ระบบจำแนกประเภทประวัติย่อของผู้สมัครงานด้วยวิธีการเรียนรู้ของเครื่อง

Resume Classification System using Machine Learning Method

วรวิทย์ สังฆทิพย์[1*] , วิษณุรัตน์ วงสีทา[2] และ พัณกร ผายเงิน[3]

Worawith Sangkatip[1*], Witsanurat Wongsritha[2], and Pannakorn Pai-ngeon[3]

ภาควิชาสื่อนฤมิต คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม[1*]

บริษัท ซีพีเอฟ ไอทีเซ็นเตอร์ จำกัด[2,3]

Department of New Media, Faculty of Informatics, Mahasarakham University[1]

CPF IT CENTER CO., LTD.[2,3]

E-Mail : worawith.s@msu.ac.th[1*], witsanurat.won@axonstech.com[2], pannakorn.pai@axonstech.com[3]

(Received: June 20, 2023; Revised: September 19, 2023; Accepted: July 19, 2023)

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) สร้างแบบจำลองและการหาประสิทธิภาพแบบจำลองการจำแนกประเภทประวัติย่อของผู้สมัครงาน 2) พัฒนาระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง และ 3) ศึกษาความพึงพอใจของผู้ใช้งานระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง กลุ่มตัวอย่าง ได้แก่ นักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ชั้นปีที่ 4 ที่จะออกปฏิบัติงานสหกิจศึกษา จำนวน 30 คน เครื่องมือที่ใช้ในการวิจัย ได้แก่ 1) ชุดข้อมูลประวัติย่อของผู้สมัครงาน 2) ระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงาน 3) แบบสอบถามความพึงพอใจผู้ใช้งานระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงาน ได้ใช้อัลกอริทึมจำแนกประเภทประวัติย่อของผู้สมัครงาน 5 วิธี ได้แก่ วิธีนาอีฟเบย์ วิธีซัพพอร์ตเวกเตอร์แมทชีน วิธีเพื่อนบ้านใกล้ที่สุด วิธีต้นไม้ตัดสินใจ วิธีเพอร์เซปตรอนแบบหลายชั้น โดยใช้สมการในการประเมินประสิทธิภาพของแบบจำลอง ได้แก่ ค่าความแม่นยำ ค่าความระลึก ค่าเอฟวันสกอร์ และค่าความถูกต้อง และสถิติที่ใช้ศึกษาความพึงพอใจ ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

ผลการวิจัย พบว่า 1) การสร้างและหาประสิทธิภาพของแบบจำลองการจำแนกประเภทประวัติย่อของผู้สมัครงาน พบว่า วิธีเพอร์เซปตรอนแบบหลายชั้น และวิธีซัพพอร์ตเวกเตอร์แมทชีน ให้ประสิทธิภาพสูงที่สุด ค่าความแม่นยำ 99.64% ค่าความระลึก 99.59% ค่าเอฟวันสกอร์ 99.56% และค่าความถูกต้อง 99.59% 2) การพัฒนาระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง ระบบสามารถทำงานได้จริงตามที่ออกแบบไว้ โดยมีองค์ประกอบของระบบ 5 ระบบ ได้แก่ ระบบสมาชิก ระบบประมวลผลไฟล์ PDF ประวัติย่อของผู้สมัครงาน ระบบประมวลผลข้อมูลร่วมกับแบบจำลอง ระบบแสดงชื่อตำแหน่งงานที่สอดคล้องกับประวัติย่อของผู้สมัครงาน และระบบแสดงตำแหน่งงานว่างจากบริษัทจัดหางาน 3) ผลการศึกษาความพึงพอใจของผู้ใช้งานต่อระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง พบว่า อยู่ในระดับมากที่สุด ค่าเฉลี่ยเท่ากับ 4.69 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.47

คำสำคัญ : การจำแนกประเภท, ประวัติย่อ, การเรียนรู้ของเครื่อง, การประมวลผลภาษาธรรมชาติ

Abstract

The objective of this research is to 1) create the model and evaluate the performance of a model for classifying job applicants' resumes, 2) develop a web application system for job position recommendations with machine learning method, and 3) study user satisfaction with the web application system for job position recommendations with machine learning method. The sample group consists of 30 fourth-year students majoring in Information Technology at Rajabhat Maha Sarakham University who will be undertaking an internship. The research tools used include

1) a dataset of job applicants' resumes, 2) a web application system for job position recommendations, and 3) a questionnaire to assess user satisfaction with the web application system for job position recommendations. The algorithms used to classify job applicants' resumes include Naïve Bayes, Support Vector Machines, Decision Trees, K-Nearest Neighbors, and Multilayer Perceptron. The evaluation metrics used for the model include Precision, Recall, F1-Score, and Accuracy. The statistics used to study satisfaction were mean and standard deviation.

The research findings revealed the following: 1) The creation and evaluation of the model for classifying job applicants' resumes showed that Multilayer Perceptron and Support Vector Machine methods achieved the highest performance with a precision of 99.64%, recall of 99.59%, F1-score of 99.596%, and accuracy of 99.59%. 2) Development of a web application system for recommending job positions with machine learning methods. The system can work as designed. It consists of five components: a membership system, a PDF file processing system for job applicants' resumes, a data processing system integrated with a model, a system that displays job titles corresponding to job applicants' resumes, and a system that shows job vacancies from recruitment companies. 3) The assessment of user satisfaction with the web application system for job position recommendations with the machine learning method indicated an extremely high satisfaction level, with an average score of 4.69 and a standard deviation of 0.47.

Keywords : Classification, Resume, Machine Learning, Natural Language Processing

บทนำ

สหกิจศึกษาและการศึกษาเชิงบูรณาการกับการทำงาน (Cooperative and Work Integrated Education : CWIE) [1] คือ หลักสูตรการเรียนการสอนในลักษณะร่วมผลิตระหว่างสถาบันอุดมศึกษาและสถานประกอบการ ได้แก่ ภาครัฐ เอกชน และชุมชน เพื่อให้ได้บัณฑิตพร้อมสู่โลกแห่งการทำงานจริงได้ทันที มีสมรรถนะตรงกับความต้องการของตลาดงาน สามารถพัฒนาอาชีพในปัจจุบันและเตรียมพร้อมรองรับตำแหน่งงานในอนาคต โดยให้นักศึกษาได้เรียนรู้ในสถาบันอุดมศึกษาควบคู่กับการปฏิบัติงานจริงในหน่วยงานภายนอก (Work-based Learning) ในทุกรูปแบบ ที่ทำให้นักศึกษามีสมรรถนะ ความรู้ (Knowledge) ทักษะ (Skills) ทัศนคติ (Attitudes) และค่านิยม (Values) และคุณลักษณะตรงกับความต้องการของตลาดงาน โดย CWIE เป็นคำที่ครอบคลุมถึงสหกิจศึกษาและการศึกษาเชิงบูรณาการกับการทำงาน ในรูปแบบต่างๆ เช่น Sandwich Course, Practicum, Postcourse Internship เป็นต้น

สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม มีการจัดการเรียนการสอนในรูปแบบของสหกิจศึกษา หรือสหกิจศึกษาและการศึกษาเชิงบูรณาการกับการทำงาน จากการดำเนินงานในขั้นตอนการคัดเลือกนักศึกษาชั้นปีที่ 4 เพื่อเตรียมออกปฏิบัติงานสหกิจศึกษา ได้มีการจับคู่กับสถานประกอบการ (Matching) โดยจะคัดเลือกนักศึกษาเบื้องต้นที่มีความรู้ ทักษะ ความสามารถที่เหมาะสมกับตำแหน่งงานที่สถานประกอบการเสนองานมา เพื่อส่งให้สถานประกอบการทำการทดสอบและสัมภาษณ์ในขั้นตอนต่อไป พบว่า ในขั้นตอนนี้อาจารย์ผู้รับผิดชอบจะทำการคัดเลือกนักศึกษาจากประวัติย่อของผู้สมัครงาน (Resume) โดยดูข้อมูลจากผลการศึกษา ความรู้ ทักษะ และประสบการณ์ในการอบรบของนักศึกษาแต่ละคน จากนั้นประเมินว่านักศึกษาคนใดจะเหมาะกับตำแหน่งงานที่สถานประกอบการต้องการ ประกอบกับนักศึกษาก็ต้องการทราบถึงความรู้ ความสามารถของตนเองมีว่าตรงกับตำแหน่งงานด้านใด ซึ่งเกิดความล่าช้า และอาจเกิดความผิดพลาดในการคัดเลือกนักศึกษาไม่ตรงกับตำแหน่งงานและความต้องการของสถานประกอบการได้

ดังนั้น ผู้วิจัยจึงได้ทำการทดลองสร้างแบบจำลอง (Model) จากชุดข้อมูลประวัติย่อของผู้สมัครงานด้วยวิธีการเรียนรู้ของเครื่อง (Machine Learning) เพื่อจำแนกประเภท (Classification) ตำแหน่งงาน และพัฒนาระบบเว็บแอปพลิเคชันประมวลผลร่วมกับแบบจำลอง เพื่อแนะนำตำแหน่งงานที่เหมาะสมจากประวัติย่อของผู้สมัครงาน ซึ่งจะช่วยให้นักศึกษา ซึ่งจะช่วยให้อาจารย์และนักศึกษาสามารถตรวจสอบตำแหน่งงานที่เหมาะสมจากประวัติย่อของผู้สมัครงานได้อย่างรวดเร็ว ส่งผลให้การคัดเลือกนักศึกษาด้านความรู้ ด้านความสามารถ ตรงกับตำแหน่งงานและตรงกับความต้องการของสถานประกอบการอย่างแม่นยำมากยิ่งขึ้น

1. วัตถุประสงค์การวิจัย

1.1 เพื่อสร้างแบบจำลองและการหาประสิทธิภาพแบบจำลองการจำแนกประเภทประวัติย่อของผู้สมัครงาน
1.2 เพื่อพัฒนาระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง
1.3 เพื่อศึกษาความพึงพอใจของผู้ใช้งานระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

Goindani, Liu, and Jijkoun [2] ได้นำเสนอ การจำแนกประเภทอุตสาหกรรมของนายจ้างโดยใช้การโพสต์รับสมัครงาน โดยใช้ชุดข้อมูลการโพสต์รับสมัคร งานวิจัยนี้สนใจอยู่สองชุดข้อมูล คือ ชุดข้อมูลการขนส่ง และชุดข้อมูลการดูแลสุขภาพ มีข้อมูลการโพสต์ 10,000 รายการ โดยการจำแนกประเภทเป็นแบบ Binary Classification ใช้อัลกอริทึมจำแนกประเภท 2 วิธี คือ Support Vector Machine (SVM) และ Gradient Boosted Decision Trees (GBDT) ผลการจำแนกประเภทพบว่า วิธี GBDT ให้ประสิทธิภาพมากกว่า SVM

Daryani et al. [3] ได้นำเสนอ ระบบคัดกรองประวัติย่อของผู้สมัครงานให้สอดคล้องกับรายละเอียดงานโดยใช้วิธีการ Natural Language Processing และวิธี Similarity โดยระบบแบ่งออกเป็น 2 ส่วน คือ 1) การสกัดข้อความจากประวัติย่อของผู้สมัครงานแบบที่เป็นข้อมูลแบบไม่มีโครงสร้าง โดยใช้เทคนิคทางด้าน NLP ได้แก่ Tokenization, Stemming, POS Tagging, Named Entity Recognition และ 2) สร้างระบบแนะนำประวัติย่อของผู้สมัครงานให้เหมาะสมกับรายละเอียดงาน โดยใช้วิธี ได้แก่ Vectorization, TF-IDF และการหาความคล้ายคลึง เช่น Cosine Distance สำหรับการทดสอบระบบใช้ข้อมูลโพสต์รายละเอียดงานจาก Amazon.com โดยสกัดข้อมูลจากประวัติย่อของผู้สมัครงานมาเก็บไว้ จากนั้นทำให้อยู่ในรูปแบบของ Vectorization เพื่อหาความคล้ายคลึงระหว่างประวัติย่อของผู้สมัครงานกับรายละเอียดงานด้วย Cosine Distance และจัดลำดับและแสดงผลคะแนนความคล้ายคลึงด้วย

Chowdhary and Bhatia [4] ได้นำเสนอ วิธีการเรียนรู้ของเครื่องสำหรับระบบแนะนำประวัติย่อของผู้สมัครงานแบบอัตโนมัติ โดยต้องการที่จะแยกหมวดหมู่ประวัติย่อของผู้สมัครงานที่ถูกส่งเข้ามาให้กับฝ่ายบุคคลและแนะนำประวัติย่อของผู้สมัครงานที่สอดคล้องกับรายละเอียดงาน (Job Description) ที่ฝ่ายบุคคลต้องการแบบอัตโนมัติ โดยใช้ชุดข้อมูลประวัติย่อของผู้สมัครงาน จาก Kaggle เพื่อทำการทดลองจำแนกประเภท แบ่งวิธีการออกเป็น 2 ขั้นตอน ได้แก่ 1) จำแนกประเภทของประวัติย่อของผู้สมัครงาน และ 2) แนะนำประวัติย่อของผู้สมัครงานที่สอดคล้องกับรายละเอียดงาน ใช้วิธีการจำแนกประเภท ได้แก่ Random Forest, Multinomial Naïve bayes, Logistic Regression, Linear Support Vector Classifier ผลการวัดประสิทธิภาพบว่า วิธี Linear Support Vector Classifier ได้ค่าความถูกต้อง (Accuracy) สูงที่สุด 78.53% และการแนะนำประวัติย่อของผู้สมัครงานให้ตรงกับรายละเอียดงานจะใช้ Cosine similarity และใช้วิธี *K*-NN ที่มีความใกล้เคียงกับรายละเอียดงาน

Javed Mehedi Shamrat et al. [5] ได้นำเสนอ วิธีการส่งประกาศตำแหน่งงานเวียนไปยังอีเมล์ให้ตรงกับความต้องการของแต่ละคน (Personalization) โดยใช้อัลกอริทึม Decision Tree ช่วยในการตัดสินใจ โดยวิเคราะห์จากโครงสร้างของ Decision Tree เพื่อพิจารณา เช่น ตำแหน่งงาน ประสบการณ์ เงินเดือนที่ต้องการ สาขาที่สนใจ

เป็นต้น ในการวัดประสิทธิภาพใช้ 12 รายการทดสอบ พบว่า ได้ค่า Accuracy เฉลี่ย 99.52% พร้อมกับทดสอบส่งอีเมล์ด้วย

Van Huynh et al. [6] ได้นำเสนอ วิธีการทำนายตำแหน่งงานจากโมเดลการเรียนรู้เชิงลึกไปจนถึงการพัฒนาแอปพลิเคชัน ใช้ชุดข้อมูลที่เกี่ยวกับตำแหน่งงานด้านไอที มีตำแหน่งงาน 25 รายการ และรายละเอียดงาน 10,000 รายการ ในการทดลองใช้โมเดลทางด้าน Deep Learning ได้แก่ Bi-GRU-CNN, Bi-GRU-LSTM-CNN และใช้ Ensemble Model ผลการทดลองพบว่า ได้ค่า F1-score ของ Ensemble Model เท่ากับ 72.59%

Tran, Vo, and Luu [7] ได้นำเสนอวิธีการทำนายชื่อตำแหน่งงานจากรายละเอียดงาน โดยใช้วิธีการจำแนกประเภทแบบหลายเลเบล ชุดข้อมูลรวบรวมจากรายละเอียดงานในเว็บไซต์ต่าง ๆ มีทั้งหมด 68 ตำแหน่งงาน และรายละเอียดงาน 22,000 รายการ ในการทดลองใช้อัลกอริทึมทางด้าน Deep Learning ได้แก่ วิธีการด้าน Pre-trained Embeddings และ วิธีการด้าน Transformer Models ผลการทดลองพบว่า วิธี BERT ให้ค่า F1-Score สูงที่สุด 62.20% กับชุดข้อมูล Development Set และได้ค่า F1-Score 47.44% กับชุด Test set

กนิษฐา อินธิชิต, ภควัฒน์ ปิยวงษ์ และสุภาพร สุขใส [8] ได้วิจัยเรื่อง การพัฒนาแอปพลิเคชันช่วยตัดสินใจในการเลือกเรียนสาขาวิชาคอมพิวเตอร์ในมหาวิทยาลัยราชภัฏศรีสะเกษบนระบบปฏิบัติการแอนดรอยด์ โดยใช้เทคนิคต้นไม้ตัดสินใจ ผลการวิจัยพบว่า แอปพลิเคชันสามารถวิเคราะห์ข้อมูลที่ผู้ใช้ตอบแบบทดสอบแล้วเก็บมาเป็นคะแนน และนำคะแนนไปเปรียบเทียบสาขาวิชาที่เหมาะสมกับผู้ใช้และแสดงผลในรูปแบบข้อความเรียงลำดับสาขาวิชาตามคะแนนที่ได้สูงสุด 3 อันดับแรก ผลการประเมินแอปพลิเคชันมีความเหมาะสมอยู่ในระดับมาก และผู้ใช้มีความพึงพอใจแอปพลิเคชันอยู่ในระดับมากที่สุด

วิธีดำเนินการวิจัย

1. เครื่องมือการวิจัย

1.1 ชุดข้อมูล (Dataset) ที่ใช้ในการทดลองสร้างแบบจำลอง (Model) ได้แก่ ชุดข้อมูลประวัติย่อของผู้สมัครงาน (Resume) [9] จากเว็บไซต์ Kaggle.com

1.2 ระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง

1.3 แบบสอบถามความพึงพอใจผู้ใช้งานระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง

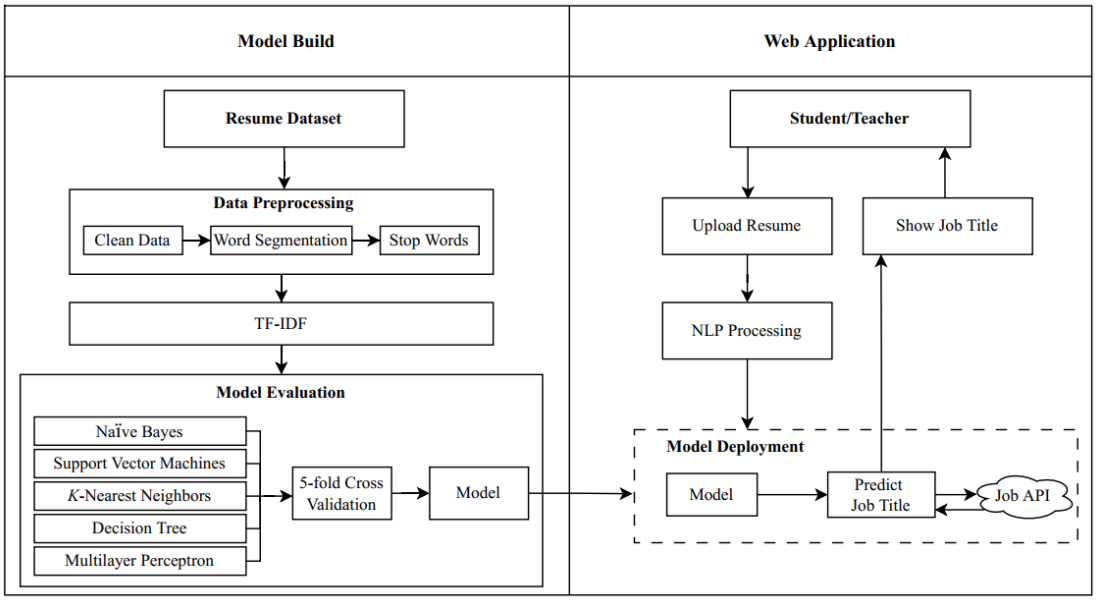
2. ประชากรและกลุ่มตัวอย่าง

2.1 ประชากร คือ นักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ชั้นปีที่ 4 ที่จะออกปฏิบัติงานสหกิจศึกษาในปีการศึกษา 2/2565 จำนวน 68 คน

2.2 กลุ่มตัวอย่าง คือ นักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ชั้นปีที่ 4 ที่จะออกปฏิบัติงานสหกิจศึกษาในปีการศึกษา 2/2565 จำนวน 30 คน โดยการสุ่มตัวอย่างแบบกลุ่ม (Cluster Sampling)

3. ขั้นตอนการดำเนินการวิจัย

ในขั้นตอนการดำเนินการวิจัย ผู้วิจัยได้ออกแบบกรอบการวิจัย เพื่อให้เห็นภาพรวมของการทดลองสร้างแบบจำลองและพัฒนาระบบ ดังภาพที่ 1

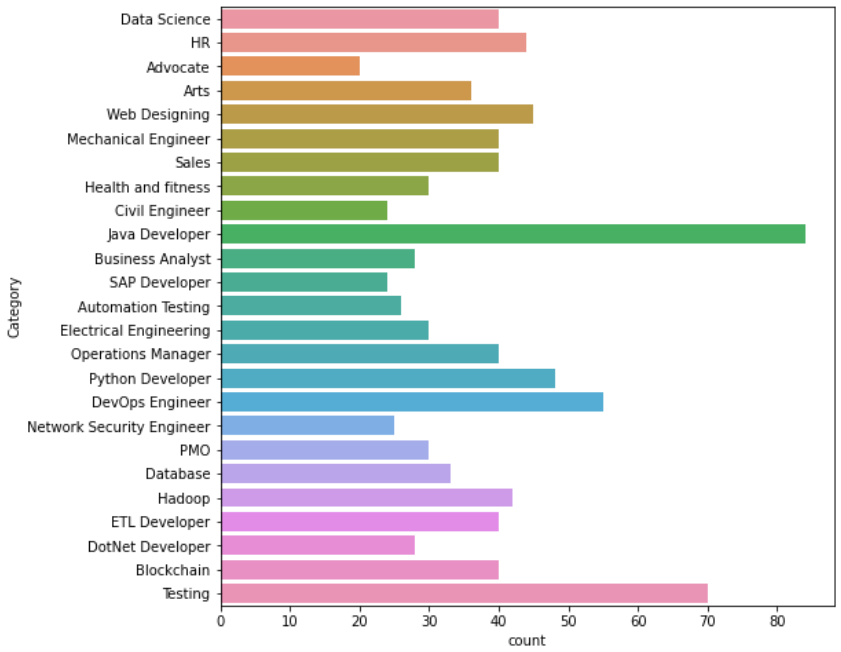


ภาพที่ 1 กรอบการวิจัยของระบบจำแนกประเภทตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง

จากภาพที่ 1 กรอบการวิจัยแบ่งออกเป็น 2 ส่วน ได้แก่ 1) การสร้างแบบจำลอง เป็นส่วนของการนำข้อมูลประวัติย่อของผู้สมัครงาน มาใช้ทดลองเพื่อสร้างแบบจำลองและประเมินประสิทธิภาพแบบจำลอง ผลลัพธ์ที่ได้จากส่วนนี้คือแบบจำลองที่มีประสิทธิภาพที่จะนำไปใช้งานกับเว็บแอปพลิเคชัน และ 2) การพัฒนาเว็บแอปพลิเคชัน เป็นส่วนของการออกแบบและพัฒนาเว็บแอปพลิเคชันเพื่อให้ผู้ใช้งานอัปโหลดไฟล์ประวัติย่อของผู้สมัครงานให้ระบบประมวลผลจากแบบจำลองที่ผ่านการเรียนรู้มาแล้ว ทำการแนะนำตำแหน่งงานที่สอดคล้องกับประวัติย่อของผู้สมัครงานของผู้ใช้งาน โดยแต่ละขั้นตอนของการดำเนินการวิจัย สามารถอธิบายได้ ดังต่อไปนี้

3.1 การเข้าถึงข้อมูลและรวบรวมข้อมูล (Data Access and Data Collection)

ชุดข้อมูลที่นำใช้ในการทดลองสร้างแบบจำลองในการแนะนำตำแหน่งงาน โดยใช้ชุดข้อมูล resume อยู่ในรูปแบบของไฟล์ CSV และชุดข้อมูล Resume เป็นภาษาอังกฤษ จากเว็บไซต์ Kaggle.com [9] ซึ่งเป็นเว็บไซต์ที่เผยแพร่ชุดข้อมูลมาตรฐานไว้สำหรับการนำมาทดลองด้านการเรียนรู้ของเครื่อง (Machine Learning) ชุดข้อมูลประกอบไปด้วยข้อมูลจำนวน 962 แถว และคอมลัมน์จำนวน 2 คอมลัมน์ โดยแต่ละคอลัมน์ประกอบไปด้วยคอลัมน์ Category เป็นคอลัมน์ที่เก็บประเภทตำแหน่งงานของประวัติย่อของผู้สมัครงานไว้ และคอลัมน์ Resume เป็นคอลัมน์ที่เก็บข้อความของประวัติย่อของผู้สมัครงาน อยู่ในรูปแบบของประโยคแบบไม่มีโครงสร้าง (Unstructured data) เมื่อวิเคราะห์ชุดข้อมูลพบว่า มีจำนวนประเภทของตำแหน่งงาน ทั้งหมด 25 ตำแหน่งงาน สามารถแสดงได้ดังภาพที่ 2 และแสดงตัวอย่างข้อความที่อยู่ในประวัติย่อของผู้สมัครงานได้ ดังภาพที่ 3



ภาพที่ 2 การกระจายของประวัติย่อของผู้สมัครงานแต่ละตำแหน่งงาน

Category	Resume
Data Science	Skills * Programming Languages: Python (pandas, numpy, scipy, scikit-learn, matplotlib), Sql, Java, JavaScript/JQuery. * Machine learning: Regression, SVM, NaÃƒÂ¯ve Bayes, KNN, Random Forest, Decision Trees, Boosting techniques, Cluster Analysis, Word Embedding, Sentiment Analysis, Natural Language processing, Dimensionality reduction, Topic Modelling (LDA, NMF), PCA & Neural Nets. * Database Visualizations: Mysql, SqlServer, Cassandra, Hbase, ElasticSearch D3.js, DC.js, Plotly, kibana, matplotlib, ggplot, Tableau. * Others: Regular Expression, HTML, CSS, Angular 6, Logstash, Kafka, Python Flask, Git, Docker, computer vision - Open CV and understanding of Deep learning.Education Details Data Science Assurance Associate Data Science Assurance Associate - Ernst & Young LLP
Data Science	Areas of Interest Deep Learning, Control System Design, Programming in-Python, Electric Machinery, Web Development, Analytics Technical Activities q Hindustan Aeronautics Limited, Bangalore - For 4 weeks under the guidance of Mr. Satish, Senior Engineer in the hangar of Mirage 2000 fighter aircraft Technical Skills Programming Matlab, Python and Java, LabView, Python WebFrameWork-Django, Flask, LTSPICE-intermediate Languages and and MIPOWER-intermediate, Github (GitBash), Jupyter Notebook, Xampp, MySQL-Basics, Python Software Packages Interpreters-Anaconda, Python2, Python3, Pycharm, Java IDE-Eclipse Operating Systems Windows, Ubuntu, Debian-Kali Linux Education Details January 2019 B.Tech. Electrical and Electronics Engineering Manipal Institute of Technology January 2015 DEEKSHA CENTER January 2013 Little Flower Public
Data Science	Skills Ã¢Â€Â¢ R Ã¢Â€Â¢ Python Ã¢Â€Â¢ SAP HANA Ã¢Â€Â¢ Tableau Ã¢Â€Â¢ SAP HANA SQL Ã¢Â€Â¢ SAP HANA PAL Ã¢Â€Â¢ MS SQL Ã¢Â€Â¢ SAP Lumira Ã¢Â€Â¢ C# Ã¢Â€Â¢ Linear Programming Ã¢Â€Â¢ Data Modelling Ã¢Â€Â¢ Advance Analytics Ã¢Â€Â¢ SCM Analytics Ã¢Â€Â¢ Retail Analytics Ã¢Â€Â¢Social Media Analytics Ã¢Â€Â¢ NLP Education Details January 2017 to January 2018 PGDM Business Analytics Great Lakes Institute of Management & Illinois Institute of Technology January 2013 Bachelor of Engineering Electronics and Communication Bengaluru, Karnataka New Horizon College of Engineering, Bangalore Visvesvaraya Technological University Data Science Consultant Consultant - Deloitte USI Skil Details

ภาพที่ 3 ตัวอย่างข้อความในประวัติย่อของผู้สมัครงาน

3.2 การเตรียมข้อมูล (Data Preprocessing)

ผู้วิจัยได้ดำเนินการเตรียมชุดข้อมูลตามขั้นตอนของกระบวนการประมวลผลภาษาธรรมชาติ (Natural Language Processing) [10] จะเห็นได้ว่ามีวิเคราะห์ชุดข้อมูลประวัติย่อของผู้สมัครงาน ในส่วนของข้อความของประวัติย่อของผู้สมัครงานจะมีตัวอักษรพิเศษ มีขั้นตอนดังต่อไปนี้

3.2.1 ทำความสะอาดข้อความ (Clean Data)

ในขั้นตอนนี้ทำการลบข้อความที่เป็นตัวอักษรพิเศษ เช่น ! € @ % # * & ~ $. เป็นต้น และลบข้อความ HTML Tag ที่ติดมากับข้อความประวัติย่อของผู้สมัครงาน เช่น , <p>, <a> เป็นต้น จะได้ดังภาพที่ 4

Category	Resume	Resume_clean
Data Science	Skills Ã¢Â€Â¢ Python Ã¢Â€Â¢ Tableau Ã¢Â€Â¢ Data Visualization Â¢Â€Â¢ R Studio Ã¢Â€Â¢ Machine Learning Ã¢Â€Â¢ Statistics IABAC Certified Data Scientist with versatile experience over 1+ years in managing business, data science consulting and leading innovation projects, bringing business ideas to working real world solutions. Being a strong advocator of augmented era, where human capabilities are enhanced by machines, Fahed is passionate about bringing business concepts in area of machine learning, AI, robotics etc., to real life solutions.Education Details January 2017 B. Tech Computer Science & Engineering Mohali, Punjab Indo Global College of Engineering Data Science Consultant Data Science Consultant - Datamites Skill Details MACHINE LEARNING- Exprience - 13 months PYTHON- Exprience - 24 months SOLUTIONS- Exprience - 24 months DATA SCIENCE- Exprience - 24 months DATA VISUALIZATION- Exprience - 24 months Tableau- Exprience - 24 monthsCompany Details company - Datamites description - Â¢Â€Â¢ Analyzed and processed complex data sets using advanced querying, visualization and analytics tools. Â¢Â€Â¢ Responsible for loading, extracting and validation of client data. Â¢Â€Â¢ Worked on manipulating, cleaning & processing data using	skills python tableau data visualization studio machine learning statistics iabac certified data scientist versatile experience years managing business data science consulting leading innovation projects bringing business ideas working real solutions strong advocator augmented era human capabilities enhanced machines fahed passionate bringing business concepts area machine learning robotics real life solutionseducation details january tech computer science engineering mohali punjab indo global college engineering data science consultant data science consultant datamites skill details machine learning exprience months python exprience months solutions exprience months data science exprience months data visualization exprience months tableau exprience monthscompany details company datamites description analyzed processed complex data sets advanced querying visualization analytics tools responsible loading extracting validation client data worked manipulating cleaning processing data python tableau data visualization company heretic solutions pvt description worked closely business identify issues data propose solutions effective decision making manipulating cleansing processing data python excel analyzed raw data drawing conclusions developing recommendations machine learning tools statistical techniques produce solutions problems

ภาพที่ 4 ข้อความที่ได้หลังทำความสะอาดข้อความ

3.2.2 การตัดคำ (Word Segmentation)

หลังจากการทำความสะอาดข้อมูลแล้วจะได้ข้อความแบบไม่มีโครงสร้าง (Unstructured data) จำเป็นต้องนำข้อความเหล่านี้มาแปลงให้เป็นข้อมูลแบบมีโครงสร้าง (Structured data) เพื่อให้สามารถนำไปใช้สร้างแบบจำลองได้ โดยกระบวนการตัดคำได้ใช้ไลบรารี่ Natural Language Toolkit (NLTK) [11] มาช่วยในการตัดคำ ดังแสดงในตารางที่ 1

ตารางที่ 1 ผลการตัดคำ

ข้อความก่อนตัดคำ	ข้อความหลังตัดคำ
Database Used: SQL Server	Used | : | SQL | Server | . |
IT SKILLS Languages: C (Basic), JAVA (Basic) Web	IT | SKILLS | Languages | : | C | (| Basic |) | , | JAVA | (| Basic |) | Web |
Skill Details BOOTSTRAP- Exprience - 24 months	Skill | Details | BOOTSTRAP | - | Experience | - | 24 | months |

3.2.3 การกำจัดคำหยุด (Stop Words)

ในขั้นตอนนี้เป็นการกำจัดคำหยุดที่ไม่มีนัยสำคัญออก ซึ่งจำนวนคำหยุดที่ปรากฏในข้อความมีความถี่ค่อนข้างสูง คำหยุดอาจเป็นคุณลักษณะที่ไม่เกี่ยวข้องกับเนื้อหา อาจส่งผลกระทบในการจำแนกประเภทข้อมูลได้ ตัวอย่างคำหยุด เช่น has, hasn't, have, haven't, having, he, he's, hello, help, how, howbeit, however, i'd, i'll, i'm, i've, ie, if, ignored, is, isn't, it, it'd, its, obviously, of, oh, ok, and, can, for, the เป็นต้น

3.2.4 การทำ TF-IDF (Term Frequency – Inverse Document Frequency)

ในขั้นตอนการทำ TF-IDF เป็นวิธีการที่พิจารณาองค์ประกอบของคำภายในประโยค หรือประโยคในเอกสารเป็นหลัก โดยจะไม่นำลำดับของคำภายในเอกสารมาใช้วิเคราะห์ มี 2 องค์ประกอบด้วยกัน คือ Term Frequency (TF) และ Inverse document Frequency (IDF) [12] สามารถอธิบายการทำงานได้ ดังนี้

1) ขั้นตอนการทำ Term-Frequency (TF) เป็นค่าที่บอกความถี่ของคำแต่ละคำที่ปรากฏในเอกสารเอกสารหนึ่ง โดยคิดคำนวณจากการนำจำนวนครั้งที่คำนั้นๆ ปรากฏในเอกสารมาหารด้วยจำนวนคำทั้งหมดในเอกสาร คำนวณได้ดังสมการ

$$TF(t,d) = \frac{count\ of\ term\ in\ document}{number\ of\ word\ in\ document} \quad (1)$$

2) ขั้นตอนการทำ Inverse Document Frequency (IDF) เป็นการคำนวณค่าน้ำหนัก (weight) ความสำคัญของแต่ละคำ โดยจะนำคำที่พบเจอได้บ่อย ๆ (ในหลายๆเอกสาร) จะมีค่า IDF ต่ำ ซึ่งบ่งบอกว่าคำเหล่านั้นจะไม่สามารถดึงเอาจุดเด่นของเอกสารที่คำเหล่านั้นปรากฏอยู่ออกมาได้ดี ค่า IDF สามารถคำนวณได้ดังสมการ

$$IDF(t) = \log\frac{N}{df(t)} \quad (2)$$

3) ขั้นตอนการทำ TF-IDF โดยการนำเอาทั้งสองค่ามารวมกัน จะได้การคำนวณ TF-IDF ดังสมการ

$$TF - IDF(t,d) = TF(t,d)\ x\ IDF(t) \quad (3)$$

โดยที่

t คือ คำที่ปรากฏในเอกสาร
d คือ เอกสารที่ใช้พิจารณา
df คือ จำนวนของเอกสารที่มีคำคำนั้นปรากฏอยู่
N คือ จำนวนเอกสารทั้งหมด

3.3 การสร้างแบบจำลอง (Model Build)

การสร้างแบบจำลองด้วยวิธีการเรียนรู้ของเครื่อง เพื่อใช้ในการจำแนกประเภทประวัติย่อของผู้สมัครงานจากชุดข้อมูล Resume โดยใช้ภาษา Python บน Google Colab [13] ในการทดลองสร้างแบบจำลองและประเมินประสิทธิภาพแบบจำลอง และงานวิจัยนี้ได้เลือกใช้วิธีการจำแนกประเภท 5 วิธี เนื่องจากเป็นวิธีที่ได้รับความสนใจจากการวิจัยด้านการเรียนรู้ของเครื่อง [14] สามารถอธิบายการทำงานแต่ละวิธีได้ ดังนี้

3.3.1 วิธีนาอีฟเบย์ (Naïve Bayes) เป็นวิธีการจำแนกประเภทข้อมูลที่ใช้ตัวชี้วัดความน่าจะเป็น (Probability) เพื่อช่วยในกระบวนการคำนวณหรือคาดการณ์ผลลัพธ์ในการจำแนกข้อมูล วิธีนี้ใช้ทฤษฎีทางสถิติและการประมาณค่าความน่าจะเป็นจากข้อมูลที่มีอยู่เพื่อตัดสินใจในการจำแนกประเภท

3.3.2 วิธีซัพพอร์ตเวกเตอร์แมทชีน (Support Vector Machines - SVM) เป็นวิธีการจำแนกประเภทข้อมูลที่มีความสามารถในการสร้างแบ่งแยกระหว่างกลุ่มข้อมูลที่ซับซ้อนหรือไม่เป็นเชิงเส้นได้อย่างมีประสิทธิภาพ วิธีการนี้ใช้หลักการหาเส้นแบ่ง (Hyper Plane) ที่มีระยะห่างระหว่างข้อมูลสองกลุ่มที่ใกล้ที่สุด (Margin) มากที่สุด เพื่อแยกกลุ่มข้อมูลออกเป็นสองส่วน ใน SVM, เส้นแบ่งที่ใช้สร้างกลุ่มข้อมูลเรียกว่าไฮเปอร์เพลน (Hyper Plane)

ซึ่งสามารถเป็นได้ในหลายมิติ และสามารถแบ่งข้อมูลเป็นสองกลุ่มได้อย่างมีประสิทธิภาพ ยิ่งมีระยะห่างระหว่างข้อมูลสองกลุ่มที่ใกล้ที่สุดมากเท่าไร ก็จะยิ่งมีความมั่นใจในการแยกและจำแนกกลุ่มข้อมูลเพิ่มขึ้น

3.3.3 วิธีเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors) เป็นวิธีที่ใช้ในการจำแนกกลุ่มข้อมูลโดยคำนึงถึงคุณลักษณะที่ใกล้เคียงกันที่สุดของ *k* กลุ่มข้อมูลที่ใกล้เคียงกันมากที่สุด วิธีการนี้ใช้การคำนวณค่าระยะห่างระหว่างจุดข้อมูลโดยใช้วิธี Euclidean Distance ซึ่งเป็นการวัดระยะห่างระหว่างข้อมูลสองจุด ผลลัพธ์ที่ได้จากการคำนวณระยะห่างจะแสดงให้เห็นถึงความคล้ายคลึงกันของข้อมูล เมื่อระยะห่างน้อยแสดงถึงความคล้ายคลึงกันของข้อมูลที่มากขึ้น

3.3.4 วิธีต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคในการแบ่งประเภทข้อมูลโดยใช้ลักษณะต่าง ๆ ของข้อมูลเป็นเกณฑ์ในการตัดสินใจ โครงสร้างของต้นไม้ประกอบด้วยโหนด (Node) ที่แบ่งประเภทข้อมูลตามลักษณะที่กำหนด โหนดล่างสุดของต้นไม้ (Leaf Node) จะเป็นผลลัพธ์หรือประเภทที่แบ่งได้จากข้อมูลที่ส่งเข้ามาในต้นไม้

3.3.5 วิธีเพอร์เซปตรอนแบบหลายชั้น (Multilayer Perceptron) เป็นวิธีหนึ่งของอัลกอริทึมโครงข่ายประสาทเทียม (Artificial Neural Networks) ที่มีโครงสร้างแบบหลายๆ ชั้น ใช้สำหรับประมวลผลหรือจำแนกประเภทข้อมูลที่มีความซับซ้อน โดยโครงสร้างประกอบด้วย Input Layer ในการรับข้อมูลมาที่ชั้นของ Hidden Layer และชั้นของ Hidden Layer จะเชื่อมต่อกับทุกโหนดไปยังชั้น Output Layer โครงข่ายประสาทเทียมมีการเรียนรู้ค่าน้ำหนัก (Weights) การคำนวณค่าน้ำหนัก และฟังก์ชันการแปลงด้วยฟังก์ชันซิกมอยด์ (Sigmoid Function) ฟังก์ชันไฮเปอร์โบลิกแทนเจนต์ (Hyperbolic Tangent Function)

3.4 การประเมินประสิทธิภาพแบบจำลอง (Model Evaluation)

การประเมินประสิทธิภาพของแบบจำลองใช้ด้วยวิธี k-fold Cross Validation [15] โดยทำการสุ่มข้อมูลออกเป็น k ส่วนเท่า ๆ กัน งานวิจัยนี้แบ่งเป็น 5 ส่วน หรือ 5-Fold เพื่อใช้สลับเป็นชุดข้อมูลเรียนรู้ (Train Data) และชุดข้อมูลการทดสอบ (Test Data) โดยการทดลองครั้งแรกกำหนดให้ส่วนที่ 1 เป็นชุดทดสอบและส่วนที่เหลือคือ k-1 เป็นชุดเรียนรู้ ทำจนกระทั่งข้อมูลทุกส่วนได้เป็นชุดเรียนรู้และชุดทดสอบจนครบ สำหรับการประมวลผลจะประมวลผลทั้งหมด k ครั้ง หรือประมวลผลจำนวน 5 ครั้ง โดยวิธีนี้จะส่งผลให้ได้แบบจำลองได้ผ่านการเรียนรู้และการทดสอบกับชุดข้อมูลทั้งหมด ในการทดลองแต่ละครั้งจะได้ประสิทธิภาพของแบบจำลอง โดยงานวิจัยนี้ได้กำหนดให้มีการประเมินประสิทธิภาพของแบบจำลองแต่ละวิธีด้วยการวัดแบบเทรชโฮลด์ (Threshold) ในรูปแบบของตารางคอนฟิวชันเมทริกซ์ (Confusion Matrix) โดยใช้เมทริกซ์การประเมินประสิทธิภาพ ได้แก่ ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าเอฟวันสกอร์ (F1-Score) และค่าความถูกต้อง (Accuracy) ดังต่อไปนี้

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

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

$$F1 = 2x\frac{Pr ecisionx\, Re call}{Pr ecision+Re call} \quad (6)$$

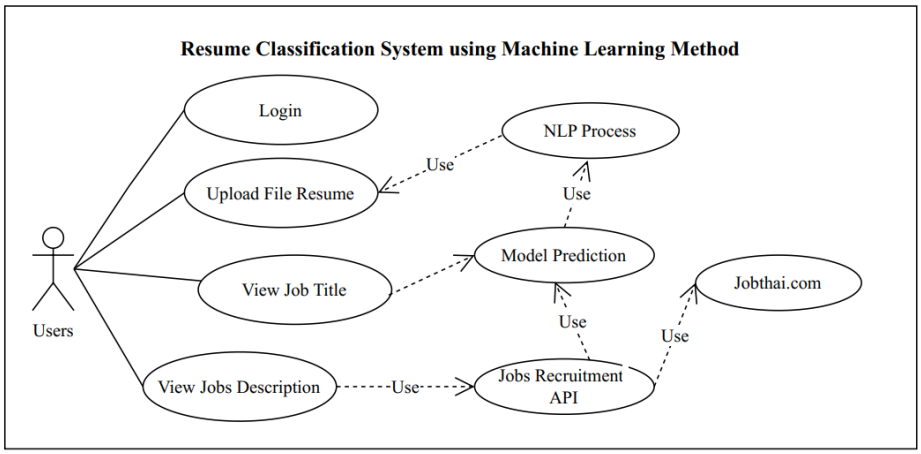
$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (7)$$

โดยที่

TP (True Positive) คือ แบบจำลองทำนายว่า จริง เทียบกับ ผลเฉลยคือ จริง
TN (True Negative) คือ แบบจำลองทำนายว่า ไม่จริง เทียบกับ ผลเฉลยคือ ไม่จริง
FN (False Negative) คือ แบบจำลองทำนายว่า ไม่จริง เทียบกับ ผลเฉลยคือ จริง
FP (False Positive) คือ แบบจำลองทำนายว่า จริง เทียบกับ ผลเฉลยคือ ไม่จริง

3.5 การพัฒนาเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง

ในส่วนของการออกแบบและพัฒนาเว็บแอปพลิเคชัน ผู้วิจัยได้ดำเนินการพัฒนาระบบตามวงจรการพัฒนาระบบ (System Development Life Cycle : SDLC) 5 ขั้นตอน ในขั้นตอนการวิเคราะห์และออกแบบระบบใช้ทฤษฎีการออกแบบแผนภาพ UML (Unified Modeling Language) และใช้แผนภาพ Use Case Diagram อธิบายการทำงานของผู้ใช้งานกับความสามารถของระบบ ดังภาพที่ 5



ภาพที่ 5 Use Case Diagram ระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง

จากภาพที่ 5 แสดง Use Case ของเว็บแอปพลิเคชันที่พัฒนาให้ผู้ใช้งานสามารถอัปโหลดไฟล์ประวัติย่อของผู้สมัครงานและระบบจะประมวลผลจากแบบจำลองที่ได้จากการเรียนรู้ของเครื่อง เพื่อแนะนำตำแหน่งงานที่สอดคล้องกับประวัติย่อของผู้สมัครงานของผู้ใช้งาน และระบบยังสามารถแสดงรายละเอียดตำแหน่งงานว่างจากบริษัทต่าง ๆ ที่สอดคล้องกับตำแหน่งงานที่ระบบแนะนำด้วย

4. สถิติที่ใช้ในการวิจัย

สถิติที่ใช้ในการวิเคราะห์ข้อมูลความพึงพอใจของผู้ใช้งานระบบแนะนำตำแหน่งงานสหกิจศึกษาและการศึกษาเชิงบูรณาการกับการทำงาน ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน โดยนำผลที่ได้เทียบกับเกณฑ์การประเมิน 5 ระดับของลิเคิร์ท (Likert Scale) [16] ดังนี้

ค่าเฉลี่ยเท่ากับ 4.51 – 5.00 หมายความว่า ระดับมากที่สุด
ค่าเฉลี่ยเท่ากับ 3.51 – 4.50 หมายความว่า ระดับมาก
ค่าเฉลี่ยเท่ากับ 2.51 – 3.50 หมายความว่า ระดับปานกลาง
ค่าเฉลี่ยเท่ากับ 1.51 – 2.50 หมายความว่า ระดับน้อย
ค่าเฉลี่ยเท่ากับ 1.01 – 1.50 หมายความว่า ระดับน้อยที่สุด

ผลการวิจัย

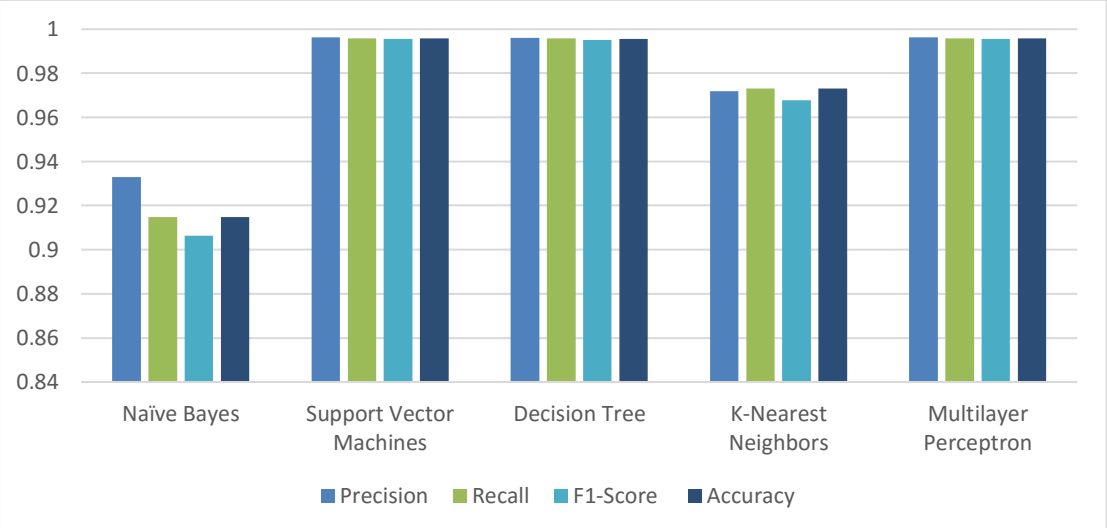
1. ผลการสร้างแบบจำลองและหาประสิทธิภาพแบบจำลองการจำแนกประเภทประวัติย่อของผู้สมัครงาน

ผู้วิจัยได้ทำการทดลองสร้างแบบจำลองจากชุดข้อมูล Resume ที่ผ่านขั้นตอนการเตรียมข้อมูลแล้ว ในขั้นตอนการสร้างแบบจำลองได้ใช้วิธีการจำแนกประเภทข้อมูล 5 วิธี ได้แก่ วิธีนาอีฟเบย์ (Naïve Bayes) วิธีซัพพอร์ตเวกเตอร์แมทชีน (Support Vector Machines) วิธีต้นไม้ตัดสินใจ (Decision Tree) วิธีเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors) และเพอร์เซปตรอนแบบหลายชั้น (Multilayer Perceptron) โดยประเมินประสิทธิภาพด้วยวิธี 5-fold Cross Validation และใช้ค่าเมทริกซ์ คือ ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าเอฟวันสกอร์ (F1-Score) และค่าความถูกต้อง (Accuracy) จากนั้นจะได้ผลการประเมินประสิทธิภาพของแต่ละวิธี ดังตารางที่ 2

ตารางที่ 2 ผลการประเมินประสิทธิภาพของวิธีการจำแนกประเภท

Method	Precision	Recall	F1-Score	Accuracy
Naïve Bayes	0.9329	0.9148	0.9064	0.9148
Support Vector Machines	0.9964	0.9959	0.9955	**0.9959**
Decision Tree	0.9961	0.9959	0.9952	0.9957
K-Nearest Neighbors	0.9719	0.9730	0.9677	0.9730
Multilayer Perceptron	0.9964	0.9959	0.9956	**0.9959**

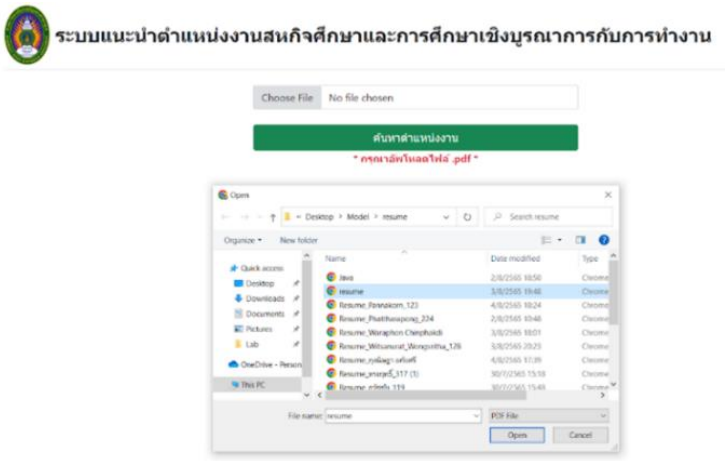
จากตารางที่ 2 ผลการการประเมินประสิทธิภาพของวิธีการจำแนกประเภทตำแหน่งงานทั้ง 5 วิธี พบว่า วิธี Multilayer Perceptron และวิธี Support Vector Machines มีค่าความถูกต้อง (Accuracy) เท่ากันที่ 99.59% มากกว่าวิธีการอื่น ๆ รองลงมาได้แก่ วิธี Decision Tree มีค่าความถูกต้อง 99.57% วิธี *K*-Nearest Neighbors มีความความถูกต้อง 97.30% และวิธี Naïve Bayes มีความความถูกต้องน้อยที่สุด 91.48% ตามลำดับ สามารถแสดงกราฟเปรียบเทียบประสิทธิภาพของแต่ละวิธีการจำแนกประเภทได้ ดังภาพที่ 6



ภาพที่ 6 เปรียบเทียบประสิทธิภาพของแต่ละวิธีการจำแนกประเภทข้อมูล

2. ผลการพัฒนาระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง

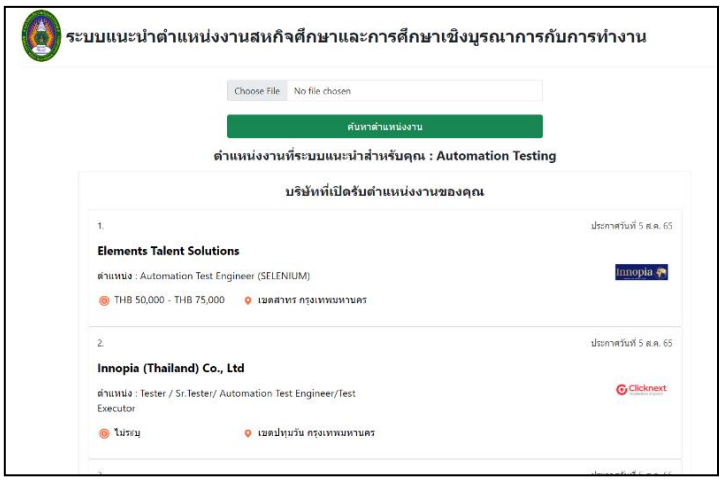
2.1 ผลการออกแบบองค์ประกอบระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงาน พบว่า ระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานสามารถทำงานได้จริงตามที่ออกแบบไว้ โดยมีองค์ประกอบของระบบ 4 ระบบ ได้แก่ ระบบสมาชิก ระบบประมวลผลไฟล์ PDF ประวัติย่อของผู้สมัครงาน ระบบประมวลผลข้อมูลร่วมกับแบบจำลอง ระบบแสดงชื่อตำแหน่งงานที่สอดคล้องกับประวัติย่อของผู้สมัครงาน และระบบแสดงตำแหน่งงานว่างจากบริษัทจัดหางาน ดังภาพที่ 7



ภาพที่ 7 หน้าอัปโหลดไฟล์ PDF ประวัติย่อของผู้สมัครงาน

จากภาพที่ 7 หน้าระบบอัปโหลดไฟล์ PDF ประวัติย่อของผู้สมัครงาน ระบบจะประมวลผลดึงข้อความจากไฟล์และนำข้อความที่ได้ไปประมวลผลร่วมกับแบบจำลองที่มีประสิทธิภาพดีที่สุด เพื่อใช้ในการจำแนกประเภทตำแหน่งงานต่อไป

2.2 ผลการพัฒนาเว็บแอปพลิเคชันตามวงจรการพัฒนาระบบ พบว่า ผู้ใช้สามารถที่จะอัปโหลดไฟล์ประวัติย่อของผู้สมัครงานเพื่อให้ระบบทำการประมวลผลข้อมูลในประวัติย่อของผู้สมัครงาน และแนะนำตำแหน่งงานที่สอดคล้องกับประวัติย่อของผู้สมัครงานให้ผู้ใช้งานได้ ดังภาพที่ 8



ภาพที่ 8 เว็บแอปพลิเคชันแนะนำตำแหน่งงานที่สอดคล้องกับประวัติย่อของผู้สมัครงาน

จากภาพที่ 8 หลังจากที่ระบบแนะนำตำแหน่งงานให้กับผู้ใช้งานแล้ว ระบบยังเชื่อมต่อกับเว็บไซต์ของบริษัทจัดหางานเพื่อดึงตำแหน่งงานว่างและข้อมูลรายละเอียดของตำแหน่งงานมาให้ผู้ใช้งานด้วย เช่น ชื่อบริษัท ชื่อตำแหน่งงานที่เปิดรับสมัคร เงินเดือน วันที่ประกาศ

3. ผลการศึกษาความพึงพอใจของผู้ใช้งานระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง

ผู้วิจัยได้นำเว็บแอปพลิเคชันที่พัฒนาขึ้นให้กลุ่มตัวอย่างใช้งาน คือ นักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ชั้นปีที่ 4 จำนวน 30 คน โดยให้ตอบแบบสอบถามความพึงพอใจของผู้ใช้งานที่มีต่อระบบ จากนั้นนำผลการตอบแบบสอบถามมาวิเคราะห์ด้วยค่าสถิติพื้นฐานเทียบกับเกณฑ์และสรุปผล ดังตารางที่ 3

ตารางที่ 3 ผลการวิเคราะห์แบบสอบถามความพึงพอใจของผู้ใช้งาน

รายการ	$\bar{X}$	SD.	ระดับความคิดเห็น
1. ความเหมาะสมในการอัปโหลดไฟล์ประวัติย่อของผู้สมัครงาน	4.77	0.43	มากที่สุด
2. ความเหมาะสมของตำแหน่งงานที่แสดงสอดคล้องกับประวัติย่อของผู้สมัครงาน	4.60	0.50	มากที่สุด
3. ความเหมาะสมของการประมวลผลแบบจำลอง (Model) ในการแสดงตำแหน่งงาน	4.77	0.43	มากที่สุด
4. ความเหมาะสมในการแสดงตำแหน่งงานว่างของสถานประกอบการที่สอดคล้องกับประวัติย่อของผู้สมัครงาน	4.63	0.49	มากที่สุด
5. ความเหมาะสมของการแสดงรายละเอียดของตำแหน่งงานว่างของสถานประกอบการ	4.56	0.75	มากที่สุด
6. การออกแบบหน้าจอมีสัดส่วนที่เหมาะสม	4.73	0.45	มากที่สุด
7. ความเหมาะสมในการใช้ขนาดตัวอักษร	4.73	0.45	มากที่สุด
8. สีของตัวอักษรมีความชัดเจนอ่านง่าย	4.67	0.48	มากที่สุด
9. สีของพื้นหลังมีความเหมาะสม	4.70	0.47	มากที่สุด
โดยรวม	**4.69**	**0.47**	**มากที่สุด**

จากตารางที่ 3 ผลการศึกษาความพึงพอใจของระบบแนะนำตำแหน่งงานสหกิจศึกษาและการศึกษาเชิบูรณาการกับการทำงาน พบว่า อยู่ในระดับ มากที่สุด ($\bar{X}$ = 4.69, SD. = 0.47) เมื่อพิจารณเป็นรายข้อพบว่าข้อที่มีผลการประเมินสูงสุด คือ ความเหมาะสมในการอัปโหลดไฟล์ประวัติย่อของผู้สมัครงาน และ ความเหมาะสมของการประมวลผลแบบจำลอง (Model) ในการแสดงตำแหน่งงาน ($\bar{X}$ = 4.77, SD. = 0.43)

อภิปรายผลการวิจัย

1. ผลการสร้างแบบจำลองและหาประสิทธิภาพแบบจำลองการจำแนกประเภทตำแหน่งงาน โดยใช้ชุดข้อมูล Resume เพื่อทำการทดลองสร้างแบบจำลอง ใช้วิธีการจำแนกประเภท 5 วิธี และประเมินประสิทธิภาพของแต่ละวิธีด้วย 5-Fold Cross Validation พบว่า วิธีที่ให้ประสิทธิภาพมากที่สุด ได้แก่ วิธีเพอร์เซปตรอนแบบหลายชั้น และวิธีซัพพอร์ตเวกเตอร์แมทชีน มีค่าความถูกต้อง เท่ากันที่ 99.59% ทั้งนี้เนื่องจากวิธีเพอร์เซปตรอนแบบหลายชั้น เป็นอัลกอริทึมโครงข่ายประสาทเทียม ที่มีโครงสร้างแบบหลายชั้น เหมาะสำหรับประมวลผลหรือจำแนกประเภทข้อมูลที่มีความซับซ้อนได้ดี จึงส่งผลให้การจำแนกประเภทตำแหน่งงานได้ค่าความถูกต้องสูง สอดคล้องกับงานวิจัยของ Pradeep Kumar Roy และคณะ [4] ได้นำเสนอ วิธีการเรียนรู้ของเครื่องสำหรับระบบแนะนำประวัติ

ย่อของผู้สมัครงานแบบอัตโนมัติได้ พบว่า วิธีซัพพอร์ตเวกเตอร์แมทชีน ได้ค่าความถูกต้อง (Accuracy) สูงที่สุด 78.53%

2. ผลการพัฒนาระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง ได้พัฒนาระบบตามวงจรการพัฒนาระบบ SDLC 5 ขั้นตอน และนำแบบจำลองที่มีประสิทธิภาพมากที่สุดที่ได้จากขั้นตอนการสร้างแบบจำลอง มาใช้งานร่วมกับเว็ปแอปพลิเคชันเพื่อแนะนำตำแหน่งงานที่สอดคล้องกับประวัติย่อของผู้สมัครงานของผู้ใช้งาน สามารถทำงานได้ตามวัตถุประสงค์ที่กำหนดไว้ สอดคล้องกับงานวิจัยของ Daryani at al [3] ได้นำเสนอระบบคัดกรองประวัติย่อของผู้สมัครงานให้สอดคล้องกับรายละเอียดงาน

3. ผลการศึกษาความพึงพอใจของระบบเว็บแอปพลิเคชันแนะนำตำแหน่งงานด้วยวิธีการเรียนรู้ของเครื่อง พบว่า อยู่ในระดับ มากที่สุด ได้ค่าเฉลี่ยเท่ากับ 4.69 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.47 ทั้งนี้อาจเนื่องจากภายในระบบมีการประมวลผลด้วยแบบจำลองที่มีประสิทธิภาพในการจำแนกประเภทของตำแหน่งงาน จึงส่งผลให้ระบบสามารถแนะนำตำแหน่งงานได้อย่างถูกต้องกับประวัติย่อของผู้สมัครงานของผู้ใช้งาน สอดคล้องกับงานวิจัยของ กนิษฐา อินธิชิต, ภควัฒน์ ปิยวงษ์ และสุภาพร สุขใส [8] ได้วิจัยเรื่อง การพัฒนาแอปพลิเคชันช่วยตัดสินใจในการเลือกเรียนสาขาวิชาคอมพิวเตอร์ในมหาวิทยาลัยราชภัฏศรีสะเกษบนระบบปฏิบัติการแอนดรอยด์ พบว่า ผู้ใช้มีความพึงพอใจแอปพลิเคชันอยู่ในระดับมากที่สุด

ข้อเสนอแนะ

การวิจัยครั้งต่อไปควรหาชุดข้อมูลประวัติย่อของผู้สมัครงานที่เป็นภาษาไทย มาทำการทดลองสร้างแบบจำลองเพิ่มเติม และเลือกใช้วิธีการเรียนรู้เชิงลึก (Deep Learning) เพื่อจำแนกประเภทให้แม่นยำมากยิ่งขึ้น

เอกสารอ้างอิง

[1] กระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม. (2565). *โครงการส่งเสริมการจัดสหกิจศึกษาและการศึกษาเชิงบูรณาการกับการทำงาน (Cooperative and Work Integrated Education หรือ CWIE).* สืบค้น 11 มกราคม 2566, จาก https://www.mhesi.go.th/index.php/flagship-project/6820-CWIE.html

[2] Goindani, M., Liu, Q., Chao, J., & Jijkoun, V. (2017). Employer Industry Classification Using Job Postings. *IEEE International Conference on Data Mining Workshops*, ICDMW, 2017-November, 183–188. Retrieved from https://doi.org/10.1109/ICDMW.2017.30

[3] Daryani, C., Chhabra, G. S., Patel, H., Chhabra, I. K., & Patel, R. (2020). AN AUTOMATED RESUME SCREENING SYSTEM USING NATURAL LANGUAGE PROCESSING AND SIMILARITY. *Topics In Intelligent Computing And Industry Design*, 2(2): 99-103. Retrieved from https://doi.org/10.26480/ETIT.02.2020.99.103

[4] Roy, P. K., Chowdhary, S. S., & Bhatia, R. (2020). A Machine Learning approach for automation of Resume Recommendation system. *Procedia Computer Science*, 167, 2318–2327. Retrieved from https://doi.org/10.1016/J.PROCS.2020.03.284

[5] Javed Mehedi Shamrat, F. M., Tasnim, Z., Ghosh, P., Majumder, A., & Hasan, M. Z. (2020). Personalization of Job Circular Announcement to Applicants Using Decision Tree Classification Algorithm. 2020 IEEE *International Conference for Innovation in Technology*, INOCON 2020. Retrieved from https://doi.org/10.1109/INOCON50539.2020.9298253

[6] van Huynh, T., van Nguyen, K., Nguyen, N. L. T., & Nguyen, A. G. T. (2020). Job Prediction: From Deep Neural Network Models to Applications. *Proceedings - 2020 RIVF International Conference on Computing and Communication Technologies,* RIVF 2020. Retrieved from https://doi.org/10.1109/RIVF48685.2020.9140760

[7] Tran, H. T., Vo, H. H. P., & Luu, S. T. (2021). Predicting Job Titles from Job Descriptions with Multi-label Text Classification. *Proceedings - 2021 8th NAFOSTED Conference on Information and Computer Science,* NICS 2021, 513–518. Retrieved from https://doi.org/10.1109/NICS54270.2021.9701541

[8] กนิษฐา อินธิชิต, ภควัฒน์ ปิยวงษ์ และสุภาพร สุขใส, (2565). การพัฒนาแอปพลิเคชันช่วยตัดสินใจในการเลือกเรียนสาขาวิชาคอมพิวเตอร์ ในมหาวิทยาลัยราชภัฏศรีสะเกษ บนระบบปฏิบัติการแอนดรอยด์ โดยใช้เทคนิคต้นไม้ตัดสินใจ. *วารสารวิชาการ การจัดการเทคโนโลยี มหาวิทยาลัยราชภัฏมหาสารคาม* , 9(2), 97–107. สืบค้นจาก https://ph02.tci-thaijo.org/index.php/itm-journal/article/view/247879/168306

[9] Kaggle. (2021). *Resume Dataset.* Retrieved January 11, 2023, from https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset

[10] Zhu, C. (2021). *The basics of natural language processing.* Machine Reading Comprehension, 27–46. Retrieved from https://doi.org/10.1016/B978-0-323-90118-5.00002-3

[11] NLTK. (2023). *nltk.tokenize.casual module.* Retrieved January 11, 2023, from https://www.nltk.org/api/nltk.tokenize.casual.html

[12] Jabri, S., Dahbi, A., Gadi, T., & Bassir, A. (2018). Ranking of text documents using TF-IDF weighting and association rules mining. *Proceedings of the 2018 International Conference on Optimization and Applications,* ICOA 2018, 1–6. Retrieved from https://doi.org/10.1109/ICOA.2018.8370597

[13] Google. (n.d.). *Colaboratory.* Retrieved January 11, 2023, from https://research.google.com/colaboratory/faq.html

[14] Shalev-Shwartz, S., & Ben-David, S. (2013). Understanding machine learning: From theory to algorithms. In Understanding Machine Learning: *From Theory to Algorithms* (Vol. 9781107057). Retrieved from https://doi.org/10.1017/CBO9781107298019

[15] Kohavi, R. (n.d.). *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection.* Retrieved January 11, 2023, from http//roboticsStanfordedu/"ronnykLikert,

[16] Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22 140, 55.