

การเพิ่มประสิทธิภาพการจัดสรรงบประมาณด้านครุภัณฑ์คอมพิวเตอร์ ด้วยการประยุกต์ใช้เทคนิคเหมืองข้อความ (Text Mining)

ENHANCING BUDGET ALLOCATION FOR COMPUTER EQUIPMENT USING TEXT MINING

พรพรรณชล มาทัก¹
ดร. ธนภัทร นังคะจิตร²

บทคัดย่อ

ระบบงานหลักของสำนักงานเทคโนโลยีสารสนเทศ สำนักงานศาลยุติธรรม คือ การวิเคราะห์ กลั่นกรอง การพิจารณา และตรวจสอบค่างบประมาณที่ร้องขอสำหรับการจัดซื้อครุภัณฑ์คอมพิวเตอร์ให้แก่ หน่วยงานในสังกัดสำนักงานศาลยุติธรรมทั่วประเทศ ซึ่งในทุกปีมีเอกสารคำขอของงบประมาณกิจกรรมจัดซื้อ ครุภัณฑ์คอมพิวเตอร์และเอกสารประกอบเพิ่มเติมเป็นจำนวนมาก ด้วยเหตุนี้จึงส่งผลให้กระบวนการพิจารณา และตรวจสอบมีความล่าช้าเนื่องจากความยุ่งยากในการจัดเก็บ การเรียกดู และการใช้ข้อมูล โดยเฉพาะอย่างยิ่ง เจ้าหน้าที่ของสำนักงานเทคโนโลยีสารสนเทศจะต้องคัดแยกและจัดประเภทเอกสารคำขอของงบประมาณที่หน่วยงาน ในสังกัดร้องขอ เพื่อนำส่งไปยังแผนกงานที่เกี่ยวข้องดำเนินการตรวจสอบและพิจารณา

ดังนั้น งานวิจัยนี้จึงมุ่งหวังที่จะนำเสนอเทคนิคการทำเหมืองข้อความ และการประมวลผล ภาษาธรรมชาติเพื่อจัดหมวดหมู่แผนกที่เกี่ยวข้องซึ่งรับผิดชอบต่อการพิจารณาอนุมัติงบประมาณตามประเภท ของครุภัณฑ์คอมพิวเตอร์ และช่วยการตัดสินใจเรื่องการอนุมัติงบประมาณตามรายละเอียดของเหตุผลและความ จำเป็นที่ปรากฏในคำขอของงบประมาณ โดยเริ่มจากขั้นตอนการจัดเก็บข้อมูลลงในระบบฐานข้อมูลและนำข้อมูล รายการครุภัณฑ์คอมพิวเตอร์และเหตุผลความจำเป็นซึ่งปรากฏในคำขอของงบประมาณที่จัดเก็บไว้เข้าสู่ กระบวนการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ด้วยวิธีการตัดคำ การแปลงคำให้เป็น รูปแบบเชิงปริมาณ และการสกัดคุณลักษณะ เพื่อนำไปใช้สร้างแบบจำลองสำหรับจัดหมวดหมู่ประเภทข้อความ (Text Classification) โดยใช้อัลกอริทึมได้แก่ ต้นไม้ตัดสินใจ (Decision Tree), ชุดต้นไม้ตัดสินใจ (Random Forest), นาอิวเบย์ (Naïve Bayes) และลอจิสติกส์รีเกรสชัน (Logistic Regression) และวัดประสิทธิภาพของ

¹ นักศึกษาหลักสูตรวิศวกรรมมหาบัณฑิต สาขาวิศวกรรมข้อมูลขนาดใหญ่

² ที่ปรึกษาสารนิพนธ์หลัก

แบบจำลองที่ได้ด้วยวิธีการ 5-Fold Cross Validation ซึ่งประสิทธิภาพของแบบจำลองการจัดหมวดหมู่แผนกที่เกี่ยวข้อง และแบบจำลองการตัดสินใจการอนุมัติงบประมาณที่ได้จากงานวิจัยนี้มีค่าความถูกต้องแม่นยำที่น่าพอใจ ดังนั้นจึงสรุปได้ว่างานวิจัยนี้สามารถเพิ่มความเร็วในกระบวนการตรวจสอบและพิจารณางบประมาณได้อย่างมีประสิทธิภาพ ด้วยการลดกระบวนการที่ไม่จำเป็นของเจ้าหน้าที่ในสำนักเทคโนโลยีสารสนเทศ และยังให้ข้อมูลสำคัญในการตัดสินใจได้อีกด้วย

คำสำคัญ: คำของงบประมาณ, เหมืองข้อความ, การประมวลผลทางธรรมชาติ, การจัดหมวดหมู่ประเภทข้อความ

Abstract

The essential task of the Office of Information Technology (OIT) is to consider, analyze and screen the budget that requests for purchasing computer hardware for agencies under the Office of the Judiciary throughout the country. Each year, a large number of budget allocation request documents have been submitted with supplementary documents. Consequently, the inspection process might delay since the difficulties in storing, retrieving, and utilizing the data. More specifically, the OIT staffs manually classify the requested documents and deliver them to the relevant OIT divisions. Thus, this work aims to apply text mining and natural language processing techniques to categorize the relevant divisions responsible for budget requests according to the type of computer equipment. Based on the details of the reasons and necessity appearing in the budget request, the approval decisions also have been made. Firstly, all relevant data are stored in the database management system. Then, the computer equipment names and the requested reasons are tokenized into words that are used as input features to the classification models. In this work, several machine learning techniques such as Decision Tree, Naïve Bayes, and Logistic Regression are utilized. The experimental results based on 5-fold cross-validation show that our models offer satisfied accuracy to classify the relevant divisions and the approval decision for the requests. Therefore, we can conclude that our work is capable of speeding up the inspection process by reducing redundant processes of the OIT staff and also providing useful decision-making information.

Keywords: Budget request, Text Mining, Natural Language Processing, Text Classification

บทนำ

สำนักเทคโนโลยีสารสนเทศ สำนักงานศาลยุติธรรม มีภารกิจงานเกี่ยวกับการพิจารณา วิเคราะห์ ถัดกรอง คำขอขบประมาณจัดซื้อครุภัณฑ์คอมพิวเตอร์ให้แก่หน่วยงานในสังกัดสำนักงานศาลยุติธรรมทั่วประเทศ โดยในแต่ละปีงบประมาณเมื่อหน่วยงานในสังกัดสำนักงานศาลยุติธรรมมีความจำเป็นต้องขอรับการ จัดสรรงบประมาณเพื่อจัดซื้อครุภัณฑ์คอมพิวเตอร์สำหรับใช้สนับสนุนการปฏิบัติงานจะดำเนินการจัดส่งคำขอ ขบประมาณกิจกรรมจัดซื้อครุภัณฑ์ประจำปี (แผนความต้องการครุภัณฑ์คอมพิวเตอร์) มายังสำนักแผนงานและ งบประมาณในรูปแบบของแบบฟอร์มเอกสารและไฟล์เอกสาร (File Type : .xls, .xlsx) ซึ่งสำนักแผนงานและ งบประมาณจะดำเนินการรวบรวมและจัดส่งข้อมูลคำขอขบประมาณดังกล่าวให้แก่สำนักเทคโนโลยีสารสนเทศ ใช้พิจารณา วิเคราะห์ ถัดกรอง โดยในแต่ละปีมีเอกสารคำขอขบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์ และเอกสารประกอบเพิ่มเติมเป็นจำนวนมาก ซึ่งก่อให้เกิดปัญหาข้อยุ่งยากในด้านการบริหารจัดการเอกสาร อาทิ เช่น การจำแนก การจัดเก็บ การสืบค้น การนำข้อมูลไปใช้ประโยชน์ เป็นต้น และเกิดความล่าช้าในขั้นตอนการ ตรวจสอบ การพิจารณาจัดสรรงบประมาณ และการแจ้งรายงานผลการพิจารณากลับไปยังสำนักแผนงานและ งบประมาณ และหน่วยงานศาลยุติธรรมที่ขอรับการ จัดสรรงบประมาณ เนื่องจากต้องให้เจ้าหน้าที่คัดแยก ประเภทคำขอขบประมาณเพื่อจัดส่งให้ในแต่ละส่วนงานที่เกี่ยวข้องดำเนินการตรวจสอบและพิจารณา ซึ่งปัญหา และข้อยุ่งยากที่เกิดขึ้นดังกล่าวมีผลกระทบต่อประสิทธิภาพการบริหารงบประมาณของหน่วยงานอีกประการ หนึ่งด้วย ทั้งนี้ ด้วยความก้าวหน้าของเทคโนโลยีการวิเคราะห์ข้อมูล (Data Analytic) โดยใช้เทคนิคเหมือง ข้อความ (Text Mining) ซึ่งเป็นกระบวนการที่สกัดเอาความรู้จากภาษาธรรมชาติที่มีลักษณะของข้อมูลแบบไม่มี โครงสร้าง (Unstructured Data) และการทำเหมืองข้อมูลจำนวนมากจะเป็นลักษณะของการนำเสนอรูปแบบ สำหรับการทำความเข้าใจข้อความที่ใช้ประโยชน์ในเอกสารและใช้ความรู้ที่ได้จากการสกัดข้อความเหล่านั้นมาเพื่อ ปรับปรุงรูปแบบและปัญหาต่างๆ ของสิ่งนั้น

จากสภาพปัญหาข้อขัดข้อง ประกอบกับความก้าวหน้าของเทคโนโลยีการวิเคราะห์ข้อมูลข้างต้น ผู้วิจัยเห็นว่า สามารถนำมาประยุกต์ใช้เพื่อแก้ปัญหาและเพิ่มความสะดวกรวดเร็วในการดำเนินงานของ หน่วยงานให้มีประสิทธิภาพมากขึ้น จึงดำเนินงานวิจัยภายใต้หัวข้อ “การเพิ่มประสิทธิภาพการจัดสรร งบประมาณด้านครุภัณฑ์คอมพิวเตอร์ด้วยการใช้เทคนิคเหมืองข้อความ (Text Mining)” โดยนำเทคนิคเหมือง ข้อความมาใช้เป็นเครื่องมือในการสกัดข้อความในคำขอขบประมาณจัดซื้อครุภัณฑ์คอมพิวเตอร์เพื่อจำแนก ประเภทคำขอขบประมาณในด้านต่างๆ เช่น ด้านระบบเครือข่าย ด้านระบบงาน และด้านเครื่องคอมพิวเตอร์และ อุปกรณ์ต่อพ่วง รวมไปถึงการสกัดหากกลุ่มคำหรือข้อความที่เป็นเหตุปัจจัยให้หน่วยงานมีความจำเป็นต้องขอรับ การจัดสรรงบประมาณโดยที่กำหนดแนวทางและวิธีการใช้เทคนิคเหมืองข้อมูลให้เกิดประโยชน์ต่อการลดภาระ

และขั้นตอนปฏิบัติงานของบุคลากรที่เกี่ยวข้อง รวมถึงพัฒนาระบบข้อมูลที่จะช่วยสนับสนุนการตัดสินใจให้แก่เจ้าหน้าที่ และผู้บริหารในกระบวนการพิจารณาจัดสรรงบประมาณให้มีประสิทธิภาพมากขึ้น ทั้งนี้ ผู้วิจัยคาดหวังเป็นอย่างยิ่งว่าผลการดำเนินงานและข้อสรุปที่ได้จากงานวิจัยจะเป็นประโยชน์ต่อการพัฒนาระบบบริหารงบประมาณของสำนักงานศาลยุติธรรม ตลอดจนสามารถนำไปเป็นต้นแบบของการนำข้อมูลที่มีอยู่ของหน่วยงานไปใช้ประโยชน์สำหรับพัฒนางานในด้านต่างๆ เพื่อสนับสนุนและขับเคลื่อนองค์กรให้บรรลุตามเป้าหมาย พันธกิจ วิสัยทัศน์ภายใต้แผนยุทธศาสตร์ศาลยุติธรรมและแผนพัฒนาดิจิทัลศาลยุติธรรมที่ต้องการปรับเปลี่ยนองค์กรสู่ดิจิทัล

วัตถุประสงค์ของงานวิจัย

1. เพื่อวิเคราะห์เอกสารที่เกี่ยวข้องกับคำของบประมาณจัดซื้อครุภัณฑ์คอมพิวเตอร์ และจำแนกประเภทคำของบประมาณตามประเภทครุภัณฑ์และส่วนงานที่เกี่ยวข้อง
2. เพื่อพัฒนาระบบที่จะช่วยสนับสนุนการตัดสินใจการจัดสรรงบประมาณจากการจำแนกข้อมูลเหตุผลความจำเป็นของหน่วยงานที่ขอรับการจัดสรรให้มีความสะดวกรวดเร็ว และเพิ่มประสิทธิภาพการบริหารงบประมาณ
3. เพื่อลดความซ้ำซ้อนในการจัดเก็บข้อมูล เพิ่มมูลค่าและใช้ประโยชน์จากข้อมูลที่มีการจัดเก็บไว้แล้วให้เกิดประโยชน์สูงสุดแก่การพัฒนาระบบบริหารงบประมาณ

ขอบเขตงานวิจัย

1. พัฒนาระบบวิเคราะห์เอกสารคำของบประมาณจัดซื้อครุภัณฑ์คอมพิวเตอร์ของหน่วยงานศาลยุติธรรมทั่วประเทศเพื่อจำแนกประเภทคำของบประมาณตามประเภทครุภัณฑ์และส่วนงานที่เกี่ยวข้อง
2. พัฒนาระบบข้อมูลสนับสนุนการตัดสินใจการจัดสรรงบประมาณของสำนักงานศาลยุติธรรมโดยการวิเคราะห์ข้อมูลเหตุผลความจำเป็นของหน่วยงานที่ขอรับการจัดสรร

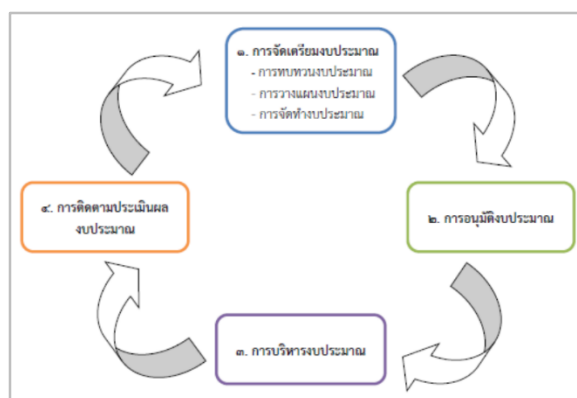
ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

1. ทฤษฎีที่เกี่ยวข้อง

1.1 ระบบบริหารงบประมาณศาลยุติธรรมและสำนักงานศาลยุติธรรม [1]

กระบวนการงบประมาณประกอบด้วย 4 ขั้นตอน คือ การจัดเตรียมงบประมาณ การอนุมัติงบประมาณ การบริหารงบประมาณ และการติดตามประเมินผลงบประมาณ ที่กล่าวมาข้างต้น สำนักงานศาลยุติธรรมได้

ดำเนินการตามกระบวนการดังกล่าวมาตั้งแต่ปี พ.ศ. 2543 จนถึงปัจจุบัน โดยที่ขั้นตอนการบริหารงบประมาณ ถือเป็นขั้นตอนที่สำคัญที่หน่วยงานในสังกัดสำนักงานศาลยุติธรรมต้องนำงบประมาณที่ได้รับอนุมัติจาก คณะกรรมการ บริหารศาลยุติธรรมไปใช้จ่ายให้ถูกต้อง เหมาะสม เกิดประโยชน์สูงสุดต่อราชการ และมีให้เกิด ความผิดพลาดในการทำงาน สำนักงานศาลยุติธรรมจึงได้ปรับปรุงเอกสารระบบบริหารงบประมาณศาลยุติธรรม และสำนักงานศาลยุติธรรมฉบับนี้ ให้สอดคล้องกับแผนยุทธศาสตร์ศาลยุติธรรม พ.ศ. 2565 -2568 และเป็นไป ตามระเบียบต่างๆ ที่มีการปรับปรุงและแก้ไขเพิ่มเติม เพื่อให้หน่วยงานในสังกัดสำนักงานศาลยุติธรรมใช้เป็น แนวทางในการบริหารงบประมาณ ได้อย่าง มีประสิทธิภาพและเป็นไปในทิศทางเดียวกัน



ภาพที่ 1 กระบวนการงบประมาณ

1.2 CRISP-DM) เป็นกระบวนการในการทำ Data Mining โดยในกระบวนการคริป-ดี เอ็ม จะ ประกอบด้วย 6 ขั้นตอน คือ 1) การทำความเข้าใจกับธุรกิจ (Business Understanding) 2) การทำความเข้าใจข้อมูล (Data Understanding) 3) การเตรียมข้อมูล (Data Preparation) 4) การสร้างแบบจำลอง (Modeling) 5) การ ประเมินผล (Evaluation) 6) การนำไปใช้ (Deployment)

1.3 เหมืองข้อความ (Text Mining) เป็นกระบวนการวิเคราะห์ข้อมูลที่เป็นข้อความ เพื่อแยกประเภท จำแนกรูปแบบและความสัมพันธ์ใหม่ ๆ ที่เกิดขึ้นระหว่างข้อมูล และช่วยค้นพบเนื้อหาสาระที่ซ่อนในเอกสาร ต่าง ๆ เพื่อนำมาใช้ประโยชน์ และใช้คำสำคัญมาทำการสกัดคุณลักษณะเพื่อดึงคุณลักษณะของคำมาเป็นตัวแทน ของเอกสารภาษาไทย โดยมีการตัดคำหยุด ซึ่งได้แก่ คำที่ไม่มีความหมายสำคัญต่อเอกสารและไม่ทำให้ใจความ ของเอกสารเปลี่ยน ตัวอย่างเช่น คำสรรพนามเป็นคำที่ใช้แทนคำนาม คำสันธานเป็นคำที่ทำหน้าที่เชื่อมคำกับคำ คำบุพบทเป็นคำที่ใช้เชื่อมคำหรือกลุ่มคำให้สัมพันธ์กันหลังจากนั้นเมื่อได้คุณลักษณะของคำแล้วจึงนำมาสร้าง

แบบจำลองในรูปแบบต่าง ๆ ได้ เช่น การจำแนกหมวดหมู่, การจัดกลุ่ม และการสร้างกฎความสัมพันธ์ (Association Rule Mining) เป็นต้น

1.4 การประมวลผลภาษาธรรมชาติ (NLP) เป็นศาสตร์ที่สำคัญทางด้าน Machine Learning โดยมันเป็นสาขาวิชาหนึ่งที่ประกอบด้วยองค์ความรู้จากหลากหลายแขนง อาทิ ภาษาศาสตร์ (Linguistics) วิทยาการคอมพิวเตอร์ (Computer Science) ปัญญาประดิษฐ์ (Artificial Intelligence: AI) รวมถึงสถิติ (Statistics) โดยมีจุดมุ่งหมายเพื่อให้คอมพิวเตอร์สามารถทำความเข้าใจ ข้อมูลที่มีลักษณะเป็นข้อความหรือคำพูด เช่นเดียวกับที่มนุษย์ทำได้ ซึ่งไม่ใช่เพียงแค่เข้าใจความหมายโดยตรงของข้อความนั้น ๆ แต่ยังรวมถึงการรับรู้ถึงความหมายโดยนัยความรู้สึของผู้เขียน

1.5 Classification Technique กระบวนการสร้างโมเดลจำแนกประเภทข้อมูลเพื่อทำนายกลุ่มของข้อมูลใหม่ตัวอย่างของกลุ่ม เช่น กลุ่มของลูกค้าที่ซื้อคอมพิวเตอร์-ไม่ซื้อคอมพิวเตอร์ กลุ่มของลูกค้าที่ฐานะดี-ปานกลาง-แย่ กลุ่มของการผลิตสินค้า ผ่านเกณฑ์-ไม่ผ่านเกณฑ์ ในที่นี้คำว่ากลุ่มจะเรียกว่าคลาส (Class) ของข้อมูล ซึ่งในคลาสเดียวกันนั้นจะต้องมีข้อมูลที่มีความเหมือนหรือคล้ายคลึงกันมากกว่าข้อมูลที่อยู่ในที่แตกต่างกัน การสร้างโมเดลจำแนกประเภทข้อมูลจะเกิดขึ้นมาจากการหาความสัมพันธ์ของข้อมูลในฐานข้อมูลขนาดใหญ่ โดยข้อมูลทั้งหมดจะมีการแบ่งออกเป็น 2 กลุ่ม คือกลุ่มข้อมูลเรียนรู้ เป็นชุดข้อมูลที่มีบทบาทในการสร้างโมเดลจำแนกประเภทข้อมูลขึ้นมา และมีกลุ่มข้อมูลทดสอบเป็นชุดข้อมูลประเมินความถูกต้องของโมเดลจำแนกประเภทข้อมูลโมเดลจำแนกประเภทข้อมูลได้ถูกนำมาประยุกต์ใช้งานหลาย ๆ ด้าน

1.6 การทดสอบประสิทธิภาพ และวัดประสิทธิภาพอัลกอริทึม มี 3 วิธีการใหญ่ ดังนี้

- 1) Self Consistency Test เป็นการใช้อ้างอิงข้อมูล Training ในการทดสอบประสิทธิภาพของโมเดล
- 2) Split Test แบ่งข้อมูลออกเป็น 2 ชุด Training data สำหรับสร้างโมเดล และ Testing data สำหรับทดสอบ
- 3) Cross-Validation แบ่งข้อมูลออกเป็น N ชุด เช่น $N = 5$ หรือ 10 ซึ่งข้อมูล $N-1$ ชุด สำหรับสร้างโมเดล และข้อมูลส่วนที่เหลือสำหรับทดสอบวนทำจนครบ N

งานวิจัยที่เกี่ยวข้อง

เฉลิมชัย พิเศษ, (2563) [2] ประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing) วิทยาการข้อมูล หลักการเรียนรู้ของเครื่อง เพื่อช่วยในการแก้ปัญหาการแยกประเภทตัวงาน โดยการวิเคราะห์คำสำคัญ และใช้เครื่องมือการตัดคำภาษาทพในขั้นตอนการประมวลผลภาษาธรรมชาติที่สามารถให้ผลทดลองมีความถูกต้องสูงถึง 91 เปอร์เซนต์

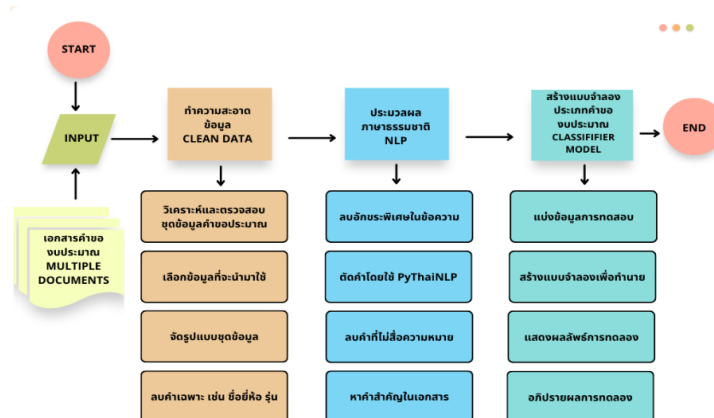
วาทัญญู ชูประจิดต์, (2562) [3] เสนอวิธีการจำแนกประเภท และวิธีการคัดเลือกคุณลักษณะสำหรับการสนับสนุนการวิเคราะห์ความรู้สึกต่อสถานที่ท่องเที่ยวไทยจากข้อมูลการแสดงความคิดเห็นต่อสถานที่ท่องเที่ยวไทยที่เก็บรวบรวมจากสื่อสังคมออนไลน์ภายในประเทศไทย การจำแนกประเภทใช้กระบวนการคัดเลือกคุณลักษณะแบบหาความถี่ค่าและสร้างตัวจำแนกประเภทด้วย 4 อัลกอริทึม คือ วิธีการหาเพื่อนบ้านใกล้สุด เค ตัว (K-Nearest Neighbors:KNN) , ต้นไม้ตัดสินใจ (Decision Tree) , นาอ์ฟเบย์ (Naïve Bayes) และ ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

Hindawi Computational Intelligence and Neuroscience, (2022) [4] บทความนี้นำเสนอระบบการจำแนกข้อความโดยใช้กระบวนการ NLP และเปรียบเทียบผลการสร้างแบบจำลองด้วยเทคนิคการจำแนกข้อความ (Text Classification) โดยใช้วิธีการกำหนดคุณลักษณะฟีเจอร์ด้วยเทคนิคต่างๆ ได้แก่ Standard Feature Selection Methods (ประกอบด้วย Filter Methods, Wrapper Methods, Embedded Methods) และ Text Feature Extraction Methods

Witchapong Darontham, (2018) [5] บทความนี้เสนอขั้นตอนการเตรียมข้อมูลประเภทข้อความภาษาไทย (Text) โดยใช้ภาษา Python ซึ่งเริ่มตั้งแต่ขั้นตอนการทำความสะอาดข้อมูล (clean text) การตัดคำ (tokenize) สร้างคลังคำศัพท์จากข้อความทั้งหมด (create bag of words) และสร้างฟีเจอร์ด้วยการนำคำ (count words) และสร้างฟีเจอร์ด้วย TFIDF

ระเบียบวิธีวิจัย

ผู้วิจัยได้กำหนดขอบเขตงานวิจัยโดยนำเสนอแบบจำลองสำหรับช่วยตัดสินใจพิจารณาคำของบประมาณของหน่วยงาน ประกอบด้วย แบบจำลองที่ 1 จำแนกส่วนงานที่เกี่ยวข้องกับคำของบประมาณ และแบบจำลองที่ 2 พิจารณา “อนุมัติ” หรือ “ไม่อนุมัติ” คำของบประมาณจากเหตุผลความจำเป็น ซึ่งมีรายละเอียดขั้นตอนการดำเนินงานวิจัยปรากฏตามภาพที่ 2



ภาพที่ 2 ภาพรวมวิธีดำเนินการวิจัย

3.1 ศึกษาวิเคราะห์ระบบการของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์ของหน่วยงานในสังกัดสำนักงานศาลยุติธรรม และวิเคราะห์ปัญหาของหน่วยงาน พบว่าเอกสารคำของบประมาณของหน่วยงานจัดเก็บในรูปแบบของเอกสารเป็นจำนวนมาก ทำให้เกิดข้อยุ่งยากในการบริหารจัดการจัดเก็บเอกสารและการตรวจสอบข้อมูล อาทิเช่น 1) การจำแนกประเภทเอกสารคำขอประมาณการจัดซื้อครุภัณฑ์คอมพิวเตอร์เพื่อดำเนินการวิเคราะห์และพิจารณาผลการจัดสรรงบประมาณให้แก่หน่วยงานที่เกี่ยวข้องมีความล่าช้า เนื่องจากต้องดำเนินการคัดแยกโดยบุคลากร 2) การตรวจสอบและการประมวลผลงบประมาณการจัดซื้อครุภัณฑ์คอมพิวเตอร์ของหน่วยงานจำแนกตามปีงบประมาณและรายการครุภัณฑ์คอมพิวเตอร์ขาดประสิทธิภาพ ไม่สามารถดำเนินการด้วยระบบโปรแกรม

3.2 การรวบรวมเอกสารข้อมูล จากการวิเคราะห์และตรวจสอบข้อมูลคำขอประมาณตามหัวข้อ 3.1 ผู้วิจัยได้จัดเก็บข้อมูลสำหรับนำไปดำเนินการตามขั้นตอนการวิจัย โดยเปลี่ยนรูปแบบคำขอประมาณจากเดิมที่จัดเก็บในรูปแบบเอกสารเปลี่ยนเป็นข้อมูลที่อยู่ในรูปแบบข้อมูลอิเล็กทรอนิกส์ (ไฟล์นามสกุล xlsx, csv) ซึ่งเอกสารคำขอประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์ที่ผู้วิจัยดำเนินการรวบรวมอยู่ในระหว่างปีงบประมาณ พ.ศ. 2563 – 2565 จำนวนทั้งสิ้น 717 ชุด และมีจำนวนคอลัมน์ 8 คอลัมน์ รายละเอียดดังภาพที่ 3 และตารางที่ 1 ที่ปรากฏด้านล่างนี้

req_id	req_year	region_name	office_name	req_item	req_detail	group	approve
322	2563	ศาลสูง	ศาลอุทธรณ์ภาค 8	เครื่องพิมพ์ Multifunction ชนิดเลเซอร์	ขอใหม่ จำนวน 1 เครื่อง เพื่อใช้ประกอบงานธุรการ	support	yes
331	2563	ภาค 4	ศาลเยาวชนและอาญา	เครื่องเล่นแผ่นเสียง	ขอทดแทน จำนวน 1 เครื่อง เนื่องจากเครื่องเก่า	support	yes
335	2563	ศาลสูง	ศาลอุทธรณ์คดีชั้นอุทธรณ์	เครื่องคอมพิวเตอร์ Server ชนิด Tower	ขอใหม่ จำนวน 1 เครื่อง ศาลอุทธรณ์คดีชั้นอุทธรณ์	network	yes
337	2563	สำนักงาน	ศูนย์ปฏิบัติการศาลยุติธรรม	กระดานอิเล็กทรอนิกส์ (e-Board)	ขอใหม่ จำนวน 1 ชุด เพื่อใช้กับข้อมูลของพื้นที่	app	yes
339	2563	ภาค 7	ศาลจังหวัดเพชรบุรี	อุปกรณ์ประชุมทางไกลผ่านจอภาพ	ขอทดแทน จำนวน 1 ชุด เนื่องจากอุปกรณ์เก่า	network	yes
341	2563	ภาค 2	ศาลเยาวชนและครอบครัวคดี	เครื่องพิมพ์ชนิด Dot Matrix Printer	ขอทดแทน จำนวน 1 เครื่อง เพื่อใช้ในการออกใบ	support	yes
342	2563	ศาลชั้นอุทธรณ์	ศาลแรงงานภาค 3 (ในศาล)	เครื่องคอมพิวเตอร์สำหรับงานประมวลผล	ขอใหม่ จำนวน 3 เครื่อง เนื่องจากเครื่องคอมพิวเตอร์	support	no
343	2563	ภาค 2	ศาลจังหวัดสระบุรี	เครื่องพิมพ์ชนิด Dot Matrix แบบสี	ขอทดแทน จำนวน 1 เครื่อง เนื่องจากเครื่องพิมพ์	support	yes
344	2563	ภาค 8	ศาลเยาวชนและครอบครัวคดี	เครื่องคอมพิวเตอร์สำหรับประมวลผล	ขอใหม่ จำนวน 1 เครื่อง เนื่องจากอุปกรณ์คอมพิวเตอร์	support	no
345	2563	สำนักงาน	สถาบันพัฒนาข้าราชการฝ่าย	Zoom Meeting BIZ Plus	ขอใหม่ จำนวน 1 ชุด เนื่องจากสถานการณ์ระบาด	network	yes
346	2563	สำนักงาน	สำนักการแพทย์	เครื่องคอมพิวเตอร์สำหรับประมวลผล	ขอใหม่ จำนวน 2 ชุด คณะกรรมการบริหารศาลยุติธรรม	support	no
347	2563	ภาค 9	ศาลจังหวัดสุรินทร์	เครื่องพิมพ์ชนิด Dot Matrix Printer	ขอใหม่ จำนวน 2 เครื่อง ปัจจุบันสำนักงานขาด	support	yes

ตารางที่ 1 ความหมายของข้อมูล

คำของบประมาณ

ลำดับ	คอลัมน์	ความหมาย
1	req_id	เลขที่คำของบประมาณ
2	req_year	ปีที่ขอ
3	region_name	สังกัดของหน่วยงาน
4	office_name	หน่วยงานที่ขอ
5	req_item	รายการครุภัณฑ์ที่ขอ
6	req_detail	รายละเอียดเหตุผลความจำเป็น
7	group	ประเภทรายการครุภัณฑ์
8	approve	ผลการพิจารณา

ภาพที่ 3 ข้อมูลคำของบประมาณในรูปแบบ Excel File

3.3 การเตรียมข้อมูล สำหรับแบบจำลองที่ 1 การจำแนกส่วนงานที่เกี่ยวข้องกับคำของบประมาณ ผู้วิจัยเลือกใช้ข้อมูลคอลัมน์ req_item (รายการครุภัณฑ์) ชนิดข้อมูลเป็น Text เป็นอินพุตฟีเจอร์ (Input Feature) และเลือกข้อมูลคอลัมน์ group (ส่วนงานที่รับผิดชอบ) ชนิดข้อมูลเป็น Text ซึ่งประกอบด้วย 2 คลาส ได้แก่ support และ network เป็นลาเบล (Label) สำหรับนำไปสร้างแบบจำลอง ดังภาพที่ 4

สำหรับแบบจำลองที่ 2 การจำแนกผลการพิจารณา “อนุมัติ” หรือ “ไม่อนุมัติ” คำของบประมาณจากเหตุผลความจำเป็น ผู้วิจัยเลือกใช้ข้อมูลคอลัมน์ req_detail (รายละเอียดเหตุผลความจำเป็น) ชนิดข้อมูลเป็น Text เป็นอินพุตฟีเจอร์ (Input Feature) และเลือกข้อมูลคอลัมน์ approve (ผลการพิจารณา) ชนิดข้อมูลเป็น Text ซึ่งประกอบด้วย 2 คลาส ได้แก่ “yes” = “อนุมัติ” และ “no” = “ไม่อนุมัติ” เป็นลาเบล (Label) สำหรับนำไปสร้างแบบจำลอง ดังภาพที่ 5

req_item	group
คอมพิวเตอร์ All In One สำหรับงานประมวลผล	support
เครื่องพิมพ์ Multifunction ชนิดเลเซอร์หรือชนิดสี	network
เครื่องเสิร์ฟเวอร์ข้อมูล (Network Attached Storage : NAS)	network
อุปกรณ์จัดเก็บข้อมูล NAS QNAP รุ่น TS-4538U-2G	network
อุปกรณ์จัดเก็บข้อมูล NAS รุ่น TS-451+ 2G	network
เครื่องเล่นแผ่นเสียง	support

ภาพที่ 4 ฟีเจอร์และลาเบลที่นำไปใช้สร้างแบบจำลองที่ 1

req_detail	approve
ขอเพิ่ม จำนวน 10 เครื่อง เนื่องจากปีติกรประจำองค์คณะผู้พิพากษา ผู้ช่วยผู้พิพากษาไปเพียงพอ	no
ขอใหม่ จำนวน 1 เครื่อง เพื่อไว้ระบบงานธุรการงานคดีมีประสิทธิภาพมากขึ้น และเป็นการประหยัด	yes
ขอใหม่ จำนวน 1 เครื่อง เนื่องจากศาลจังหวัดนนทบุรียังไม่เคยได้รับการจัดสรรเครื่องสำรองข้อมูล (yes)	
ขอใหม่ จำนวน 1 ชุด เนื่องจากสถาบันพัฒนาข้าราชการฝ่ายตุลาการศาลยุติธรรมไม่มีเครื่องเก็บข้อมูล	no
ขอทดแทน จำนวน 1 ชุด เนื่องจากอุปกรณ์ของเดิมชำรุด เพื่อให้บุคลากรภายในหน่วยงานใช้	yes
ขอทดแทน จำนวน 1 เครื่อง เนื่องจากเครื่องสแกนลายนิ้วมือที่ติดตั้งอยู่ที่ส่วนช่วยอำนวยความสะดวก	yes
ขอใหม่ จำนวน 1 เครื่อง ศาลอุทธรณ์คดีชั้นอุทธรณ์ไม่มีเครื่องคอมพิวเตอร์ Server สำหรับใช้ใน	yes
ขอใหม่ จำนวน 1 ชุด เพื่อจัดเก็บข้อมูลของพิชิตภัทศ์ศาลไทยและห้องหมายเหตุ (ตั้งอยู่ที่อาคาร	yes
ขอทดแทน จำนวน 1 ชุด เนื่องจากอุปกรณ์ชุดความคุมกล้องเสีย และมีอายุการใช้งานเกินระยะเวลา	yes
ขอเพิ่ม จำนวน 2 เครื่อง เพื่อให้ศูนย์ข้อมูลความ ศูนย์ประสานงานพยาน งานระบบนิติการภาคต่อ	yes
ขอทดแทน จำนวน 1 เครื่อง เพื่อใช้ในการออกใบเสร็จประจำวันทุกวัน เนื่องจากเครื่องเดิมชำรุดเสีย	yes
ขอเพิ่ม จำนวน 3 เครื่อง เนื่องจากเครื่องคอมพิวเตอร์สำหรับงานประมวลผล แบบที่ 2 (จอแสดงภาพ	no
ขอทดแทน จำนวน 1 เครื่อง เนื่องจากเครื่องพิมพ์ชนิด Dot Matrix แบบแคชยาว ชำรุด เนื่องจากใช้	yes
ขอเพิ่ม จำนวน 1 เครื่อง เนื่องจากปัจจุบันเครื่องคอมพิวเตอร์ไม่เพียงพอต่อการใช้งาน โดยปัจจุบัน	no
ขอใหม่ จำนวน 1 ชุด เนื่องจากสถานการณ์ระบาดของโรคติดเชื้อไวรัสโคโรนา ๒๐๑๙ (COVID-19)	yes
ขอเพิ่ม จำนวน 2 ชุด คณะกรรมการบริหารศาลยุติธรรมมีมติในการประชุมครั้งที่ 6/256 เมื่อวันที่ 22	no
ขอใหม่ จำนวน 2 เครื่อง มีจำนวนส่วนช่วยพิจารณาคดี ยังไม่มีเครื่องพิมพ์ในการออกใบเสร็จรับเงิน	yes
ขอเพิ่ม จำนวน 17 เครื่อง สำหรับพิมพ์งานช่วยพิจารณาคดี	no
ขอเพิ่ม จำนวน 1 เครื่อง สำหรับงานประจำกลุ่มงานโฆษกและวิเคราะห์ข่าว โดยกองสารนิเทศและ	no
ขอเพิ่ม จำนวน 9 เครื่องสแกนเนอร์ : เนื่องจากปัจจุบันศาลเยาวชนและครอบครัวจังหวัดเชียงใหม่ยัง	no

ภาพที่ 5 พีเจอร์และลาเบลที่นำไปใช้สร้างแบบจำลองที่ 2

โดยขั้นตอนในการเตรียมข้อมูลประเภท Text ซึ่งมีรายละเอียดดังนี้

(1) ลบคำเฉพาะ (Text Cleaning) เช่น ชื่อยี่ห้อ รุ่น เช่น “TS-4538U-2G” เป็นต้น และลบเครื่องหมายต่างๆ ที่เราไม่ต้องการ (!@%#*&~) ซึ่งคำเหล่านี้ไม่สื่อความความหมายและมีผลต่อการตัดคำในกระบวนการตัดคำ ซึ่งจะทำให้จำนวนคำที่ได้มีจำนวนมาก คำเหล่านี้จะต้องถูกลบออกจากชุดข้อมูลเพื่อลดจำนวนของคำที่จะนำไปใช้เป็นคุณลักษณะพีเจอร์ (Features) ในการสร้างแบบจำลองทำนาย ทั้งนี้ในขั้นตอนนี้กรองเอาเฉพาะตัวอักษรและสระที่เป็นภาษาไทย และตัวอักษรภาษาอังกฤษเท่านั้น เครื่องมือที่ใช้คือไลบรารีที่ชื่อว่า RegEx (Reggular Expression) ซึ่งกำหนดรูปแบบได้เป็น `[\u0E00-\u0E7Fa-zA-Z' ]|'|'` และเขียนด้วยภาษาไพธอน (Python) ดังที่แสดงไว้ในภาพที่ 6 - 7

0	คอมพิวเตอร์AllInOneสำหรับงานประมวลผล
1	เครื่องพิมพ์Multifunctionชนิดเลเซอร์หรือชนิดสี
2	เครื่องสำรองข้อมูลNetworkAttachedStorageNAS
3	อุปกรณ์จัดเก็บข้อมูลNAS
4	อุปกรณ์จัดเก็บข้อมูลNAS

ภาพที่ 6 ตัวอย่างข้อความที่กรองด้วย Regular Expression แบบจำลองที่ 1

0	ขอเพิ่ม จำนวน เครื่อง เนื่องจากปีติกรประจำองค์...
1	ขอใหม่ จำนวน เครื่อง เพื่อไว้ระบบงานธุรการงาน...
2	ขอใหม่ จำนวน เครื่อง เนื่องจากศาลจังหวัดนนทบุรี...
3	ขอใหม่ จำนวน ชุด เนื่องจากสถาบันพัฒนาข้าราชการ...
4	ขอทดแทน จำนวน ชุด เนื่องจากอุปกรณ์ของเดิมชำรุด...
...	...
712	ขอเพิ่ม จำนวน เครื่อง สำหรับจัดสรรให้ข้าราชการ...
713	ขอเพิ่ม จำนวน เครื่อง สำหรับรองรับการใช้งานขอ...
714	ขอเพิ่ม จำนวน ชุด สำหรับรองรับอัตราค่าจ้างข้าราชการ...
715	ขอใหม่ เครื่อง สำหรับใช้งานประจำวันศูนย์ข้อมูลแ...
716	ขอทดแทน จำนวน เครื่อง เนื่องจากเป็นเครื่องที่...

ภาพที่ 7 ตัวอย่างข้อความที่กรองด้วย Regular Expression แบบจำลองที่ 2

(2) การตัดคำ (Tokenize)

ในขั้นตอนนี้ ผู้วิจัยได้นำความรู้เรื่องการประมวลผลภาษาธรรมชาติ โดยประยุกต์ใช้ไลบรารีที่ชื่อว่า PyThaiNLP ซึ่งมี Module “word_tokenize” และ engine ที่ชื่อว่า “newmm” เข้ามาช่วยในการตัดคำ และในขั้นตอนนี้ยังเพิ่มเติมเรื่องการตัด stop words หรือ คำที่เจอบ่อยๆ แต่ไม่สื่อความหมายอะไรมา เช่น “การ”, “ความ”, “คือ”, “ที่”, “ซึ่ง” เป็นต้น ซึ่งการตัด stop words จะมีข้อดี คือทำให้คลังคำศัพท์ (Bag of Words) ไม่ใหญ่เกินไปโดยไม่จำเป็น ซึ่งจะช่วยให้แบบจำลองหรือโมเดลของเรามีความถูกต้องแม่นยำมากขึ้น และในงานวิจัยนี้ใช้ไลบรารีที่ชื่อว่า “thai_stopwords” ซึ่งเป็นกลุ่มของคำภาษาไทยที่ไม่สื่อความหมายมาลบออกจากข้อความในคำของบประมาณด้วย ดังแสดงไว้ตามภาพที่ 8 - 9

```
0 คอมพิวเตอร์ All In One สำหรับ งาน ประมวลผล
1 เครื่องพิมพ์ Multifunction ขีด เลเซอร์ หรือ ...
2 เครื่อง สารอง ข้อมูล Network Attached Storage NAS
3 อุปกรณ์ จัดเก็บ ข้อมูล NAS
4 อุปกรณ์ จัดเก็บ ข้อมูล NAS
...
712 เครื่องคอมพิวเตอร์ สำหรับ งาน ประมวลผล แบบ ที่
713 เครื่องพิมพ์ เลเซอร์ หรือ LED ขาว ดำ ขีด netw...
714 คอมพิวเตอร์ สำหรับ งาน ประมวลผล แบบ ที่
715 สแกนเนอร์ สำหรับ งาน เก็บ เอกสาร ระดับ ศูนย์ บ...
716 เครื่อง สแกน ลายนิ้วมือ ขีด มันทัก เวลา เข้า ...
```

ภาพที่ 8 ตัวอย่างคำของบประมาณที่ตัดคำด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 1

```
0 ขอ เพิ่ม จำนวน เครื่อง เนื่องจาก นิติกร ประจำ ...
1 ขอ ใหม่ จำนวน เครื่อง เพื่อให้ ระบบ งานธุรการ ...
2 ขอ ใหม่ จำนวน เครื่อง เนื่องจาก ศาลจังหวัด นนท...
3 ขอ ใหม่ จำนวน ชุด เนื่องจาก สถาบัน พัฒนา ข้าราชการ...
4 ขอ ทดแทน จำนวน ชุด เนื่องจาก อุปกรณ์ ของ เดิม ...
...
712 ขอ เพิ่ม จำนวน เครื่อง สำหรับ จัดสรร ไฟ ข้าราชการ...
713 ขอ เพิ่ม จำนวน เครื่อง สำหรับ รองรับ การใช้งาน...
714 ขอ เพิ่ม จำนวน ชุด สำหรับ รองรับ อัตรา กำลัง ข...
715 ขอ ใหม่ เครื่อง สำหรับ ใช้ งานประจำ ศูนย์ บัด ...
716 ขอ ทดแทน จำนวน เครื่อง เนื่องจาก เป็น ครุภัณฑ์...
```

ภาพที่ 9 ตัวอย่างคำของบประมาณที่ตัดคำด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 2

(3) การสร้างคลังคำศัพท์ (Bag of Words)

สำหรับขั้นตอนนี้ คือ การแปลงคำแต่ละคำ (ที่ไม่ซ้ำกัน) เป็น ID ซึ่งขั้นตอนนี้ใช้ไลบรารี “sklearn” ซึ่งจะรวมไว้กับขั้นตอนการหาจำนวนคำในคำของบประมาณที่ได้จากการตัดคำเพื่อให้ได้เฉพาะออกมาจากทุกข้อความในคำของบประมาณ ซึ่งในงานวิจัยนี้ใช้ไลบรารีที่ชื่อว่า “CountVectorizer” ในภาษาไพธอนเป็นเครื่องมือในการหาจำนวนคำในเอกสาร และจัดเก็บในรูปแบบ Vector (จะคล้ายๆ กับลักษณะของพจนานุกรมที่มีการระบุตัวเลข index ของแต่ละคำ) และได้ผลการนับคำในเอกสารทั้งหมดเป็นคลังคำศัพท์ ดังภาพที่ 10

```
cvec = CountVectorizer(analyzer=lambda x:x.split(' '))
c_feat = cvec.fit_transform(tokens_list)

cvec.vocabulary_

{'คอมพิวเตอร์': 95,
 'All': 2,
 'In': 22,
 'One': 42,
 'สำหรับ': 200,
 'งาน': 98,
 'ประมวลผล': 151,
 'เครื่องพิมพ์': 230,
 'MultiFunction': 37,
 'ชนิด': 109,
 'เลเซอร์': 238,
```

ภาพที่ 10 ตัวอย่างคลังคำศัพท์ (Bag of Words) จากการนับคำในเอกสารคำของบประมาณ

(4) สร้างฟีเจอร์ (Feature) จากการนับคำ (Count Words)

สำหรับขั้นตอนนี้เป็นการสร้าง Feature จากการนับคำ คือ การนับจำนวนครั้งที่แต่ละ word id ที่แต่ละ word id ทั้งหมดปรากฏในแต่ละ text (ไม่มี คือ 0) โดยแสดงได้ดังภาพที่ 11

```
train_bow = cvec.transform(tokens_list)
pd.DataFrame(train_bow.toarray(), columns=cvec.get_feature_names_out(), index=tokens_list)
print( pd.DataFrame(train_bow.toarray(), columns=cvec.get_feature_names_out(), index=tokens_list))
```

final_word_token	A	Access	All	Attached	\
คอมพิวเตอร์ All In One สำหรับ งาน ประมวลผล	0	0	1	0	
เครื่องพิมพ์ Multifunction ชนิด เลเซอร์ หรือ ชน...	0	0	0	0	
เครื่อง สารอง ข้อมูล Network Attached Storage NAS	0	0	0	1	
อุปกรณ์ จัดเก็บ ข้อมูล NAS	0	0	0	0	
อุปกรณ์ จัดเก็บ ข้อมูล NAS	0	0	0	0	
...	...	...	...	...	...
เครื่องคอมพิวเตอร์ สำหรับ งาน ประมวลผล แบบ ที่	0	0	0	0	
เครื่องพิมพ์ เลเซอร์ หรือ LED ขาว ดำ ชนิด netwo...	0	0	0	0	
คอมพิวเตอร์ สำหรับ งาน ประมวลผล แบบ ที่	0	0	0	0	
สแกนเนอร์ สำหรับ งาน เก็บ เอกสาร ระดับ ศูนย์ นร...	0	0	0	0	
เครื่อง สแกน ลายนิ้วมือ ชนิด บันทึก เวลา เข้า อ...	0	0	0	0	

ภาพที่ 11 ตัวอย่างฟีเจอร์ (Feature) ที่สร้างจากการนับคำ

3.4 สร้างแบบจำลอง

(1) แบ่งข้อมูลชุดข้อมูล การสร้างแบบจำลองเพื่อการทำนาย ซึ่งถ้าข้อมูลที่ใช้ในการทำนายเป็นชุดเดียวกันจะทำให้เกิดปัญหาที่เรียกว่า Overfitting เพื่อที่จะหลีกเลี่ยงปัญหานี้ ผู้วิจัยได้แบ่งชุดข้อมูลออกเป็น ส่วนย่อยๆ (Fold) โดยชุดข้อมูลที่ถูกแบ่งจะสามารถกำหนดจำนวนครั้ง (k) สำหรับเป็นชุดข้อมูลที่ใช้สร้างแบบจำลอง (Training set) และชุดข้อมูลสำหรับทดสอบ (Testing set) ซึ่งในงานวิจัยนี้ ผู้วิจัยเลือกใช้เทคนิคที่ชื่อ

ว่า K-Fold Cross Validation โดยใช้ไลบรารีชื่อ “StratifiedKFold” สำหรับช่วยในการแบ่งข้อมูล และกำหนดจำนวนชุดข้อมูลที่ใช้สร้างแบบจำลองและทดสอบ จำนวน 5 ชุด

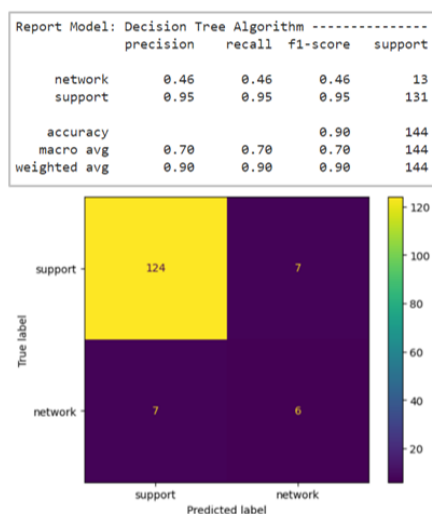
(2) สร้างแบบจำลอง เมื่อแบ่งชุดข้อมูลตามขั้นตอนข้างต้นเรียบร้อยแล้ว จึงนำข้อมูลชุดดังกล่าวไปสร้างแบบจำลองโดยใช้อัลกอริทึมต่างๆ ซึ่งในงานวิจัยนี้ เลือกใช้อัลกอริทึมจำแนกข้อมูลแบบมีผู้สอน หรือการเรียนรู้แบบมีผู้สอน (Supervised Learning) ได้แก่ ต้นไม้ตัดสินใจ (Decision Tree), ชุดของต้นไม้ตัดสินใจ (Ensemble of Decision Trees), นาอิว เบย์ (Naïve Bayes) และลอจิสติกเรเกรสชัน (Logistic Regression)

3.5 การทดสอบและประเมินประสิทธิภาพของแบบจำลอง ในงานวิจัยนี้ ดำเนินการทดสอบ

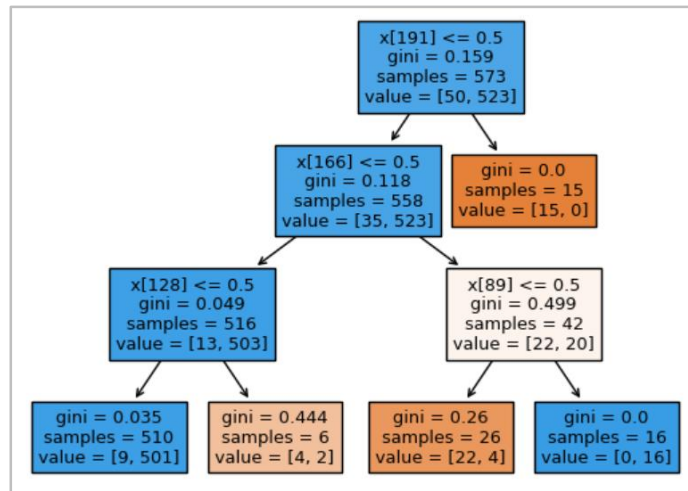
ประสิทธิภาพแบบจำลองโดยการแบ่งชุดข้อมูล testing data ด้วยเทคนิค 5-Fold Cross Validation รวมถึงใช้วิธีการวัดประสิทธิภาพของแบบจำลองด้วยค่า Accuracy, Precision, Recall, Macro F1 โดยแสดงค่าการทำนายผลของแบบจำลองด้วยตาราง Confusion Matrix เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมระหว่างค่าการทำนายและค่าจริงในรูปแบบของตาราง

ผลการวิจัย

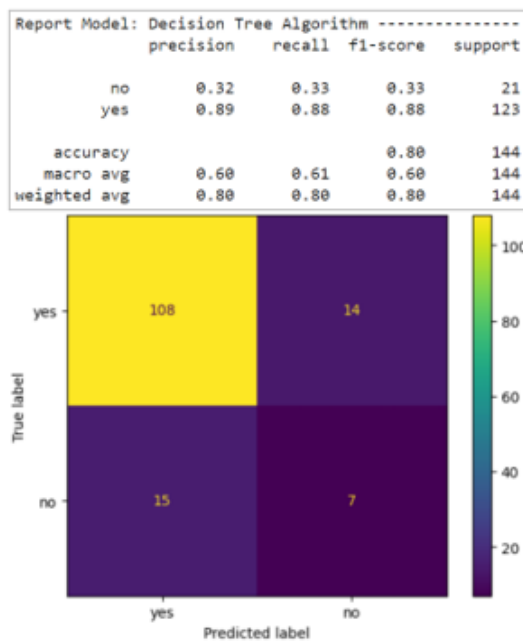
การเลือกเทคนิคสำหรับใช้สร้างแบบจำลองที่ 1 จำแนกกลุ่มงานที่รับผิดชอบค่าของงบประมาณ และแบบจำลองที่ 2 จำแนกผลการพิจารณาค่าของงบประมาณจากเหตุผลความจำเป็น ผู้วิจัยใช้อัลกอริทึมต่างๆ ของเทคนิค Text Classification ได้แก่ 1) Decision Tree 2) Random Forest 3) Naïve Bayes และ 4) Logistic Regression และทดสอบประสิทธิภาพของแบบจำลองด้วยเทคนิค k-Fold Cross Validation สำหรับแบ่งข้อมูลเป็น Training data และ Testing data จำนวน 5 ชุด รวมถึงใช้วิธีการวัดประสิทธิภาพของแบบจำลองแบบ Classification ด้วยค่า Confusion Matrix ซึ่งมีผลการประเมินประสิทธิภาพของแบบจำลองที่ 1 และแบบจำลองที่ 2 ดังภาพที่ 12 – 15 และแสดงค่า Accuracy ซึ่งเป็นค่าเฉลี่ยที่ได้จากการทดสอบประสิทธิภาพแบบจำลองด้วยเทคนิค 5-Folds Cross Validation ไว้ในตารางที่ 2



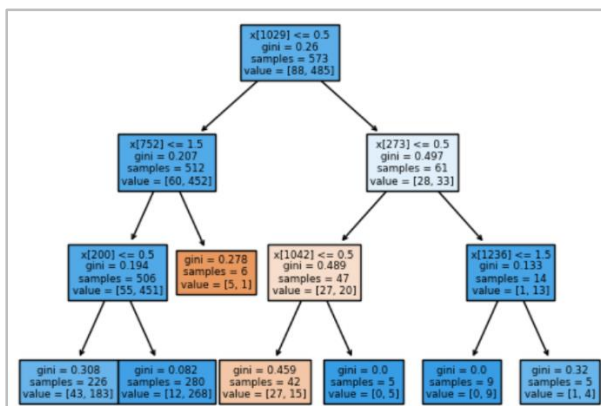
ภาพที่ 12 ประสิทธิภาพของแบบจำลองที่ 1 ด้วยอัลกอริทึม Decision Tree



ภาพที่ 13 Decision Tree แบบจำลองที่ 1



ภาพที่ 14 ประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Decision Tree



ภาพที่ 15 ตัวอย่าง Decision Tree แบบจำลองที่ 2 (ข้อมูลชุดที่ 1)

ตารางที่ 2 การประเมินประสิทธิภาพของแบบจำลองที่ 1 จำแนกส่วนงานที่เกี่ยวข้องกับคำของบประมาณ และแบบจำลองที่ 2 จำแนกผลการพิจารณาคำของบประมาณ

ชุดแบบจำลอง	ค่าความถูกต้อง (Accuracy)			
	Decision Tree	Random Forest	Naïve Bayes	Logistic Regression
แบบจำลองที่ 1	95 %	96 %	95 %	95 %
แบบจำลองที่ 2	82 %	82 %	69 %	78 %

สรุปผลการวิจัย

ผลการวิจัยหัวข้อ “การเพิ่มประสิทธิภาพการจัดสรรงบประมาณด้านครุภัณฑ์คอมพิวเตอร์ด้วยการประยุกต์ใช้เทคนิคเหมืองข้อความ (Text Mining)” สามารถสร้างแบบจำลองที่ช่วยตัดสินใจการพิจารณาคำงบประมาณ ได้ตามขอบเขตงานที่กำหนดไว้ จำนวน 2 โมเดล ประกอบด้วย

(1) แบบจำลองที่ 1 จำแนกส่วนงานที่เกี่ยวข้องกับคำของบประมาณ ได้แก่ support และ network โดยใช้ข้อมูลรายการคำของบประมาณ เป็นชุดฝึกเพื่อสร้างแบบจำลอง และทดสอบแบบจำลอง

(2) แบบจำลองที่ 2 จำแนกผลการพิจารณาคำของบประมาณ ได้แก่ “อนุมัติ=yes” และ “ไม่อนุมัติ=no” โดยใช้ชุดข้อมูลรายละเอียดเหตุผลความจำเป็น เป็นฝึกเพื่อสร้างแบบจำลอง และทดสอบแบบจำลอง

สำหรับงานวิจัยนี้ ใช้กระบวนการทำความเข้าใจข้อมูล การตัดคำ การค้นหาคำสำคัญในข้อความ เพื่อสร้างคลังคำศัพท์ และสร้างฟีเจอร์ด้วยการประยุกต์ใช้ความรู้การประมวลผลภาษาธรรมชาติ และใช้อัลกอริทึมประเภท Classification สำหรับสร้างและประเมินประสิทธิภาพของแบบจำลอง ทั้งนี้ ผู้วิจัยเลือกใช้อัลกอริทึม ได้แก่ Decision Tree, Random Forest, Naïve Bayes และ Logistic Regression ซึ่งผลการทดลองพบว่าอัลกอริทึมที่มีประสิทธิภาพสูงสุด ได้แก่ Random Forest รองลงมาได้แก่ Decision Tree Logistic Regression และ Naïve Bayes ตามลำดับ

บรรณานุกรม

สำนักแผนงานและงบประมาณ สำนักงานศาลยุติธรรม. ระบบบริหารงบประมาณศาลยุติธรรมและสำนักงานศาลยุติธรรม, 2565.

เฉลิมชัย พิเศษ. ระบบจำแนกตัวงานอัจฉริยะด้วยเทคนิควิเคราะห์ข้อความไทย – อังกฤษ และการเรียนรู้ของเครื่อง. สารนิพนธ์ปริญญาโทบริหารธุรกิจ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ, 2562.

วาทัญญู ชูประจิตต์. การจำแนกประเภทและวิธีการคัดเลือกคุณลักษณะสำหรับการสนับสนุนการวิเคราะห์ความรู้สึกต่อสถานที่ท่องเที่ยวไทย. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยทักษิณ, 2563.

Hindawi Computational Intelligence and Neuroscience. **Advances in Computational Intelligence Techniques for Next Generation Internet of Things**, 2022. ค้นวันที่ 1 สิงหาคม 2565 จาก <https://www.hindawi.com/journals/cin/2022/1883698>.

Witchapong Darontham. ขั้นตอนการเตรียมข้อมูลประเภท Text ภาษาไทย แบบง่ายๆ โดยใช้ Python, 2018. ค้นวันที่ 1 สิงหาคม 2565 จาก <https://medium.com>.