแต่พิจารณาจากค่า Validation Accuracy ที่ได้จากข้อมูลทดสอบ (test data) เท่านั้น การที่โมเดลหน่วยความจำระยะยาว-ระยะสั้น เกิดการ overfit อาจแสดงถึงการที่โมเดลเรียนรู้จากข้อมูลการฝึกฝน (training data) ได้ดีเกินไปจนไม่สามารถทำนายข้อมูลใหม่ (test data) ได้ดีเท่าที่ควร ดังนั้นควรเลือกโมเดลโครงข่ายประสาทแบบคอนโวลูชัน ที่มีความสามารถในการทำนายข้อมูลใหม่ได้ดี ไม่เพียงแต่ข้อมูลการฝึกฝน ซึ่งสอดคล้องกับทฤษฎี Model Generalization (Kawaguchi, Kaelbling and Bengio, 2022) จากเหตุดังกล่าว โมเดลการเรียนรู้เชิงลึกที่สร้างด้วยอัลกอริทึม โครงข่ายประสาทแบบคอนโวลูชันเหมาะสมที่จะนำไปพัฒนาระบบทำนายการยืนยันสิทธิเข้าศึกษาต่อของนักศึกษามหาวิทยาลัย

## References


Cheewaprakobkit, P. (2019). Improving the Performance of an Image Classification with Convolutional Neural Network Model by Using Image Augmentations Technique (in Thai). **TNI Journal of Engineering and Technology**, 7(1), 59-64.

Jitboonyapinit, C., Maneerat, P. and Chirawichitchai, N. (2022). Development of Sentiment Analysis Model Based on Thai Social Media Using Deep Learning Technique (in Thai). **Huachiew Chalermprakiet Science and Technology Journal**, 8(2), 8-18.

Kawaguchi, K., Kaelbling, L., and Bengio, Y. (2022). **Mathematical Aspects of Deep Learning**. United Kingdom: Cambridge University Press.

Nandal, P. (2020). Deep Learning in diverse Computing and Network Applications Student Admission Predictor using Deep Learning. **Proceedings of the International Conference on Innovative Computing & Communications (ICICC) 2020** (pp. 1-5). India: University of Delhi.

Palwisut, P. (2016). Improving Decision Tree Technique in Imbalanced Data Sets Using SMOTE for Internet Addiction Disorder Data (in Thai). **Information Technology Journal**, 12(1), 54-63.

Paphat, A., Sriurai, W. and Ditcharoen, N. (2022). University student dropout prediction improved by feature selection with Multilayer Perceptron Neural Network (in Thai). **Journal of Science and Science Education**, 5(1), 39-48.

Suepitak, S., Nissaidee, B. and Buathong, W. (2021). Classification Techniques Comparison for Predicting Graduate Trend (in Thai). **PKRU SciTech Journal**, 5(2), 42-50.

Tongpuang, P. and Sanrach, C. (2021). The Comparison of Data Classification Efficiency to Predict Scholarship Awarded for Undergraduate Students by Data Mining Techniques (in Thai). **Journal of Professional Routine to Research**, 8(2), 44-52.

Tosasukul, J. and Kuttipon, S. (2022). An Efficiency Comparison of Forecasting Models for the Amount of Water in Dam by Data Mining Techniques (in Thai). **Journal of Accountancy and Management**, 14(4), 1-17.