การประยุกต์เทคนิคเหมืองข้อมูลเพื่อพยากรณ์ผลการประเมินบทความจากวารสารนวัตกรรมสังคมและการเรียนรู้ตลอดชีวิต

Application of Data Mining Techniques for Article Evaluation Results Prediction from Journal of Social Innovation and Lifelong Learning

สืบพงษ์ พงษ์สวัสดิ์[1], พงศ์กร จันทราช[2]

[1]หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูลและนวัตกรรมดิจิทัล, มหาวิทยาลัยฟาร์อีสเทอร์น, subpong@feu.edu
[2]คณะนวัตกรรม เทคโนโลยี และการสร้างสรรค์, มหาวิทยาลัยฟาร์อีสเทอร์น, pongkorn@feu.edu

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) เพื่อสร้างโมเดลการพยากรณ์ผลการประเมินบทความจากข้อมูลในระบบวารสาร และ 2) เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลที่ใช้จากข้อมูลการประเมินบทความของวารสารนวัตกรรมสังคมและการเรียนรู้ตลอดชีวิต จำนวน 170 บทความ โดยใช้เทคนิค Naive Bayes Decision Tree และ K-Nearest Neighbors ด้วยภาษาไพธอนบนแพลตฟอร์ม Google Colab ผลการวิจัย พบว่า การคัดเลือกคุณลักษณะที่สำคัญต่อการทำนายโดยใช้ Information Gain จากคุณลักษณะทั้งหมด 19 คุณลักษณะ พบว่า 10 คุณลักษณะมีความสำคัญ ได้แก่ การใช้ภาษา รูปแบบตารางและภาพประกอบ ไฟล์บทความและเอกสารประกอบ ความชัดเจนและรายละเอียดของบทคัดย่อและ Abstract การกำหนดวัตถุประสงค์การวิจัยในบทคัดย่อและ Abstract การระบุวิธีการวิจัยในบทคัดย่อและ Abstract ผลการวิจัยในบทคัดย่อและ Abstract โจทย์วิจัยชัดเจน วิธีการวิจัย และการอภิปรายผลการวิจัย จากนั้นใช้เทคนิค Naive Bayes Decision Tree และ K-Nearest Neighbors ในการสร้างแบบจำลองทำนาย พบว่า Naive Bayes มีประสิทธิภาพต่ำที่สุด (20.66%) ในขณะที่ Nearest Neighbors มีความแม่นยำโดยรวมดีที่สุด (82.46%) แต่ Decision Tree ให้ผลลัพธ์ที่สม่ำเสมอมากกว่า ผลการวิจัยนี้สามารถนำไปใช้ในการพัฒนาระบบสนับสนุนการประเมินบทความ และเป็นแนวทางสำหรับผู้เขียนในการปรับปรุงคุณภาพบทความเพื่อเพิ่มโอกาสในการตีพิมพ์

คำหลัก: เหมืองข้อมูล การประเมินบทความ การทำนาย

Abstract

This research aims to (1) develop a predictive model for article evaluation results based on data from the Journal of Social Innovation and Lifelong Learning, and (2) compare the performance of data mining techniques used. The study analyzed evaluation data from 170 articles using Naive Bayes, Decision Tree, and K-Nearest Neighbors techniques implemented in

Python on the Google Colab platform. Feature selection using Information Gain from 19 attributes identified 10 key features important for prediction: language use, table and illustration formats, article and supplementary file quality, clarity and detail of the abstract and summary, specification of research objectives in the abstract and summary, identification of research methods in the abstract and summary, research results in the abstract and summary, clarity of research questions, research methods, and discussion of research findings. The predictive models were then built using Naive Bayes, Decision Tree, and K-Nearest Neighbors. The results showed that Naive Bayes had the lowest accuracy at 20.66%, while K-Nearest Neighbors achieved the highest overall accuracy at 82.46%. However, the Decision Tree produced more consistent results. The findings of this research can be applied to develop a support system for article evaluation and serve as a guideline for authors to improve article quality, thereby increasing their chances of publication.

***Keywords**:* Data Mining, Article Evaluation, Prediction

ความเป็นมาและความสำคัญของปัญหา

การตีพิมพ์บทความในวารสารวิชาการเป็นดัชนีชี้วัดสำคัญของความก้าวหน้าทางวิชาการและการวิจัย ทั้งสำหรับนักวิจัย สถาบันการศึกษา และหน่วยงานวิจัยต่างๆ (Björk & Solomon, 2013) อย่างไรก็ตาม กระบวนการประเมินและคัดเลือกบทความเพื่อตีพิมพ์มีความซับซ้อนและมีปัจจัยหลายประการที่ส่งผลต่อการพิจารณา ซึ่งนำไปสู่อัตราการปฏิเสธบทความที่สูงในหลายสาขาวิชา โดยเฉพาะในวารสารที่มีค่า Impact Factor สูง (Hargens, 2019) การทำความเข้าใจถึงปัจจัยที่ส่งผลต่อการปฏิเสธการตีพิมพ์จึงมีความสำคัญอย่างยิ่งสำหรับนักวิจัยที่ต้องการเพิ่มโอกาสในการเผยแพร่ผลงานของตน

ในปัจจุบัน เทคนิคการทำเหมืองข้อมูล (Data Mining) ได้เข้ามามีบทบาทสำคัญในการวิเคราะห์ข้อมูลขนาดใหญ่และซับซ้อน โดยสามารถค้นหารูปแบบ (Patterns) ความสัมพันธ์ที่ซ่อนอยู่ และสร้างโมเดลการทำนายที่มีประสิทธิภาพ (Chen et al., 2022) การประยุกต์ใช้เทคนิคเหล่านี้ในการวิเคราะห์ปัจจัยที่ส่งผลต่อการปฏิเสธการตีพิมพ์บทความจึงเป็นแนวทางที่มีศักยภาพในการสร้างความเข้าใจที่ลึกซึ้งและแม่นยำยิ่งขึ้น งานวิจัยนี้มุ่งศึกษาและวิเคราะห์ปัจจัยที่ส่งผลต่อความเสี่ยงในการปฏิเสธการตีพิมพ์บทความวิชาการ โดยใช้เทคนิคการทำเหมืองข้อมูลหลากหลายรูปแบบ ได้แก่ เทคนิคนาอีฟเบย์ (Naive Bayes) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคเพื่อนบ้านใกล้สุด (K-nearest Neighbor) และวัดประสิทธิภาพของตัวแบบการทำนายด้วย ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure)

จากการทบทวนวรรณกรรมที่เกี่ยวข้อง พบว่ามีการศึกษาเกี่ยวกับปัจจัยที่ส่งผลต่อการตีพิมพ์บทความในหลายมิติ เช่น คุณภาพของงานวิจัย (Calcagno et al., 2012) ความสอดคล้องกับขอบเขตของวารสาร (Wang et al., 2016) และปัจจัยด้านภาษาและการเขียน (Santos et al., 2020) อย่างไรก็ตาม ยังขาดการวิเคราะห์ในเชิงลึก

ด้วยเทคนิคการทำเหมืองข้อมูลที่ทันสมัย บทความนี้นำเสนอแนวทางใหม่ในการวิเคราะห์ปัจจัยที่ส่งผลต่อการปฏิเสธการตีพิมพ์บทความ โดยใช้เทคนิคการทำเหมืองข้อมูลเพื่อสกัดความรู้จากข้อมูลขนาดใหญ่ โดยปัจจัยที่นำมาพิจารณาประกอบด้วย ขอบเขตเนื้อหาของบทความ ชื่อเรื่องตรงกับประเด็นปัญหาที่ศึกษา การใช้ภาษา คุณภาพการนำเสนอ จำนวนหน้า รูปแบบและขนาดตัวอักษร รูปแบบการพิมพ์ รูปแบบการอ้างอิง รูปแบบตารางและภาพประกอบ ไฟล์บทความและเอกสารประกอบ ความชัดเจนและรายละเอียดของบทคัดย่อและ Abstract การกำหนดวัตถุประสงค์การวิจัยในบทคัดย่อและ Abstract การระบุวิธีการวิจัยในบทคัดย่อและ Abstract ผลการวิจัยในบทคัดย่อและ Abstract โจทย์วิจัยชัดเจน การทบทวนวรรณกรรมที่เกี่ยวข้อง วิธีการวิจัย ผลการวิจัย การอภิปรายผลการวิจัย และการสรุปผลการวิจัยและให้ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์

ผลการวิจัยนี้จะช่วยให้นักวิจัยเข้าใจถึงปัจจัยสำคัญที่ส่งผลต่อการพิจารณาบทความ และสามารถนำไปประยุกต์ใช้ในการเตรียมบทความเพื่อเพิ่มโอกาสในการตีพิมพ์ นอกจากนี้ ยังเป็นประโยชน์ต่อกองบรรณาธิการวารสารในการพัฒนากระบวนการประเมินบทความให้มีความเป็นธรรมและมีประสิทธิภาพมากยิ่งขึ้น

วัตถุประสงค์

1. เพื่อสร้างโมเดลการพยากรณ์ผลการประเมินบทความจากข้อมูลในระบบวารสาร
2. เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลที่ใช้

ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถทำนายผลการประเมินบทความได้อย่างแม่นยำและรวดเร็ว ช่วยลดภาระงานในการประเมินด้วยมือ และเพิ่มความสม่ำเสมอและความเป็นกลางในการตัดสินใจ นอกจากนี้ยังช่วยสนับสนุนการตัดสินใจเชิงกลยุทธ์ในการจัดการบทความและพัฒนาคุณภาพของวารสารได้อย่างมีประสิทธิภาพ
2. เทคนิคที่เหมาะสมที่สุดสำหรับการพยากรณ์ผลการประเมินบทความในระบบวารสาร โดยการวิเคราะห์ข้อดีข้อเสียและความแม่นยำของแต่ละเทคนิค จะช่วยให้สามารถเลือกใช้วิธีการที่มีประสิทธิภาพสูงสุดในการประมวลผลข้อมูลและพยากรณ์ผลได้อย่างถูกต้องและรวดเร็ว ซึ่งจะส่งผลให้กระบวนการประเมินบทความมีความน่าเชื่อถือและประสิทธิผลมากยิ่งขึ้น

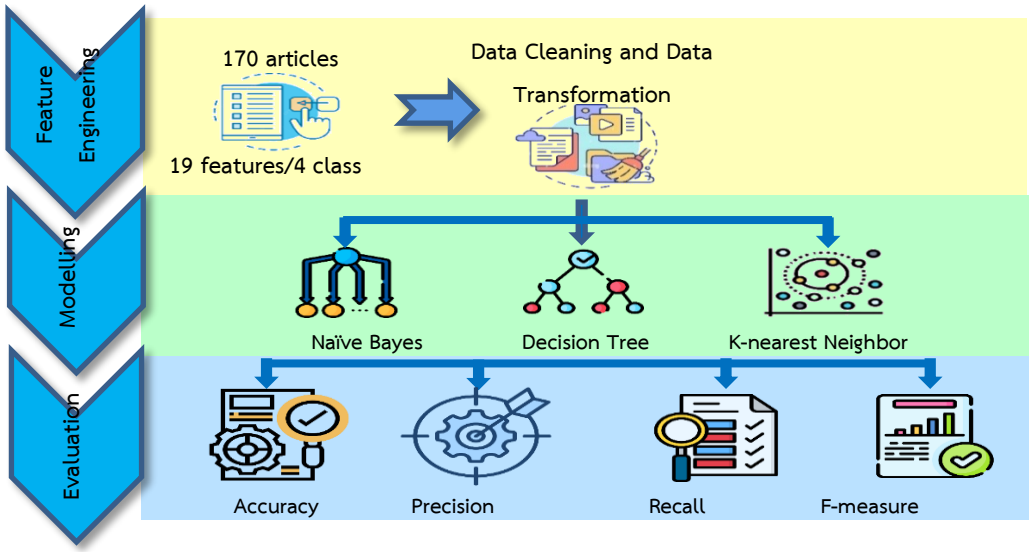
ทบทวนวรรณกรรม

เทคนิคการทำเหมืองข้อมูลมีการนำมาประยุกต์ใช้อย่างหลากหลาย ในด้านการศึกษามีการทำนายคุณสมบัติของนักศึกษาเพื่อให้อาจารย์ที่ปรึกษาจะสามารถดูแลนักศึกษาได้ตรงกับกลุ่มเป้าหมาย ดังการวิจัยของสุธีรา วงศ์อนันทรัพย์ (2559) ใช้เทคนิคเหมืองข้อมูลในการประเมินความรู้ และหาความถนัดเพื่อพัฒนาศักยภาพของนักศึกษา ผลการศึกษาพบว่า งานวิจัยนี้นำผลการทำแบบทดสอบประเมินความรู้ก่อนเรียน และแบบทดสอบวัดความถนัดของกลุ่มตัวอย่างมาวิเคราะห์โดยใช้เทคนิคการทำเหมืองข้อมูลผลการวิเคราะห์แบบประเมินความรู้

ก่อนเรียน ผลลัพธ์ที่ได้สามารถจำแนกนักศึกษา โดยใช้เทคนิคการจัดกลุ่มด้วยอัลกอริธึมส์ K-mean และงานวิจัยของ ชนะวงศ์ คงสอน และ ณัฐวุฒิ วิสุทธิพิเนตร (2559) ได้พัฒนาระบบสนับสนุนการตัดสินใจในการเลือกสาขาการเรียนต่อระดับประกาศนียบัตรวิชาชีพ สังกัดอาชีวศึกษาจังหวัดพระนครศรีอยุธยา ผลการศึกษาพบว่า การเปรียบเทียบประสิทธิภาพของโมเดลการจำแนกข้อมูลทั้ง 3 ได้แก่ K-nearest Neighbors, Decision Tree และ Rule based ซึ่งผลการเปรียบเทียบประสิทธิภาพโมเดลที่ได้ จากเทคนิค K-nearest Neighbors (K=6) มีความถูกต้องมากที่สุดร้อยละ 76.62

การประยุกต์เหมืองข้อมูลในด้านภูมิปัญญาท้องถิ่น ดังการศึกษาของ นพมาศ ปักเข็ม และคณะ (2566) ที่ได้จำแนกประเภทภูมิปัญญาท้องถิ่นแบบอัตโนมัติด้วยวิธีการทางเหมืองข้อมูลเพื่อให้สามารถจำแนกประเภทภูมิปัญญาท้องถิ่นจากข้อมูลแบบพรรณนาโวหารด้วยการประมวลผลของคอมพิวเตอร์ได้ช่วยลดเวลาในการจับใจความรวมถึงแก้ปัญหาการจำแนกประเภทโดยมนุษย์ โดยใช้ Decision Tree K-Nearest Neighbor และ Naïve Bayes พบว่า อัลกอริทึม Naive Bayes ให้ค่าความแม่นยำสูงที่สุด จากนั้นทำการพัฒนาการใช้งานโมเดลผ่านทาง Web Application โดยการพัฒนาด้วยภาษา PHP ร่วมกบั WEKA API ทำให้สามารถนาพัฒนาไปสู่ระบบแหล่งรวมสารสนเทศทางภูมิปัญญาท้องถิ่นที่มีประสิทธิภาพได้ นอกจากนี้ยังมีการประยุกต์ด้านวิทยาศาสตร์การแพทย์ ดังการศึกษาของ นพรัตน์ นนท์ศิริ และคณะ (2566) ที่ได้สร้างแบบจำลองการจำแนกข้อมูลเพื่อวินิจฉัยความเสี่ยงของการเป็นโรคเบาหวานโดยใช้วิธี Naïve Bayes Support Vector Machine K-Nearest Neighbor และ Decision Tree ผลการวิจัยพบว่า วิธีต้นไม้ตัดสินใจมีประสิทธิภาพในการสร้างแบบจำลองมากที่สุด เนื่องจากเป็นวิธีที่ไม่มีการแจกแจงหรือไม่ใช้พารามิเตอร์ซึ่งไม่ได้ขึ้นอยู่กับสมมติฐานการแจกแจงความน่าจะเป็น สามารถจัดการกับข้อมูลที่มีมิติสูงได้อย่างแม่นยำ และสามารถนำมาใช้เป็นแนวทางในการสนับสนุนการตัดสินใจทางการแพทย์เพื่อวินิจฉัยความเสี่ยงการเป็นโรคเบาหวาน

กรอบแนวคิด



ภาพ 1 กรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

1. ข้อมูลที่ใช้ในการวิจัยนี้ใช้ข้อมูลบทความจากวารสารนวัตกรรมสังคมและการเรียนรู้ตลอดชีวิต ปีที่ 12 ฉบับที่ 1 พ.ศ. 2561 ถึง ปีที่ 18 ฉบับที่ 1 พ.ศ. 2567 จำนวน 170 บทความ

2. การวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลเพื่อทำนายผลการประเมินบทความของผู้ทรงคุณวุฒิประจำสาขา จากผลการประเมินคุณภาพบทความจากผู้ทรงคุณวุฒิประจำกองบรรณาธิการ ด้วยการใช้เทคนิคนาอีฟเบย์ (Naive Bayes) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคเพื่อนบ้านใกล้สุด (K-nearest Neighbor) และวัดประสิทธิภาพของตัวแบบการทำนายด้วย ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure) โดยใช้ภาษาไพธอนในการวิเคราะห์และใช้ Google Colab เป็นเครื่องมือ

การทำนายผลการประเมินคุณภาพบทความจากผู้ทรงคุณวุฒิประจำกองบรรณาธิการประกอบด้วย 19 คุณลักษณะ (Feature) ดังนี้

- ขอบเขตเนื้อหาของบทความ
- ชื่อเรื่องตรงกับประเด็นปัญหาที่ศึกษา
- การใช้ภาษา
- คุณภาพการนำเสนอ
- จำนวนหน้า รูปแบบและขนาดตัวอักษร
- รูปแบบการพิมพ์
- รูปแบบการอ้างอิง
- รูปแบบตารางและภาพประกอบ
- ไฟล์บทความและเอกสารประกอบ
- ความชัดเจนและรายละเอียดของบทคัดย่อและ Abstract
- การกำหนดวัตถุประสงค์การวิจัยในบทคัดย่อและ Abstract
- การระบุวิธีการวิจัยในบทคัดย่อและ Abstract
- ผลการวิจัยในบทคัดย่อและ Abstract
- โจทย์วิจัยชัดเจน
- การทบทวนวรรณกรรมที่เกี่ยวข้อง
- วิธีการวิจัย
- ผลการวิจัย
- การอภิปรายผลการวิจัย
- การสรุปผลการวิจัยและให้ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์

ในแต่ละคุณลักษณะมีผลการประเมินได้ 2 แบบ ได้แก่ ผ่าน และ ไม่ผ่าน การทำนายผลการประเมินคุณภาพบทความจากผู้ทรงคุณวุฒิประจำสาขาประกอบด้วยผลการประเมิน 4 แบบ ดังนี้ รับตีพิมพ์บทความโดยไม่ต้องแก้ไข (Accept Submission) แก้ไขบทความโดยให้บรรณาธิการพิจารณาต่อ (Revisions Required) แก้ไขบทความโดยส่งให้ผู้ประเมินตรวจสอบอีกครั้ง (Resubmit for Review) และยังไม่สมควรนำบทความนี้ลงตีพิมพ์ (Decline Submission)

3. การเตรียมข้อมูลสำหรับการวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลเพื่อทำนายผลการประเมินบทความของผู้ทรงคุณวุฒิประจำสาขา จากผลการประเมินคุณภาพบทความจากผู้ทรงคุณวุฒิประจำกองบรรณาธิการ จำนวน 170 บทความ ตั้งแต่ปีที่ 12 ฉบับที่ 1 พ.ศ. 2561 ถึง ปีที่ 18 ฉบับที่ 1 พ.ศ. 2567 ประกอบด้วย 19 คุณลักษณะ ซึ่งเป็นหัวข้อของการประเมินคุณภาพบทความโดยผู้ทรงคุณวุฒิประจำกองบรรณาธิการ โดยทำการเปลี่ยนชื่อคุณลักษณะให้เป็นชื่อภาษาอังกฤษเพื่อความสะดวกในการวิเคราะห์ ดังนี้

ตาราง 1 ชื่อคุณลักษณะและชื่อฟิลด์

ชื่อคุณลักษณะ (Feature Name)	**ชื่อฟิลด์ (Column Name)**
ขอบเขตเนื้อหาของบทความ	Item1
ชื่อเรื่องตรงกับประเด็นปัญหาที่ศึกษา	Item2
การใช้ภาษา	Item3
คุณภาพการนำเสนอ	Item4
จำนวนหน้า รูปแบบและขนาดตัวอักษร	Item5
รูปแบบการพิมพ์	Item6
รูปแบบการอ้างอิง	Item7
รูปแบบตารางและภาพประกอบ	Item8
ไฟล์บทความและเอกสารประกอบ	Item9
ความชัดเจนและรายละเอียดของบทคัดย่อและ Abstract	Item10
การกำหนดวัตถุประสงค์การวิจัยในบทคัดย่อและ Abstract	Item11
การระบุวิธีการวิจัยในบทคัดย่อและ Abstract	Item12
ผลการวิจัยในบทคัดย่อและ Abstract	Item13
โจทย์วิจัยชัดเจน	Item14
การทบทวนวรรณกรรมที่เกี่ยวข้อง	Item15
วิธีการวิจัย	Item16
ผลการวิจัย	Item17
การอภิปรายผลการวิจัย	Item18
การสรุปผลการวิจัยและให้ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์	Item19

การเตรียมข้อมูลได้ทำการส่งออกข้อมูลผลการประเมินคุณภาพบทความจากผู้ทรงคุณวุฒิประจำกองบรรณาธิการทั้งหมด 19 คุณลักษณะ โดยแต่ละคุณลักษณะมีผลการประเมินเป็น ผ่าน และ ไม่ผ่าน ส่วนผลการประเมินของผู้ทรงคุณวุฒิประจำสาขา เป็น target ประกอบด้วยแก้ไขบทความโดยให้บรรณาธิการพิจารณาต่อ (Revisions Required) แก้ไขบทความโดยส่งให้ผู้ประเมินตรวจสอบอีกครั้ง (Resubmit for Review) และยังไม่สมควรนำบทความนี้ลงตีพิมพ์ (Decline Submission) จากนั้นทำการแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมสำหรับการประมวลผลได้ โดยการเปลี่ยนผลการประเมินจาก ผ่าน และ ไม่ผ่าน ให้เป็น 1 และ 0 พร้อมกับเปลี่ยนชื่อคุณลักษณะให้เป็นภาษาอังกฤษ ดังนี้

ตาราง 2 ข้อมูลที่พร้อมสำหรับการวิเคราะห์ข้อมูล

item1	item2	item3	item4	item5	item6	item7	item8	item9	item10	item11	item12	item13	item14	item15	item16	item17	item18	item19	label
1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	Revisions Required
1	1	1	1	1	0	0	1	1	0	1	0	1	1	0	1	0	1	1	Revisions Required
1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	Revisions Required
1	0	0	0	1	1	1	0	1	0	1	1	0	1	1	0	0	0	1	Resubmit for Review
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Resubmit for Review
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Revisions Required
1	1	0	1	1	1	1	1	1	0	0	1	1	1	0	1	1	0	0	Revisions Required
1	0	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	Revisions Required
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	Revisions Required
1	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	Revisions Required
1	1	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	Resubmit for Review
1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	Revisions Required
1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	1	0	1	0	Revisions Required
1	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1	Revisions Required
1	1	1	0	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	Revisions Required
1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	Revisions Required
1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	1	1	0	0	Revisions Required

ผลการวิจัย

1. ผลการคัดเลือกคุณลักษณะ

เป็นการคัดเลือกคุณลักษณะที่มีความสำคัญกับแบบจำลองมากที่สุดเพื่อลดความซับซ้อนของแบบจำลอง ปรับปรุงประสิทธิภาพของแบบจำลอง และลดระยะเวลาการคำนวณ โดยใช้ค่า Information Gain ดังนี้

```python
import pandas as pd
import numpy as np
X = data[['item1', 'item2', 'item3', 'item4', 'item5',
          'item6', 'item7', 'item8', 'item9', 'item10',
          'item11', 'item12', 'item13', 'item14', 'item15',
          'item16', 'item17', 'item18', 'item19']]
y = data['label']

def calculate_entropy(y):
    probabilities = y.value_counts(normalize=True)
    entropy = -np.sum(probabilities * np.log2(probabilities + 1e-9))
    return entropy

def calculate_information_gain(X, y, feature):
    original_entropy = calculate_entropy(y)
    feature_values = X[feature].unique()

    weighted_entropy = 0
    for value in feature_values:
        subset_y = y[X[feature] == value]
        weighted_entropy += (len(subset_y) / len(y)) * calculate_entropy(subset_y)

    information_gain = original_entropy - weighted_entropy
    return information_gain

for feature in X.columns:
    entropy = calculate_entropy(y)
    print(f'entropy {feature}: {entropy}')

for feature in X.columns:
    ig = calculate_information_gain(X, y, feature)
    print(f'Information Gain for {feature}: {ig}')

```

feature name	Information Gain
item1	0.00000000
item2	0.00984889
item3	0.06561602
item4	0.02326376
item5	0.00647358
item6	0.02131568
item7	0.01449483
item8	0.08703697
item9	0.03507365
item10	0.04119300
item11	0.05992732
item12	0.03242172
item13	0.04176766
item14	0.03952998
item15	0.00995376
item16	0.04195875
item17	0.02729627
item18	0.05476281
item19	0.02410549

ภาพ 2 การคำนวณค่า Information Gain ด้วยภาษาไพธอน

จะเห็นได้ว่ามีคุณลักษณะที่มีความสำคัญโดยพิจารณาค่า Information Gain มากกว่า 0.03 จำนวน 10 คุณลักษณะ ประกอบด้วย item3 item8 item9 item10 item11 item12 item13 item14 item16 และ item18 นั่นคือ การใช้ภาษา รูปแบบตารางและภาพประกอบ ไฟล์บทความและเอกสารประกอบ ความชัดเจนและรายละเอียดของบทคัดย่อและ Abstract การกำหนดวัตถุประสงค์การวิจัยในบทคัดย่อและ Abstract การระบุวิธีการวิจัยในบทคัดย่อและ Abstract ผลการวิจัยในบทคัดย่อและ Abstract โจทย์วิจัยชัดเจน วิธีการวิจัย และการอภิปรายผลการวิจัย

2. ผลการทำนายผลการประเมินบทความของผู้ทรงคุณวุฒิประจำสาขาจากผลการประเมินคุณภาพบทความของผู้ทรงคุณวุฒิประจำกองบรรณาธิการ

การทำนายผลการประเมินบทความของผู้ทรงคุณวุฒิประจำสาขา จากผลการประเมินคุณภาพบทความจากผู้ทรงคุณวุฒิประจำกองบรรณาธิการ ได้ใช้เทคนิคนาอีฟเบย์ (Naive Bayes) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคเพื่อนบ้านใกล้สุด (K-nearest Neighbor) และวัดประสิทธิภาพของตัวแบบการทำนายด้วย ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure) โดยกำหนดให้ตัวแปร X แทน ชุดข้อมูลคุณลักษณะ (Feature Set) หรือตัวแปรต้นโดยเป็นผลประเมินจากผู้ทรงคุณวุฒิประจำกองบรรณาธิการ ที่ใช้ในการทำนายหรือจำแนก ตัวแปร y แทน ชุดข้อมูลป้ายกำกับ (Label

Set) หรือตัวแปรตาม ที่เป็นผลลัพธ์ที่ต้องการทำนาย ซึ่งเป็นผลการประเมินจากผู้ทรงคุณวุฒิประจำสาขา ใช้ฟังก์ชัน train_test_split จาก scikit-learn เพื่อแบ่งข้อมูล X และ y ออกเป็นสองส่วน คือ X_train และ y_train เป็นชุดข้อมูลสำหรับการเรียนรู้ (Training Set) 80% ของข้อมูลทั้งหมด ส่วน X_test และ y_test เป็นชุดข้อมูลสำหรับทดสอบประสิทธิภาพของโมเดล (Test Set) กำหนดสัดส่วนของข้อมูลที่จะใช้เป็นชุดทดสอบ คือ 20% ของข้อมูลทั้งหมด และกำหนดค่าเริ่มต้นในการสุ่มแบ่งข้อมูลเท่ากับ 10 เพื่อให้ได้ผลลัพธ์ที่เหมือนเดิมทุกครั้งที่รันโค้ด (Reproducibility) โดยได้ทำนายจากคุณลักษณะทั้งหมด 10 คุณลักษณะ

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.metrics import confusion_matrix
data = pd.read_csv('journal1.csv')
X = data[['item3', 'item8', 'item9', 'item10', 'item11', 'item12', 'item13', 'item14', 'item16', 'item18']]
y = data['label']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=10)
model = GaussianNB()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')
recall = recall_score(y_test, y_pred, average='weighted')
f1 = f1_score(y_test, y_pred, average='weighted')
print("Accuracy: %.4f" %accuracy)
print("Precision: %.4f" %precision)
print("Recall: %.4f" %recall)
print("F1 Score: %.4f" %f1)

new_data = pd.DataFrame([[1, 0, 0, 0, 0, 1, 1, 1, 1, 0]])
predicted_label = model.predict(new_data)
print("Predicted label:", predicted_label)
```

```
Accuracy: 0.3235
Precision: 0.1809
Recall: 0.3235
F1 Score: 0.2066
Predicted label: ['Resubmit for Review']
```

การทำนายด้วย Naïve Bayes

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
data = pd.read_csv('journal1.csv')
X = data[['item3', 'item8', 'item9', 'item10', 'item11', 'item12', 'item13', 'item14', 'item16', 'item18']]
y = data['label']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = DecisionTreeClassifier()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')
recall = recall_score(y_test, y_pred, average='weighted')
f1 = f1_score(y_test, y_pred, average='weighted')
print("Accuracy: %.4f" %accuracy)
print("Precision: %.4f" %precision)
print("Recall: %.4f" %recall)
print("F1 Score: %.4f" %f1)

new_data = pd.DataFrame([[1, 0, 0, 0, 0, 1, 1, 1, 1, 0]])
predicted_label = model.predict(new_data)
print("Predicted label:", predicted_label)
```

```
Accuracy: 0.7353
Precision: 0.7225
Recall: 0.7353
F1 Score: 0.7257
Predicted label: ['Resubmit for Review']
```

การทำนายด้วย Decision Tree

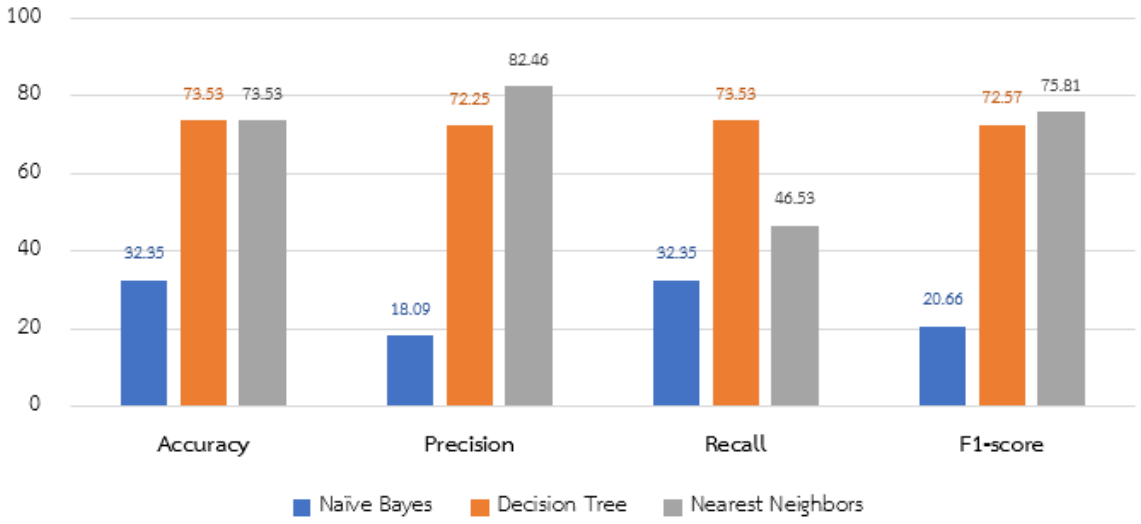
```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
data = pd.read_csv('journal1.csv')
X = data[['item3', 'item8', 'item9', 'item10', 'item11', 'item12', 'item13', 'item14', 'item16', 'item18']]
y = data['label']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(X_train, y_train)
y_pred = knn.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')
recall = recall_score(y_test, y_pred, average='weighted')
f1 = f1_score(y_test, y_pred, average='weighted')
print("Accuracy: %.4f" %accuracy)
print("Precision: %.4f" %precision)
print("Recall: %.4f" %recall)
print("F1 Score: %.4f" %f1)

new_data = pd.DataFrame([[1, 0, 0, 0, 0, 1, 1, 1, 1, 0]])
predicted_label = knn.predict(new_data)
print("Predicted label:", predicted_label)
```

```
Accuracy: 0.7353
Precision: 0.8246
Recall: 0.7353
F1 Score: 0.7581
Predicted label: ['Revisions Required']
```

การทำนายด้วย K-Nearest Neighbor

ภาพ 3 ภาษาไพธอนสำหรับการทำนายผลการประเมินบทความ

การทำนายผลการประเมินบทความได้ใช้ภาษาไพธอน โดยเรียกใช้ไลบรารี่ Scikit-learn ทำให้ได้ประสิทธิภาพของโมเดล ดังนี้



ภาพ 4 เปรียบเทียบประสิทธิภาพของโมเดล

จากภาพ 4 สามารถอธิบายได้ว่า Naive Bayes มีประสิทธิภาพต่ำที่สุดในทุกด้าน ไม่ว่าจะเป็น Accuracy Precision Recall และ F1-score ในขณะที่ Decision Tree มีประสิทธิภาพดีกว่า Naive Bayes อย่างเห็นได้ชัด โดยมีค่า Accuracy Precision Recall และ F1-score ที่สูงกว่า สำหรับ Nearest Neighbors โดยรวมมีประสิทธิภาพดีที่สุดในแง่ของ Accuracy แต่มี Precision และ Recall ที่แตกต่างกันไปในแต่ละชุดคุณลักษณะ และสามารถแสดง Confusion Matrix ได้ดังนี้

ตาราง 3 Confusion Matrix ของโมเดล Naïve Bayes

		Predicted			
		Decline Submission	Resubmit for Review	Revisions Required	total
Actual	Decline Submission	8	1	0	9
	Resubmit for Review	2	3	0	5
	Revisions Required	20	0	0	20
	total	30	4	0	34

จากตาราง 3 โมเดล Naïve Bayes มีประสิทธิภาพในการทำนาย Decline Submission ได้ดีพอสมควร แต่มีปัญหาในการทำนาย Resubmit for Review และ Revisions Required โดยเฉพาะอย่างยิ่งกับ Revisions Required ที่โมเดลทำนายผิดพลาดทั้งหมด แสดงให้เห็นว่า โมเดล Naïve Bayes ที่ใช้ในข้อมูลชุดนี้ มีปัญหาในการจำแนกประเภท “Revisions Required” อย่างมาก สามารถวิเคราะห์ประสิทธิภาพของการทำนายได้ ดังนี้

- Decline Submission ทำนายได้แม่นยำ เท่ากับ 8/(8+2+20) = 26.67%

- Resubmit for Review ทำนายได้แม่นยำ เท่ากับ 3/(1+3+0) = 75.00%
- Revisions Required ทำนายได้แม่นยำ เท่ากับ 0/(0+0+0) = 00.00%

ตาราง 4 Confusion Matrix ของโมเดล Decision Tree

		Predicted			
		Decline Submission	Resubmit for Review	Revisions Required	total
Actual	Decline Submission	1	0	2	3
	Resubmit for Review	2	1	2	5
	Revisions Required	1	2	23	26
	total	4	3	27	34

จากตาราง 4 โมเดล Decision Tree มีประสิทธิภาพในการทำนาย Revisions Required ได้ดีพอสมควร โมเดลมีปัญหาในการทำนาย Decline Submission และ Resubmit for Review โดยเฉพาะอย่างยิ่งกับ Resubmit for Review ที่โมเดลทำนายผิดพลาดมากที่สุด แต่ยังคงมีปัญหาในการจำแนกประเภท Decline Submission และ Resubmit for Review เช่นกัน สามารถวิเคราะห์ประสิทธิภาพของการทำนายได้ ดังนี้

- Decline Submission ทำนายได้แม่นยำ เท่ากับ 1/(1+2+1) = 25.00%
- Resubmit for Review ทำนายได้แม่นยำ เท่ากับ 1/(0+1+2) = 33.33%
- Revisions Required ทำนายได้แม่นยำ เท่ากับ 23/(2+2+23) = 85.19%

ตาราง 5 Confusion Matrix ของโมเดล K-Nearest Neighbors

		Predicted			
		Decline Submission	Resubmit for Review	Revisions Required	total
Actual	Decline Submission	2	0	1	3
	Resubmit for Review	4	1	0	5
	Revisions Required	3	1	22	26
	total	9	2	23	34

โมเดล Nearest Neighbors มีประสิทธิภาพในการทำนาย Revisions Required ได้ดีพอสมควร แต่มีปัญหาในการทำนาย Resubmit for Review โดยเฉพาะอย่างยิ่งที่โมเดลทำนายผิดพลาดโดยส่วนมากทำนายว่าเป็น Decline Submission สามารถวิเคราะห์ประสิทธิภาพของการทำนายได้ ดังนี้

- Decline Submission ทำนายได้แม่นยำ เท่ากับ 2/(2+4+3) = 22.22%

- Resubmit for Review ทำนายได้แม่นยำ เท่ากับ 1/(0+1+1) = 50.00%
- Revisions Required ทำนายได้แม่นยำ เท่ากับ 22/(1+0+22) = 95.65%

จากการเปรียบเทียบประสิทธิภาพของ 3 อัลกอริทึม ได้แก่ Naive Bayes Decision Tree และ Nearest Neighbors โดยใช้ชุดข้อมูลจากการคัดเลือก 10 คุณลักษณะ สามารถสรุปประสิทธิภาพได้ดังนี้

ตาราง 6 ภาพรวมของการวิเคราะห์ข้อมูล

โมเดล	Accuracy	Precision	Recall	F1-score
Naïve Bayes	32.35	18.09	32.35	20.66
Decision Tree	73.53	72.25	73.53	72.57
Nearest Neighbors	73.53	82.46	43.53	75.81

จากตาราง 6 สามารถอธิบายได้ว่า Naive Bayes มีประสิทธิภาพต่ำที่สุดในทุกด้าน ไม่ว่าจะเป็น Accuracy Precision Recall และ F1-score ในขณะที่ Decision Tree มีประสิทธิภาพดีกว่า Naive Bayes อย่างเห็นได้ชัด โดยมีค่า Accuracy Precision Recall และ F1-score ที่สูงกว่า สำหรับ Nearest Neighbors โดยรวมมีประสิทธิภาพดีที่สุดในแง่ของ Accuracy แต่มี Precision และ Recall ที่แตกต่างกันไปในแต่ละชุดคุณลักษณะ

สรุปผลการวิจัย

การทำนายผลการประเมินบทความของผู้ทรงคุณวุฒิประจำสาขา จากผลการประเมินคุณภาพบทความจากผู้ทรงคุณวุฒิประจำกองบรรณาธิการประกอบด้วยคุณลักษณะทั้งหมด 19 คุณลักษณะ โดยใช้ค่า Information Gain เป็นตัวคัดเลือกคุณลักษณะ ทำให้ได้ 10 คุณลักษณะที่มีความสำคัญต่อการทำนาย ได้แก่ การใช้ภาษา รูปแบบตารางและภาพประกอบ ไฟล์บทความและเอกสารประกอบ ความชัดเจนและรายละเอียดของบทคัดย่อและ Abstract การกำหนดวัตถุประสงค์การวิจัยในบทคัดย่อและ Abstract การระบุวิธีการวิจัยในบทคัดย่อและ Abstract ผลการวิจัยในบทคัดย่อและ Abstract โจทย์วิจัยชัดเจน วิธีการวิจัย และการอภิปรายผลการวิจัย

สำหรับประสิทธิภาพของผลการทำนาย พบว่า Naive Bayes มีประสิทธิภาพต่ำที่สุดในทุกด้าน สำหรับ Nearest Neighbors โดยรวมมีประสิทธิภาพดีที่สุดในแง่ของ Accuracy แต่มี Precision และ Recall ที่แตกต่างกันไปในแต่ละชุดคุณลักษณะ ในขณะที่ Decision Tree ถึงแม้จะมีค่า Precision และ F1-score ที่ต่ำกว่า แต่ประสิทธิภาพของการทำนายมีความคงเส้นคงวาหรือมีความสม่ำเสมอมากกว่า

อภิปรายผล

การค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล เป็นกระบวนการที่ช่วยให้มีการนำข้อมูลมาวิเคราะห์ ตีความ และทำนายกลุ่มของข้อมูลชุดใหม่ โดยมีการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ชุดข้อมูลสำหรับการเรียนรู้ (Training Data) และชุดข้อมูลทดสอบ (Test Data) การคัดเลือกคุณลักษณะโดยใช้ค่า Information Gain ทำให้สามารถเลือกคุณลักษณะที่มีความสำคัญต่อการทำนาย ส่งผลให้การทำนายมีความถูกต้องมากขึ้น สอดคล้องกับ กันต์ ศิระพรธนารัชต์ และ ชุติพนธ์ ศรีสวัสดิ์ (2568) ที่พบว่าการเลือกคุณลักษณะสามารถเพิ่มประสิทธิภาพของโมเดลให้ดีขึ้นได้ ในบางกรณีอาจมีการคัดเลือกคุณลักษณะที่แตกต่างออกไป ดังการศึกษาของ นพมาศ ปักเข็ม และคณะ (2566) ที่ใช้ข้อมูลที่มีปรากฏบ่อยที่สุดเป็นตัวแทน 5 ลำดับแรก ทั้งนี้เนื่องมาจากชุดข้อมูลที่เป็นข้อความ ดังนั้นการคัดเลือกคุณลักษณะต้องใช้วิธีการที่มีความเหมาะสมกับประเภทของข้อมูลด้วย

สำหรับประสิทธิภาพของโมเดลพบว่า Decision Tree แม้จะมีค่า Precision และ F1-score จะต่ำกว่า แต่ความสม่ำเสมอของประสิทธิมีความใกล้เคียงกัน กล่าวคือ มีค่า Accuracy = 73.53 ค่า Precision = 72.25 ค่า Recall = 73.53 และ F1-score = 72.57 ในขณะที่ Naïve Bayes และ Nearest Neighbors มีค่าที่แตกต่างกันอย่างมาก แสดงถึงความไม่สม่ำเสมอของประสิทธิภาพการทำนาย เช่นเดียวกับการศึกษาของ กันต์ ศิระพรธนารัชต์ และ ชุติพนธ์ ศรีสวัสดิ์ (2568) ที่พบว่า Decision Tree มีความถูกต้องสูงที่สุด เนื่องจากให้ค่า Precision และ F1-score สูงที่สุด นอกจากนี้การพยากรณ์ด้วย Naïve Bayes มีประสิทธิภาพต่ำที่สุด สอดคล้องกับการศึกษาของ จิรโรจน์ ตอสะสุกุล (2564) ที่พบว่าประสิทธิภาพของ Decision Tree เป็นโมเดลสำหรับการพยากรณ์ที่เหมาะสมที่สุด ในขณะที่โมเดล Naïve Bayes มีประสิทธิภาพต่ำที่สุด นอกจากนี้ การศึกษาของ เจษฎาพร ปาคำวัง และคณะ (2563) ยังพบว่าโมเดล Decision Tree มีความเหมาะสมกับการพยากรณ์ แต่การเลือกใช้โมเดลต้องศึกษาข้อดีและข้อจำกัดของแต่ละโมเดล เพื่อให้เหมาะสมกับประเภทของข้อมูลและวัตถุประสงค์ของการหาคำตอบ ซึ่งแต่ละโมเดลอาจให้ประสิทธิภาพที่แตกต่างกันไปในแต่ละชุดข้อมูลที่มีความแตกต่างกันด้วย ดังการศึกษาของ นพมาศ ปักเข็ม และคณะ (2560) ที่พบว่าโมเดล Naïve Bayes มีประสิทธิภาพสูงที่สุด รองลงมาเป็นโมเดล Decision Tree และ Nearest Neighbor ตามลำดับ อย่างไรก็ตาม จากผลการวิจัยนี้ยังมีข้อจำกัดในด้านลักษณะของข้อมูล ซึ่งอาจมีลักษณะของข้อมูลที่ไม่ครอบคลุมขอบเขตของวารสารหลากหลายประเภท หรือการประเมินบทความอาจมีความลำเอียง (Bias) ได้

ข้อเสนอแนะ

ข้อเสนอแนะสำหรับการนำผลการวิจัยไปใช้ประโยชน์

1. ผู้เขียนบทความควรตรวจสอบความครบถ้วนโดยเฉพาะองค์ประกอบที่มีความสำคัญ ได้แก่ การใช้ภาษา รูปแบบตารางและภาพประกอบ ไฟล์บทความและเอกสารประกอบ ความชัดเจนและรายละเอียดของบทคัดย่อและ Abstract การกำหนดวัตถุประสงค์การวิจัยในบทคัดย่อและ Abstract การระบุวิธีการวิจัยใน

บทคัดย่อและ Abstract ผลการวิจัยในบทคัดย่อและ Abstract โจทย์วิจัยชัดเจน วิธีการวิจัย และการอภิปรายผลการวิจัย เพื่อลดความเสี่ยงของการถูกปฏิเสธการตีพิมพ์บทความ

2. พัฒนาการใช้ AI และ Machine Learning ในการวิเคราะห์ข้อมูลและตรวจสอบความสมบูรณ์ของบทความ โดยเฉพาะการตรวจสอบความครบถ้วนขององค์ประกอบวิจัยตามเกณฑ์ที่กำหนด

ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. ควรเพิ่มชุดข้อมูลให้มีจำนวนเพิ่มมากขึ้น เพื่อให้มีชุดข้อมูลสำหรับการเรียนรู้ที่มากขึ้นซึ่งอาจส่งผลต่อประสิทธิภาพของโมเดลที่สูงขึ้น

2. ทดลองใช้กับข้อมูลจากวารสารอื่น หรือเพิ่มจำนวนคุณลักษณะ (Feature) เช่น การวิเคราะห์ข้อความในบทความโดยตรง (Text Mining)

3. ควรใช้โมเดลประเภทอื่นเพื่อให้ผลการพยากรณ์ที่มีความครอบคลุมมากขึ้น เช่น Random Forest และ Support Vector Machine

เอกสารอ้างอิง

กันต์ ศิระพรธนารัชต์ และชุติพนธ์ ศรีสวัสดิ์. (2568). การใช้เทคนิคเหมืองข้อมูล เพื่อพยากรณ์การสำเร็จการศึกษา. *วารสารเทคโนโลยีสารสนเทศ มจพ.*, *21*(1), 33-43.

จิรโรจน์ ตอสะสุกุล. (2564). แบบจำลองการพยากรณ์การระบาดของโรคไข้เลือดออกโดยใช้เทคนิคการทำเหมืองข้อมูล. *วารสารวิชาการซายน์เทค มรภ.ภูเก็ต*, *5*(2), 51-60.

เจษฎาพร ปาคำวัง, กาญจน์ คุ้มทรัพย์ และวรชัย ศรีเมือง. (2563). การเปรียบเทียบแบบจำลองจำแนกประเภทเพื่อการพัฒนาการท่องเที่ยว อำเภอเขาค้อ จังหวัดเพชรบูรณ์ด้วยเทคนิคเหมืองข้อมูล. *Life Sciences and Environment Journal*, *Pibulsongkram Rajabhat University*, *21*(1), 213-223.

ชนะวงศ์ คงสอน และณัฐวุฒิ วิสุทธิพิเนตร. (2559). *การพัฒนาระบบสนับสนุนการตัดสินใจในการ เลือกสาขาการเรียนต่อระดับประกาศนียบัตรวิชาชีพ สังกัดอาชีวศึกษาจังหวัด พระนครศรีอยุธยา*. [วิทยานิพนธ์ปริญญามหาบัณฑิต]. มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ.

นพรัตน์ นนท์ศิริ, ราตรี มนัสศิลา และกริช สมกันธา. (2566). การจำแนกข้อมูลเพื่อวินิจฉัยความเสี่ยงการเป็นโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูล. *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, 33(2), 538-547.

นพมาศ ปักเข็ม, ชนิดา จันมณีย์ และศิวกร อุยสุย. (2566). การจำแนกประเภทภูมิปัญญาท้องถิ่นไทยแบบอัตโนมัติโดยวิธีการทางเหมืองข้อมูล. *วารสารมหาวิทยาลัยทักษิณ*, *20*(ฉบับพิเศษ), 300-307.

สุธีรา วงศ์อนันทรัพย์. (2559). การใช้เทคนิคเหมืองข้อมูลในการประเมินความรู้ และหาความถนัด เพื่อพัฒนาศักยภาพของนักศึกษา. *วารสารสังคมศาสตร์*, 5(1), 12-19.

Björk, B. C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetric*, *7*(4), 914-923. https://doi.org/10.1016/j.joi.2013.09.001

Calcagno, V., Demoinet, E., Gollner, K., Guidi, L., Ruths, D., & De Mazancourt, C. (2012). Flows of research manuscripts among scientific journals reveal hidden submission patterns. *Science*, *338*(6110), 1065-1069. https://doi.org/10.1126/science.1227833

Chen, X., Xie, H., & Wang, F. L. (2022). A bibliometric analysis of data mining research in higher education from 2000 to 2019. *Studies in Higher Education*, *47*(2), 223-239. https://doi.org/10.1080/03075079.2020.1750584

Hargens, L. L. (2019). Rejection rates and journal impact factors. *Journal of the American Society for Information Science and Technology*, *70*(11), 1145-1151. https://doi.org/10.1002/asi.24200

Santos, J. A. C., Santos, M. C., & Pereira, J. R. (2020). Predicting article impact with text features: Evidence from top management and economics journals. *Scient metrics*, *124*(2), 1007-1029. https://doi.org/10.1007/s11192-020-03489-3

Wang, W., Kong, X., Zhang, J., Chen, Z., Xia, F., & Wang, X. (2016). Editorial behaviors in peer review. *Scient metrics*, *109*(1), 357-368. https://doi.org/10.1007/s11192-016-2020-4