DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

การเพิ่มประสิทธิภาพการจัดสรรงบประมาณด้านครุภัณฑ์คอมพิวเตอร์ด้วยการประยุกต์ใช้เทคนิคเหมืองข้อความ (Text Mining)

พรพรรษชล มาทัพ

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
ปีการศึกษา 2565

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ENHANCING BUDGET ALLOCATION FOR COMPUTER EQUIPMENT USING TEXT MINING

PHONPATSACHON MATAP

A Thematic Paper Submitted in Partial Fulfillment of the
Requirements for the Degree of Master of Big Data Engineering,
College of Innovative Technology and Engineering,
Dhurakij Pundit University
Academic Year 2022

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ใบรับรองสารนิพนธ์

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์

วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่

หัวข้อสารนิพนธ์ การเพิ่มประสิทธิภาพการจัดสรรงบประมาณด้านครุภัณฑ์คอมพิวเตอร์ด้วยการประยุกต์ใช้เทคนิคเหมืองข้อความ (Text Mining)

เสนอโดย พรพรรษชล มาทัพ

สาขาวิชา วิศวกรรมข้อมูลขนาดใหญ่

อาจารย์ที่ปรึกษาสารนิพนธ์ ดร.ธนภัทร มังคะจิตร

ได้พิจารณาเห็นชอบโดยคณะกรรมการสอบสารนิพนธ์แล้ว

(ผู้ช่วยศาสตราจารย์ ดร.ณัฐพัชร์ อารีรัชกุลกานต์) ประธานกรรมการ

(ดร.ธนภัทร มังคะจิตร) กรรมการที่ปรึกษาสารนิพนธ์

(ผู้ช่วยศาสตราจารย์ ดร.ดวงใจ จิตคงชื่น) กรรมการ

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ รับรองแล้ว

(ดร.ชัยพร เขมะภาตะพันธ์) คณบดีวิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์

วันที่ 31 เดือน กรกฎาคม พ.ศ. 2566

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

หัวข้อสารนิพนธ์ การเพิ่มประสิทธิภาพการจัดสรรงบประมาณด้านครุภัณฑ์คอมพิวเตอร์ด้วยการประยุกต์ใช้เทคนิคเหมืองข้อความ (Text Mining)

ชื่อผู้เขียน พรพรรษชล มาทัพ

อาจารย์ที่ปรึกษา ดร.ธนภัทร ฆังคะจิตร

หลักสูตร วิศวกรรมข้อมูลขนาดใหญ่

ปีการศึกษา 2565

บทคัดย่อ

ระบบงานหลักของสำนักเทคโนโลยีสารสนเทศ สำนักงานศาลยุติธรรม คือ การพิจารณา และตรวจสอบคำงบประมาณที่ร้องขอสำหรับการจัดซื้อครุภัณฑ์คอมพิวเตอร์ให้แก่หน่วยงานในสังกัดสำนักงานศาลยุติธรรมทั่วประเทศ ซึ่งในทุกปีมีเอกสารคำของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์และเอกสารประกอบเพิ่มเติมเป็นจำนวนมาก ด้วยเหตุนี้จึงส่งผลให้กระบวนการพิจารณาและตรวจสอบมีความล่าช้าเนื่องจากความยุ่งยากในการจัดเก็บ การเรียกดู และการใช้ข้อมูล โดยเฉพาะอย่างยิ่งเจ้าหน้าที่ของสำนักเทคโนโลยีสารสนเทศจะต้องคัดแยกและจัดประเภทเอกสารคำของบประมาณที่หน่วยงานในสังกัดร้องขอ เพื่อนำส่งไปยังแผนกงานที่เกี่ยวข้องดำเนินการตรวจสอบและพิจารณา

ดังนั้น งานวิจัยนี้จึงมุ่งหวังที่จะนำเสนอเทคนิคการทำเหมืองข้อความ และการประมวลผลภาษาธรรมชาติเพื่อจัดหมวดหมู่แผนกที่เกี่ยวข้องซึ่งรับผิดชอบต่อการพิจารณาอนุมัติงบประมาณตามประเภทของครุภัณฑ์คอมพิวเตอร์ และช่วยการตัดสินใจเรื่องการอนุมัติงบประมาณตามรายละเอียดของเหตุผลและความจำเป็นที่ปรากฏในคำของบประมาณ โดยเริ่มจากขั้นตอนการจัดเก็บข้อมูลลงในระบบฐานข้อมูล และนำข้อมูลรายการครุภัณฑ์คอมพิวเตอร์และเหตุผลความจำเป็นซึ่งปรากฏในคำของบประมาณที่จัดเก็บไว้เข้าสู่กระบวนการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ด้วยวิธีการตัดคำ การแปลงคำให้เป็นรูปแบบเชิงปริมาณ และการสกัดคุณลักษณะ เพื่อนำไปใช้สร้างแบบจำลองสำหรับจัดหมวดหมู่ประเภทข้อความ (Text Classification) โดยใช้อัลกอริทึมได้แก่ ต้นไม้ตัดสินใจ (Decision Tree), ชุดต้นไม้ตัดสินใจ (Random Forest), นาอีฟเบย์ (Naïve Bayes) และลอจิสติกส์รีเกรสชัน (Logistic Regression) และวัดประสิทธิภาพของแบบจำลองที่ได้ด้วยวิธีการ 5-Fold Cross Validation ซึ่งประสิทธิภาพของแบบจำลองการจัดหมวดหมู่แผนกที่เกี่ยวข้อง และแบบจำลองการตัดสินใจการอนุมัติงบประมาณที่ได้จากงานวิจัยนี้มีค่าความถูกต้องแม่นยำที่น่าพอใจ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ดังนั้นจึงสรุปได้ว่างานวิจัยนี้สามารถเพิ่มความเร็วในกระบวนการตรวจสอบและพิจารณางบประมาณได้อย่างมีประสิทธิภาพ ด้วยการลดกระบวนงานที่ไม่จำเป็นของเจ้าหน้าที่ในสำนักเทคโนโลยีสารสนเทศ และยังให้ข้อมูลสำคัญในการตัดสินใจได้อีกด้วย

คำสำคัญ : คำของบประมาณ, เหมืองข้อความ, การประมวลผลทางธรรมชาติ, การจัดหมวดหมู่ประเภทข้อความ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Thematic Paper Title	ENHANCING BUDGET ALLOCATION FOR COMPUTER EQUIPMENT USING TEXT MINING
Author	PHONPATSACHON MATAP
Thematic Paper Advisor	Dr. Thanapat Kangkachit
Program	Big Data Engineering
Academic Year	2022

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.

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

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

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

กิตติกรรมประกาศ

ในงานวิจัยฉบับนี้สำเร็จลุล่วงได้อย่างสมบูรณ์ ด้วยความช่วยเหลือจาก ดร.ธนภัทร ฆังคะจิตร อาจารย์ที่ปรึกษา ซึ่งท่านได้ให้คำแนะนำ และให้ข้อคิดเห็นต่างๆ อันเป็นประโยชน์อย่างยิ่งในการดำเนินงานวิจัย อีกทั้งยังช่วยแก้ไขปัญหาต่างๆ ที่เกิดขึ้นในระหว่างการดำเนินงาน ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาและขอขอบพระคุณท่านเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบคุณ คณาจารย์วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ และเพื่อนนักศึกษาปริญญาโท สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่ ที่ได้กรุณาให้คำแนะนำช่วยเหลือ และปรับปรุงงานวิจัยจนสำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณ ผู้บังคับบัญชา สำนักงานศาลยุติธรรมที่ให้โอกาสในด้านการศึกษาในระดับที่สูงขึ้นแก่ผู้วิจัย รวมถึงเพื่อนร่วมงานในสำนักเทคโนโลยีสารสนเทศที่สนับสนุนข้อมูลที่ใช้สำหรับการดำเนินงานวิจัยนี้

สุดท้ายนี้ ผู้วิจัยขอขอบพระคุณมารดาซึ่งเป็นที่รักยิ่ง ที่คอยให้กำลังใจและสนับสนุนในทุกด้านแก่ผู้วิจัยเป็นอย่างดีเสมอมา

พรพรรษชล มาทัพ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... ฉ

กิตติกรรมประกาศ..... ซ

สารบัญ..... ฌ

สารบัญตาราง..... ฎ

สารบัญภาพ..... ฏ

บทที่

1. บทนำ..... 1
   - 1.1 ที่มาและความสำคัญของปัญหา..... 1
   - 1.2 วัตถุประสงค์ของการศึกษาหรือวิจัย..... 2
   - 1.3 ขอบเขตของงานวิจัย..... 2
   - 1.4 ประโยชน์ที่คาดว่าจะได้รับ..... 2
   - 1.5 ขั้นตอนการทำวิจัย..... 3
   - 1.6 นิยามศัพท์..... 4
2. แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง..... 6
   - 2.1 ทฤษฎีที่เกี่ยวข้อง..... 6
   - 2.2 งานวิจัยที่เกี่ยวข้อง..... 17
3. ระเบียบวิธีวิจัย..... 20
   - 3.1 แบบจำลองที่ 1 จำแนกส่วนงานที่เกี่ยวข้องกับคำของบประมาณ 20
   - 3.2 แบบจำลองที่ 2 จำแนกผลการพิจารณาคำของบประมาณจากเหตุผลความจำเป็น 34
4. ผลการวิจัย..... 46
   - 4.1 การทำความสะอาดข้อมูล..... 46
   - 4.2 การประมวลผลภาษาธรรมชาติ..... 46
   - 4.3 การสร้างแบบจำลอง และประเมินประสิทธภาพของแบบจำลอง..... 48
   - 4.4 การวิเคราะห์ผลการทำนายของแบบจำลอง..... 60

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญ (ต่อ)

บทที่ **หน้า**

5. สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ..... 63
5.1 สรุปผลการวิจัย..... 63
5.2 อภิปรายผล..... 63
5.3 ข้อเสนอแนะ..... 64
บรรณานุกรม..... 66
ภาคผนวก..... 68
ก เอกสารคำของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์..... 69
ประวัติผู้เขียน..... 81

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญตาราง

ตารางที่ **หน้า**

3.1 ความหมายของข้อมูลคำของบประมาณ..... 22
3.2 ความหมายของ Regular Expression..... 24
3.3 ค่าพารามิเตอร์ที่ใช้ในอัลกอริทึม Decision Tree แบบจำลองที่ 1..... 30
3.4 ค่าพารามิเตอร์ที่ใช้ในอัลกอริทึม Random Forest แบบจำลองที่ 1..... 31
3.5 ค่าพารามิเตอร์ที่ใช้ในอัลกอริทึม Decision Tree แบบจำลองที่ 2..... 41
3.6 ค่าพารามิเตอร์ที่ใช้ในอัลกอริทึม Random Forest แบบจำลองที่ 2..... 42
4.1 ตัวอย่างคลังคำศัพท์ และฟีเจอร์ แบบจำลองที่ 1 และแบบจำลองที่ 2..... 48
4.2 การประเมินประสิทธิภาพของแบบจำลองที่ 1..... 59
4.3 การประเมินประสิทธิภาพของแบบจำลองที่ 2..... 60

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญภาพ

ภาพที่ **หน้า**

2.1 กระบวนการงบประมาณ..... 10
2.2 ขั้นตอนในกระบวนการ CRISP – DM..... 12
3.1 ภาพรวมวิธีดำเนินการวิจัย..... 20
3.2 เอกสารคำของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์..... 21
3.3 ข้อมูลคำของบประมาณในรูปแบบ Excel File..... 22
3.4 ฟีเจอร์ และลาเบลที่นำไปใช้สร้างแบบจำลองที่ 1..... 23
3.5 การกรองข้อความที่กรองด้วย Regular Expression..... 23
3.6 ตัวอย่างข้อความที่กรองด้วย Regular Expression แบบจำลองที่ 1..... 24
3.7 การตัดด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 1..... 25
3.8 ตัวอย่างคำของบประมาณที่ตัดด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 1..... 25
3.9 ภาพรวมของข้อความในคำของบประมาณ..... 26
3.10 กราฟแท่งแสดงจำนวนข้อมความในคำของบประมาณ..... 26
3.11 ภาพรวมของข้อมูลหน่วยงานที่รับผิดชอบพิจารณาคำของบประมาณ..... 27
3.12 กราฟแท่งแสดงกลุ่มงานที่รับผิดชอบพิจารณาคำของบประมาณ..... 27
3.13 ตัวอย่างคลังคำศัพท์ (Bag of Words) จากการนับค่าคำในเอกสารคำของบประมาณ..... 28
3.14 ตัวอย่างฟีเจอร์ (Feature) ที่สร้างจากการนับคำ..... 28
3.15 การแบ่งชุดข้อมูลสำหรับสร้างแบบจำลองและทดสอบประสิทธิภาพ แบบจำลองที่ 1..... 29
3.16 ตัวอย่างชุดข้อมูล Training set และ Testing set ชุดที่ 1 แบบจำลองที่ 1..... 29
3.17 ตัวอย่างชุดข้อมูล Training set และ Testing set ชุดที่ 5 แบบจำลองที่ 1..... 29
3.18 เครื่องมือสำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Decision Tree แบบจำลองที่ 1.. 30
3.19 เครื่องมือสำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Random Forest..... 31
3.20 เครื่องมือสำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Naïve Bayes..... 31
3.21 เครื่องมือสำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Logistic Regression..... 32
3.22 เครื่องมือที่ใช้ในการทดสอบอัลกอริทึมของแบบจำลองที่ 1..... 32
3.23 ตัวอย่างชุดข้อมูลที่ใช้ในการทดสอบอัลกอริทึมของแบบจำลองที่ 1..... 33
3.24 เครื่องมือสำหรับประเมินประสิทธิภาพแบบจำลองที่ 1..... 33

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

3.25 ตัวอย่าง Confusion Matrix แสดงผลประสิทธิภาพและผลการทำนายของแบบจำลองที่ 1..... 34
3.26 ฟีเจอร์และลาเบลที่นำไปใช้สร้างแบบจำลองที่ 2..... 35
3.27 การกรองข้อความด้วย Regular Expression แบบจำลองที่ 2..... 36
3.28 ตัวอย่างข้อความที่กรองด้วย Regular Expression แบบจำลองที่ 2..... 36
3.29 การตัดคำด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 2..... 36
3.30 ตัวอย่างคำของบประมาณที่ตัดด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 2..... 37
3.31 ภาพรวมของข้อความในรายละเอียดเหตุผลความจำเป็น..... 37
3.32 กราฟแท่งแสดงจำนวนข้อความในรายละเอียดเหตุผลความจำเป็น..... 38
3.33 ภาพรวมของข้อมูลผลการพิจารณาคำของบประประมาณ..... 38
3.34 กราฟแท่งแสดงผลการพิจารณาคำของบประมาณ..... 38
3.35 ตัวอย่างคลังคำศัพท์ (Bag of Words) จากการนับค่าคำในข้อความเหตุผลความจำเป็น. 39
3.36 ตัวอย่างฟีเจอร์ (Feature) ที่สร้างจากการนับคำในข้อความเหตุผลความจำเป็น..... 39
3.37 การแบ่งชุดข้อมูลสำหรับสร้างแบบจำลองและทดสอบประสิทธิภาพแบบจำลองที่ 2..... 40
3.38 ตัวอย่างชุดข้อมูล Training set และ Testing set ชุดที่ 1 แบบจำลองที่ 2..... 40
3.39 ตัวอย่างชุดข้อมูล Training set และ Testing set ชุดที่ 5 แบบจำลองที่ 2..... 41
3.40 เครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Decision Tree..... 42
3.41 เครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Random Forest..... 42
3.42 เครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Naïve Bayes..... 43
3.43 เครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Logistic Regression..... 43
3.44 เครื่องมือที่ใช้ในการทดสอบอัลกอริทึมของแบบจำลองที่ 2..... 44
3.45 ตัวอย่างชุดข้อมูลที่ใช้ในการทดสอบอัลกอริทึมของแบบจำลองที่ 2..... 44
3.46 เครื่องมือสำหรับประเมินประสิทธิภาพแบบจำลองที่ 2..... 44
3.47 ตัวอย่าง Confusion Matrix แสดงประสิทธิภาพและผลการทำนายของแบบจำลองที่ 2..... 45
4.1 เปรียบเทียบการทำความสะอาดข้อมูลโดยใช้ไลบรารี RegEx..... 47
4.2 เปรียบเทียบการตัดคำโดยใช้ engin newmm..... 47

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

4.3 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 1 ด้วยอัลกอริทึม Decision Tree (ข้อมูลชุดที่ 1)..... 49
4.4 ตัวอย่าง Decision Tree แบบจำลองที่ 1 (ข้อมูลชุดที่ 1)..... 50
4.5 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 1 ด้วยอัลกอริทึม Decision Tree (ข้อมูลชุดที่ 5)..... 50
4.6 ตัวอย่าง Decision Tree แบบจำลองที่ 1 (ข้อมูลชุดที่ 5..... 51
4.7 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 1ด้วยอัลกอริทึม Random Forest (ข้อมูลชุดที่ 1)..... 51
4.8 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 1ด้วยอัลกอริทึม Random Forest (ข้อมูลชุดที่ 5)..... 52
4.9 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 1 ด้วยอัลกอริทึม Naïve Bayes (ข้อมลชุดที่ 1)..... 52
4.10 ตัวอย่างผลประสิทธิภาพแบบจำลองที่1 ด้วยอัลกอริทึม Naïve Bayes (ข้อมูลชุดที่ 5)..... 53
4.11 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 1 ด้วยอัลกอริทึม Logistic Regression (ข้อมูลชุดที่ 1)..... 53
4.12 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 1 ด้วยอัลกอริทึม Logistic Regression ข้อมูลชุดที่ 5)..... 54
4.13 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Decision Tree (ข้อมูลชุดที่ 1)..... 54
4.14 ตัวอย่าง Decision Tree แบบจำลองที่ 2 (ข้อมูลชุดที่ 1)..... 55
4.15 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Decision Tree (ข้อมูลชุดที่ 5)..... 55
4.16 ตัวอย่าง Decision Tree แบบจำลองที่ 2 (ข้อมูลชุดที่ 5)..... 56
4.17 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 2 ด้วยอัลกอริทึม Random Forest (ข้อมูลชุดที่ 1)..... 56
4.18 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 2 ด้วยอัลกอริทึม Random Forest (ข้อมูลชุดที่ 5)..... 57
4.19 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 2 ด้วยอัลกอริทึม Naïve Bayes (ข้อมูลชุดที่ 1)..... 57
4.20 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 2 ด้วยอัลกอริทึม Naïve Bayes (ข้อมูลชุดที่ 5)..... 58
4.21 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 2 ด้วยอัลกอริทึม Logistic Regression (ข้อมูลชุดที่ 1)..... 58
4.22 ตัวอย่างผลประสิทธิภาพแบบจำลองที่ 1 ด้วยอัลกอริทึม Logistic Regression (ข้อมูลชุดที่ 5)..... 59

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

4.23 ตัวอย่างชุดข้อมูลที่ทำนายผลผิดพลาดด้วยอัลกอริทึม Random Forest ของแบบจำลองที่ 1..... 61

4.24 ตัวอย่างชุดข้อมูลที่ทำนายผลผิดพลาดด้วยอัลกอริทึม Random Forest ของแบบจำลองที่ 2..... 62

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

สำนักเทคโนโลยีสารสนเทศ สำนักงานศาลยุติธรรม มีภารกิจงานเกี่ยวกับการพิจารณา วิเคราะห์ กลั่นกรอง คำของบประมาณจัดซื้อครุภัณฑ์คอมพิวเตอร์ให้แก่หน่วยงานในสังกัดสำนักงานศาลยุติธรรมทั่วประเทศ โดยในแต่ละปีงบประมาณเมื่อหน่วยงานในสังกัดสำนักงานศาลยุติธรรมมีความจำเป็นต้องขอรับการจัดสรรงบประมาณเพื่อจัดซื้อครุภัณฑ์คอมพิวเตอร์สำหรับใช้สนับสนุนการปฏิบัติงานจะดำเนินการจัดส่งคำของบประมาณกิจกรรมจัดซื้อครุภัณฑ์ประจำปี (แผนความต้องการครุภัณฑ์คอมพิวเตอร์) มายังสำนักแผนงานและงบประมาณในรูปแบบของแบบฟอร์มเอกสารและไฟล์เอกสาร (File Type : .xls, .xlsx) ซึ่งสำนักแผนงานและงบประมาณจะดำเนินการรวบรวมและจัดส่งข้อมูลคำของบประมาณดังกล่าวให้แก่สำนักเทคโนโลยีสารสนเทศใช้พิจารณา วิเคราะห์ กลั่นกรอง โดยในแต่ละปีมีเอกสารคำของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์และเอกสารประกอบเพิ่มเติมเป็นจำนวนมาก ซึ่งก่อให้เกิดปัญหาข้อยุ่งยากในด้านการบริหารจัดการเอกสาร อาทิเช่น การจำแนก การจัดเก็บ การสืบค้น การนำข้อมูลไปใช้ประโยชน์ เป็นต้น และเกิดความล่าช้าในขั้นตอนการตรวจสอบ การพิจารณาจัดสรรงบประมาณ และการแจ้งรายงานผลการพิจารณากลับไปยังสำนักแผนงานและงบประมาณ และหน่วยงานศาลยุติธรรมที่ขอรับการจัดสรรงบประมาณ เนื่องจากต้องให้เจ้าหน้าที่คัดแยกประเภทคำของบประมาณเพื่อจัดส่งให้ในแต่ละส่วนงานที่เกี่ยวข้องดำเนินการตรวจสอบและพิจารณา ซึ่งปัญหาและข้อยุ่งยากที่เกิดขึ้นดังกล่าวมีผลกระทบต่อประสิทธิภาพการบริหารงบประมาณของหน่วยงานอีกประการหนึ่งด้วย

ด้วยความก้าวหน้าของเทคโนโลยีการวิเคราะห์ข้อมูล (Data Analytic) โดยใช้เทคนิคเหมืองข้อความ (Text Mining) ซึ่งเป็นกระบวนการที่สกัดเอาความรู้จากภาษาธรรมชาติที่มีลักษณะของข้อมูลแบบไม่มีโครงสร้าง (Unstructured Data) ทั้งนี้ การทำเหมืองข้อมูลจำนวนมากจะเป็นลักษณะของการนำเสนอรูปแบบสำหรับการทำเหมืองข้อความที่ใช้ประโยชน์ในเอกสารและใช้ความรู้ที่ได้จากการสกัดข้อความเหล่านั้นมาเพื่อปรับปรุงรูปแบบและปัญหาต่างๆ ของสิ่งนั้น

จากสภาพปัญหาข้อขัดข้อง ประกอบกับความก้าวหน้าของเทคโนโลยีการวิเคราะห์ข้อมูลข้างต้น ผู้วิจัยเห็นว่า สามารถนำมาประยุกต์ใช้เพื่อแก้ปัญหาและเพิ่มความสะดวกรวดเร็วในการดำเนินงานของหน่วยงานให้มีประสิทธิภาพมากขึ้น จึงดำเนินงานวิจัยภายใต้หัวข้อ “การเพิ่มประสิทธิภาพการจัดสรรงบประมาณด้านครุภัณฑ์คอมพิวเตอร์ด้วยการใช้เทคนิคเหมืองข้อความ (Text Mining) ” โดยนำเทคนิคเหมืองข้อความมาใช้เป็นเครื่องมือในการสกัดข้อความในคำของบประมาณจัดซื้อครุภัณฑ์คอมพิวเตอร์เพื่อจำแนกประเภทคำขอประมาณในด้านต่างๆ เช่น ด้านระบบเครือข่าย ด้านระบบงาน และด้านเครื่องคอมพิวเตอร์และอุปกรณ์ต่อพ่วง รวมไปถึงการสกัดหากลุ่มคำหรือข้อความที่เป็นเหตุปัจจัยให้หน่วยงานมีความ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

จำเป็นต้องขอรับการจัดสรรงบประมาณโดยที่กำหนดแนวทางและวิธีการใช้เทคนิคเหมืองข้อมูลให้เกิดประโยชน์ต่อการลดภาระและขั้นตอนปฏิบัติงานของบุคลากรที่เกี่ยวข้อง รวมถึงพัฒนาระบบข้อมูลที่ช่วยสนับสนุนการตัดสินใจให้แก่เจ้าหน้าที่ และผู้บริหารในกระบวนการพิจารณาจัดสรรงบประมาณให้มีประสิทธิภาพมากขึ้น ทั้งนี้ ผู้วิจัยคาดหวังเป็นอย่างยิ่งว่าผลการดำเนินงานและข้อสรุปที่ได้จากงานวิจัยจะเป็นประโยชน์ต่อการพัฒนาระบบบริหารงบประมาณของสำนักงานศาลยุติธรรม ตลอดจนสามารถนำไปเป็นต้นแบบของการนำข้อมูลที่มีอยู่ของหน่วยงานไปใช้ประโยชน์สำหรับพัฒนางานในด้านต่างๆ เพื่อสนับสนุนและขับเคลื่อนองค์กรให้บรรลุตามเป้าหมาย พันธกิจ วิสัยทัศน์ภายใต้แผนยุทธศาสตร์ศาลยุติธรรมและแผนพัฒนาดิจิทัลศาลยุติธรรมที่ต้องการปรับเปลี่ยนองค์กรสู่ดิจิทัล

1.2 วัตถุประสงค์ของการศึกษาหรือวิจัย

1.2.1 เพื่อวิเคราะห์เอกสารที่เกี่ยวข้องกับคำของบประมาณจัดซื้อครุภัณฑ์คอมพิวเตอร์ และจำแนกประเภทคำของบประมาณตามประเภทครุภัณฑ์และส่วนงานที่เกี่ยวข้อง

1.2.2 เพื่อพัฒนาระบบที่ช่วยสนับสนุนการตัดสินใจการจัดสรรงบประมาณจากการจำแนกข้อมูลเหตุผลความจำเป็นของหน่วยงานที่ขอรับการจัดสรรให้มีความสะดวกรวดเร็ว และเพิ่มประสิทธิภาพการบริหารงบประมาณ

1.2.3 เพื่อลดความซ้ำซ้อนในการจัดเก็บข้อมูล เพิ่มมูลค่าและใช้ประโยชน์จากข้อมูลที่มีการจัดเก็บไว้แล้วให้เกิดประโยชน์สูงสุดแก่การพัฒนาระบบบริหารงบประมาณ

1.3 ขอบเขตของงานวิจัย

1.3.1 พัฒนาระบบวิเคราะห์เอกสารคำของบประมาณจัดซื้อครุภัณฑ์คอมพิวเตอร์ของหน่วยงานศาลยุติธรรมทั่วประเทศเพื่อจำแนกประเภทคำของบประมาณตามประเภทครุภัณฑ์และส่วนงานที่เกี่ยวข้อง

1.3.2 พัฒนาระบบข้อมูลสนับสนุนการตัดสินใจการจัดสรรงบประมาณของสำนักงานศาลยุติธรรมโดยการวิเคราะห์ข้อมูลเหตุผลความจำเป็นของหน่วยงานที่ขอรับการจัดสรร

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 สำนักงานศาลยุติธรรมมีระบบข้อมูลที่ใช้ในการสนับสนุนการพิจารณาจัดสรรงบประมาณได้อย่างสะดวกรวดเร็ว ถูกต้อง และมีมาตรฐานเดียวกัน

1.4.2 เจ้าหน้าที่ผู้ปฏิบัติงานสามารถลดขั้นตอนและระยะเวลาในการตรวจสอบ พิจารณา วิเคราะห์กลั่นกรองการพิจารณาจัดสรรงบประมาณ

1.4.3 ผู้บริหารศาลยุติธรรมมีระบบสารสนเทศที่ใช้สนับสนุนการตัดสินใจในการบริหารจัดการงบประมาณให้มีประสิทธิภาพมากขึ้น

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

1.5 ขั้นตอนการทำวิจัย

1.5.1 การศึกษาความรู้ที่เกี่ยวข้องและวิเคราะห์ปัญหาการจัดสรรงบประมาณจัดซื้อครุภัณฑ์

วิเคราะห์ปัญหา และศึกษาขั้นตอนการจัดทำคำของบประมาณจัดซื้อครุภัณฑ์ การพิจารณาจัดสรรงบประมาณที่ดำเนินงานในปัจจุบัน

1.5.2 การเก็บรวบรวมเอกสาร

รวบรวมข้อมูลเอกสารคำของบประมาณจัดซื้อครุภัณฑ์คอมพิวเตอร์ของหน่วยงานศาลยุติธรรมทั่วประเทศตั้งแต่ปี พ.ศ. 2563 – 2565 โดยในแต่ละคำของบประมาณจะมีการระบุชนิดของครุภัณฑ์คอมพิวเตอร์ จำนวนเงินงบประมาณ และรายการเหตุผลความจำเป็นที่ขอรับการจัดสรรงบประมาณ

1.5.3 การเตรียมข้อมูล

โดยเริ่มจากการจัดเก็บข้อมูลที่ได้ลงในระบบฐานข้อมูล และนำข้อมูลที่ได้ลงในฐานข้อมูล และนำข้อมูลที่เก็บไว้มาเข้าสู่การประมวลผลทางธรรมชาติ (Natural Language Processing) กระบวนการตัดคำออกจากข้อความ (Tokenization) การแปลงคำให้เป็นรูปแบบเชิงปริมาณ (Word Embedding) ทำการสกัดคุณลักษณะ (Feature extraction) เพื่อนำไปใช้สร้างตัวแบบ (Model) สำหรับการจำแนกประเภท (Classification)

1.5.4 การสร้างตัวแบบ

ในขั้นตอนนี้เป็นการนำอัลกอริทึมที่ได้ศึกษามาทำงานกับกลุ่มตัวอย่างในการเรียนรู้ เพื่อนำมาสร้างตัวแบบ (Model) และทดสอบความเหมาะสมของตัวแบบ โดยจะเลือกแบบที่ให้ผลแม่นยำที่สุดมาใช้ในการแยกประเภทคำของบประมาณครุภัณฑ์ และเหตุผลความจำเป็น ได้แก่ อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึมนาอีฟเบย์ (Naïve Bayes) และอัลกอริทึมการถดถอยลอจิสติกส์ (Logistic Regression)

1.5.5 การประเมินผล

เป็นการประเมินผลประสิทธิภาพของตัวแบบหรือโมเดล โดยนำผลลัพธ์ที่ได้จากเทคนิคที่ใช้ในการสร้างตัวแบบมาเปรียบเทียบเพื่อวิเคราะห์หาเทคนิคที่ให้ผลได้ถูกต้องแม่นยำที่สุดเพื่อใช้จำแนกประเภทคำของบประมาณให้แก่ส่วนงานที่รับผิดชอบและจำแนกเหตุผลผลความจำเป็นเพื่อช่วยสนับสนุนผลการพิจารณาการจัดสรรงบประมาณ

1.5.6 พัฒนาต้นแบบระบบสารสนเทศที่ใช้สนับสนุนการตัดสินใจการพิจารณาจัดสรรงบประมาณ

นำข้อมูลที่ได้จากการสร้างตัวแบบจำแนกประเภทคำของบประมาณครุภัณฑ์ และเหตุผลความจำเป็นไปพัฒนาระบบช่วยการตัดสินใจการพิจารณาจัดสรรงบประมาณของเจ้าหน้าที่ที่เกี่ยวข้องโดยการออกแบบระบบโปรแกรมที่สามารถประมวลผลข้อมูลได้อัตโนมัติ

1.5.7 สรุปผลการวิจัย และปัจจัยที่ส่งผลต่อการพิจารณาจัดสรรงบประมาณ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

1.6 นิยามศัพท์

1.6.1 **เทคนิคเหมืองข้อมูล (Data Mining Technique)** คือ กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์ รวมทั้งในด้านเศรษฐกิจและสังคม การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูลจากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล ดังนี้

(1) กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง

(2) การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย

(3) การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล

(4) การวางแผนทรัพยากรขององค์กร โดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

ขั้นตอนการทำ Data Mining

Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป

Data Integration เป็นขั้นตอนการรวบรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชนิดเดียวกัน

Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้

Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน

Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่

Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล

Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ

1.6.2 **เทคนิคเหมืองข้อความ (Text Mining)** คือ เหมืองข้อความ คือ การค้นหาข้อมูลสารสนเทศที่มีประโยชน์จากภาษาธรรมชาติ (Natural Language) เป็นการสกัดความรู้เอาข้อมูลที่ไม่เคยรู้มาก่อน และสามารถนำไปใช้ประโยชน์ได้ตามวัตถุประสงค์ ข้อมูลที่มีอยู่ในโลกส่วนใหญ่เป็นข้อมูลแบบไม่มีโครงสร้าง (Unstructured) หรือเป็นข้อมูลกึ่งโครงสร้าง (Semi-Structured) ซึ่งเป็นข้อมูลที่มีหลากหลายมิติ มีความซับซ้อนและมีความไม่ชัดเจน การทำเหมืองข้อความเป็นการใช้เทคนิคในส่วนของคอมพิวเตอร์ คือ เหมืองข้อมูล (Data Mining) และส่วนของการวิเคราะห์ทางภาษามนุษย์ คือ การประมวลผลภาษาทางธรรมชาติ (Natural Language Processing) รวมถึงการใช้หลักการของการค้นคืนสารสนเทศ (Information retrieval) การเรียนรู้ของเครื่อง (Machine Learning) และสถิติ (Statistics) เข้าด้วยกัน

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

1.6.3 **การประมวลผลทางธรรมชาติ (Natural Language Processing)** เป็นเทคโนโลยีแมชชีนเลิร์นนิ่งที่ช่วยให้คอมพิวเตอร์สามารถตีความ จัดการ และทำความเข้าใจภาษามนุษย์ได้ องค์กรในปัจจุบันมีข้อมูลเสียงและข้อความจำนวนมากจากช่องทางการสื่อสารต่างๆ เช่น อีเมล ข้อความ ฟีดข่าวโซเชียลมีเดีย วิดีโอ เสียง และอื่นๆ ซึ่งสามารถใช้ซอฟต์แวร์ NLP เพื่อประมวลผลข้อมูลนี้โดยอัตโนมัติ วิเคราะห์เจตนาหรือความเชื่อมั่นในข้อความ และตอบสนองการสื่อสารของมนุษย์แบบเรียลไทม์

1.6.4 **เทคนิคการจำแนกประเภท (Classification)** เป็นกระบวนการสร้างตัวแบบ (Model) เพื่อจัดการข้อมูลให้อยู่ในกลุ่มที่เรียกว่า Class เพื่อใช้ในการทำนายข้อมูลต่อไป เช่น การจัดกลุ่มของพนักงานในบริษัทว่า ดีมาก ดี ปานกลาง ไม่ดี โดยพิจารณาจากข้อมูลที่มีอยู่ โดยมีอัลกอริทึมที่ใช้ในการจำแนกประเภท (Classifier Algorithm) อาทิเช่น Decision Tree, Naïve Bayes, Support Vector Machine และ Logistic Regression เป็นต้น

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 2
ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ระบบบริหารงบประมาณศาลยุติธรรมและสำนักงานศาลยุติธรรม [1]

โครงสร้างแผนงานงบประมาณของศาลยุติธรรมและสำนักงานศาลยุติธรรม ระบบงบประมาณของรัฐทำหน้าที่หลักในการบริหารรายจ่ายสาธารณะ 3 ประการ คือ 1) ควบคุมวินัยทางการคลัง 2) บริหารทรัพยากรให้เกิดประสิทธิภาพสูงสุด 3) จัดสรรงบประมาณเพื่อประโยชน์สูงสุดของประชาชนพลเมืองและสังคมส่วนรวม

ระบบงบประมาณของรัฐมีพัฒนาการมาหลายรูปแบบ เช่น ระบบงบประมาณ แบบแสดงรายการ (Line Item) ระบบงบประมาณแบบแผนงาน (PPB) ปัจจุบันเป็นระบบงบประมาณแบบมุ่งเน้นผลงานตามยุทธศาสตร์ (SPBB) ซึ่งมีความคล้ายคลึงกับระบบงบประมาณ แบบแผนงานที่มีการจัดโครงสร้างแผนงาน แต่มีจุดเน้นในเรื่องการกำหนดเป้าประสงค์ของหน่วยงานและวัตถุประสงค์ของแผนงาน การวัดผลผลิตและผลลัพธ์ประสิทธิภาพ และคุณภาพ โดยการจัดสรรงบประมาณไม่ได้เพียงให้ความสำคัญกับความเชื่อมโยงระหว่างแผนงานเท่านั้น แต่ยังให้ความสำคัญต่อการกำหนดเกณฑ์การวัดผลการปฏิบัติงานเพื่อพิสูจน์ความสำเร็จตามที่คาดหวัง สำนักงานศาลยุติธรรมได้คำนึงถึงการบริหารงบประมาณให้สอดคล้องกับ ระบบงบประมาณของรัฐข้างต้น จึงได้ใช้แผนยุทธศาสตร์ศาลยุติธรรมเป็นกรอบกำหนดโครงสร้างแผนงานงบประมาณมาอย่างต่อเนื่อง โดยระบบบริหารงบประมาณ ศาลยุติธรรมและสำนักงานศาลยุติธรรมฉบับนี้ สำนักงานศาลยุติธรรมได้ใช้ยุทธศาสตร์ตามแผนยุทธศาสตร์ศาลยุติธรรม พ.ศ. 2565 -2568 เป็นกรอบกำหนดโครงสร้างแผนงานงบประมาณ

กระบวนการงบประมาณ

ลำดับขั้นตอนเกี่ยวกับการกำหนดแผนความต้องการทรัพยากรสำหรับการจัดทำงบประมาณรายจ่ายประจำปี ตั้งแต่การทบทวนผลการดำเนินงานและผลการบริหารงบประมาณในปีงบประมาณที่ผ่านมา การจัดทำกรอบวงเงินงบประมาณ ในภาพรวมขององค์กร การกำหนดแผนงาน (งาน กิจกรรม และโครงการ) และเป้าหมายตามยุทธศาสตร์ขององค์กรที่สอดคล้องกับยุทธศาสตร์ชาติ นโยบาย หรือแผนแม่บทอื่นๆ ระดับชาติ การเสนอของบประมาณรายจ่ายประจำปีขององค์กร จนถึงการเสนอรัฐสภาพิจารณาอนุมัติและตราเป็นพระราชบัญญัติประกาศเป็นกฎหมาย เพื่อใช้เป็นกรอบในการบริหารงบประมาณและติดตามประเมินผลงบประมาณรายจ่ายประจำปี กระบวนการงบประมาณประกอบด้วยขั้นตอนหลักที่สำคัญ 4 ขั้นตอน ดังนี้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ขั้นตอนที่ 1 การจัดเตรียมงบประมาณ (Budget Preparation)

เป็นขั้นตอนแรกของกระบวนการงบประมาณ ซึ่งฝ่ายบริหาร คือ รัฐบาลมีหน้าที่ในการเสนองบประมาณรายจ่ายประจำปีต่อฝ่ายนิติบัญญัติเพื่อพิจารณาอนุมัติ โดยมีส่วนราชการผู้ใช้งบประมาณประกอบด้วย ส่วนราชการ รัฐวิสาหกิจ และหน่วยงานอื่นของรัฐ เป็นผู้มีบทบาทในการจัดทำงบประมาณรายจ่ายประจำปีโดยตรง และสำนักงบประมาณมีอำนาจหน้าที่ในการวิเคราะห์คำขอตั้งงบประมาณรายจ่ายของส่วนราชการ รัฐวิสาหกิจ และหน่วยงานอื่นของรัฐ และจัดทำเป็นร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปีเสนอต่อนายกรัฐมนตรี เพื่อให้คณะรัฐมนตรีพิจารณาและเสนอต่อรัฐสภาต่อไป

การจัดเตรียมงบประมาณประกอบด้วยกระบวนการสำคัญ 3 ประการ คือ การทบทวนงบประมาณ การวางแผนงบประมาณ และการจัดทำงบประมาณ ดังนี้

1) การทบทวนงบประมาณ (Budget Revision) กระบวนการทบทวนงบประมาณ มีวัตถุประสงค์หลักเพื่อวางกรอบในการปรับเปลี่ยนบทบาท ภารกิจ และแผนดำเนินงานของหน่วยงานให้สอดคล้องกับลำดับความสำคัญทางนโยบาย ศักยภาพของหน่วยงาน และสภาพแวดล้อมที่เปลี่ยนแปลงไป การทบทวนงบประมาณจะให้ความสำคัญกับองค์ประกอบต่างๆ ดังนี้

- ผลการดำเนินงานที่ผ่านมา
- ความพร้อมและศักยภาพของหน่วยงานที่รับนโยบายไปปฏิบัติ
- กฎหมายรองรับและระเบียบต่างๆ ที่เกี่ยวข้อง
- สภาพแวดล้อมทางสังคม เศรษฐกิจ และการเมืองที่เปลี่ยนแปลงไป
- นโยบายหรือยุทธศาสตร์การพัฒนาของรัฐบาลใหม่ที่รัฐบาลประกาศในปีที่ผ่านมาจนถึงปัจจุบัน
- แนวโน้มและสถิติย้อนหลังของวงเงินงบประมาณที่ได้รับจัดสรร

2) การวางแผนงบประมาณ (Budget Planning) การวางแผนงบประมาณ เป็นการสร้างความเชื่อมโยงของเศรษฐกิจในภาพรวม แผนพัฒนาการเศรษฐกิจและสังคมแห่งชาติ แผนการบริหารราชการแผ่นดิน แผนปฏิบัติราชการของหน่วยงาน เพื่อกำหนดเป็นนโยบายงบประมาณ วงเงินงบประมาณรายจ่ายประจำปี และยุทธศาสตร์การจัดสรรงบประมาณรายจ่ายประจำปี

การวางแผนงบประมาณที่สำคัญประกอบด้วย

1. การคาดการณ์ทางเศรษฐกิจและการคลังมหภาค เป็นการวิเคราะห์และประเมินผลกระทบของการใช้จ่ายภาครัฐที่มีต่อสถานการณ์ทางเศรษฐกิจ เพื่อจัดทำข้อเสนอทางเศรษฐกิจมหภาค โดยการพิจารณาร่วมกันของ 4 หน่วยงานหลัก ได้แก่ สำนักงบประมาณ สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ กระทรวงการคลัง และธนาคารแห่งประเทศไทย โดยมีการดำเนินงาน ดังนี้

การคาดการณ์แนวโน้มเศรษฐกิจทั่วไป

การกำหนดเครื่องชี้วัดระดับมหภาค เช่น GDP เงินเฟ้อ อัตราการเจริญเติบโตทางเศรษฐกิจ อัตราการว่างงาน

การวิเคราะห์ภาระงบประมาณอันเนื่องจากนโยบายสำคัญ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

2. การประมาณการรายได้

กระทรวงการคลัง โดยสำนักงานเศรษฐกิจการคลังจะทำหน้าที่รายงานประมาณการรายได้ต่อที่ประชุม 4 หน่วยงานหลัก เพื่อให้ความเห็นชอบร่วมกันและนำไปเป็นกรอบในการกำหนดวงเงินงบประมาณรายจ่ายประจำปีต่อไป

3. การประมาณการหนี้สาธารณะ

เป็นการจัดทำประมาณการภาระหนี้ภาครัฐประจำปี และประมาณการล่วงหน้า รวมทั้งการพิจารณาแผนการก่อหนี้ต่างประเทศ

4. การกำหนดนโยบายงบประมาณ และวงเงินงบประมาณรายจ่ายประจำปี

หน่วยงานหลัก 4 หน่วยงาน คือ สำนักงบประมาณ กระทรวงการคลัง สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ และธนาคารแห่งประเทศไทย จะร่วมกันกำหนดนโยบายและวงเงินงบประมาณที่เหมาะสมกับสภาพเศรษฐกิจและนโยบายของรัฐบาล ซึ่งสำนักงบประมาณจะจัดทำข้อเสนอวงเงินงบประมาณดังกล่าวเสนอต่อคณะรัฐมนตรีพิจารณาให้ความเห็นชอบก่อนที่จะจัดทำงบประมาณในขั้นรายละเอียด

3) การจัดทำงบประมาณ (Budget Formulation)

สำนักงบประมาณจะนำเสนอปฏิทินงบประมาณเพื่อให้คณะรัฐมนตรี ให้ความเห็นชอบและแจ้งส่วนราชการและรัฐวิสาหกิจ เพื่อเป็นการเตรียมความพร้อมในการจัดทำงบประมาณรายจ่ายประจำปี และเมื่อคณะรัฐมนตรีให้ความเห็นชอบนโยบายงบประมาณ ยุทธศาสตร์การจัดสรรงบประมาณ และวงเงินงบประมาณรายจ่ายประจำปีแล้ว สำนักงบประมาณจะมีหนังสือแจ้งแนวทางการดำเนินงานในการจัดทำงบประมาณรายจ่ายประจำปี และคู่มือปฏิบัติเกี่ยวกับการจัดทำคำของบประมาณรายจ่ายประจำปี

ในส่วนของหน่วยงาน เมื่อหน่วยงานระดับกระทรวงได้รับปฏิทินงบประมาณและคู่มือจัดทำคำขอจากสำนักงบประมาณแล้ว กระทรวงจะแจ้งไปยังหน่วยงานในสังกัด เพื่อจัดทำคำของบประมาณส่งให้สำนักงบประมาณพิจารณา และจัดทำเป็นร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปี พร้อมทั้งเอกสารงบประมาณเสนอต่อนายกรัฐมนตรี เพื่อนำเสนอคณะรัฐมนตรีพิจารณาและนำเสนอต่อรัฐสภา เป็นเวลาอย่างน้อย 2 เดือน ก่อนวันเริ่มปีงบประมาณนั้นตามที่กำหนดไว้ในมาตรา 15 ของพระราชบัญญัติวิธีการงบประมาณ พ.ศ. 2502 และที่แก้ไขเพิ่มเติม

สำหรับเอกสารงบประมาณที่นำเสนอต่อรัฐสภาตามพระราชบัญญัติ วิธีการงบประมาณ พ.ศ. 2502 และที่แก้ไขเพิ่มเติม ได้กำหนดให้รัฐบาลเสนอคำแถลงของนายกรัฐมนตรี พร้อมด้วยข้อมูลรายละเอียดต่างๆ เช่น ฐานะการเงินการคลังของประเทศ รายรับรายจ่ายเปรียบเทียบปีปัจจุบัน ปีที่ขอตั้งงบประมาณรายจ่ายและปีที่ล่วงมาแล้ว หนี้ของรัฐบาลที่มีอยู่และที่จะขอกู้ใหม่ เป็นต้น ซึ่งโดยทั่วไปจะนำเสนอร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปี พร้อมเอกสารประกอบ จำนวนประมาณ 12 เล่ม และเรียกเอกสารฉบับนี้ว่าเอกสารงบประมาณ (Budget Document)

ขั้นตอนที่ 2 การอนุมัติงบประมาณ (Budget Adoption) หมายถึง การพิจารณาอนุมัติร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปี พร้อมทั้งเอกสารประกอบงบประมาณซึ่งฝ่ายบริหารหรือรัฐบาล

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

เสนอต่อฝ่ายนิติบัญญัติหรือรัฐสภา ซึ่งประกอบด้วย 2 สภา คือ สภาผู้แทนราษฎร และวุฒิสภา โดยร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปี จะผ่านการพิจารณาจากสภาผู้แทนราษฎรก่อน 3 วาระ แล้วจึงเสนอวุฒิสภาพิจารณา ก่อนนำทูลเกล้าถวายพระบาทสมเด็จพระเจ้าอยู่หัวทรงลงพระปรมาภิไธยประกาศใช้เป็นกฎหมายต่อไป

กระบวนการอนุมัติงบประมาณจะเริ่มขึ้นภายหลังจากที่ร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปีผ่านการพิจารณาโดยคณะรัฐมนตรีแล้ว สำนักงบประมาณจะเตรียมเอกสารงบประมาณเพื่อให้นายกรัฐมนตรีเสนอต่อรัฐสภาเพื่อพิจารณา ซึ่งการพิจารณา ร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปีของสภาผู้แทนราษฎร ประกอบด้วย 3 วาระ ดังนี้

วาระที่ 1 เป็นการพิจารณารับหลักการแห่งร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปี รวมถึงการแต่งตั้งคณะกรรมาธิการวิสามัญพิจารณาร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปี

วาระที่ 2 เป็นการพิจารณารายละเอียดแห่งร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปีโดยคณะกรรมาธิการ และการพิจารณาเต็มสภาเรียงตามมาตรา

วาระที่ 3 เป็นการพิจารณาอนุมัติ และประธานสภาผู้แทนราษฎร จะเป็นผู้เสนอร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปีต่อวุฒิสภาเพื่อพิจารณาต่อไป

วุฒิสภาจะดำเนินการพิจารณาร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปี โดยตั้งคณะกรรมาธิการเพื่อศึกษาร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปีเช่นเดียวกัน และต้องพิจารณาให้ความ “เห็นชอบ” หรือ “ไม่เห็นชอบ” ให้แล้วเสร็จภายใน 20 วัน นับแต่วันที่ร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปีมาถึงวุฒิสภา ทั้งนี้ เว้นแต่ สภาผู้แทนราษฎรได้ลงมติให้ขยายเวลาออกไปเป็นกรณีพิเศษ หากวุฒิสภาไม่สามารถพิจารณา ให้เสร็จภายในกำหนดเวลา ให้ถือว่าวุฒิสภาได้ให้ความเห็นชอบตามร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปีนั้น

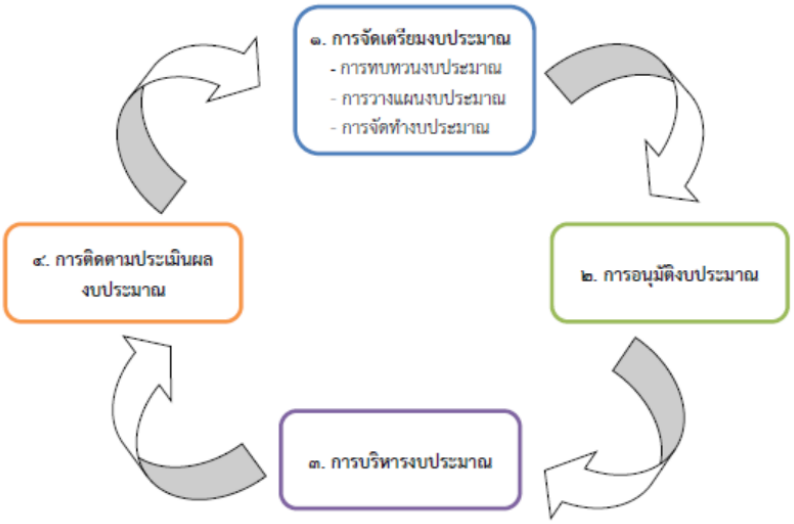
หากร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปี ได้รับการเห็นชอบจากวุฒิสภาแล้ว นายกรัฐมนตรีนำร่างพระราชบัญญัติงบประมาณรายจ่ายประจำปีขึ้นทูลเกล้าถวายเพื่อพระมหากษัตริย์ทรงลงพระปรมาภิไธย ประกาศในราชกิจจานุเบกษาใช้บังคับเป็นกฎหมายต่อไป

ขั้นตอนที่ 3 การบริหารงบประมาณ (Budget Execution) เป็นขั้นตอนการนำงบประมาณไปใช้จ่ายหรือก่อหนี้ผูกพัน เมื่อพระราชบัญญัติงบประมาณรายจ่ายประจำปีประกาศเป็นกฎหมายใช้บังคับแล้ว ส่วนราชการ รัฐวิสาหกิจ และหน่วยงานอื่นของรัฐ ผู้ใช้งบประมาณ และหน่วยงานที่เกี่ยวข้องกับการบริหารงบประมาณรายจ่ายจะต้องวางแผนการปฏิบัติงานและแผนการใช้จ่ายงบประมาณให้สอดคล้องกับเป้าหมายการให้บริการของกระทรวง ผลผลิต/โครงการ ตามวงเงินงบประมาณรายจ่ายที่ระบุ ไว้ในเอกสารประกอบพระราชบัญญัติงบประมาณรายจ่ายประจำปี ซึ่งการบริหารงบประมาณเกี่ยวข้องกับ การอนุมัติจัดสรรงบประมาณ การโอนเปลี่ยนแปลงรายการ และการกันเงินเหลื่อมปี

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ทั้งนี้ หน่วยงานจะต้องมีการกำกับ ติดตามการดำเนินงานเพื่อเร่งรัดการดำเนินงาน ตามนโยบายรัฐบาลและตามเป้าหมายการเบิกจ่ายงบประมาณที่คณะรัฐมนตรีกำหนดไว้ ตามมาตรการเร่งรัดการใช้จ่ายงบประมาณหรือมาตรการเพิ่มประสิทธิภาพการใช้จ่าย

ขั้นตอนที่ 4 การติดตามประเมินผลงบประมาณ (Budget Monitoring and Evaluation) การติดตามประเมินผลงบประมาณ หมายถึง ขั้นตอนการควบคุมติดตาม การปฏิบัติงาน และการใช้จ่ายงบประมาณของส่วนราชการ รัฐวิสาหกิจ และหน่วยงานอื่นของรัฐ ให้เป็นไปตามวัตถุประสงค์ เป้าหมายการให้บริการของกระทรวง ผลผลิต/โครงการ ที่ปรากฏ ในพระราชบัญญัติงบประมาณรายจ่ายประจำปี และเอกสารงบประมาณรายจ่ายประจำปี และตามกฎหมาย ระเบียบ และข้อบังคับที่เกี่ยวกับการใช้จ่ายเงินแผ่นดิน รวมทั้งการประเมินผลสัมฤทธิ์ของการดำเนินงานของหน่วยงานที่ผ่านมา



ภาพที่ 2.1 กระบวนการงบประมาณ

จากกระบวนการงบประมาณทั้ง 4 ขั้นตอน คือ การจัดเตรียมงบประมาณ การอนุมัติงบประมาณ การบริหารงบประมาณ และการติดตามประเมินผลงบประมาณ ที่กล่าวมาข้างต้น สำนักงานศาลยุติธรรมได้ดำเนินการตามกระบวนการดังกล่าวมาตั้งแต่ปี พ.ศ. 2543 จนถึงปัจจุบัน โดยที่ขั้นตอนการบริหารงบประมาณถือเป็นขั้นตอนที่สำคัญที่หน่วยงานในสังกัดสำนักงานศาลยุติธรรมต้องนำงบประมาณที่ได้รับอนุมัติจากคณะกรรมการ บริหารศาลยุติธรรมไปใช้จ่ายให้ถูกต้อง เหมาะสม เกิดประโยชน์สูงสุดต่อราชการ และมิให้เกิดความผิดพลาดในการทำงาน สำนักงานศาลยุติธรรมจึงได้ปรับปรุงเอกสารระบบบริหารงบประมาณศาลยุติธรรมและสำนักงานศาลยุติธรรมฉบับนี้ ให้สอดคล้องกับแผนยุทธศาสตร์ศาลยุติธรรม พ.ศ. 2565 -2568 และเป็นไปตามระเบียบต่างๆ ที่มีการปรับปรุงและแก้ไขเพิ่มเติม เพื่อให้หน่วยงานในสังกัดสำนักงานศาลยุติธรรมใช้เป็นแนวทางในการบริหารงบประมาณ ได้อย่างมีประสิทธิภาพและเป็นไปในทิศทางเดียวกัน

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

2.1.2 CRISP-DM [2]

คริป-ดี เอ็ม CRIPS-DM (Cross Reference Industry Standard for Data Mining) โดยในกระบวนการคริป-ดี เอ็ม จะประกอบด้วย 6 ขั้นตอน ดังนี้

• การทำความเข้าใจกับธุรกิจ (Business Understanding) เป็นขั้นตอนแรกในกระบวนการคริป-ดี เอ็ม ซึ่งเน้นไปที่การเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูลทางดาต้าไมน์นิง พร้อมทั้งวางแผนในการดำเนินการคร่าว ๆ

• การทำความเข้าใจข้อมูล (Data Understanding) ขั้นตอนนี้เริ่มจากการเก็บรวบรวมข้อมูล หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

• การเตรียมข้อมูล (Data Preparation) ขั้นตอนนี้เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมข้อมูลดิบให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง เช่น การแปลงข้อมูลให้อยู่ในช่วงขนาดเดียวกัน หรือการเติมข้อมูลที่ขาดหายไป เป็นต้น โดยขั้นตอนนี้จะเป็นขั้นตอนที่ใช้เวลามากที่สุดของกระบวนการคริป-ดีเอ็ม

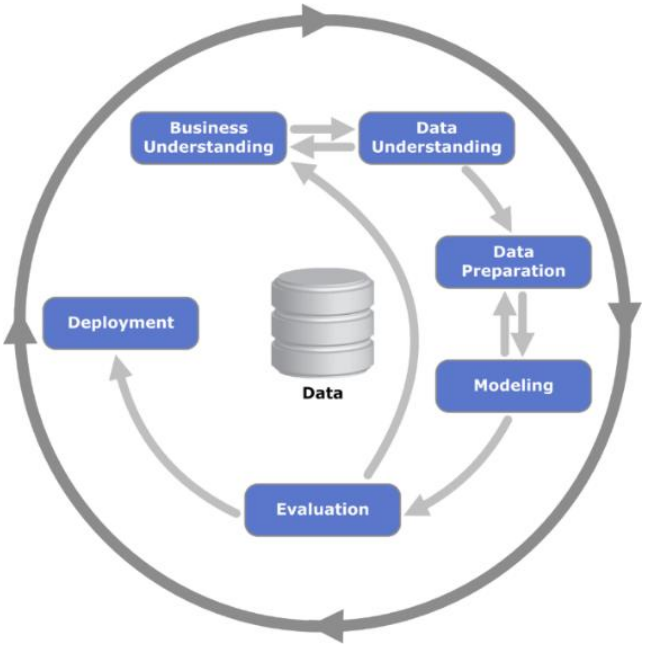
• การสร้างแบบจำลอง (Modeling) ขั้นตอนนี้จะเป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิง เช่น การจำแนกประเภทข้อมูล หรือการแบ่งกลุ่มข้อมูล ซึ่งในขั้นตอนนี้หลายเทคนิคจะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด ดังนั้นในบางครั้งอาจจะต้องมีการย้อนกลับไปที่ขั้นตอนการเตรียมข้อมูล เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย ตัวอย่างเทคนิคในการวิเคราะห์ข้อมูลต่าง ๆ เช่น การแบ่งกลุ่มข้อมูล, การหากฎความสัมพันธ์, การจำแนกประเภทข้อมูล ตัวอย่างเช่น (1) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) (2)เทคนิคนาอีฟเบย์ (Naïve Bayes) (3) วิธีการเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors) (4) เทคนิคซับพอร์ตเวคเตอร์แมชชีน Support Vector Machines (SVM)

• การประเมินผล (Evaluation) ในขั้นตอนนี้จะได้ผลการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิงแล้ว แต่ก่อนที่จะนำผลลัพธ์ที่ได้ไปใช้งานต่อไปก็จะต้องมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก หรือมีความน่าเชื่อถือมากน้อยเพียงใด ซึ่งอาจจะย้อนกลับไปยังขั้นตอนก่อนหน้าเพื่อเปลี่ยนแปลงแก้ไขเพื่อให้ได้ผลลัพธ์ตามที่ต้องการได้ สำหรับการสร้างโมเดลด้วยเทคนิคจำแนกประเภทข้อมูล มีการทดสอบประสิทธิภาพของโมเดลอยู่ 3 แบบใหญ่ คือ เซฟ-คอนซิสเทนซีเทส (Self-Consitency Test), การแบ่งข้อมูลทดสอบ (Split Test) และครอส-วาลิเดชั่นเทส(Cross-Validation Test)

• การนำไปใช้ (Deployment) ในกระบวนการทำงานของคริป-ดีเอ็มนั้นไม่ได้หยุดเพียงแค่ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้า ไมน์นิงเท่านั้น แม้ว่าผลลัพธ์ที่ได้จะแสดงถึงองค์ความรู้ที่มีประโยชน์แต่จะต้องนำองค์ความรู้ที่ได้เหล่านี้ไปใช้ได้จริงในองค์กรหรือบริษัท แต่ละขั้นตอนในรูปจะเป็นขั้นตอนที่ต่อเนื่องกัน นั่นคือขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้าซึ่งแสดงด้วยลูกศรที่เชื่อมระหว่างกล่องสี่เหลี่ยมแต่ละกล่อง ตัวอย่างเช่นเมื่อได้ผลลัพธ์จากขั้นตอนการเตรียมข้อมูลแล้วจะนำไปสร้างโมเดล

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

จำแนกประเภทข้อมูลในขั้นการสร้างแบบจำลองและหลังจากนั้นอาจจะย้อนกลับมาเปลี่ยนแปลงข้อมูลให้ถูกต้องมากขึ้นเพื่อหวังว่าจะโมเดลที่ให้ความถูกต้องมากขึ้นก็ได้เป็นต้น



ภาพที่ 2.2 ขั้นตอนในกระบวนการ CRISP-DM

ที่มา: https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

2.2.3 เหมืองข้อความ (Text Mining) [3]

การทำเหมืองข้อความเป็นกระบวนการวิเคราะห์ข้อมูลที่เป็นข้อความ เพื่อแยกประเภทจำแนกรูปแบบและความสัมพันธ์ใหม่ ๆ ที่เกิดขึ้นระหว่างข้อมูล และช่วยค้นพบเนื้อหาสาระที่ซ่อนในเอกสารต่าง ๆ เพื่อนำมาใช้ประโยชน์ และใช้คำสำคัญมาทำการสกัดคุณลักษณะเพื่อดึงคุณลักษณะของคำมาเป็นตัวแทนของเอกสารภาษาไทย โดยมีการตัดคำหยุด ซึ่งได้แก่ คำที่ไม่มีความหมายสำคัญต่อเอกสารและไม่ทำให้ใจความของเอกสารเปลี่ยน ตัวอย่างเช่น คำสรรพนามเป็นคำที่ใช้แทนคำนาม คำสันธานเป็นคำที่ทำหน้าที่เชื่อมคำกับคำ คำบุพบทเป็นคำที่ใช้เชื่อมคำหรือกลุ่มคำให้สัมพันธ์กันหลังจากนั้นเมื่อได้คุณลักษณะของคำแล้วจึงนำมาสร้างแบบจำลองในรูปแบบต่าง ๆได้ เช่น การจำแนกหมวดหมู่, การจัดกลุ่ม และการสร้างกฎความสัมพันธ์ (Association RuleMining) เป็นต้น

2.2.4 การประมวลผลภาษาธรรมชาติ **(**Naural Laguage Processing**)** [4]

การประมวลผลภาษาธรรมชาติ (NLP) เป็นศาสตร์ที่สำคัญทางด้าน Machine Learning โดยมันเป็นสาขาวิชาหนึ่งที่ประกอบด้วยองค์ความรู้จากหลากหลายแขนง อาทิ ภาษาศาสตร์ (Linguistics) วิทยาการ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

คอมพิวเตอร์ (Computer Science) ปัญญาประดิษฐ์ (Artificial Intelligence: AI) รวมถึงสถิติ (Statistics) โดยมีจุดมุ่งหมายเพื่อให้คอมพิวเตอร์สามารถทำความ “เข้าใจ” ข้อมูลที่มีลักษณะเป็นข้อความหรือคำพูดเฉกเช่นเดียวกับที่มนุษย์ที่ทำได้ ซึ่งไม่ใช่เพียงแค่เข้าใจความหมายโดยตรงของข้อความนั้น ๆ แต่ยังรวมถึงการรับรู้ถึงความหมายโดยนัย ความรู้สึกของผู้เขียน ความแตกต่างทางบริบทของภาษา รวมถึงสามารถทำการวิเคราะห์ในรูปแบบต่าง ๆ ได้อีกด้วย กระบวนการโดยทั่วไปจะประกอบไปด้วย [5]

(1) การทำความสะอาดข้อความ (Text Cleansing) ประโยคหรือข้อความในภาษาของมนุษย์ไม่ได้มีแค่ตัวอักษรหรือสระ (ในภาษาไทย)แต่จะมีเครื่องหมายหรือสัญลักษณ์ต่าง ๆ ประกอบอยู่ด้วย เช่น เครื่องหมายตกใจ (!) เครื่องหมายแอด (@) เครื่องหมายคำพูด (“ ”) เป็นต้น เครื่องหมายเหล่านี้ไม่มีความหมายในการประมวลผลภาษาธรรมชาติ และจะถูกนำออกไปจากข้อความเพื่อลดจำนวนของคำ

(2) การตัดคำในข้อความ (Tokenization) การแบ่งคำในประโยคหรือข้อความออกมาเป็นคำ ๆ หรือการกำหนดขอบเขตของคำอย่างถูกต้องตามหลักของภาษานั้นๆ เช่น ข้อความภาษาอังกฤษ “I am very happy” ถูกตัดคำออกมาได้เป็น “I”, “am”, “very”, “happy” หรือข้อความภาษาไทย “ฉันมีความสุขมาก” ถูกตัดคำออกมำได้เป็น “ฉัน”, “มี”, “ความสุข”, “มาก” เป็นต้น.

(3) การตัดส่วนท้ายของคำ (Stemming) เป็นตัดส่วนส่วนท้ายของคำแบบหยาบ ๆ ทำให้คำลดสั้นลงเหลือแต่ส่วนหน้าของคำเช่น “universe”, “university”, “universal” ได้ผลลัพธ์คือ “universe” บางครั้งผลลัพธ์ของคำ ทีผ่านการตัดส่วนท้ายอาจจะผิดจากความหมายเดิมได้ และไม่มีในดิกชันนารี ตัวอย่าง เช่น “fly”,“flying” ได้ผลลัพธ์คือ “fri” เป็นต้น

(4) การแปลงคำให้เป็นรากศัพท์ (Lemmatization) การแปลงคำต่าง ๆ ให้อยู่ในรูปของรากศัพท์ของคำนั้น ๆ วิธีนี้จะได้ผลลัพธ์ที่ดีกว่าการทำ Stemming เพราะคำที่ได้จะมีอยู่ในดิกชันนารี เช่น “is”, “am”, “are” ได้ผลลัพธ์คือ “be” เป็นต้น

(5) การหาคำสำคัญในข้อความ เทคนิคที่ชื่อว่า TF-IDF (Term Frequency - Inverse Document Frequency) (Qaiser & Ali, 2018)ซึ่งการหาคำสำคัญสามารถหาได้จากนำผลคูณของทั้งสองค่า คือ ความถี่ของคำ (TF) กับความถี่เอกสารผกผัน (IDF) และจะได้ผลลัพธ์เป็นน้ำหนักของคำสำคัญออกมาตามสูตรดังนี้

$$w_{t,d} = tf_{t,d} \times idf_t$$

เมื่อ tf = ความถี่ของคำในเอกสาร
idf = การผกผันในความถี่ของเอกสาร
w (weight) = ค่าน้ำหนักของคำ
t (term) = คำในเอกสาร
d (document) = เอกสาร

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

เทคนิค Bag of Words (BoW) [6]

โมเดลที่ใช้กันแพร่หลายในงานจัดแบ่งประเภทข้อความ Text Classification ในโมเดลของ BOW กลุ่มของคำจะถูกอธิบายด้วยกระเป๋าคำ Bag of words หรือกลุ่มรวมของคำ โดยไม่ได้คำนึงถึงหลักไวยากรณ์ ความถี่ที่พบ และลำดับของคำ โดยนำมาใช้เป็น Feature ในการเทรนตัวจัดแบ่งข้อความ Classifier

2.2.5 Clasification Technic [7]

การจำแนกประเภทข้อมูลคือกระบวนการสร้างโมเดลจำแนกประเภทข้อมูลเพื่อทำนายกลุ่มของข้อมูลใหม่ตัวอย่างของกลุ่ม เช่น กลุ่มของลูกค้าที่ซื้อคอมพิวเตอร์-ไม่ซื้อคอมพิวเตอร์ กลุ่มของลูกค้าที่ฐานะดี-ปานกลาง-แย่ กลุ่มของการผลิตสินค้า ผ่านเกณฑ์-ไม่ผ่านเกณฑ์ ในที่นี้คำว่ากลุ่มจะเรียกว่าคลาส (Class) ของข้อมูล ซึ่งในคลาสเดียวกันนั้นจะต้องมีข้อมูลที่มีความเหมือนหรือคล้ายคลึงกันมากกว่าข้อมูลที่อยู่ในที่แตกต่างกัน การสร้างโมเดลจำแนกประเภทข้อมูลจะเกิดขึ้นมาจากการหาความสัมพันธ์ของข้อมูลในฐานข้อมูลขนาดใหญ่ โดยข้อมูลทั้งหมดจะมีการแบ่งออกเป็น 2 กลุ่ม คือกลุ่มข้อมูลเรียนรู้ เป็นชุดข้อมูลที่มีบทบาทในการสร้างโมเดลจำแนกประเภทข้อมูลขึ้นมา และมีกลุ่มข้อมูลทดสอบเป็นชุดข้อมูลประเมินความถูกต้องของโมเดลจำแนกประเภทข้อมูลโมเดลจำแนกประเภทข้อมูลได้ถูกนำมาประยุกต์ใช้งานหลาย ๆ ด้าน ไม่ว่าจะเป็นการวิเคราะห์หุ้น เพื่อหาว่าหุ้นแต่ละบริษัทมีคุณภาพเป็นอย่างไร เมื่อมีปัจจัยที่เกี่ยวข้อง ไม่ว่าจะเป็นการเติบโตของรายได้ ความสามารถในการควบคุมต้นทุน ความผันผวนของรายได้และกำไร และผู้บริหาร หรือจะเป็นการพยากรณ์อากาศ การจัดสรรกฎหมายที่เหมาะสมในการพิจารณาคดีความการจัดการความสัมพันธ์ของลูกค้าและอื่นๆ

อัลกอริทึมสำหรับการจำแนกประเภท เช่น (1) ต้นไม้ตัดสินใจ (Decition Tree) (2) นาอีฟเบย์ (Naïve Bayes) (3) การถดถอยโลจิสติก (Logistic Regression)

ต้นไม้ตัดสินใจ (Decition Tree) เป็นอัลกอริทึมหนึ่งในกลุ่มของการจำแนกประเภท โดยจะมีลักษณะเป็นการทำงานเหมือนโครงสร้างต้นไม้ ที่แต่ละโหนดแสดงลักษณะประจำ ที่ใช้ทดสอบข้อมูล แต่ละกิ่งแสดงผลในการทดสอบและลีฟโหนด แสดงกลุ่มหรือคลาสที่กำหนดไว้ ซึ่งต้นไม้การตัดสินใจนี้ง่ายต่อการเข้าใจและปรับเปลี่ยนเป็นกฎการจำแนก อัลกอริทึมพื้นฐานของการสร้างต้นไม้ตัดสินใจคือ อัลกอริทึมละโมบ (Greedy Algorithm) โดยจะสร้างต้นไม้จากบนลงล่างแบบวนซ้ำ ด้วยวิธีการแบ่งปัญหาใหญ่เป็นปัญหาย่อย ๆ ซึ่งรูปแบบของต้นไม้จะประกอบด้วยโหนดแรกสุด (Root Node) ก็จะแตกออกเป็นโหนดลูกและที่โหนดลูกก็จะมีลูกของตัวเองซึ่งโหนดในระดับสุดท้ายจะเรียกว่าลีฟโหนด (Leaf Node) แทนเซตของคำตอบทั้งหมดที่เป็นไปได้

ในการสร้างต้นไม้การตัดสินใจ เงื่อนไขที่สำคัญคือ การคัดเลือกแอตทริบิวต์ ที่เหมาะสมในแต่ละตำแหน่งของโหนดโดยทั่วไปใช้ อินฟอร์เมชันเกน, เกนเรโช และ เกนอินเด็กซ์

นาอีฟเบย์ (Naïve Bayes) [8] เป็นการจำแนกประเภทโดยใช้หลักสถิติในการพยากรณ์ความน่าจะเป็นของสมาชิก โดยใช้ทฤษฎีของเบย์ ซึ่งเป็นการเรียนรู้เพิ่มได้ โดยตัวอย่างที่ได้มาถูกนำมาปรับเปลี่ยนการแจกแจงซึ่งมีผลต่อการเพิ่ม/ลดความน่าจะเป็น ทำให้มีการเรียนรู้ที่เปลี่ยนไป วิธีการนี้ตัวแบบจะถูก

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ปรับเปลี่ยนไปตามตัวอย่างใหม่ที่ได้ โดยผนวกกับความรู้เดิมที่มี และมีการทำนายค่าคลาสเป้าหมายของตัวอย่างโดยใช้ความน่าจะเป็นมากที่สุดของทุกสมมติฐาน จากทฤษฎีของเบย์ สามารถคำนวณความน่าจะเป็นของสมมติฐานต่าง ๆ โดยใช้สมการที่ (1) ดังนี้

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)} \quad (1)$$

โดยที่ D แทนข้อมูลที่นามาใช้ในการคานวณการแจกแจงความน่าจะเป็น Posteriori Probability ของสมมติฐาน h คือ P(h|D) ตามทฤษฎี

P(h) คือ ความน่าจะเป็นก่อนหน้าของสมมติฐาน h

P(D) คือ ความน่าจะเป็นก่อนหน้าของชุดข้อมูลตัวอย่าง D

P(h|D) คือ ความน่าจะเป็นของ h เมื่อรู้ D

P(D|h) คือ ความน่าจะเป็นของ D เมื่อรู้ h

การถดถอยโลจิสติก (Logistic Regression) [9] เป็นเทคนิคที่ใช้ในการวิเคราะห์ข้อมูลเชิงคุณภาพ (Qualitative data) ซึ่งลักษณะของข้อมูลนั้นจะเป็นประเภทหมวดหมู่การวิเคราะห์การถดถอยโลจิสติกมักถูกใช้กับงานด้านการจำแนกกลุ่มข้อมูล เพื่อทำนายความน่าจะเป็นของการเกิดของเหตุการณ์ต่าง ๆ โดยสามารถแบ่งประเภทการวิเคราะห์ได้ 3 ประเภทหลักๆ ดังนี้ (1) การวิเคราะห์การถดถอยโลจิสติกแบบไบนารี(Binary Logistic Regression Analysis) คือการวิเคราะห์ข้อมูลโดยที่ค่าของคำตอบ(Labels) มีความน่าจะเป็นเพียง 2 กลุ่ม (2) การวิเคราะห์การถดถอยโลจิสติกพหุกลุ่ม(Multinomial Logistic Regression Analysis) เป็นการวิเคราะห์ข้อมูลโดยที่ค่าของคำตอบมีความน่าจะเป็นมากกว่า 2 กลุ่ม ซึ่งเป็นค่าที่ไม่สามารถจัดลำดับได้ (3) การวิเคราะห์การถดถอยโลจิสติกเชิงลำดับ (Ordinal Logistic Regression) เป็นการวิเคราะห์ข้อมูลโดยที่ค่าของคำตอบมีความน่าจะเป็นมากกว่า 2 กลุ่ม ซึ่งเป็นค่าที่สามารถจัดลำดับได้ โดยหลักการทำงานจะต้องมีการกำหนดขอบเขตการตัดสินใจ (Decision Boundaries) เพื่อระบุว่าข้อมูลที่ทำนายควรจะอยู่กลุ่มใด จากนั้นจะทำการประมาณค่าสัมประสิทธิ์การถดถอยด้วยเทคนิคต่าง ๆ โดยเทคนิคที่นิยมได้แก่ Maximum Likelihood เป็นต้น

การวัดประสิทธิภาพของอัลกอริทึม [4]

(1) Accuracy คือการวัดค่าความถูกต้องโดยรวมของระบบระหว่างค่าจริงและค่าการทำนาย ถ้าหากค่า Accuracy มีค่ามาก นั่นหมายถึงค่าการทำนายนั้นสามารถทำนายได้ถูกต้องใกล้เคียงกับค่าจริง ดังสมการที่ (2)

$$\text{Accuracy} = \frac{TP+TN}{TP + TN + FP + FN} \quad (2)$$

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

จากสมการ

Accuracy คือ ค่าความถูกต้อง

TP คือ ค่าการทำนายที่ทำนายว่าจริงซึ่งตรงกับค่าจริง

TN คือ ค่าการทำนายที่ทำนายว่าไม่จริงซึ่งตรงกับค่าจริง

FP คือ ค่าการทำนายที่ทำนายว่าไม่จริงซึ่งไม่ตรงกับค่าจริง

FN คือ ค่าการทำที่ทำนายว่าจริงซึ่งไม่ตรงกับค่าจริง

(2) Precision คือการวัดค่าความแม่นยำ โดยวัดจากความซ้ำเดิมของค่าการทำนายที่ทำนายได้ถูกต้องตรงกับค่าจริง หากค่า Pricision มีค่ามาก นั่นหมายถึงค่าการทำนายนั้นสามารถทำนายได้แม่นยำใกล้เคียงกับค่าจริง ดังสมการที่ (3)

$$\text{Precision} = \frac{\mathbf{TP}}{\mathbf{TP + FP}} \tag{3}$$

จากสมการ

Precision คือ ค่าความแม่นยำ

TP คือ ค่าการทำนายที่ทำนายว่าจริงซึ่งตรงกับค่าจริง

FP คือ ค่าการทำนายที่ทำนายว่าไม่จริงซึ่งไม่ตรงกับค่าจริง

(3) Recal คือการวัดค่าความครบถ้วน ซึ่งหมายถึงอัตราส่วนการวัดค่าการทำนายที่ทำนายได้ถูกต้องตรงกับค่าจริงจากจำนวนค่าจริงทั้งหมด หากค่า Recal มีค่ามาก นั่นหมายถึงค่าการทำนายนั้นสามารถทำนายได้อย่างถูกต้องครบถ้วนใกล้เคียงค่าจริง ดังสมการที่ (4)

$$\text{Recal} = \frac{\mathbf{TP}}{\mathbf{TP + FN}} \tag{4}$$

จากสมการ

Precision คือ ค่าความแม่นยำ

TP คือ ค่าการทำนายที่ทำนายว่าจริงซึ่งตรงกับค่าจริง

FN คือ ค่าการทำที่ทำนายว่าจริงซึ่งไม่ตรงกับค่าจริง

(4) Macro F1 คือ การวัดค่าประสิทธิภาพโดยเฉลี่ยของค่า F1 โดยที่ค่า F1 คือการวัดค่าเฉลี่ยการทำนายระหว่างค่า Precision และค่า Recall ดังสมการที่ (5)

$$\text{Macro F1} = \left(\frac{\mathbf{2 * Precision * Recall}}{\mathbf{Precision + Recall}}\right)/\mathrm{N} \tag{5}$$

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

จากสมการ

Micro F1	คือ ค่าประสิทธิภาพโดยเฉลี่ยของค่า F1
Precision	คือ ค่าความแม่นยำ
Recall	คือ ค่าความครบถ้วน

(5) Confusion Matrix คือ การเปรียบเทียบประสิทธิภาพของอัลกอริทึมระหว่างค่าการทำนายและค่าจริงในรูปแบบดังนี้

	Actually Positive	Actually Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

จากตาราง

Predicted Positive	คือ ค่าการทำนายเชิงบวก
Predicted Negative	คือ ค่าการทำนายเชิงลบ
Actually Positive	คือ ค่าจริงเชิงบวก
Actually Negative	คือ ค่าจริงเชิงลบ
TP	คือ ค่าการทำนายที่ทำนายว่าจริงซึ่งตรงกับค่าจริง
TN	คือ ค่าการทำนายที่ทำนายว่าไม่จริงซึ่งตรงกับค่าจริง
FP	คือ ค่าการทำนายที่ทำนายว่าไม่จริงซึ่งไม่ตรงกับค่าจริง
FN	คือ ค่าการทำที่ทำนายว่าจริงซึ่งไม่ตรงกับค่าจริง

2.2 งานวิจัยที่เกี่ยวข้อง

เฉลิมชัย พิเดช, (2563) ประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing) วิทยาการข้อมูล หลักการเรียนรู้ของเครื่อง เพื่อช่วยในการแก้ปัญหาการแยกประเภทตั๋วงาน โดยการวิเคราะห์คำสำคัญ และใช้เครื่องมือการตัดคำภาษาไทยในขั้นตอนการประมวลผลภาษาธรรมชาติที่สามารถให้ผลทดลองมีความถูกต้องสูงถึง 91 เปอร์เซ็นต์

วทัญญู ชูประจิตต์, (2562) เสนอวิธีการจำแนกประเภท และวิธีการคัดเลือกคุณลักษณะสำหรับการสนับสนุนการวิเคราะห์ความรู้สึกต่อสถานที่ท่องเที่ยวไทยจากข้อมูลการแสดงความคิดเห็นต่อสถานที่ท่องเที่ยวไทยที่เก็บรวบรวมจากสื่อสังคมออนไลน์ภายในประเทศไทยการจำแนกประเภทใช้กระบวนการคัดเลือกคุณลักษณะแบบหาความถี่คำและสร้างตัวจำแนกประเภทด้วย 4 อัลกอริทึม คือ วิธีการหาเพื่อนบ้านใกล้สุด

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

(K-Nearest Neighbors:KNN) , ต้นไม้ตัดสินใจ (Decision Tree) , นาอีฟเบย์ (Naïve Bayes) และ ซัพพอร์ตเวกเตอร์แมชชีน (Support Vecter Machine)

ขวัญตา ศิลปะไพบูลย์พานิช, (2562) เสนอการใช้เทคโนโลยีการเรียนรู้ของเครื่องในการทำนายผลการเรียนรายวิชาคณิตศาสตร์ วิทยาศาสตร์ และภาษาอังกฤษ โดยงานวิจัยนี้มุ่งหวังที่จะศึกษาและเปรียบเทียบสมรรถนะของแบบจำลองการเรียนรู้ของเครื่องจำนวน 5 แบบจำลอง ได้แก่ Random Forest, Logistic Regression, Support Vector Machine, K-Nearest Neighbor และ Extreme Gradient Boosting โดยใช้ข้อมูลผลการเรียนในอดีตของนักเรียนระดับชั้นมัธยมศึกษาตอนต้น ซึ่งผลการทดลองพบว่าแบบจำลองที่มีประสิทธิภาพเฉลี่ยสะสมดีที่สุดในการท านายผลการเรียนรายวิชาคณิตศาสตร์ คือแบบจำลอง K-Nearest Neighbor ที่ค่า accuracy 69.25% macro f1 67.00%

พรรณนาภรณ์ เกตุภู่พงษ์, (2561) นำเสนอแบบจำลองสำหรับจำแนกประเภทโรคจากอาการ โดยการประยุกต์ใช้การทำเหมืองข้อความ เพื่อช่วยแพทย์ในการวินิจฉัยโรคและจำแนกรหัสไอซีเทนซีเอ็มได้ด้วยข้อมูลอาการของผู้ป่วย ซึ่งการสร้างแบบจำลองในงานวิจัยนี้จะเลือกใช้ตัวจำแนกประเภทที่นิยมใช้ในการทำเหมืองข้อความ ได้แก่ ต้นไม้ตัดสินใจ การเรียนรู้เบส์อย่างง่าย ซัพพอร์ตเวกเตอร์แมชชีน และโครงข่ายประสาทเทียม มาเปรียบเทียบกันโดยใช้ระยะเวลาที่ใช้ในการสร้างแบบจำลองระยะเวลาที่แบบจำลองใช้ในการทำนายกราฟเส้นโค้งอาร์โอซี อัตราผลบวกจริง อัตราผลบวกเท็จ ค่าความเที่ยง และค่าความแม่นมาเป็นตัวชี้วัด ซึ่งผลลัพธ์ที่ได้พบว่าการใช้โครงข่ายประสาทเทียมเป็นนตัวจำแนกประเภทในการสร้างแบบจำลองมีความเหมาะสมที่สุดสำหรับงานวิจัยนี้ เนื่องจากให้อัตราผลบวกจริงสูงสุดที่ร้อยละ 89.03 และมีพื้นที่ใต้เส้นโค้งของกราฟเส้นโค้งอาร์โอซีมากที่สุด

สันติ สุขเกษม, (2561) นำเสนอวิธีการวิเคราะห์รูปแบบความคิดเห็นของข้อมูลโรงแรม โดยใช้เทคนิคการตัดคำแบบผมผสานเพื่อวิเคราะห์รูปแบบความคิดเห็นของข้อมูลโรงแรม และเพื่อประเมินประสิทธิภาพการวิเคราะห์ความคิดเห็นโดยการจำแนกแบบตัดสินใจต้นไม้ แบบนาอีฟ เบย์ แบบเคเนียเรสเนเบอร์ โดยรวบรวมข้อมูลจากเว็บไซต์อโกด้าเป็นข้อมูลการแสดงความคิดเห็นของผู้ใช้บริการที่เข้าพักโรงแรม นำมาวิเคราะห์ความรู้สึกโดยหลักการวิธีการตัดคำแบบผสมผสาน

Hindawi Computational Intelligence and Neuroscience, (2022), บทความนี้นำเสนอระบบการจำแนกข้อความโดยใช้กระบวนการ NLP และเปรียบเทียบผลการสร้างแบบจำลองด้วยเทคนิคการจำแนกข้อความ (Text Classification) โดยใช้วิธีการกำหนดคุณลักษณะฟีเจอร์ด้วยเทคนิคต่างๆ ได้แก่ Standard Feature Selection Methods (ประกอบด้วย Filter Methods, Wraper Methods, Embeded Methods) และText Feature Extraction Methods

Yih-Shan Sheu, (2021), งานวิจัยนี้นำเสนอการใช้เทคนิคเหมืองข้อความและเทคโนโลยีการวิเคราะห์เพื่อเพิ่มประสิทธิภาพการวางแผนและการวิเคราะห์ทางการเงิน โดยมีวัตถุประสงค์เพื่อจัดการข้อมูลที่

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

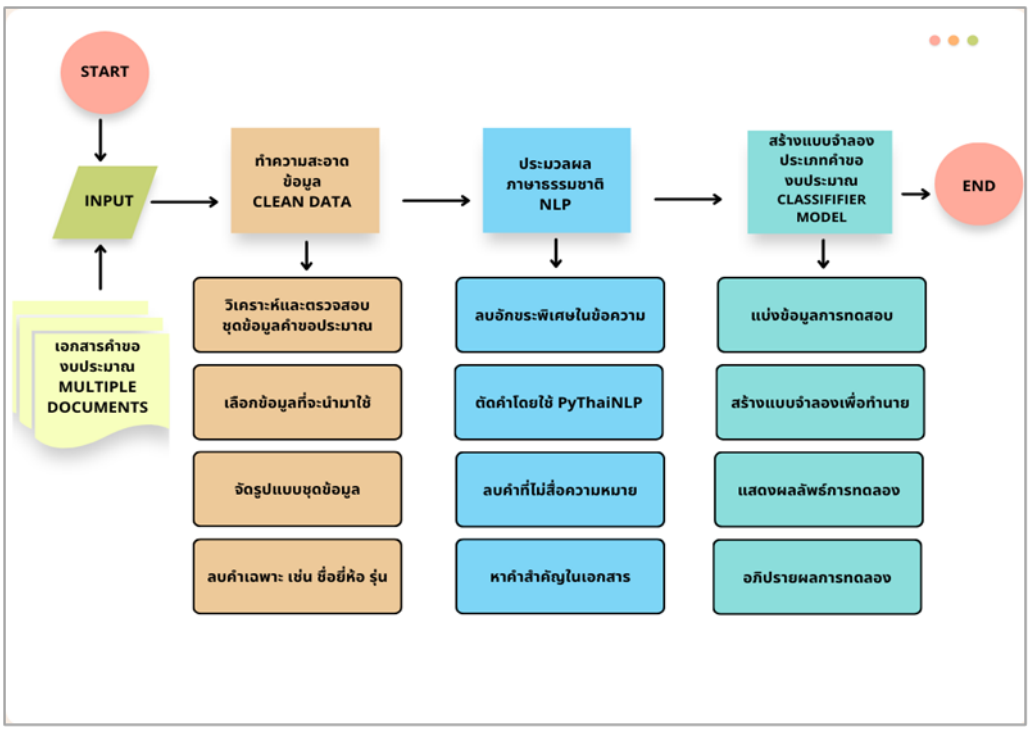
ไม่มีโครงสร้างสำหรับค้นหาปัจจัยที่ส่งกระทบต่อประสิทธิภาพทางการเงิน และใช้เทคโนโลยีแมชชีนเลิร์นนิ่งสำหรับวิเคราะห์ข้อมูลเพื่อทำนายแนวโน้มทางการตลาดให้มีประสิทธิภาพมากขึ้น

Witchapong Darontham, (2018), Simple Thai text preprocessing using Python บทความนี้นำเสนอขั้นตอนการเตรียมข้อมูลประเภทข้อความภาษาภาษาไทย (Text) โดยใช้ภาษา Python ซึ่งเริ่มตั้งแต่ขั้นตอนการทำความสะอาดข้อมูล (clean text) การตัดคำ (tokenize) สร้างคลังคำศัพท์จากข้อความทั้งหมด (create bag of words) และสร้างฟีเจอร์ด้วยการนำคำ (count words) และสร้างฟีเจอร์ด้วย TFIDF

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 3
วิธีวิจัย

ผู้วิจัยได้กำหนดขอบเขตงานวิจัยโดยนำเสนอแบบจำลองสำหรับช่วยตัดสินใจพิจารณาคำของบประมาณของหน่วยงาน ประกอบด้วย แบบจำลองที่ 1 จำแนกส่วนงานที่เกี่ยวข้องกับคำของบประมาณ และแบบจำลองที่ 2 พิจารณา “อนุมัติ” หรือ “ไม่อนุมัติ” คำของบประมาณจากเหตุผลความจำเป็น ซึ่งมีรายละเอียดขั้นตอนการดำเนินงานวิจัย ดังนี้ รายละเอียดปรากฏตามภาพที่ 3.1

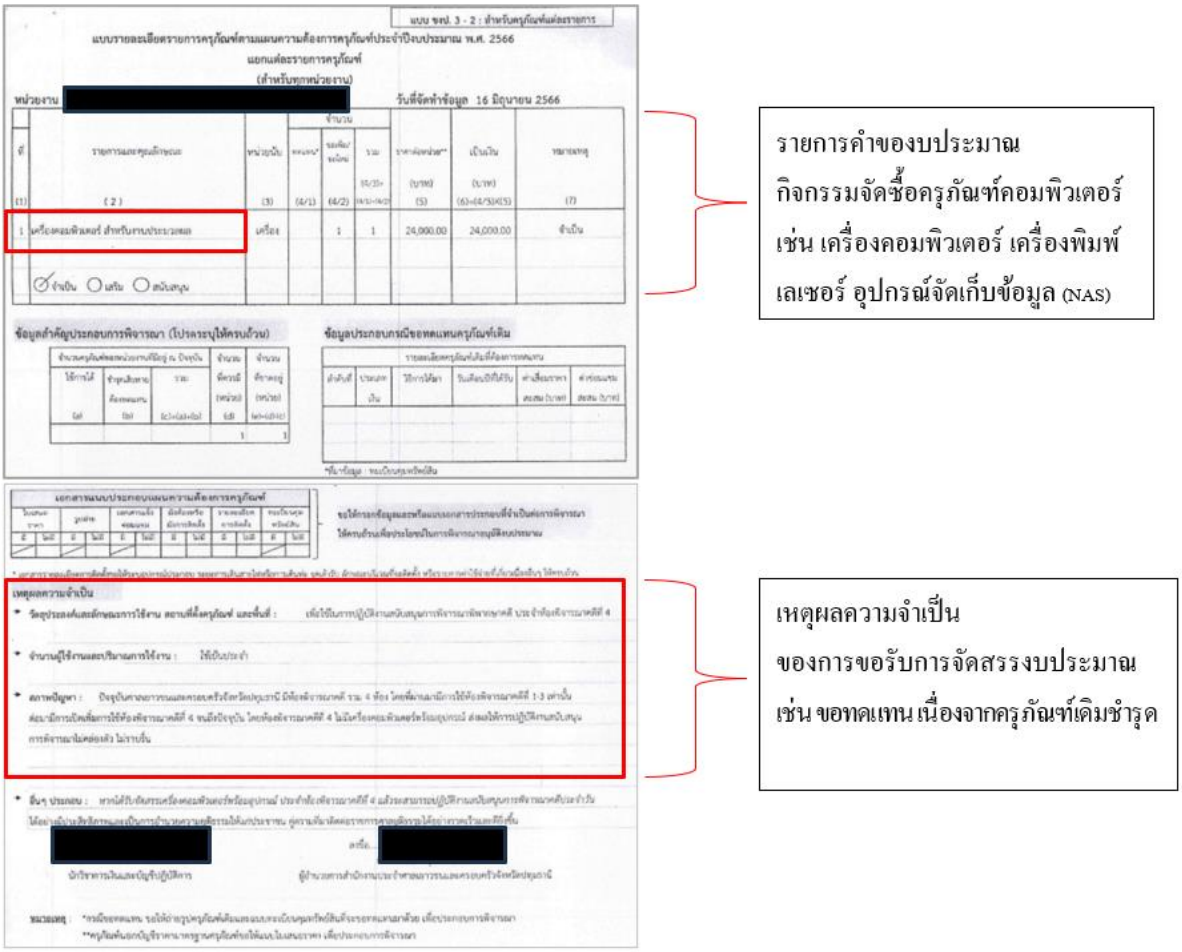


ภาพที่ 3.1 ภาพรวมวิธีดำเนินการวิจัย

3.1 แบบจำลองที่ 1 จำแนกส่วนงานที่เกี่ยวข้องกับคำของบประมาณ

3.1.1 ศึกษาวิเคราะห์ระบบการของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์ของหน่วยงานในสังกัดสำนักงานศาลยุติธรรม และวิเคราะห์ปัญหาของหน่วยงานพบว่า เอกสารคำของบประมาณของหน่วยงานจัดเก็บในรูปแบบของเอกสารเป็นจำนวนมาก ทำให้เกิดข้อยุ่งยากในการบริหารจัดการจัดเก็บเอกสารและการตรวจสอบข้อมูล อาทิเช่น (1) การจำแนกประเภทเอกสารคำของบประมาณการจัดซื้อครุภัณฑ์คอมพิวเตอร์เพื่อดำเนินการวิเคราะห์และพิจารณผลการจัดสรรงบประมาณให้แก่หน่วยงานที่เกี่ยวข้องมีความล่าช้าเนื่องจากต้องดำเนินการคัดแยกโดยบุคลากร (2) การตรวจสอบและการประมวลผลงบประมาณการจัดซื้อครุภัณฑ์คอมพิวเตอร์ของหน่วยงานจำแนกตามปีงบประมาณและรายการครุภัณฑ์คอมพิวเตอร์ขาด

ประสิทธิภาพ ไม่สามารถดำเนินการด้วยระบบโปรแกรม ซึ่งรูปแบบเอกสารคำของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์ มีรายละเอียดปรากฏดังรูปภาพที่ 3.2



ภาพที่ 3.2 เอกสารคำของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์

3.1.2 การรวบรวมเอกสารข้อมูล

จากการวิเคราะห์และตรวจสอบข้อมูลคำของบประมาณตามหัวข้อ 3.1.1 ผู้วิจัยได้จัดเก็บข้อมูลสำหรับนำไปดำเนินการตามขั้นตอนการวิจัย โดยเปลี่ยนรูปแบบคำของบประมาณจากเดิมที่จัดเก็บในรูปแบบเอกสารเปลี่ยนเป็นข้อมูลที่อยู่ในรูปแบบข้อมูลอิเล็กทรอนิกส์ (ไฟล์นามสกุล xlsx, csv) ซึ่งเอกสารคำของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์ที่ผู้วิจัยดำเนินการรวบรวมอยู่ในระหว่างปีงบประมาณ พ.ศ. 2563 – 2565 จำนวนทั้งสิ้น 717 ชุด และมีจำนวนคอลัมน์ 8 คอลัมน์ รายละเอียดดังภาพที่ 3.3 และตารางที่ 3.1 ที่ปรากฏด้านล่างนี้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

req_id	req_year	region_name	office_name	req_item	req_detail	group	approve
322	2563	ศาลสูง	ศาลอุทธรณ์ภาค 8	เครื่องพิมพ์ Multifunction ชนิดเลเ	ขอใหม่ จำนวน 1 เครื่อง เพื่อให้ระบบงานธุรการ	support	yes
331	2563	ภาค 4	ศาลแขวงอุดรธานี	เครื่องสแกนลายนิ้วมือ	ขอทดแทน จำนวน 1 เครื่อง เนื่องจากเครื่องสแก	support	yes
335	2563	ศาลสูง	ศาลอุทธรณ์คดีชำนัญพิเศษ	เครื่องคอมพิวเตอร์ Server ชนิด Tow	ขอใหม่ จำนวน 1 เครื่อง ศาลอุทธรณ์คดีชำนัญพิ	network	yes
337	2563	ส่วนกลาง	ศูนย์วิทยบริการศาลยุติธรรม	กระดานอิเล็กทรอนิกส์ (e-Board)	ขอใหม่ จำนวน 1 ชุด เพื่อจัดเก็บข้อมูลของพิพิธ	app	yes
339	2563	ภาค 7	ศาลจังหวัดเพชรบุรี	อุปกรณ์ประชุมทางไกลผ่านจอภาพ	ขอทดแทน จำนวน 1 ชุด เนื่องจากอุปกรณ์ชุดคว	network	yes
341	2563	ภาค 2	ศาลเยาวชนและครอบครัวจังห	เครื่องพิมพ์ชนิด Dot Matrix Printer	ขอทดแทน จำนวน 1 เครื่อง เพื่อใช้ในการออกใบ	support	yes
342	2563	ศาลชำนัญพิเศษ	ศาลแรงงานภาค 3 (นครราชส	เครื่องคอมพิวเตอร์สำหรับงานประมวล	ขอเพิ่ม จำนวน 3 เครื่อง เนื่องจากเครื่องคอมพิวเ	support	no
343	2563	ภาค 2	ศาลจังหวัดสระแก้ว	เครื่องพิมพ์ชนิด Dot Matrix แบบแ	ขอทดแทน จำนวน 1 เครื่อง เนื่องจากเครื่องพิมพ์	support	yes
344	2563	ภาค 8	ศาลเยาวชนและครอบครัวจังห	เครื่องคอมพิวเตอร์ สำหรับประมวลผล	ขอเพิ่ม จำนวน 1 เครื่อง เนื่องจากปัจจุบันเครื่องค	support	no
345	2563	ส่วนกลาง	สถาบันพัฒนาข้าราชการฝ่า	Zoom Meeting BIZ Plus	ขอใหม่ จำนวน 1 ชุด เนื่องจากสถานการณ์ระบา	network	yes
346	2563	ส่วนกลาง	สำนักการแพทย์	เครื่องคอมพิวเตอร์สำหรับประมวลผล	ขอเพิ่ม จำนวน 2 ชุด คณะกรรมการบริหารศาลยุ	support	no
347	2563	ภาค 9	ศาลจังหวัดตรัง	เครื่องพิมพ์ชนิด Dot Matrix Printer	ขอใหม่ จำนวน 2 เครื่อง ปัจจุบันส่วนช่วยพิจารณ	support	yes

ภาพที่ 3.3 ข้อมูลคำของบประมาณในรูปแบบ Excel File

ตารางที่ 3.1 ความหมายของข้อมูลคำของบประมาณ

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

3.1.3 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 3.4

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

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

ภาพที่ 3.4 ฟีเจอร์และลาเบลที่นำไปใช้สร้างแบบจำลองที่ 1

โดยขั้นตอนในการเตรียมข้อมูลประเภท Text ซึ่งมีรายละเอียดดังนี้

(1) ลบคำเฉพาะ (Text Cleaning) เช่น ชื่อยี่ห้อ รุ่น เช่น “TS-4538U-2G” เป็นต้น และลบเครื่องหมายต่างๆ ที่เราไม่ต้องการ (!€@%#*&~) ซึ่งคำเหล่านี้ไม่สื่อความความหมายและมีผลต่อการตัดคำในกระบวนการตัดคำ ซึ่งจะทำให้จำนวนคำที่ได้มีจำนวนมาก คำเหล่านี้จะต้องถูกลบออกจากชุดข้อมูลเพื่อลดจำนวนของคำที่จะนำไปใช้เป็นคุณลักษณะฟีเจอร์ (Features) ในการสร้างแบบจำลองทำนาย ทั้งนี้ในขั้นตอนนี้กรองเอาเฉพาะตัวอักษรและสระที่เป็นภาษาไทย และตัวอักษรภาษาอังกฤษเท่านั้น เครื่องมือที่ใช้คือไลบรารีที่ชื่อว่า RegEx (Reggular Expression) ซึ่งกำหนดรูปแบบได้เป็น [^\u0E00-\u0E7Fa-zA-Z']|['']|[|] และเขียนด้วยภาษาไพธอน (Python) ดังที่แสดงไว้ในภาพที่ 3.5 ส่วนความหมายของเครื่องหมายอธิบายไว้ในตารางที่ 3.2 โดยข้อความรายการคำของบปะมาณที่ทำความสะอาดแล้ว แสดงไว้ในภาพที่ 3.6

```
# regular expression
def preprocess_text(raw_text):
  pattern = re.compile(r"[^\u0E00-\u0E7Fa-zA-Z' ]|['']|[|]|[?]|[.]|[,]|[;]|[:]|[!]|[ฯ]|[ๆ]|")
  cleaned_text= re.sub(pattern,'',raw_text)
  return cleaned_text
df['cleaned_text_item'] = df['req_item'].apply(preprocess_text)
df['cleaned_text_item']
```

ภาพที่ 3.5 การกรองข้อความด้วย Regular Expression

ตารางที่ 3.2 ความหมายของ Regular Expression

นิพจน์ทั่วไป	ความหมายของนิพจน์ทั่วไป
[^]	ลบกลุ่มของคำที่ตรงในวงเล็บออก
\u0E00-\u0E7F	Unicode แทนอักษรและสระทั้งหมดของภาษาไทย
a-zA-Z	อักษรภาษาอังกฤษ ตัวเลข ตัวใหญ่ ทั้งหมด
[]	ช่องว่าง (Space)
|	เงื่อนไข “หรือ” เช่น ใช่|ไม่ใช่ “ใช่ หรือ ไม่ใช่” เป็นต้น
[']	ตัดเครื่องหมายอัญประกาศ
[|]	ตัดเครื่องหมายเส้นตรง
[?]	ตัดเครื่องหมายปรัศนี
[.]	ตัดเครื่องหมายมหัพภาค
[,]	ตัดเครื่องหมายจุลภาค
[:]	ตัดเครื่องหมายทวิภาค
[!]	ตัดเครื่องหมายอัศเจรีย์
[ฯ]	ตัดเครื่องหมายไปยาลน้อย
[ๆ]	ตัดเครื่องหมายไม้ยมก

```
0                    คอมพิวเตอร์AllInOneสำหรับงานประมวลผล
1          เครื่องพิมพ์Multifunctionชนิดเลเซอร์หรือชนิดสี
2             เครื่องสำรองข้อมูลNetworkAttachedStorageNAS
3                                  อุปกรณ์จัดเก็บข้อมูลNAS
4                                  อุปกรณ์จัดเก็บข้อมูลNAS
```

ภาพที่ 3.6 ตัวอย่างข้อความที่กรองด้วย Regular Expression แบบจำลองที่ 1

(2) การตัดคำ (Tokenize) ในขั้นตอนนี้ ผู้วิจัยได้นำความรู้เรื่องการประมวลผลภาษาธรรมชาติ โดยประยุกต์ใช้ไลบรารีที่ชื่อว่า PyThaiNLP ซึ่งมีModule “word_tokenize” และ engine ที่ชื่อว่า “newmm”เข้ามาช่วยในการตัดคำ และในขั้นตอนนี้ยังเพิ่มเติมเรื่องการตัด stop words หรือ คำที่เจอบ่อยๆ แต่ไม่สื่อความหมายอะไรมาก เช่น “การ”, “ความ”, “คือ”, “ที่”, “ซึ่ง” เป็นต้น ซึ่งการตัด stop words จะมีข้อดี คือทำให้คลังคำศัพท์ (Bag of Words) ไม่ใหญ่เกินโดยไม่จำเป็น ซึ่งจะช่วยให้แบบจำลองหรือโมเดลของเรามีความถูกต้องแม่นยำมากขึ้น และในงานวิจัยนี้ใช้ไลบรารีที่ชื่อว่า “ thai_stopwords” ซึ่งเป็นกลุ่มของคำภาษาไทยที่ไม่สื่อความหมายมาลบออกจากข้อความในคำของบประมาณด้วย ดังแสดงไว้ตามภาพที่ 3.7-3.8

```
def text_process(text):
    final = " ".join(u for u in text)
    final = word_tokenize(final)
    final = " ".join(word for word in final)
    final = " ".join(word for word in final.split()
                     if word.lower not in thai_stopwords)
    return final
df['text_tokens'] = df['textToken'].apply(text_process)
df['text_tokens']
```

ภาพที่ 3.7 การตัดคำด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 1

```
0              คอมพิวเตอร์ All In One สำหรับ งาน ประมวลผล
1      เครื่องพิมพ์ Multifunction ชนิด เลเซอร์ หรือ ช...
2      เครื่อง สำรอง 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 ชนิด netw...
714               คอมพิวเตอร์ สำหรับ งาน ประมวลผล แบบ ที่
715    สแกนเนอร์ สำหรับ งาน เก็บ เอกสาร ระดับ ศูนย์ บ...
716    เครื่อง สแกน ลายนิ้วมือ ชนิด บันทึก เวลา เข้า ...
```

ภาพที่ 3.8 ตัวอย่างคำของบประมาณที่ตัดคำด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 1

เมื่อข้อความรายการคำของบประมาณทั้งหมดถูกตัดคำและลบคำที่ไม่สื่อความหมายแล้ว ผู้วิจัยได้นำเอาผลลัพธ์ที่ได้มาวิเคราะห์ข้อมูลทางสถิติของข้อความในรายการคำของบประมาณ (req_item) จำนวนคำในรายการคำของบประมาณ (final_word_token) และข้อมูลทางสถิติของกลุ่มงานที่รับผิดชอบพิจารณาคำของบประมาณ (group) ด้วยไลบรารี “Dataprep.eda” เพื่อทำความเข้าใจกับชุดข้อมูลรายการคำของบประมาณให้ดีมากยิ่งขึ้น ดังภาพที่ 3.9-3.12

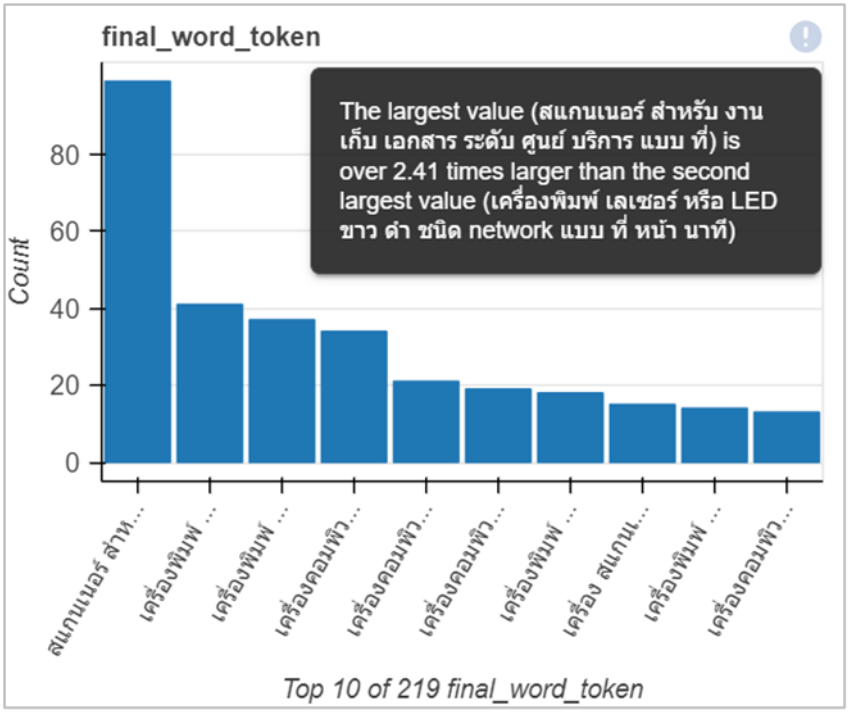
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

final_word_token
categorical

Show Details

Approximate Distinct Count	219
Approximate Unique (%)	30.5%
Missing	0
Missing (%)	0.0%
Memory Size	131885

ภาพที่ 3.9 ภาพรวมของข้อความในคำของบประมาณ



ภาพที่ 3.10 กราฟแท่งแสดงจำนวนข้อความในคำของบประมาณ

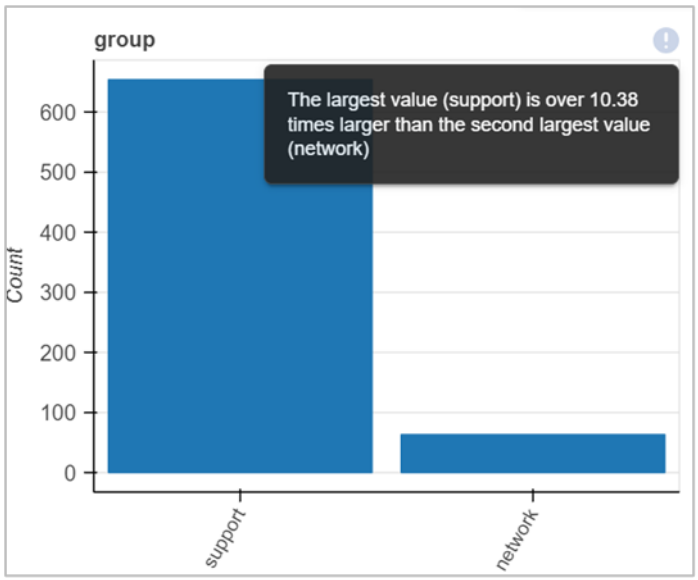
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

group
categorical

Show Details

Approximate Distinct Count	2
Approximate Unique (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory Size	51624

ภาพที่ 3.11 ภาพรวมของข้อมูลหน่วยงานที่รับผิดชอบพิจารณาคำของบประมาณ



ภาพที่ 3.12 กราฟแท่งแสดงกลุ่มงานที่รับผิดชอบพิจารณาคำของบประมาณ

(3) การสร้างคลังคำศัพท์ (Bag of Words) สำหรับขั้นตอนนี้ คือ การแปลงคำแต่ละคำ (ที่ไม่ซ้ำกัน) เป็น ID ซึ่งขั้นตอนนี้ใช้ไลบรารี “sklearn” ซึ่งจะรวมไว้กับขั้นตอนการหาจำนวนคำในคำของบประมาณที่ได้จากการตัดคำเพื่อให้ได้คำเฉพาะออกมาจากทุกข้อความในคำของบประมาณ ซึ่งในงานวิจัยนี้ใช้ไลบรารีที่ชื่อว่า “CountVectorizer” ในภาษาไพธอนเป็นเครื่องมือในการหาจำนวนคำในเอกสาร และจัดเก็บในรูปแบบ Vector (จะคล้ายๆ กับลักษณะของพจนานุกรมที่มีการระบุตัวเลข index ของแต่ละคำ) และได้ผลการนับคำในเอกสารทั้งหมดเป็นคลังคำศัพท์ ดังภาพที่ 3.13

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

```
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 3.13 ตัวอย่างคลังคำศัพท์ (Bag of Words) จากการนับค่าคำในเอกสารคำของบประมาณ

(4) สร้างฟีเจอร์ (Feature) จากการนับคำ (Count Words) สำหรับขั้นตอนนี้เป็นการสร้าง Feature จากการนับคำ คือ การนับจำนวนครั้งที่แต่ละ word id ที่แต่ละ word id ทั้งหมดปรากฏในแต่ละ text (ไม่มีคือ 0) โดยแสดงได้ดังภาพที่ 3.14

```
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))
```

```
                                                A  Access  All  Attached  \
final_word_token
คอมพิวเตอร์ All In One สำหรับ งาน ประมวลผล          0       0    1         0
เครื่องพิมพ์ Multifunction ชนิด เลเซอร์ หรือ ชน...  0       0    0         0
เครื่อง สำรอง 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 บร...  0       0    0         0
เครื่อง สแกน ลายนิ้วมือ ชนิด บันทึก เวลา เข้า อ...  0       0    0         0
```

ภาพที่ 3.14 ตัวอย่างคลังคำศัพท์ (Bag of Words) จากการนับค่าคำในเอกสารคำของบประมาณ

3.1.4 สร้างแบบจำลองที่ 1 จำแนกกลุ่มงานที่รับผิดชอบพิจารณาคำของบประมาณ

(1) แบ่งข้อมูลชุดข้อมูล การสร้างแบบจำลองเพื่อการทำนาย ซึ่งถ้าข้อมูลที่ใช้ในการทำนายเป็นชุดเดียวกันจะทำให้เกิดปัญหาที่เรียกว่า Overfitting เพื่อที่จะหลีกเลี่ยงปัญหานี้ ผู้วิจัยได้แบ่งชุดข้อมูลออกเป็นส่วนย่อยๆ (Fold) โดยชุดข้อมูลที่ถูกแบ่งจะสามารถกำหนดจำนวนครั้ง (k) สำหรับเป็นชุดข้อมูลที่ใช้สร้างแบบจำลอง (Training set) และชุดข้อมูลสำหรับทดสอบ (Testing set) ซึ่งในงานวิจัยนี้ ผู้วิจัยเลือกใช้เทคนิคที่

ชื่อว่า K-Fold Cross Validation โดยใช้ไลบรารีชื่อ “*StratifiedKFold*” สำหรับช่วยในการแบ่งข้อมูล และกำหนดจำนวนชุดข้อมูลที่ใช้สร้างแบบจำลองและทดสอบ จำนวน 5 ชุด ดังที่แสดงไว้ในภาพที่ 3.15 – 3.17

```
#split data k-fold
from sklearn.model_selection import StratifiedKFold
from sklearn.feature_extraction.text import CountVectorizer
X = df[['final_word_token']]
y = df['group']

skf = StratifiedKFold(n_splits=5, random_state=None, shuffle=False)
for i,(train_index, test_index) in enumerate(skf.split(X, y)):
        X_train, X_test = X.loc[train_index], X.loc[test_index]
        y_train, y_test = y.loc[train_index], y.loc[test_index]
```

ภาพที่ 3.15 การแบ่งชุดข้อมูลสำหรับสร้างแบบจำลองและทดสอบประสิทธิภาพ แบบจำลองที่ 1

```
Fold 1:
TRAIN DATA: (573,)
TEST DATA: (144,)
X_train:
                                        final_word_token
49              อุปกรณ์ ไมโครโฟน ชนิด ดัด เสียงสะท้อน
50    ระบบ สื่อสาร ทางไกล ผ่าน จอภาพ video conference
51  กล้อง วีดีโอ คอนเฟอเรนซ์ พร้อม อุปกรณ์ ประกอบ ...
53                          เครื่องคอมพิวเตอร์ แม่ ข่าย
56                          เครื่องคอมพิวเตอร์ แม่ ข่าย
y_train:
 49     network
50     network
51     network
53     network
56     network
Name: group, dtype: object
```

ภาพที่ 3.16 ตัวอย่างชุดข้อมูล Training set และ Testing set ชุดที่ 1 แบบจำลองที่ 1

```
Fold 5:
TRAIN DATA: (574,)
TEST DATA: (143,)
X_train:
                                        final_word_token
0          คอมพิวเตอร์ All In One สำหรับ งาน ประมวลผล
1  เครื่องพิมพ์ Multifunction ชนิด เลเซอร์ หรือ ช...
2  เครื่อง สำรอง ข้อมูล Network Attached Storage NAS
3                          อุปกรณ์ จัดเก็บ ข้อมูล NAS
4                          อุปกรณ์ จัดเก็บ ข้อมูล NAS
y_train:
 0     support
1     support
2     network
3     network
4     network
Name: group, dtype: object
```

ภาพที่ 3.17 ตัวอย่างชุดข้อมูล Training set และ Testing set ชุดที่ 5 แบบจำลองที่ 1

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

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

(2.1) ต้นไม้ตัดสินใจ (Decision Tree) ผู้วิจัยได้ใช้ไลบรารีที่ชื่อว่า “sklearn” ซึ่งมีโมดุล “tree” สำหรับสร้างแบบจำลองและทดลองกำหนดค่าพารามิเตอร์ด้วยค่าต่างๆ รายละเอียดอธิบายไว้ในตารางที่ 3.3 และแสดงเครื่องมือที่ใช้สำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Decision Tree ดังภาพที่ 3.18

ตารางที่ 3.3 ค่าพารามิเตอร์ที่ใช้ในอัลกอริทึม Decision Tree

พารามิเตอร์	คำอธิบาย
critetion = gini	ฟังก์ชันที่ใช้ในการวัดประสิทธิภาพของการแยกโหนดของ Decision Tree สามารถเลือกได้ระหว่าง gini (Gini Impurity) หรือ Entropy (Information Gain)
max_depth = 3	จำนวนชั้นของต้นไม้ ถ้า max_depth = 3 ชั้นความลึกของต้นไม้
min_sample_leaf = 5	จำนวนตัอย่างข้อมูลขั้นต่ำที่จะแสดงใน leaf node

```
#------------สร้างแบบจำลอง Decision Tree
from sklearn import tree
clf = tree.DecisionTreeClassifier(criterion='gini',max_depth=3,min_samples_leaf=5)
clf = clf.fit(train_bow, y_train)
#-----------ทดสอบแบบจำลอง Decision Tree
from sklearn.metrics import confusion_matrix, classification_report
test_bow = cvec.transform(X_test['final_word_token'])
test_predictions = clf.predict(test_bow)
print('\n')
print('Report Model: Decision Tree Algorithm ---------------')
print(classification_report(test_predictions, y_test))
plt.figure()
tree.plot_tree(clf,filled=True)
plt.savefig('decisiontree.png',format='png',bbox_inches = "tight", dpi=300)
myconfusion_matrix(y_test,test_predictions,file_name="Decision Tree_confusion.png")
```

ภาพที่ 3.18 เครื่องมือสำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Decision Tree

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

2.2) ชุดต้นไม้ตัดสินใจ (Ensemble of Decision Trees) ผู้วิจัยได้ใช้ไลบรารีที่ชื่อว่า “sklearn” ซึ่งมีโมดุล “RandomForestClassifier” สำหรับสร้างแบบจำลองและทดลองกำหนดค่าพารามิเตอร์ด้วยค่าต่างๆ รายละเอียดอธิบายไว้ในตารางที่ **3.4** และแสดงเครื่องมือที่ใช้สำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Random Forest ดังภาพที่ 3.19

ตารางที่ 3.4 ค่าพารามิเตอร์ที่ใช้ในอัลกอริทึม Random Forest

พารามิเตอร์	คำอธิบาย
n_estimators = 100	จำนวนต้นไม้ตัดสินใจ (Decision Tree) ที่ใช้ในป่านี้
random_state = 0	การกำหนดตัวเลขในการสุ่มข้อมูล
max_features = 5	การกำหนดจำนวนคุณลักษณะในการทดสอบต้นไม้แต่ละต้น

```
#------------สร้างแบบจำลอง Random Forest Algorithm
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(n_estimators=100, random_state=0, max_features=5)
rf.fit(train_bow, y_train)
#-----------ทดสอบแบบจำลอง Random Forest Algorithm
from sklearn.metrics import confusion_matrix, classification_report
test_bow = cvec.transform(X_test['final_word_token'])
test_predictions = rf.predict(test_bow)
print('Report Model: Random Forest Algorithm ---------------')
print(classification_report(test_predictions, y_test))
myconfusion_matrix(y_test,test_predictions,file_name="Random Forest_confusion.png")
```

ภาพที่ 3.19 เครื่องมือสำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Random Forest

(2.3) นาอีฟ เบย์ (Naïve Bayes) ผู้วิจัยได้ใช้ไลบรารีที่ชื่อว่า “sklearn” ซึ่งมีโมดุล “MultimomialNb” สำหรับสร้างแบบจำลอง และแสดงเครื่องมือสำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Naïve Bayes ดังภาพที่ 3.20

```
#------------สร้างแบบจำลอง Naïve Bayes Algorithm
from sklearn.naive_bayes import MultinomialNB
nb = MultinomialNB()
nb.fit(train_bow, y_train)
#-----------ทดสอบแบบจำลอง Naïve Bayes Algorithm
from sklearn.metrics import confusion_matrix, classification_report
test_bow = cvec.transform(X_test['final_word_token'])
test_predictions = nb.predict(test_bow)
print('Report Model: Naïve Bayes Algorithm ---------------')
print(classification_report(test_predictions, y_test))
myconfusion_matrix(y_test,test_predictions,file_name="Naïve Bayes_confusion.png")
```

ภาพที่ 3.20 เครื่องมือสำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Naïve Bayes

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

(2.4) ลอจิสติกรีเกรสชัน (Logistic Regression) ผู้วิจัยได้ใช้ไลบรารีที่ชื่อว่า “sklearn” ซึ่งมีโมดุล “LogisticRegression” สำหรับสร้างแบบจำลอง และแสดงเครื่องมือสำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Logistic Regression ดังภาพที่ 3.21

```
#-------------สร้างแบบจำลอง Logistic Regression
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
lr.fit(train_bow, y_train)
#-------------ทดสอบแบบจำลอง Logistic Regression
from sklearn.metrics import confusion_matrix, classification_report
test_bow = cvec.transform(X_test['final_word_token'])
test_predictions = lr.predict(test_bow)
print('Report Model: Logistic Regression ----------')
print(classification_report(test_predictions, y_test))
myconfusion_matrix(y_test,test_predictions,file_name="Logistic Regression_confusion.png")
```

ภาพที่ 3.21 เครื่องมือสำหรับสร้างแบบจำลองที่ 1 ด้วยอัลกอริทึม Logistic Regression

3.1.5 การทดสอบและประเมินประสิทธิภาพของแบบจำลอง

ในงานวิจัยนี้ ดำเนินการทดสอบประสิทธิภาพแบบจำลองโดยการแบ่งชุดข้อมูล testing data ด้วยเทคนิค 5-Fold Cross Validation รวมถึงใช้วิธีการวัดประสิทธิภาพของแบบจำลองด้วยค่า Accuracy, Precision, Recall, Macro F1 โดยแสดงค่าการทำนายผลของแบบจำลองด้วยตาราง Confusion Matrix เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมระหว่างค่าการทำนายและค่าจริงในรูปแบบของตาราง ซึ่งเครื่องมือและตัวอย่างชุดข้อมูลที่ใช้ในการทดสอบ แสดงไว้ดังภาพที่ 3.22 – 3.23 ส่วนเครื่องมือและตัวอย่างผลประสิทธิภาพและผลการทำนายของแบบจำลอง ปรากฏดังภาพ ที่ 3.24 – 3.25

```
#split data k-fold
skf = StratifiedKFold(n_splits=5, random_state=None, shuffle=False)
for i,(train_index, test_index) in enumerate(skf.split(X, y)):
        X_train, X_test = X.loc[train_index], X.loc[test_index]
        y_train, y_test = y.loc[train_index], y.loc[test_index]
        print('\n')
        print(f"Fold {i+1}:")
        print("TRAIN DATA:", train_index)
        print("TEST DATA:", test_index)
```

ภาพที่ 3.22 เครื่องมือที่ใช้ในการทดสอบอัลกอริทึมของแบบจำลองที่ 1

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์


```
TEST DATA: [  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35
  36  37  38  39  40  41  42  43  44  45  46  47  48  52  54  55  58  59
  60  61  64  66  69  70  71  72  73  74  75  76  77  78  80  82  83  85
  86  87  88  89  90  91  92  93  94  95  96  97  98  99 101 102 103 104
 105 106 107 108 109 112 113 114 115 116 117 118 119 120 121 122 124 125
 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 143 144
 145 146 147 148 149 150 151 152 153 154 155 156 157 158 160 161 162 163]
```

```
X_Test:
                                          final_word_token
0          คอมพิวเตอร์ All In One สำหรับ งาน ประมวลผล
1  เครื่องพิมพ์ Multifunction ชนิด เลเซอร์ หรือ ช...
2  เครื่อง สำรอง ข้อมูล Network Attached Storage NAS
3                          อุปกรณ์ จัดเก็บ ข้อมูล NAS
4                          อุปกรณ์ จัดเก็บ ข้อมูล NAS
y_Test:
 0    support
1    support
2    network
3    network
4    network
```

ภาพที่ 3.23 ตัวอย่างชุดข้อมูลที่ใช้ในการทดสอบอัลกอริทึมของแบบจำลองที่ 1

```python
#define function
def myconfusion_matrix(y_true,y_pred,label_list=["support", "network"],file_name="confusion_result.png"):
    result = confusion_matrix(y_true, y_pred, labels=label_list)
    print(result)
    cm_display = ConfusionMatrixDisplay(confusion_matrix=result, display_labels=label_list)
    cm_display.plot()
    plt.show()
    cm_display.figure_.savefig(file_name, dpi=300)
```

ภาพที่ 3.24 เครื่องมือสำหรับประเมินประสิทธิภาพแบบจำลองที่ 1

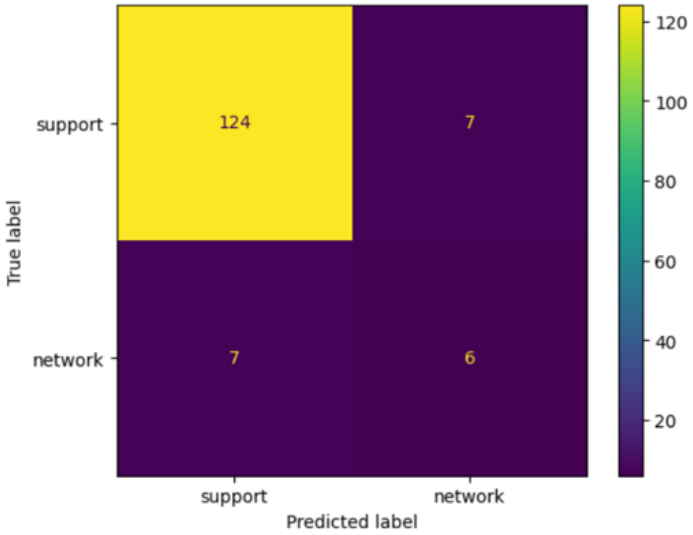
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

```
Report Model: Decision Tree Algorithm ---------------
              precision    recall  f1-score   support

     network       0.46      0.46      0.46        13
     support       0.95      0.95      0.95       131

    accuracy                           0.90       144
   macro avg       0.70      0.70      0.70       144
weighted avg       0.90      0.90      0.90       144

[[124   7]
 [  7   6]]
```



ภาพที่ 3.25 ตัวอย่าง Confusion Matrix แสดงผลประสิทธิภาพและผลการทำนายของแบบจำลองที่ 1

3.2 แบบจำลองที่ 2 จำแนกผลการพิจารณาคำของบประมาณจากเหตุผลความจำเป็น

3.2.1 ศึกษาวิเคราะห์ระบบการของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์ของหน่วยงานในสังกัดสำนักงานศาลยุติธรรม และวิเคราะห์ปัญหาของหน่วยงาน

การดำเนินงานวิจัยในแบบจำลองที่ 2 ผู้วิจัยได้อาศัยความรู้ที่ได้จากการศึกษาวิเคราะห์จากขั้นตอนที่ 3.1.1 ในแบบจำลองที่ 1

3.2.2 การรวบรวมเอกสารข้อมูล

ในขั้นตอนการรวบรวมเอกสารข้อมูลที่ใช้สำหรับแบบจำลองที่ 2 ผู้วิจัยได้ดำเนินการต่อยอดการดำเนินงานโดยใช้ข้อมูลชุดเดียวกันกับที่ได้รวบรวมไว้จากหัวข้อ 3.1.2 ในแบบจำลองที่ 1 ซึ่งมีข้อมูลตัวอย่างจำนวน 717 ชุด

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.2.3 การเตรียมข้อมูล

สำหรับแบบจำลองที่ 2 การจำแนกผลการพิจารณา “อนุมัติ” หรือ “ไม่อนุมัติ” คำของบประมาณจากเหตุผลความจำเป็น ผู้วิจัยเลือกใช้ข้อมูลคอลัมน์ req_detail (รายละอียดเหตุผลความจำเป็น) ชนิดข้อมูลเป็น Text เป็นอินพุตฟีเจอร์ (Input Feature) และเลือกข้อมูลคอลัมน์ approve (ผลการพิจารณา) ชนิดข้อมูลเป็น Text ซึ่งประกอบด้วย 2 คลาส ได้แก่ “yes” = “อนุมัติ” และ “no” = “ไม่อนุมัติ” เป็นลาเบล (Label) สำหรับนำไปสร้างแบบจำลอง ดังภาพที่ 3.26

req_detail	approve
ขอเพิ่ม จำนวน 10 เครื่อง เนื่องจากนิติกรประจำองค์คณะผู้พิพากษา ผู้ช่วยผู้พิพากษามีไม่เพียงพอ	no
ขอใหม่ จำนวน 1 เครื่อง เพื่อให้ระบบงานธุรการงานคดีมีประสิทธิภาพมากยิ่งขึ้น และเป็นการนำเทค	yes
ขอใหม่ จำนวน 1 เครื่อง เนื่องจากศาลจังหวัดนนทบุรียังไม่เคยได้รับการจัดสรรเครื่องสำรองข้อมูล (I	yes
ขอใหม่ จำนวน 1 ชุด เนื่องจากสถาบันพัฒนาข้าราชการฝ่ายตุลาการศาลยุติธรรมไม่มีเครื่องจัดเก็บข	no
ขอทดแทน จำนวน 1 ชุด เนื่องจากอุปกรณ์ของเดิมชำรุด เพื่อให้บุคลากรภายในหน่วยงานใช้	yes
ขอทดแทน จำนวน 1 เครื่อง เนื่องจากเครื่องสแกนลายนิ้วมือที่ติดตั้งอยู่ที่ส่วนช่วยอำนวยการใช้สำห	yes
ขอใหม่ จำนวน 1 เครื่อง ศาลอุทธรณ์คดีชำนัญพิเศษไม่มีเครื่องคอมพิวเตอร์ Server สำหรับใช้ในก	yes
ขอใหม่ จำนวน 1 ชุด เพื่อจัดเก็บข้อมูลของพิพิธภัณฑ์ศาลไทยและหอจดหมายเหตุ (ตั้งอยู่ที่อาคาร	yes
ขอทดแทน จำนวน 1 ชุด เนื่องจากอุปกรณ์ชุดควบคุมกล้องเสีย และมีอายุการใช้งานเกินระยะเวลาที่	yes
ขอเพิ่ม จำนวน 2 เครื่อง เพื่อให้ศูนย์นัดความ ศูนย์ประสานงานพยาน งานระบบนัดพิจารณาคดีต่อเนื่	yes
ขอทดแทน จำนวน 1 เครื่อง เพื่อใช้ในการออกใบเสร็จประจำวันทุกวัน เนื่องจากครุภัณฑ์เดิมชำรุดเสี	yes
ขอเพิ่ม จำนวน 3 เครื่อง เนื่องจากเครื่องคอมพิวเตอร์สำหรับงานประมวลผล แบบที่ 2 (จอแสดงภาพ	no
ขอทดแทน จำนวน 1 เครื่อง เนื่องจากเครื่องพิมพ์ชนิด Dot Matrix แบบแคร่ยาว ชำรุด เนื่องจากใช้	yes
ขอเพิ่ม จำนวน 1 เครื่อง เนื่องจากปัจจุบันเครื่องคอมพิวเตอร์ไม่เพียงพอต่อการใช้งาน โดยปัจจุบันไ	no
ขอใหม่ จำนวน 1 ชุด เนื่องจากสถานการณ์ระบาดของโรคติดเชื้อไวรัสโคโรนา ๒๐๑๙ (COVID-19)	yes
ขอเพิ่ม จำนวน 2 ชุด คณะกรรมการบริหารศาลยุติธรรมมีมติในการประชุมครั้งที่ 6/256 เมื่อวันที่ 22 ม	no
ขอใหม่ จำนวน 2 เครื่อง ปัจจุบันส่วนช่วยพิจารณาคดี ยังไม่มีเครื่องพิมพ์ในการออกใบเสร็จรับเงินใน	yes
ขอเพิ่ม จำนวน 17 เครื่อง สำหรับกลุ่มงานช่วยพิจารณาคดี	no
ขอเพิ่ม จำนวน 1 เครื่อง สำหรับงานประจำกลุ่มงานโฆษกและวิเคราะห์ข่าว โดยกองสารนิเทศและป	no
ขอเพิ่ม จำนวน 9 เครื่องสแกนเนอร์ : เนื่องจากปัจจุบันศาลเยาวชนและครอบครัวจังหวัดเชียงรายยัง	no

ภาพที่ 3.26 ฟีเจอร์และลาเบลที่นำไปใช้สร้างแบบจำลองที่ 2

โดยขั้นตอนในการเตรียมข้อมูลประเภท Text ซึ่งมีรายละเอียดดังนี้

(1) ลบเครื่องหมายต่างๆ ที่เราไม่ต้องการ (!€@%#*&~) ซึ่งคำเหล่านี้ไม่สื่อความความหมายและมีผลต่อการตัดคำในกระบวนการตัดคำ ซึ่งจะทำให้จำนวนคำที่ได้มีจำนวนมาก คำเหล่านี้จะต้องถูกลบออกจากชุดข้อมูลเพื่อลดจำนวนของคำที่จะนำไปใช้เป็นคุณลักษณะฟีเจอร์ (Features) ในการสร้างแบบจำลองทำนาย ทั้งนี้ในขั้นตอนนี้กรองเอาเฉพาะตัวอักษรและสระที่เป็นภาษาไทย และตัวอักษรภาษาอังกฤษเท่านั้น เครื่องมือที่ใช้คือไลบรารีที่ชื่อว่า RegEx (Reggular Expression) ซึ่งกำหนดรูปแบบได้เป็น [^\u0E00-\u0E7Fa-zA-Z']|['']|[|] และเขียนด้วยภาษาไพธอน (Python) ดังที่แสดงไว้ในภาพที่ 29 ส่วนความหมายของเครื่องหมายอธิบายไว้ในตารางที่ 2 ซึ่งเป็นรูปแบบที่ใช้ทำความสะอาดข้อมูลในแบบจำลองที่ 1 ข้างต้น ทั้งนี้ข้อความรายละเอียดเหตุผลความจำเป็นที่ทำความสะอาดแล้ว แสดงไว้ในภาพที่ 3.27

```
# regular expression
def preprocess_text(raw_text):
  pattern = re.compile(r"[^\u0E00-\u0E7Fa-zA-Z' ]|[']|[|]|[?]|[.]|[,]|[;]|[:]|[!]|[ฯ]|[ๆ]|")
  cleaned_text= re.sub(pattern,'',raw_text)
  return cleaned_text
df['cleaned_text_item'] = df['req_detail'].apply(preprocess_text)
df['cleaned_text_item']
```

ภาพที่ 3.27 การกรองข้อความด้วย Regular Expression แบบจำลองที่ 2

```
0      ขอเพิ่ม จำนวน  เครื่อง เนื่องจากนิติกรประจำองค...
1      ขอใหม่ จำนวน  เครื่อง เพื่อให้ระบบงานธุรการงาน...
2      ขอใหม่ จำนวน  เครื่อง เนื่องจากศาลจังหวัดนนทบุ...
3      ขอใหม่ จำนวน  ชุด เนื่องจากสถาบันพัฒนาข้าราชกา...
4      ขอทดแทน จำนวน  ชุด เนื่องจากอุปกรณ์ของเดิมชำรุ...
                          ...
712    ขอเพิ่ม จำนวน  เครื่อง สำหรับจัดสรรให้ข้าราชกา...
713    ขอเพิ่ม จำนวน  เครื่อง สำหรับรองรับการใช้งานขอ...
714    ขอเพิ่ม จำนวน  ชุด สำหรับรองรับอัตรากำลังข้ารา...
715    ขอใหม่  เครื่อง สำหรับใช้งานประจำศูนย์นัดความแ...
716    ขอทดแทน จำนวน  เครื่อง เนื่องจากเป็นครุภัณฑ์ที...
```

ภาพที่ 3.28 ฟีเจอร์และลาเบลที่นำไปใช้สร้างแบบจำลองที่ 2

(2) การตัดคำ (Tokenize) ในขั้นตอนนี้ ผู้วิจัยได้นำความรู้เรื่องการประมวลผลภาษาธรรมชาติ โดยประยุกต์ใช้ไลบรารีที่ชื่อว่า PyThaiNLP ซึ่งมีModule “word_tokenize” และ engine ที่ชื่อว่า “newmm”เข้ามาช่วยในการตัดคำ และในขั้นตอนนี้ยังเพิ่มเติมเรื่องการตัด stop words หรือ คำที่เจอบ่อยๆ แต่ไม่สื่อความหมายอะไรมาก เช่น “การ”, “ความ”, “คือ”, “ที่”, “ซึ่ง” เป็นต้น ซึ่งการตัด stop words จะมีข้อดี คือทำให้คลังคำศัพท์ (Bag of Words) ไม่ใหญ่เกินโดยไม่จำเป็น ซึ่งจะช่วยให้แบบจำลองหรือโมเดลของเรามีความถูกต้องแม่นยำมากขึ้น และในงานวิจัยนี้ใช้ไลบรารีที่ชื่อว่า “ thai_stopwords” ซึ่งเป็นกลุ่มของคำภาษาไทยที่ไม่สื่อความหมายมาลบออกจากข้อความในคำของบประมาณด้วย ดังแสดงไว้ตามภาพที่ 3.29 – 3.30

```
def text_process(text):
    final = " ".join(u for u in text)
    final = word_tokenize(final)
    final = " ".join(word for word in final)
    final = " ".join(word for word in final.split()
                     if word.lower not in thai_stopwords)
    return final
df['text_tokens'] = df['textToken'].apply(text_process)
df['text_tokens']
```

ภาพที่ 3.29 การตัดคำด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 2

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

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 ข้ารา...
4	ขอ ทดแทน จำนวน ชุด เนื่องจาก อุปกรณ์ ของ เดิม ..
	...
712	ขอ เพิ่ม จำนวน เครื่อง สำหรับ จัดสรร ให้ ข้ารา...
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 นัด ...
716	ขอ ทดแทน จำนวน เครื่อง เนื่องจาก เป็น ครุภัณฑ์...

ภาพที่ 3.30 ตัวอย่างคำของบประมาณที่ตัดคำด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 2

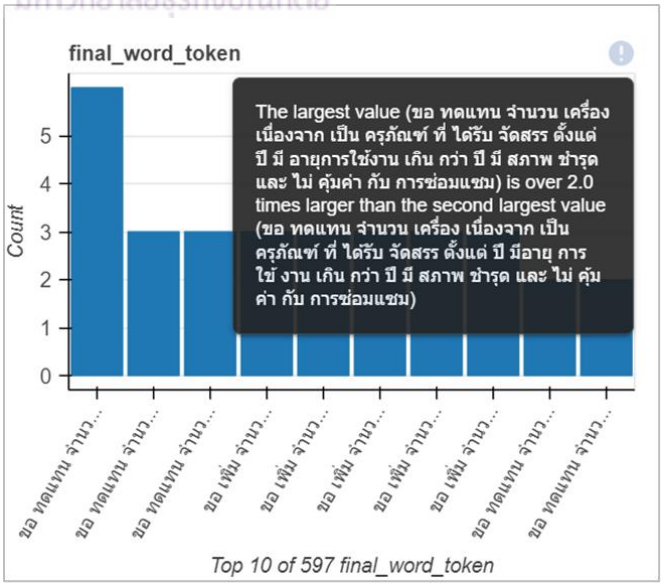
เมื่อข้อความเหตุผลความจำเป็นทั้งหมดถูกตัดคำและลบคำที่ไม่สื่อความหมายแล้ว ผู้วิจัยได้นำเอาผลลัพธ์ที่ได้มาวิเคราะห์ข้อมูลทางสถิติของข้อความในเหตุผลความจำเป็น (req_detail) จำนวนคำรายละเอียดเหตุผลความจำเป็น (final_word_token) และข้อมูลทางสถิติของผลการพิจารณาคำของบประมาณ (approve) ด้วยไลบรารี “Dataprep.eda” เพื่อทำความเข้าใจกับชุดข้อมูลรายการคำของบประมาณให้ดีมากยิ่งขึ้น ดังภาพที่ 3.31 – 3.32

final_word_token
categorical
Show Details

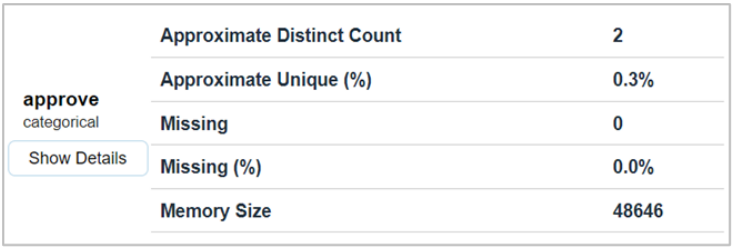
Approximate Distinct Count	597
Approximate Unique (%)	83.3%
Missing	0
Missing (%)	0.0%
Memory Size	877831

ภาพที่ 3.31 ตัวอย่างคำของบประมาณที่ตัดคำด้วยไลบรารี word_tokenize และ thai_stopwords แบบจำลองที่ 2

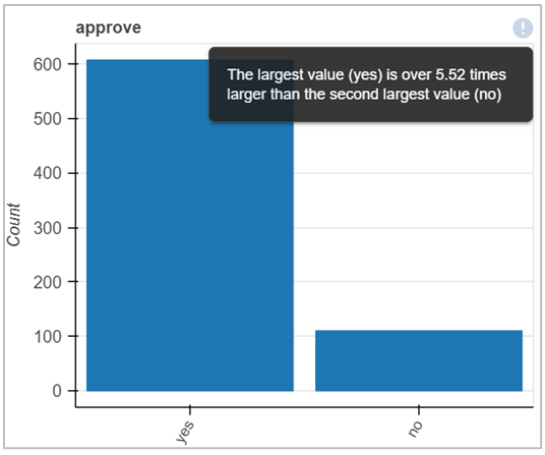
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 3.32 กราฟแท่งแสดงจำนวนข้อความในรายละเอียดเหตุผลความจำเป็น



ภาพที่ 3.33 ภาพรวมของข้อมูลผลการพิจารณาคำของบประมาณ



ภาพที่ 3.34 ภาพรวมของข้อมูลผลการพิจารณาคำของบประมาณ

(3) การสร้างคลังคำศัพท์ (Bag of Words) สำหรับขั้นตอนนี้ คือ การแปลงคำแต่ละคำ (ที่ไม่ซ้ำกัน) เป็น ID ซึ่งขั้นตอนนี้ใช้ไลบรารี “sklearn” ซึ่งจะรวมไว้กับขั้นตอนการหาจำนวนคำในข้อความเหตุผลความจำเป็นที่ได้จากการตัดคำเพื่อให้ได้คำเฉพาะออกมาจากทุกข้อความในเหตุผลความจำเป็น ซึ่งในงานวิจัยนี้ใช้ไลบรารีที่ชื่อว่า “CountVectorizer” ในภาษาไพธอนเป็นเครื่องมือในการหาจำนวนคำในเอกสาร และจัดเก็บในรูปแบบ Vector (จะคล้ายๆ กับลักษณะของพจนานุกรมที่มีการระบุตัวเลข index ของแต่ละคำ) และได้ผลการนับคำในเอกสารทั้งหมดเป็นคลังคำศัพท์ ดังภาพที่ 3.35

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

cvec.vocabulary_

{'ขอ': 195,
 'เพิ่ม': 1336,
 'จำนวน': 373,
 'เครื่อง': 1260,
 'เนื่องจาก': 1306,
 'นิติกร': 623,
 'ประจำ': 689,
 'องค์': 1172,
 'คณะ': 229,
 'ผู้พิพากษา': 759,
 'ผู้ช่วย': 749,
```

ภาพที่ 3.35 ตัวอย่างคลังคำศัพท์ (Bag of Words) จากการนับค่าคำในข้อความเหตุผลความจำเป็น

(4) สร้างฟีเจอร์ (Feature) จากการนับคำ (Count Words) สำหรับขั้นตอนนี้เป็นการสร้าง Feature จากการนับคำ คือ การนับจำนวนครั้งที่แต่ละ word id ที่แต่ละ word id ทั้งหมดปรากฏในแต่ละ text (ไม่มี คือ 0) โดยแสดงได้ดังภาพที่ 3.36

```
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	ไม่	ไม่ต้อง	ไร้	ไวรัส	\
ขอ เพิ่ม จำนวน เครื่อง เนื่องจาก นิติกร ประจำ อ...	1	0	0	0	
ขอ ใหม่ จำนวน เครื่อง เพื่อให้ ระบบ งานธุรการ ง...	1	0	0	0	
ขอ ใหม่ จำนวน เครื่อง เนื่องจาก ศาลจังหวัด นนทบ...	1	0	0	0	
ขอ ใหม่ จำนวน ชุด เนื่องจาก สถาบัน พัฒนา ข้าราช...	3	0	0	0	
ขอ ทดแทน จำนวน ชุด เนื่องจาก อุปกรณ์ ของ เดิม ช...	0	0	0	0	
...	...	...	...	...	...
ขอ เพิ่ม จำนวน เครื่อง สำหรับ จัดสรร ให้ ข้าราช...	1	0	0	0	
ขอ เพิ่ม จำนวน เครื่อง สำหรับ รองรับ การใช้งาน ...	1	0	0	0	
ขอ เพิ่ม จำนวน ชุด สำหรับ รองรับ อัตรา กำลัง บ้...	1	0	0	0	
ขอ ใหม่ เครื่อง สำหรับ ใช้ งานประจำ ศูนย์ นัด ค...	1	0	0	0	
ขอ ทดแทน จำนวน เครื่อง เนื่องจาก เป็น ครุภัณฑ์ ...	1	0	0	0	

ภาพที่ 3.36 ตัวอย่างฟีเจอร์ (Feature) ที่สร้างจากการนับคำในข้อความเหตุผลความจำเป็น

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.2.4 สร้างแบบจำลองที่ **2** จำแนกผลการพิจารณาคำของบประมาณตามเหตุผลความจำเป็น

(1) แบ่งข้อมูลชุดข้อมูล การสร้างแบบจำลองเพื่อการทำนาย ซึ่งถ้าข้อมูลที่ใช้ในการทำนายเป็นชุดเดียวกันจะทำให้เกิดปัญหาที่เรียกว่า Overfitting เพื่อที่จะหลีกเลี่ยงปัญหานี้ ผู้วิจัยได้แบ่งชุดข้อมูลออกเป็นส่วนย่อยๆ (Fold) โดยชุดข้อมูลที่ถูกแบ่งจะสามารถกำหนดจำนวนครั้ง (k) สำหรับเป็นชุดข้อมูลที่ใช้สร้างแบบจำลอง (Training set) และชุดข้อมูลสำหรับทดสอบ (Testing set) ซึ่งในงานวิจัยนี้ ผู้วิจัยเลือกใช้เทคนิคที่ชื่อว่า K-Fold Cross Validation โดยใช้ไลบรารีชื่อ “*StratifiedKFold*” สำหรับช่วยในการแบ่งข้อมูล และกำหนดจำนวนชุดข้อมูลที่ใช้สร้างแบบจำลองและทดสอบ จำนวน 5 ชุด ดังที่แสดงไว้ในภาพที่ 3.37 – 3.39

```
#split data k-fold
from sklearn.model_selection import StratifiedKFold
from sklearn.feature_extraction.text import CountVectorizer
X = df[['final_word_token']]
y = df['approve']

skf = StratifiedKFold(n_splits=5, random_state=None, shuffle=False)
for i,(train_index, test_index) in enumerate(skf.split(X, y)):
        X_train, X_test = X.loc[train_index], X.loc[test_index]
        y_train, y_test = y.loc[train_index], y.loc[test_index]
```

ภาพที่ 3.37 การแบ่งชุดข้อมูลสำหรับสร้างแบบจำลองและทดสอบประสิทธิภาพ แบบจำลองที่ 2

```
Fold 1:
TRAIN DATA: (573,)
TEST DATA: (144,)
X_train:
                                        final_word_token
59  ขอ เพิ่ม จำนวน เครื่อง เพื่อ รองรับ สำหรับ โปร...
61  ขอ เพิ่ม จำนวน เครื่อง ใช้ ใน การ พิมพ์ รายงาน...
65  ขอ เพิ่ม จำนวน ชุด เนื่อง จก ศาลจังหวัด ล่าพูน ...
68  ขอ ทดแทน จำนวน เครื่อง เนื่องจาก ของ เดิม ที่ ...
72  ขอ ทดแทน จำนวน เครื่อง ใช้ พิมพ์ เอกสาร เป็นปร...
y_train:
 59    no
61    no
65    no
68    no
72    no
Name: approve, dtype: object
```

ภาพที่ 3.38 ตัวอย่างชุดข้อมูล Training set และ Testing set ชุดที่ 1 แบบจำลองที่ 2

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์


```
Fold 5:
TRAIN DATA: (574,)
TEST DATA: (143,)
X_train:
                                              final_word_token
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 ข้ารา...
4  ขอ ทดแทน จำนวน ชุด เนื่องจาก อุปกรณ์ ของ เดิม ...
y_train:
 0     no
1    yes
2    yes
3     no
4    yes
Name: approve, dtype: object
```

ภาพที่ 3.39 ตัวอย่างชุดข้อมูล Training set และ Testing set ชุดที่ 5 แบบจำลองที่ 2

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

(2.1) ต้นไม้ตัดสินใจ (Decision Tree) ผู้วิจัยได้ใช้ไลบรารีที่ชื่อว่า “sklearn” ซึ่งมีโมดุล “tree” สำหรับสร้างแบบจำลองและทดลองกำหนดค่าพารามิเตอร์ด้วยค่าต่างๆ รายละเอียดอธิบายไว้ในตารางที่ 3.5 และแสดงเครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Decision Tree ดังภาพที่ 3.40

ตารางที่ 3.5 ค่าพารามิเตอร์ที่ใช้ในอัลกอริทึม Decision Tree

พารามิเตอร์	คำอธิบาย
critetion = gini	ฟังก์ชันที่ใช้ในการวัดประสิทธิภาพของการแยกโหนดของ Decision Tree สามารถเลือกได้ระหว่าง gini (Gini Impurity) หรือ Entropy (Information Gain)
max_depth = 3	จำนวนชั้นของต้นไม้ ถ้า max_depth = 3 ชั้นความลึกของต้นไม้
min_sample_leaf = 5	จำนวนตัอย่างข้อมูลขั้นต่ำที่จะแสดงใน leaf node

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์


```
#------------สร้างแบบจำลอง Decision Tree
from sklearn import tree
clf = tree.DecisionTreeClassifier(criterion='gini',max_depth=3,min_samples_leaf=5)
clf = clf.fit(train_bow, y_train)
#-----------ทดสอบแบบจำลอง Decision Tree
from sklearn.metrics import confusion_matrix, classification_report
test_bow = cvec.transform(X_test['final_word_token'])
test_predictions = clf.predict(test_bow)
print('\n')
print('Report Model: Decision Tree Algorithm ---------------')
print(classification_report(test_predictions, y_test))
plt.figure()
tree.plot_tree(clf,filled=True)
plt.savefig('decisiontree.png',format='png',bbox_inches = "tight", dpi=300)
myconfusion_matrix(y_test,test_predictions,file_name="Decision Tree_confusion.png")
```

ภาพที่ 3.40 ตัวอย่างชุดข้อมูล Training set และ Testing set ชุดที่ 5 แบบจำลองที่ 2

(2.2) ชุดต้นไม้ตัดสินใจ (Ensemble of Decision Trees) ผู้วิจัยได้ใช้ไลบรารีที่ชื่อว่า “sklearn” ซึ่งมีโมดุล “RandomForestClassifier” สำหรับสร้างแบบจำลองและทดลองกำหนดค่าพารามิเตอร์ด้วยค่าต่างๆ รายละเอียดอธิบายไว้ในตารางที่ 3.6 และแสดงเครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Random Forest ดังภาพที่ 3.41

ตารางที่ 3.6 ค่าพารามิเตอร์ที่ใช้ในอัลกอริทึม Random Forest

พารามิเตอร์	คำอธิบาย
n_estimators = 100	จำนวนต้นไม้ตัดสินใจ (Decision Tree) ที่ใช้ในป่านี้
random_state = 0	การกำหนดตัวเลขในการสุ่มข้อมูล
max_features = 5	การกำหนดจำนวนคุณลักษณะในการทดสอบต้นไม้แต่ละต้น

```
#------------สร้างแบบจำลอง Random Forest Algorithm
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(n_estimators=100, random_state=0, max_features=5)
rf.fit(train_bow, y_train)
#-----------ทดสอบแบบจำลอง Random Forest Algorithm
from sklearn.metrics import confusion_matrix, classification_report
test_bow = cvec.transform(X_test['final_word_token'])
test_predictions = rf.predict(test_bow)
print('Report Model: Random Forest Algorithm ---------------')
print(classification_report(test_predictions, y_test))
myconfusion_matrix(y_test,test_predictions,file_name="Random Forest_confusion.png")
```

ภาพที่ 3.41 เครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Random Forest

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

(2.3) นาอีฟ เบย์ (Naïve Bayes) ผู้วิจัยได้ใช้ไลบรารีที่ชื่อว่า “sklearn” ซึ่งมีโมดุล “MultimomialNb” สำหรับสร้างแบบจำลอง และเครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Naïve Bayes ดังภาพที่ 3.42

```
#------------สร้างแบบจำลอง Naïve Bayes Algorithm
from sklearn.naive_bayes import MultinomialNB
nb = MultinomialNB()
nb.fit(train_bow, y_train)
#-----------ทดสอบแบบจำลอง Naïve Bayes Algorithm
from sklearn.metrics import confusion_matrix, classification_report
test_bow = cvec.transform(X_test['final_word_token'])
test_predictions = nb.predict(test_bow)
print('Report Model: Naïve Bayes Algorithm ---------------')
print(classification_report(test_predictions, y_test))
myconfusion_matrix(y_test,test_predictions,file_name="Naïve Bayes_confusion.png")
```

ภาพที่ 3.42 เครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Naïve Bayes

(2.4) ลอจิสติกรีเกรสชัน (Logistic Regression) ผู้วิจัยได้ใช้ไลบรารีที่ชื่อว่า “sklearn” ซึ่งมีโมดุล “LogisticRegression” สำหรับสร้างแบบจำลอง และแสดงเครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Logistic Regression ภาพที่ 3.43

```
#-------------สร้างแบบจำลอง Logistic Regression
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
lr.fit(train_bow, y_train)
#------------ทดสอบแบบจำลอง Logistic Regression
from sklearn.metrics import confusion_matrix, classification_report
test_bow = cvec.transform(X_test['final_word_token'])
test_predictions = lr.predict(test_bow)
print('Report Model: Logistic Regression ----------')
print(classification_report(test_predictions, y_test))
myconfusion_matrix(y_test,test_predictions,file_name="Logistic Regression_confusion.png")
```

ภาพที่ 3.43 เครื่องมือสำหรับสร้างแบบจำลองที่ 2 ด้วยอัลกอริทึม Logistic Regression

3.2.5 การทดสอบและประเมินประสิทธิภาพของแบบจำลอง

ในงานวิจัยนี้ ดำเนินการทดสอบประสิทธิภาพแบบจำลองโดยการแบ่งชุดข้อมูล testing data ด้วยเทคนิค 5-Fold Cross Validation รวมถึงใช้วิธีการวัดประสิทธิภาพของแบบจำลองด้วยค่า Accuracy, Precision, Recall, Macro F1 โดยแสดงค่าการทำนายผลของแบบจำลองด้วยตาราง Confusion Matrix เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมระหว่างค่าการทำนายและค่าจริงในรูปแบบของตาราง ซึ่งเครื่องมือและตัวอย่างชุดข้อมูลที่ใช้ในการทดสอบ แสดงไว้ดังภาพที่ 3.44-3.45 ส่วนเครื่องมือและตัวอย่างผลประสิทธิภาพและผลการทำนายของแบบจำลอง ปรากฏดังภาพ ที่ 3.46-3.47

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์


```
#split data k-fold
skf = StratifiedKFold(n_splits=5, random_state=None, shuffle=False)
for i,(train_index, test_index) in enumerate(skf.split(X, y)):
        X_train, X_test = X.loc[train_index], X.loc[test_index]
        y_train, y_test = y.loc[train_index], y.loc[test_index]
        print('\n')
        print(f"Fold {i+1}:")
        print("TRAIN DATA:", train_index)
        print("TEST DATA:", test_index)
```

ภาพที่ 3.44 เครื่องมือที่ใช้ในการทดสอบอัลกอริทึมของแบบจำลองที่ 2

```
TEST DATA: [ 59  61  65  68  72  73  75  77  80  92 101 103 105 106 107 108 111 121
 126 128 129 130 189 190 192 193 196 197 198 199 200 201 202 203 204 205
 206 207 208 209 210 211 212 213 214 215 216 217 224 230 231 232 233 236
 237 238 239 240 242 244 245 246 247 248 249 250 251 253 254 258 259 261
 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298
 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316
 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334]
```

```
X_Test:
                                                final_word_token
59  ขอ เพิ่ม จำนวน เครื่อง เพื่อ รองรับ สำหรับ โปร...
61  ขอ เพิ่ม จำนวน เครื่อง ใช้ ใน การ พิมพ์ รายงาน...
65  ขอ เพิ่ม จำนวน ชุด เนื่อง จก ศาลจังหวัด ลำพูน ...
68  ขอ ทดแทน จำนวน เครื่อง เนื่องจาก ของ เดิม ที่ ...
72  ขอ ทดแทน จำนวน เครื่อง ใช้ พิมพ์ เอกสาร เป็นปร...
y_Test:
 59    no
61    no
65    no
68    no
72    no
```

ภาพที่ 3.45 ตัวอย่างชุดข้อมูลที่ใช้ในการทดสอบอัลกอริทึมของแบบจำลองที่ 2

```
#define function
def myconfusion_matrix(y_true,y_pred,label_list=["yes", "no"],file_name="confusion_result.png"):
    result = confusion_matrix(y_true, y_pred, labels=label_list)
    print(result)
    cm_display = ConfusionMatrixDisplay(confusion_matrix=result, display_labels=label_list)
    cm_display.plot()
    plt.show()
    cm_display.figure_.savefig(file_name, dpi=300)
```

ภาพที่ 3.46 เครื่องมือสำหรับประเมินประสิทธิภาพแบบจำลองที่ 2

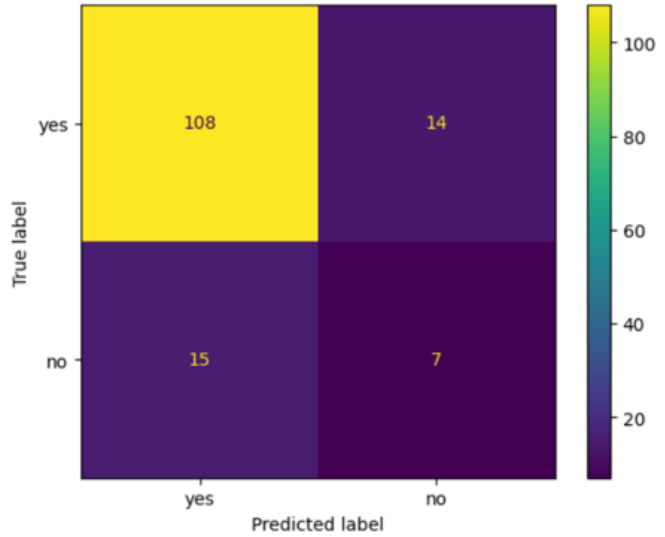
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์


```
Report Model: Decision Tree Algorithm ----------------
              precision    recall  f1-score   support

          no       0.32      0.33      0.33        21
         yes       0.89      0.88      0.88       123

    accuracy                           0.80       144
   macro avg       0.60      0.61      0.60       144
weighted avg       0.80      0.80      0.80       144

[[108  14]
 [ 15   7]]
```



ภาพที่ 3.47 ตัวอย่าง Confusion Matrix แสดงผลประสิทธิภาพและผลการทำนายของแบบจำลองที่ 2

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 4
ผลการศึกษา

การประยุกต์ใช้ความรู้เกี่ยวกับการทำเหมืองข้อความ การประมวลภาษาธรรมชาติ ผ่านกระบวนการ CRISP-DM ซึ่งเป็นเครื่องมือที่ช่วยกำหนดแนวทางและวิธีการดำเนินงานวิจัยให้ถูกต้องครบถ้วนในแต่ละขั้นตอน โดยเริ่มตั้งแต่การทำความเข้าใจระบบงาน วิเคราะห์ปัญหาของหน่วยยงานที่ต้องดำเนินการแก้ไขหรือปรับปรุงให้มีประสิทธิภาพมากขึ้น รวมถึงพิจารณาข้อมูลที่มีการรวบรวมและจัดเก็บไว้ของหน่วยงานซึ่งเป็นสิ่งสำคัญที่จะนำมาใช้สำหรับจัดทำแบบจำลองหรือโมเดลที่ใช้แก้ไขปัญหาหรือปรับปรุงประสิทธิภาพของระบบงานตามที่ได้กำหนดเป็นขอบเขตงานไว้ จากนั้นจึงอาศัยความก้าวหน้าและทันสมัยของระบบซอฟต์แวร์และภาษาโปรแกรมที่มีการพัฒนาเป็นอัลกอริทึมต่างๆ ที่ช่วยให้สามารถจัดทำแบบจำลองและทดสอบประสิทธิภาพแบบจำลองเพื่อนำไปพัฒนาต่อยอดเป็นระบบต้นแบบในการแก้ไขปัญหาและปรับปรุงประสิทธิภาพของงาน ทั้งนี้ หลังจากที่ผู้วิจัยได้ดำเนินการทดลองสร้างแบบจำลองโดยใช้ความรู้และเทคนิคที่กล่าวข้างต้นเรียบร้อยแล้ว มีผลการดำเนินงานตามขั้นตอนที่ได้ดำเนินการในบทที่ 3 (ระเบียบวิธีวิจัย) ซึ่งผู้วิจัยเห็นว่า ต้องนำมาพิจารณาและทบทวนเพื่อให้งานวิจัยนี้มีคุณภาพและประสิทธิภาพมากขึ้น ซึ่งสามารถสรุปรายละเอียดผลการดำเนินงานวิจัยได้ ดังนี้

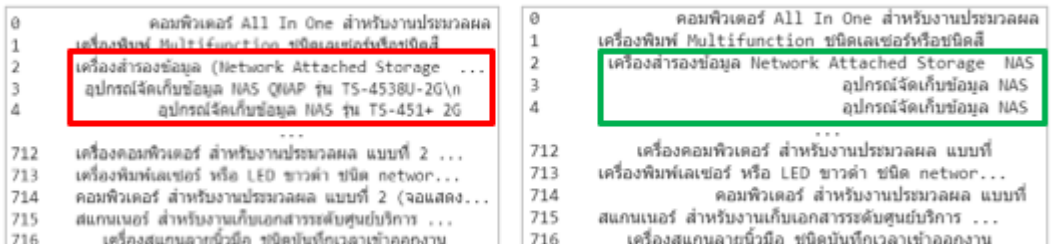
4.1 การทำความสะอาดข้อมูล

การดำเนินงานวิจัยในขั้นตอนนี้ ในระยะแรกผู้วิจัยใช้วิธีทำความสะอาดข้อมูลโดยการลบคำเฉพาะ อาทิ เช่น ชื่อยี่ห้อ รุ่น จำนวนหน้า เป็นต้น และลบเครื่องหมายต่างๆ ที่เราไม่ต้องการ เช่น (!€@%#*&~) ซึ่งข้อมูลคำของบประมาณและรายละเอียดเหตุผลความจำเป็นที่จะนำไปใช้ในขั้นตอนการตัดคำยังคงมีเครื่องหมายและอักขระพิเศษในภาษาปะปนมาด้วย

4.2 การประมวลผลภาษาธรรมชาติ

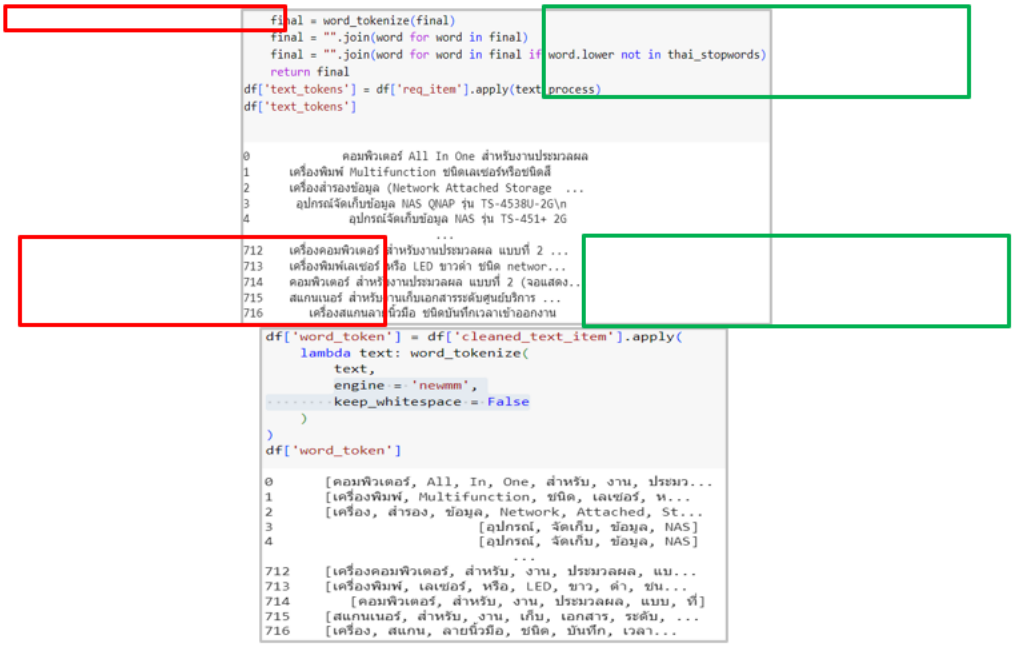
(1) การลบอักขระพิเศษ

ผลสืบเนื่องจากการทำความสะอาดข้อมูลจากในขั้นตอน 4.1 ข้อความที่จะนำไปสร้างแบบจำลองยังคงมีเครื่องหมาย และอักขระพิเศษปะปนในข้อความอยู่ ทำให้ผู้วิจัยหาจำเป็นต้องหาวิธีการแก้ไขปัญหาดังกล่าว โดยนำเครื่องมือในภาษาไพธอนซึ่งมีไลบรารีที่ชื่อว่า RegEx (Reggular Expression) ซึ่งสามารถกำหนดรูปแบบการกรองข้อความได้หลายรูปแบบ เช่น [^\u0E00-\u0E7Fa-zA-Z']|['']|[|] เข้ามาช่วยในการทำความสะอาดข้อมูล ปรากฏว่าสามารถกรองคำจากข้อมูลคำของบประมาณและรายละเอียดเหตุผลความจำเป็นที่มีทั้งตัวอักษรภาษาไทย ตัวเลข และภาษาอังกฤษได้ดีขึ้น ทำให้ช่วยลดจำนวนคำที่จะนำไปกำหนดเป็นฟีเจอร์ในการสร้างแบบจำลองได้ตามไปด้วย ดังภาพที่ 50



ภาพที่ 4.1 เปรียบเทียบการทำความสะอาดข้อมูลโดยใช้ไลบรารี RegEx

(2) การตัดคำ ขั้นตอนนี้ ผู้วิจัยได้ใช้โมดุล word_tokenize ในไลบรารีของ sklearn มาใช้ในการตัดคำ โดยการทดลองในระยะแรกไม่ได้กำหนดพารามิเตอร์ให้กับเครื่องมือนี้ ทำให้ผลลัพธ์ของคำที่ตัดได้ในข้อความคำของบประมาณและเหตุผลความจำเป็นมีลักษณะของคำหลายคำที่อยู่ติดกันเกือบจะเป็นประโยคยาวๆ ยกตัวอย่างเช่น “สำหรับงานเก็บเอกสารระดับศูนย์บริการ” ซึ่งน่าจะเป็นผลลัพธ์ที่ไม่ถูกต้อง ผู้วิจัยจึงต้องศึกษาข้อมูลเพิ่มเติมเกี่ยวกับการใช้เทคนิคนี้ และสามารถใช้วิธีกำหนดค่าพารามิเตอร์ของเทคนิค word_tokenize ให้ครบถ้วน ด้วยการใช้ engine = “newmm” ที่สามารถตัดข้อความได้เป็นคำๆ ได้ดีขึ้น ดังภาพที่ 4.2



ภาพที่ 4.2 เปรียบเทียบการตัดคำโดยใช้ engine “newmm”

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

(3) การสร้างคลังคำศัพท์ (Bag of Words) และสร้างฟีเจอร์ (Feature) ในขั้นตอนนี้ผู้วิจัยใช้วิธีการแปลงคำที่ตัดได้จากกระบวนการตัดคำมาสร้างคลังคำศัพท์ โดยการแปลงคำที่ไม่ซ้ำกันเป็นค่า word id และจัดเก็บในรูปแบบ vector ด้วยโมดุล CountVectorizer ในไลบรารีของ sklearn และสร้างฟีเจอร์ด้วยการนับจำนวนครั้งที่แต่ละ word id ปรากฏในแต่ละข้อความของรายการคำของบประมาณในแบบจำลองที่ 1 และรายการเหตุผลความจำเป็นในแบบจำลองที่ 2

ทั้งนี้ เนื่องจากข้อความในรายการคำของบประมาณที่ใช้สร้างคลังคำศัพท์และฟีเจอร์สำหรับสร้างแบบจำลองที่1 และข้อความเหตุผลความจำเป็นที่ใช้สร้างคลังคำศัพท์และสร้างฟีเจอร์ในแบบจำลองที่ 2 มีความหลากหลายเป็นจำนวนมาก ซึ่งผู้วิจัยพบว่า ฟีเจอร์ที่ใช้สำหรับสร้างแบบจำลองที่ 1 มีจำนวนเท่ากับ 272 ฟีเจอร์ และแบบจำลองที่ 2 มีจำนวนเท่ากับ 1,540 ฟีเจอร์ ซึ่งแสดงตัวอย่างคลังคำศัพท์และฟีเจอร์ที่ได้ในตารางที่ 4.1

ตารางที่ 4.1 ตัวอย่างคลังคำศัพท์ และฟีเจอร์ แบบจำลองที่ 1 และแบบจำลองที่ 2

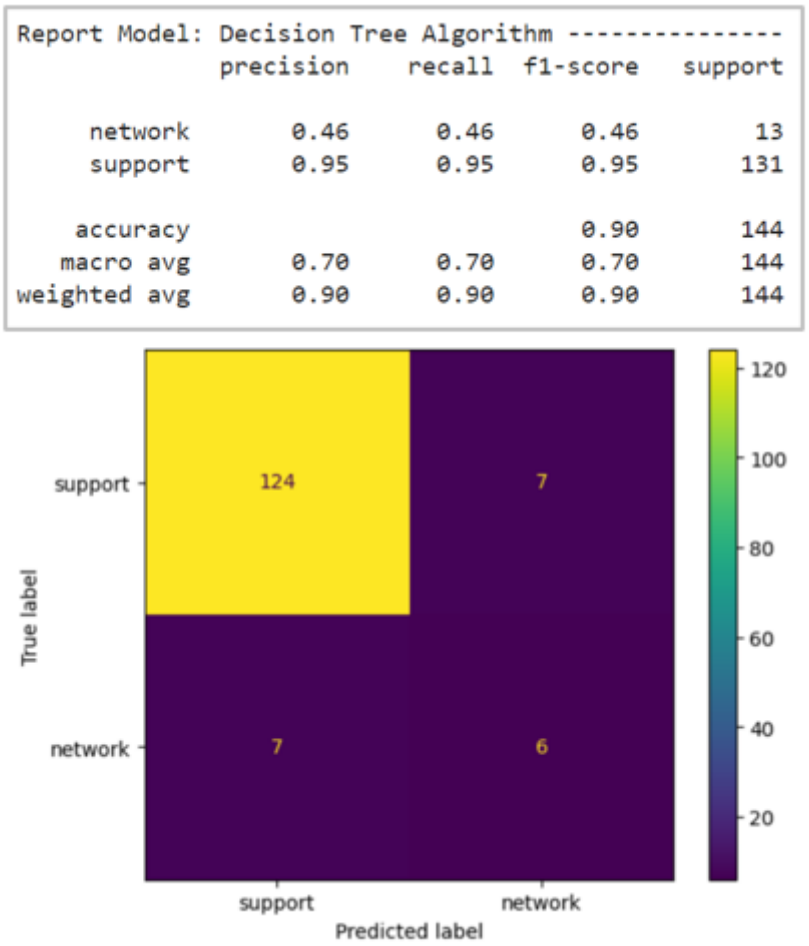
แบบจำลอง	(กลุ่มตัวอย่าง) คลังคำศัพท์ และฟีเจอร์
แบบจำลองที่ 1	{'คอมพิวเตอร์': 95, 'All': 2, 'In': 22, 'One': 42, 'สำหรับ': 200, 'งาน': 98, 'ประมวลผล': 151, 'เครื่องพิมพ์': 230, 'Multifunction': 37, 'ชนิด': 109, 'เลเซอร์': 238, 'หรือ': 210, 'สี': 201, 'เครื่อง': 228, 'สำรอง': 199, 'ข้อมูล': 85, 'Network': 39, 'Attached': 3, 'Storage': 51, 'NAS': 38, 'อุปกรณ์': 221, 'จัดเก็บ': 106, 'สแกน': 205, 'ลายนิ้วมือ': 184, 'เครื่องคอมพิวเตอร์': 229, 'Server': 50, 'Tower': 55, 'กระดาน': 75, 'อิเล็กทรอนิกส์': 219, 'eBoard': 63, 'ประชุม': 150, 'ทางไกล': 130, 'ผ่าน': 156, 'จอภาพ': 101}
แบบจำลองที่ 2	{'ขอ': 195, 'เพิ่ม': 1336, 'จำนวน': 373, 'เครื่อง': 1260, 'เนื่องจาก': 1306, 'นิติกร': 623, 'ประจำ': 689, 'องค์': 1172, 'คณะ': 229, 'ผู้พิพากษา': 759, 'ผู้ช่วย': 749, 'มี': 851, 'ไม่': 1534, 'เพียงพอ': 1343, 'จึง': 379, 'นำ': 613, 'เทคโนโลยี': 1300, 'มา': 839, 'ใช้': 1492, 'สนับสนุน': 1034, 'การ': 113, 'ค้นหา': 324, 'แนว': 1420, 'สั่ง': 1062, 'ออก': 1185, 'คำพิพากษา': 288, 'ฎีกา': 424, 'กฎหมาย': 78, 'ที่': 571, 'เกี่ยวข้อง': 1237, 'รวมถึง': 889, 'เอกสาร': 1391}

4.3 การสร้างแบบจำลอง และประเมินประสิทธิภาพของแบบจำลอง

การเลือกเทคนิคสำหรับใช้สร้างแบบจำลองที่ 1 จำแนกกลุ่มงานที่รับผิดชอบคำของบประมาณ และแบบจำลองที่ 2 จำแนกผลการพิจารณาคำของบประมาณจากเหตุผลความจำเป็น ผู้วิจัยใช้ อัลกอริทึมต่างๆ 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 รวมถึงใช้วิธีการวัด

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ประสิทธิภาพของแบบจำลองแบบ classification ด้วยค่า confusion matrix ซึ่งมีผลการประเมินประสิทธิภาพของแบบจำลองที่ 1 และแบบจำลองที่ 2 ในแต่ละอัลกอริทึม ดังภาพที่ 4.3 – 4.22

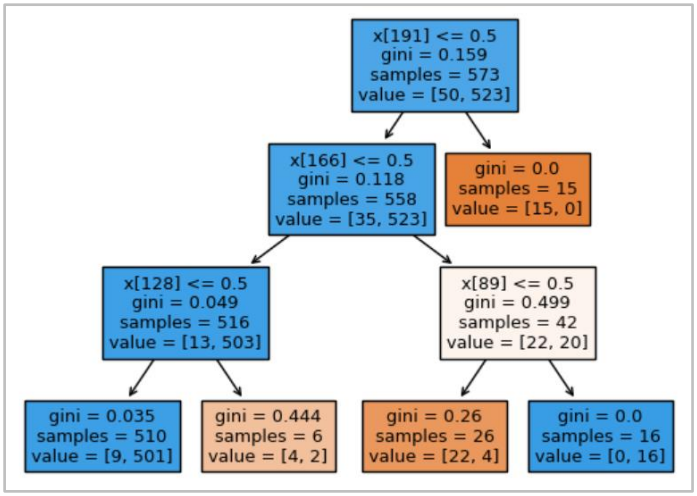


Report Model: Decision Tree Algorithm ----------

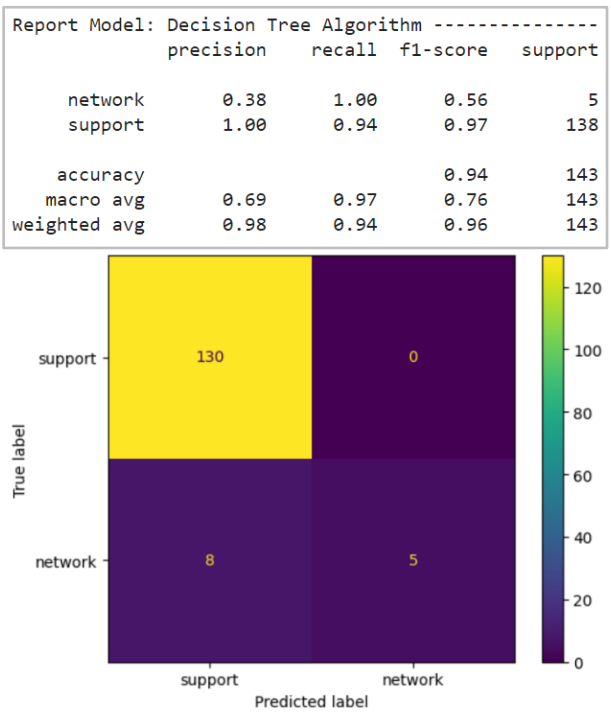
	precision	recall	f1-score	support
network	0.46	0.46	0.46	13
support	0.95	0.95	0.95	131
accuracy			0.90	144
macro avg	0.70	0.70	0.70	144
weighted avg	0.90	0.90	0.90	144

ภาพที่ 4.3 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 1 ด้วยอัลกอริทึม Decision Tree (ข้อมูลชุดที่1)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 4.4 ตัวอย่าง Decision Tree แบบจำลองที่ 1 (ข้อมูลชุดที่1)

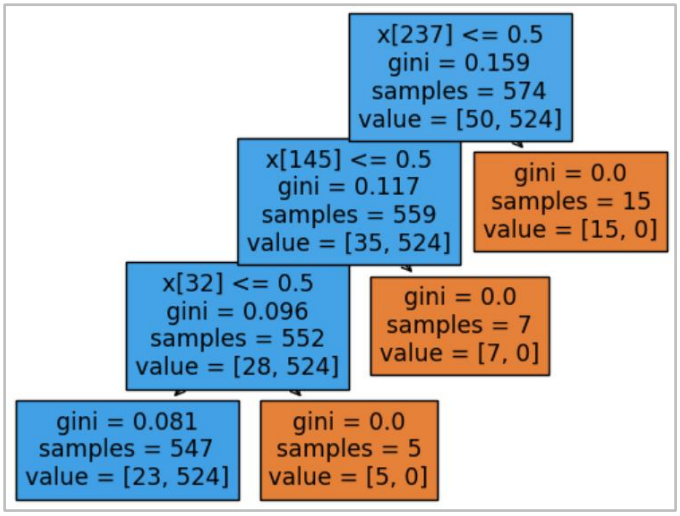


Report Model: Decision Tree Algorithm ----------

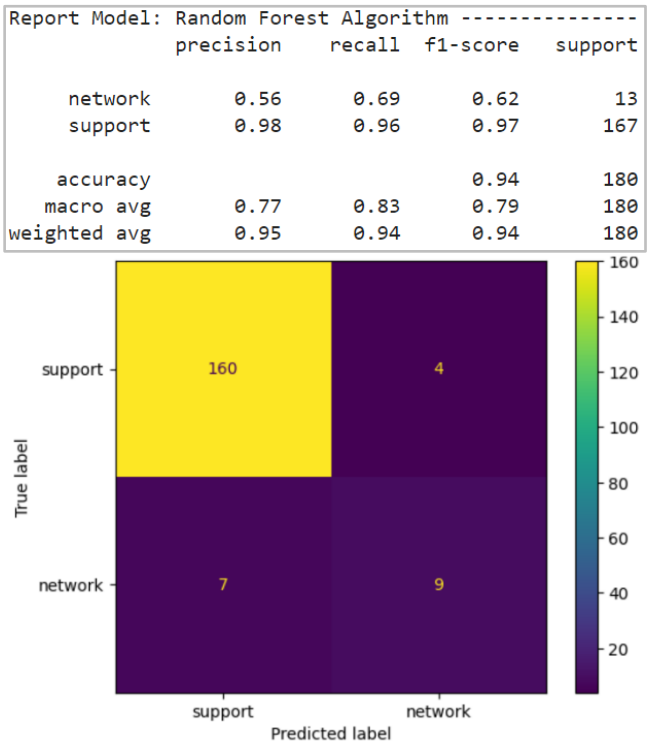
	precision	recall	f1-score	support
network	0.38	1.00	0.56	5
support	1.00	0.94	0.97	138
accuracy			0.94	143
macro avg	0.69	0.97	0.76	143
weighted avg	0.98	0.94	0.96	143

ภาพที่ 4.5 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 1 ด้วยอัลกอริทึม Decision Tree (ข้อมูลที่ 5)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 4.6 ตัวอย่าง Decision Tree แบบจำลองที่ 1 (ข้อมูลชุดที่ 5)



Report Model: Random Forest Algorithm ----------

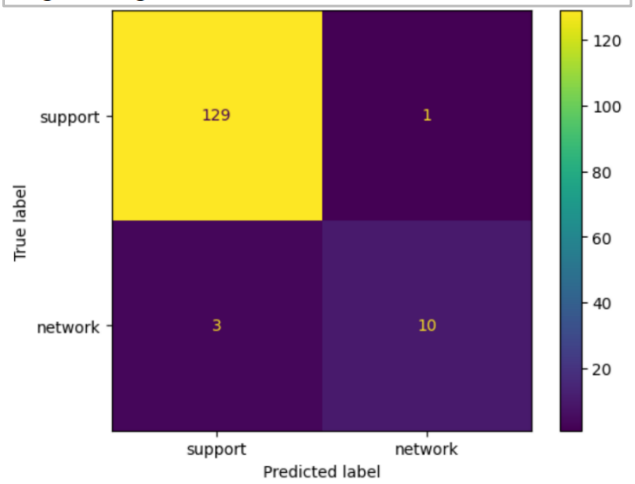
	precision	recall	f1-score	support
network	0.56	0.69	0.62	13
support	0.98	0.96	0.97	167
accuracy			0.94	180
macro avg	0.77	0.83	0.79	180
weighted avg	0.95	0.94	0.94	180

ภาพที่ 4.7 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 1 ด้วยอัลกอริทึม Random Forest (ข้อมูลชุดที่ 1)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Report Model: Random Forest Algorithm ----------

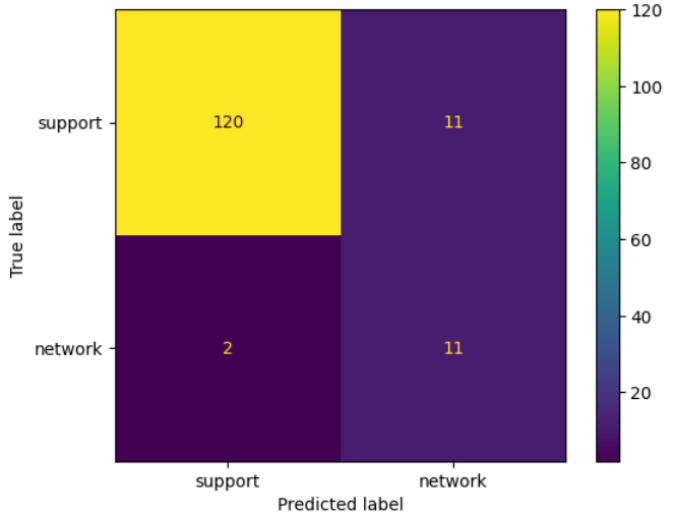
	precision	recall	f1-score	support
network	0.77	0.91	0.83	11
support	0.99	0.98	0.98	132
accuracy			0.97	143
macro avg	0.88	0.94	0.91	143
weighted avg	0.98	0.97	0.97	143



ภาพที่ 4.8 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 1 ด้วยอัลกอริทึม Random Forest (ข้อมูลชุดที่ 5)

Report Model: Naïve Bayes Algorithm ----------

	precision	recall	f1-score	support
network	0.88	0.50	0.64	28
support	0.91	0.99	0.95	152
accuracy			0.91	180
macro avg	0.89	0.74	0.79	180
weighted avg	0.91	0.91	0.90	180

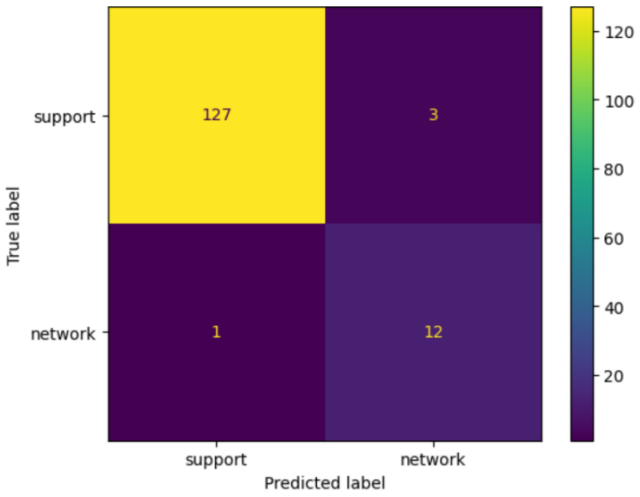


ภาพที่ 4.9 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 1 ด้วยอัลกอริทึม Naïve Bayes (ข้อมูลชุดที่ 1)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Report Model: Naïve Bayes Algorithm ----------

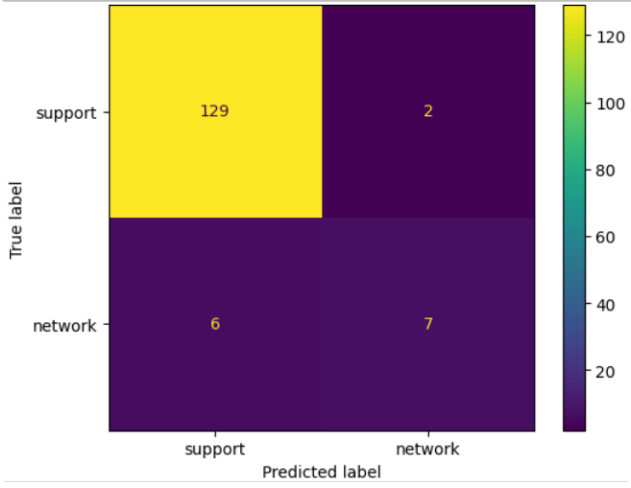
	precision	recall	f1-score	support
network	0.92	0.80	0.86	15
support	0.98	0.99	0.98	128
accuracy			0.97	143
macro avg	0.95	0.90	0.92	143
weighted avg	0.97	0.97	0.97	143



ภาพที่ 4.10 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 1 ด้วยอัลกอริทึม Naïve Bayes (ข้อมูลชุดที่ 5)

Report Model: Logistic Regression ----------

	precision	recall	f1-score	support
network	0.54	0.78	0.64	9
support	0.98	0.96	0.97	135
accuracy			0.94	144
macro avg	0.76	0.87	0.80	144
weighted avg	0.96	0.94	0.95	144

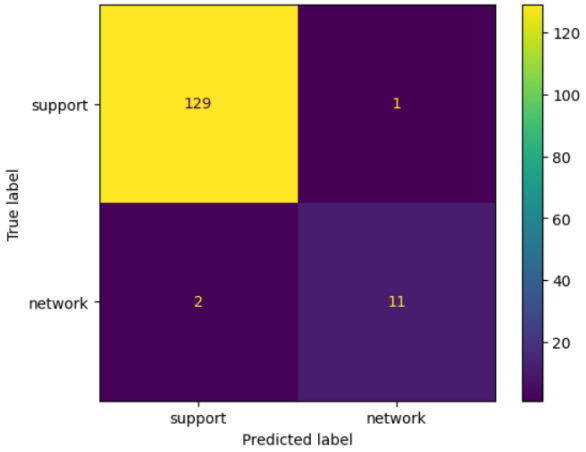


ภาพที่ 4.11 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 1 ด้วยอัลกอริทึม Logistic Regression (ข้อมูลชุดที่ 1)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Report Model: Logistic Regression ----------

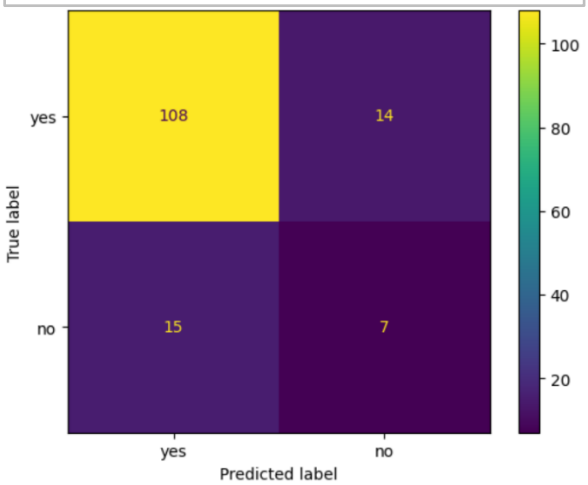
	precision	recall	f1-score	support
network	0.85	0.92	0.88	12
support	0.99	0.98	0.99	131
accuracy			0.98	143
macro avg	0.92	0.95	0.93	143
weighted avg	0.98	0.98	0.98	143



ภาพที่ 4.12 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 1 ด้วยอัลกอริทึม Logistic Regression (ข้อมูลชุดที่ 5)

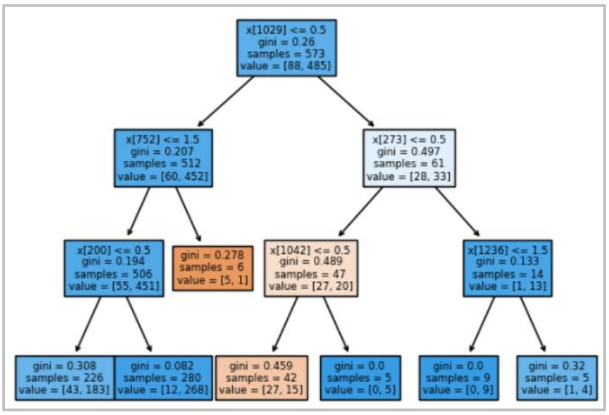
Report Model: Decision Tree Algorithm ----------

	precision	recall	f1-score	support
no	0.32	0.33	0.33	21
yes	0.89	0.88	0.88	123
accuracy			0.80	144
macro avg	0.60	0.61	0.60	144
weighted avg	0.80	0.80	0.80	144

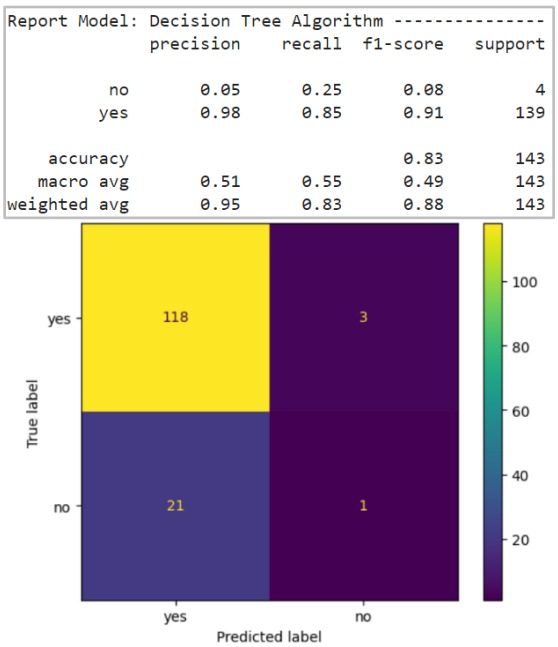


ภาพที่ 4.13 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Decision Tree (ข้อมูลชุดที่1)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

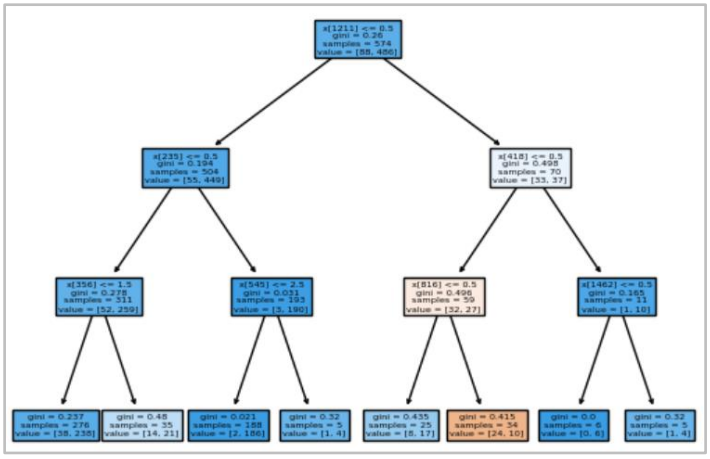


ภาพที่ 4.14 ตัวอย่าง Decision Tree แบบจำลองที่ 2 (ข้อมูลชุดที่1)



ภาพที่ 4.15 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Decition Tree (ข้อมูลที่ 5)

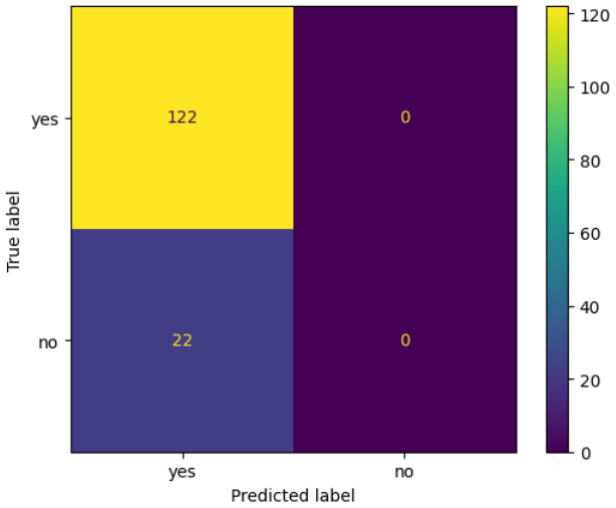
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 4.16 ตัวอย่าง Decision Tree แบบจำลองที่ 2 (ข้อมูลชุดที่ 5)

Report Model: Random Forest Algorithm ----------

	precision	recall	f1-score	support
no	0.00	0.00	0.00	0
yes	1.00	0.85	0.92	144
accuracy			0.85	144
macro avg	0.50	0.42	0.46	144
weighted avg	1.00	0.85	0.92	144

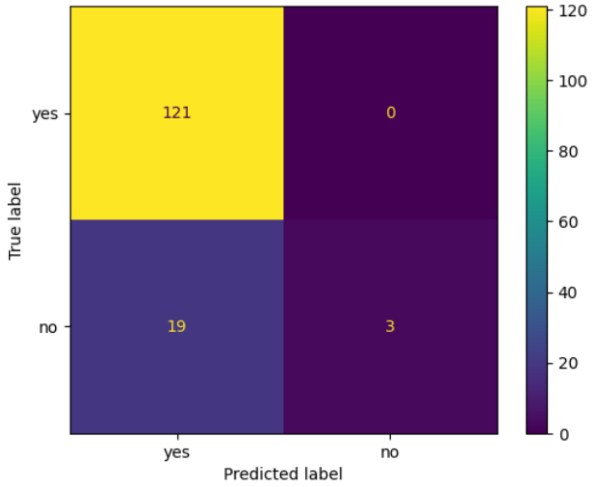


ภาพที่ 4.17 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Random Forest (ชุดข้อมูลที่ 1)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Report Model: Random Forest Algorithm ----------

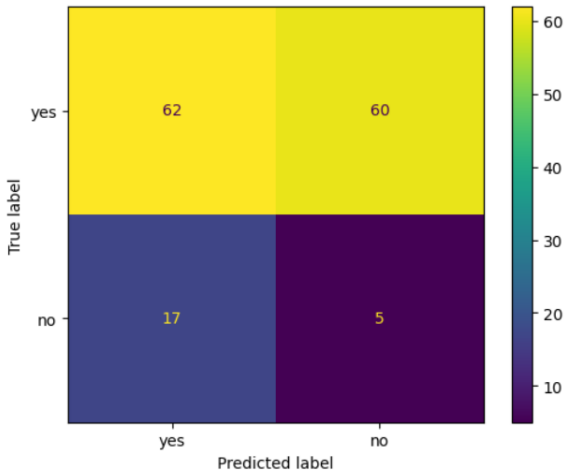
	precision	recall	f1-score	support
no	0.14	1.00	0.24	3
yes	1.00	0.86	0.93	140
accuracy			0.87	143
macro avg	0.57	0.93	0.58	143
weighted avg	0.98	0.87	0.91	143



ภาพที่ 4.18 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Random Forest (ชุดข้อมูลที่ 5)

Report Model: Naïve Bayes Algorithm ----------

	precision	recall	f1-score	support
no	0.23	0.08	0.11	65
yes	0.51	0.78	0.62	79
accuracy			0.47	144
macro avg	0.37	0.43	0.37	144
weighted avg	0.38	0.47	0.39	144

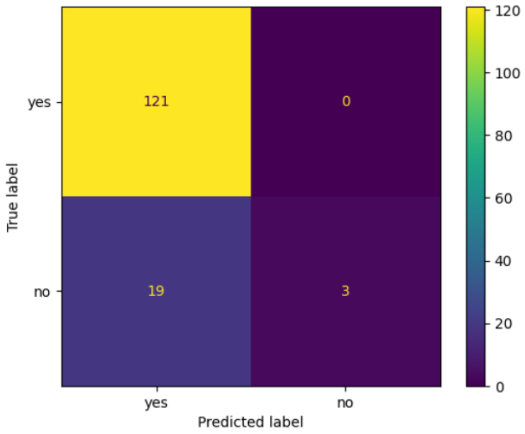


ภาพที่ 4.19 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Naïve Bayes (ชุดข้อมูลที่ 1)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Report Model: Naïve Bayes Algorithm ----------

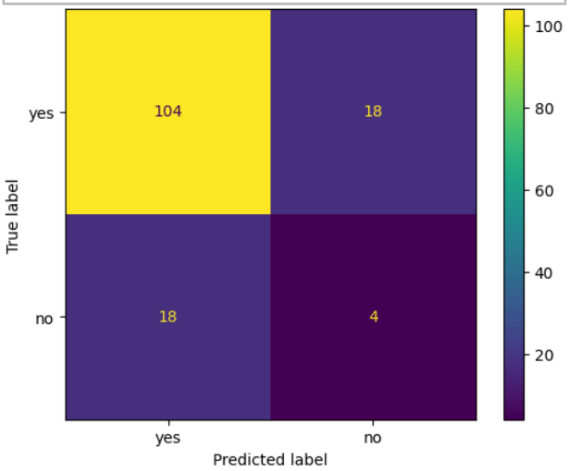
	precision	recall	f1-score	support
no	0.14	1.00	0.24	3
yes	1.00	0.86	0.93	140
accuracy			0.87	143
macro avg	0.57	0.93	0.58	143
weighted avg	0.98	0.87	0.91	143



ภาพที่ 4.20 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Naïve Bayes (ชุดข้อมูลที่ 5)

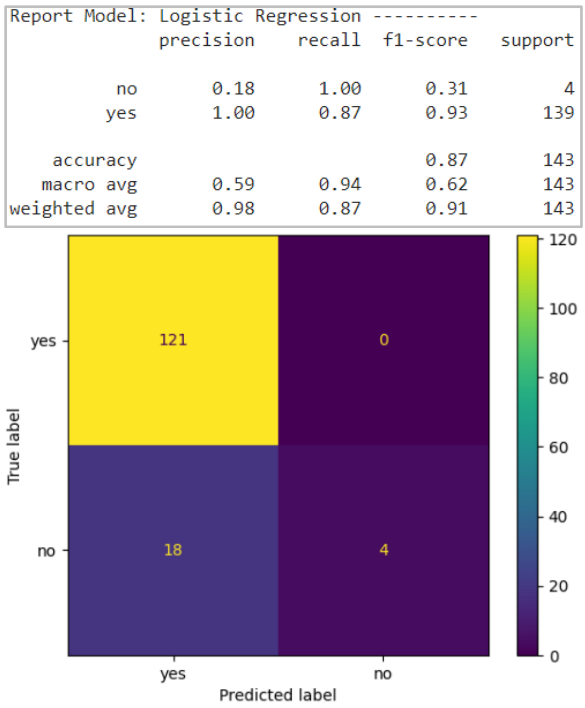
Report Model: Logistic Regression ----------

	precision	recall	f1-score	support
no	0.18	0.18	0.18	22
yes	0.85	0.85	0.85	122
accuracy			0.75	144
macro avg	0.52	0.52	0.52	144
weighted avg	0.75	0.75	0.75	144



ภาพที่ 4.21 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Logistic Regression (ข้อมูลชุดที่ 1)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 4.22 ตัวอย่างผลประสิทธิภาพของแบบจำลองที่ 2 ด้วยอัลกอริทึม Logistic Regression (ข้อมูลชุดที่ 5)

ทั้งนี้ มีผลประเมินประสิทธิของแบบจำลองที่ 1 และแบบที่ 2 ในภาพรวม ดังที่แสดงในตารางที่ 4.2-4.3 (ค่า Accuracy และค่า Macro F1-Score ที่แสดงในตาราง เป็นค่าเฉลี่ยที่ได้จากการทดสอบ 5-Folds Cross Validation)

ตารางที่ 4.2 การประเมินประสิทธิภาพของแบบจำลองที่ 1

ตัวชี้วัดประสิทธิภาพ	อัลกอริทึม			
	Decition Tree	Random Forest	Naïve Bayes	Logistic Regression
Accuracy	95 %	96 %	95 %	95 %
Macro F1-Score	83 %	87 %	86 %	90 %

จากข้อมูลในตารางที่ 4.2 จะเห็นว่า แบบจำลองที่ 1 อัลกอริทึมที่มีค่าประสิทธิภาพความถูกต้อง (Accuracy = ค่าที่โมเดลทายถูกทั้งหมด/ค่าทั้งหมด) สูงที่สุด คือ Random Forest เท่ากับ 96 % รองลงมา

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ได้แก่ Decision Tree, Naïve Bayes และ Logistic Regression แต่เมื่อพิจารณาค่าเฉลี่ยความถูกต้องแม่นยำ (Macro F1-Score = ค่าเฉลี่ยที่ได้จากการเอาค่า precision และ recall มาคำนวณรวมกัน) พบว่า อัลกอริทึม Logistic Regression มีค่าสูงที่สุด เท่ากับ 90 % ทั้งนี้ ในส่วนของการกำหนดค่าพารามิเตอร์ของแต่ละอัลกอริทึม ผู้วิจัยได้ทำการทดลองโดยใช้ค่าตั้งต้นที่กำหนดไว้ในไลบรารีของ sklearn และทดลองปรับค่าพารามิเตอร์ โดยแสดงไว้ในบทที่ 3 ระเบียบวิธีวิจัย

ตารางที่ 4.3 การประเมินประสิทธิภาพของแบบจำลองที่ 2

ตัวชี้วัดประสิทธิภาพ	อัลกอริทึม			
	Decition Tree	Random Forest	Naïve Bayes	Logistic Regression
Accuracy	82 %	82 %	69 %	78 %
Macro F1-Score	56 %	50 %	53 %	56 %

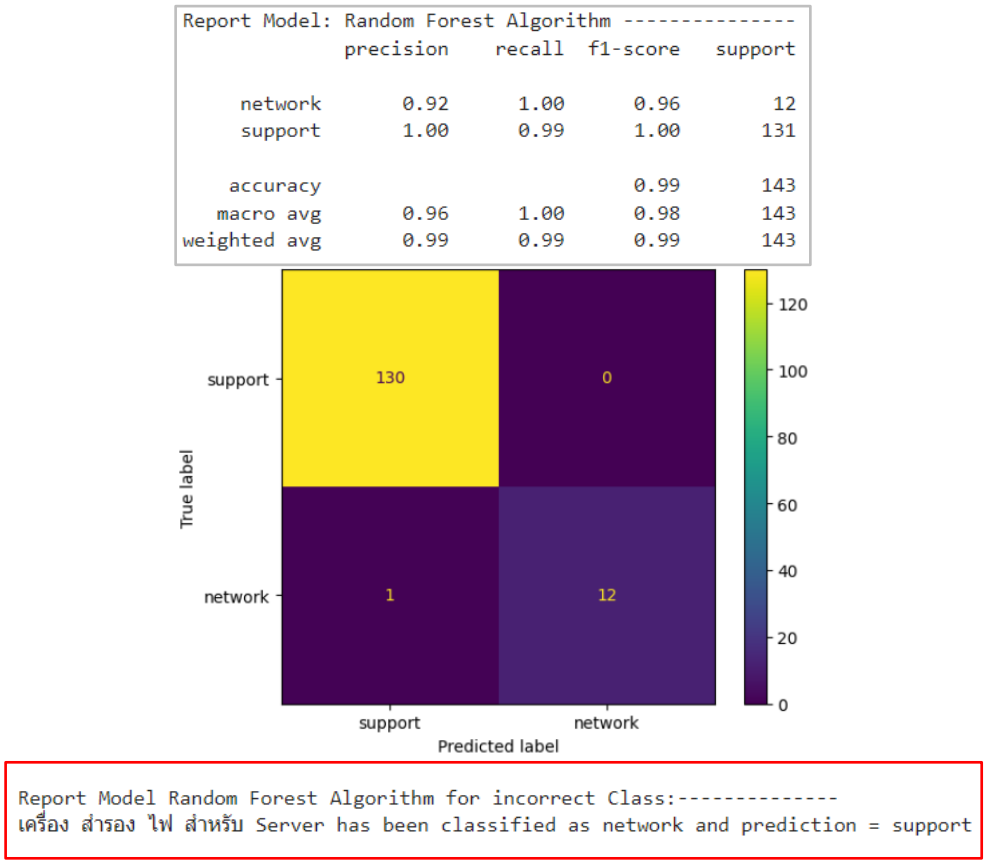
จากข้อมูลในตารางที่ 4.3 จะเห็นว่า แบบจำลองที่ 2 อัลกอริทึมที่มีค่าประสิทธิภาพความถูกต้อง (Accuracy) สูงที่สุด คือ Random Forest และ Decision Tree เท่ากับ 82 % รองลงมาได้แก่ Logistic Regression และ Naïve Bayes แต่เมื่อพิจารณาค่าเฉลี่ยความถูกต้องแม่นยำ (Macro F1-Score) พบว่า อัลกอริทึม Logistic Regression มีค่าสูงที่สุด เท่ากับ 56 %

ทั้งนี้ จากการสังเกตผลการประเมินประสิทธิภาพของแบบจำลองที่ 1 และ แบบจำลองที่ 2 ผู้วิจัยเห็นว่า ค่าความถูกต้องของแบบจำลอง 2 ต่ำกว่าแบบจำลองที่ 1 แม้ว่าจะใช้เทคนิคอัลกอริทึมเดียวกัน ซึ่งปัจจัยที่น่าจะส่งผลมากที่สุด คือ จำนวนฟีเจอร์ชุดข้อมูล Training data ของแบบจำลองที่ 2 ซึ่งมีมากกว่าแบบจำลองที่ 1 หลายเท่าตัว (แบบจำลองที่ 1 มีจำนวนฟีเจอร์มากที่สุดในชุดข้อมูล Training data = 272 ฟีเจอร์ และแบบจำลองที่ 2 มีจำนวนฟีเจอร์มากที่สุดในชุดข้อมูล Training data = 1,540 ฟีเจอร์)

4.4 การวิเคราะห์ผลการทำนายของแบบจำลอง

แม้ว่าการประเมินสิทธิภาพความถูกต้องของแบบจำลองจะให้ค่าความถูกต้องที่มีเกณฑ์อยู่ในระดับสูง แต่เพื่อนำแบบจำลองที่ได้ไปพัฒนาต่อยอดให้เกิดประโยชน์ในระบบงานจริง ผู้วิจัยได้รับคำแนะนำจากอาจารย์ที่ปรึกษาให้ตรวจสอบค่าของชุดข้อมูลที่ใช้สร้างแบบจำลองหรือโมเดลทำนายผลผิดพลาด โดยเลือก

อัลกอริทึม Random Forest ที่ให้ค่าความถูกต้องสูงที่สุด ซึ่งมีผลสรุปข้อมูลซึ่งแบบจำลองทำนายผลผิดพลาด ดังภาพที่ 4.23-4.24



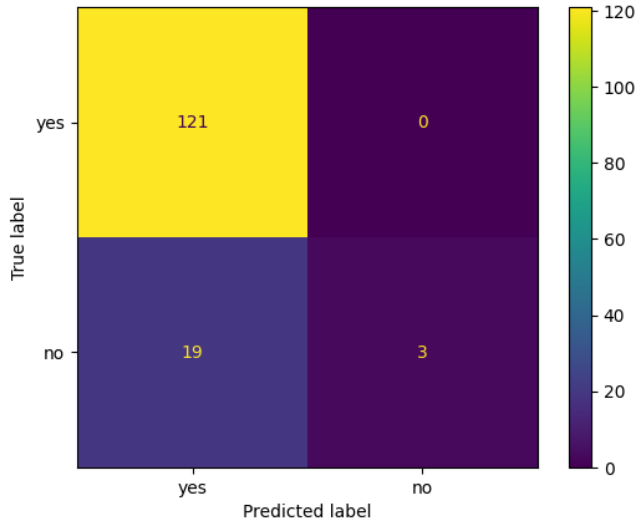
ภาพที่ 4.23 ตัวอย่างชุดข้อมูลที่ทำนายผลผิดพลาดด้วยอัลกอริทึม Random Forest ของแบบจำลองที่ 1

จากการตรวจสอบพบว่า ชุดข้อมูล Training data ที่นำไปสร้างแบบจำลองด้วยอัลกอริทึม Random Forest เป็นชุดข้อมูลที่มีคลาสคำตอบ support เป็นจำนวนมาก โดยชุดข้อมูลที่มีคลาสคำตอบ network มีจำนวนน้อยมาก และแต่ละชุดข้อมูล Training data ที่มีฟีเจอร์คำว่า “เครื่อง สำรอง ไฟฟ้า” จะมีคลาสคำตอบเป็น support และไม่พบชุดข้อมูล Training data ที่มีฟีเจอร์คำว่า “เครื่อง สำรอง ไฟ สำหรับ Server จึงมีความน่าจะเป็นที่ส่งผลให้แบบจำลองโมเดลทำนายผลผิดพลาดเป็นคลาสคำตอบ support

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Report Model: Random Forest Algorithm ----------

	precision	recall	f1-score	support
no	0.14	1.00	0.24	3
yes	1.00	0.86	0.93	140
accuracy			0.87	143
macro avg	0.57	0.93	0.58	143
weighted avg	0.98	0.87	0.91	143



Report Model Random Forest Algorithm for incorrect Class:----------

ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ พิจารณาคดี ผ่าน จอภาพ เครื่อง รองรับ สถานการณ์ การ แพร่ระบาด ของ โรค โควิด เนื่องจาก ปัจจุบัน ได้ นำ เครื่องคอมพิวเตอร์ ของ หัวหน้าส่วน ใกล้เกลี้ย และ ประชุม ข้อพิพาท และ ห้อง เก็บ สำนวน คดี ที่ ใช้ สืบค้น สำนวน คดี มา ใช้งาน ดังกล่าว
ขอ ทดแทน จำนวน เครื่อง เนื่องจาก เป็น ครุภัณฑ์ ที่ ได้รับ จัดสรร ตั้งแต่ ปี มี อายุการใช้งาน เกิน กว่า ปี ใช้ ประจำ กลุ่ม งาน ใกล้เกลี้ย และ ประชุม ข้อพิพาท มี สภาพ ชำรุด และ ไม่ คุ้มค่า กับ การซ่อมแซม และ ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ ประจำ กลุ่ม เจ้าพนักงาน คดี เนื่องจา
ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ งานประจำ ส่วน เจ้าพนักงาน คดี เนื่องจาก เป็น ส่วน งาน ตั้งขึ้น ใหม่ และ ปัจจุบัน ยัง ไม่ มี ครุภัณฑ์ ดังกล่าว ใช้งาน has been classified as no and prediction = yes
ขอ ทดแทน จำนวน เครื่อง เนื่องจาก เป็น ครุภัณฑ์ ที่ ได้รับ จัดสรร ตั้งแต่ ปี มี อายุการใช้งาน เกิน กว่า ปี ใช้ งานประจำ ส่วน งาน บริการ ประชาชน และ ประชาสัมพันธ์ ปัจจุบัน มี สภาพ ชำรุด และ ไม่ คุ้มค่า กับ การซ่อมแซม has been classified as no and prediction = yes
ขอ ทดแทน จำนวน เครื่อง เนื่องจาก เป็น ครุภัณฑ์ ที่ ได้รับ จัดสรร ตั้งแต่ ปี มี อายุการใช้งาน เกิน กว่า ปี ปัจจุบัน มี สภาพ ชำรุด และ ไม่ คุ้มค่า กับ การซ่อมแซม has been classified as no and prediction = yes
ขอ ทดแทน จำนวน เครื่อง เนื่องจาก เป็น ครุภัณฑ์ ที่ ได้รับ จัดสรร ตั้งแต่ ปี และ มี อายุการใช้งาน เกิน กว่า ปี มี สภาพ ชำรุด และ ไม่ คุ้มค่า กับ การซ่อมแซม has been classified as no and prediction = yes
ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ งานประจำ ส่วน บริหาร จัดการ คดี และ ส่วน ช่วย พิจารณาคดี เพื่อ ใช้ ใน สแกน คำพิพากษา ศาลั่ง สำนวน คดี เป็นต้น เพื่อ รองรับ สารบบ ค่าคู่ความ และ ผล การ ส่ง ค่าคู่ความ อิเล็กทรอนิกส์ เนื่องจาก มี ครุภัณฑ์ ดังกล่าว ไม่ เพียง พอกับ การใช้งาน h
ขอ เพิ่ม จำนวน ชุด สำหรับ ใช้ เป็น เครื่องคอมพิวเตอร์ จัดเก็บ ข้อมูล สแกน ลายนิ้วมือ ลงเวลา เข้าออก เนื่องจาก หน่วยงาน ได้ นำ ครุภัณฑ์ ที่ ได้รับ การทดแทน แล้ว มา ใช้งาน has been classified as no and prediction = yes
ขอ ทดแทน จำนวน เครื่อง เนื่องจาก เป็น ครุภัณฑ์ ที่ ได้รับ จัดสรร ตั้งแต่ ปี มี อายุการใช้งาน เกิน กว่า ปี ใช้ งานประจำ ห้อง พิจารณาคดี ห้อง ที่ มี สภาพ ชำรุด และ ไม่ คุ้มค่า กับ การซ่อมแซม has been classified as no and prediction = yes
ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ งานประจำ กลุ่ม ช่วย พิจารณาคดี เจ้าหน้าที่ หน้า ปลัดลัง เพื่อ ใช้ สแกน คำพิพากษา ศาลั่ง อัพโหลด เข้า ระบบ การ จัดเก็บ คำพิพากษา ระหว่าง ศาล ทั่วประเทศ เนื่องจาก มี ครุภัณฑ์ ดังกล่าว ไม่ เพียง พอกับ การใช้งาน has been classified as no
ขอ เพิ่ม จำนวน เครื่อง สำหรับ จัดสรร ให้ แก่ เจ้าพนักงาน ตำรวจ ศาล ใช้งาน เนื่องจาก หน่วยงาน ได้รับ อนุมัติ กรอบ อัตรา กำลัง ดังกล่าว จำนวน อัตรา แต่ ปัจจุบัน ยัง ไม่ มี เครื่องคอมพิวเตอร์ สำหรับ ใช้งาน ปฏิบัติงาน has been classified as no and prediction = yes
ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ งานประจำ หน้า ห้อง พิจารณาคดี ห้อง ที่ และ ห้อง งาน ศูนย์ ติดตาม พยาน เพื่อ ใช้ สแกน คำพิพากษา และ สำนวน คดี รองรับ ระบบ ECase เนื่องจาก มี ครุภัณฑ์ ไม่ เพียง พอกับ การใช้งาน has been classified as no and prediction = yes
ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ ประจำ กลุ่ม บริหาร จัดการ คดี เครื่อง งาน รับฟ้อง แพง เครื่อง งาน รับฟ้อง เครื่อง งาน ฝ่ายยัง เครื่อง และ งาน ออก หมาย สี่ เครื่อง เนื่องจาก มี ครุภัณฑ์ ดังกล่าว ไม่ เพียง พอกับ การใช้งาน has been classified as no and prediction = yes
ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ ประจำ เคาน์เตอร์ กลุ่ม งาน ประชาสัมพันธ์ และ เคาน์เตอร์ กลุ่ม ช่วย พิจารณาคดี เพื่อ ใช้ เป็น คอมพิวเตอร์ ให้บริการ แก่ ทนายความ และ ประชาชน ที่มา ติดต่อ ราชการ ใช้ ใน การ ค้นหา ข้อ มูลคดี เช่น เลข คดี ผล หมาย วัน นัด เป็นต้น เพื่อ ให้การ
ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ ประจำ งาน เก็บ คดีแดง เครื่อง งานสารบรรณ เครื่อง งาน ฝ่ายยัง เครื่อง งาน หมาย เครื่อง งาน ส่ง ค่าคู่ความ เครื่อง ห้องทำงาน หัวหน้ากลุ่ม งาน เครื่อง และ ห้อง พิจารณาคดี เครื่อง เพื่อ ใช้ ใน การ สแกน เอกสาร และ สำนวน คดี ต่าง ให้ เป็น เอกสาร
ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ งานประจำ ส่วน ช่วย พิจารณาคดี ส่วน บริหาร จัดการ คดี ห้อง เก็บ สำนวน คดี และ ส่วน ใกล้ เลี้ย และ ประชุม ข้อพิพาท ส่วน งาน ละ เครื่อง เนื่องจาก มี ครุภัณฑ์ ดังกล่าว ไม่ เพียง พอกับ การใช้งาน has been classified as no and predictio
ขอ เพิ่ม จำนวน เครื่อง สำหรับ ใช้ งานประจำ ส่วน ช่วย พิจารณาคดี และ ส่วน บริหาร จัดการ คดี เพื่อ ใช้ สแกน สำนวน คดี อัพโหลด เข้า ระบบ CIOS รวมถึง ระบบ งาน อื่น ซึ่ง ใน ปัจจุบัน ส่วน งาน มี ภารกิจ ต้อง สแกน เอกสาร และ สำนวน คดี เพิ่มมากขึ้น เนื่องจาก มี สแกนเนอร์ ไม่
ขอ เพิ่ม จำนวน เครื่อง สำหรับ จัดสรร ให้ ข้าราชการ ตาลยุติธรรม ที่ หน่วยงาน ได้รับ อนุมัติ เพิ่ม จำนวน อัตรา ได้แก่ เจ้าพนักงาน ตำรวจ ศาล และ นักวิชาการ เงิน และ บัญชี ระดับ ชำนาญ การ พิเศษ หัวหน้ากลุ่ม งาน คลัง เนื่องจาก ยัง ไม่ มี เครื่อง คอ ทพ วิ เดอร์ รองรับ อัตรา กำลัง
ขอ เพิ่ม จำนวน ชุด สำหรับ รองรับ อัตรา กำลัง ข้าราชการ ศาลยุติธรรม ตำแหน่ง นักวิชาการ เงิน และ บัญชี ชำนาญ การ พิเศษ เพิ่ม เนื่องจาก ปัจจุบัน หน่วยงาน ยัง ไม่ มี ครุภัณฑ์ คอมพิวเตอร์ รองรับ กับ อัตรา กำลัง ดังกล่าว has been classified as no and prediction = yes

ภาพที่ 4.24 ตัวอย่างชุดข้อมูลที่ทำนายผลผิดพลาดด้วยอัลกอริทึม Random Forest ของแบบจำลองที่ 2

จากการตรวจสอบพบว่า ชุดข้อมูล Training data ที่นำไปสร้างแบบจำลองด้วยอัลกอริทึม Random Forest เป็นชุดข้อมูลที่มีคลาสคำตอบ yes เป็นจำนวนมาก โดยชุดข้อมูลที่มีคลาสคำตอบ no มีจำนวนน้อยมาก จึงมีความน่าจะเป็นที่ส่งผลให้แบบจำลองโมเดลทำนายผลผิดพลาดเป็นคลาสคำตอบ yes

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

5.1 สรุปผลการศึกษา

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

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

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

สำหรับงานวิจัยนี้ ใช้กระบวนการทำความสะอาดข้อมูล การตัดคำ การค้นหาคำสำคัญในข้อความเพื่อสร้างคลังคำศัพท์ 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

5.2 อภิปรายผล

การสร้างแบบจำลองในงานวิจัยนี้ ผู้วิจัยเห็นว่าจำเป็นต้องใช้เครื่องมือและเทคนิคสำคัญต่างๆ ในแต่ละกระบวนหรือขั้นตอน เพื่อให้แบบจำลองที่ 1 และแบบจำลองที่ 2 มีประสิทธิภาพสามารถทำนายผลได้ถูกต้องแม่นยำในระดับที่น่าพอใจ รวมถึงสามารถนำไปพัฒนาต่อยอดในระบบงานจริงของหน่วยงานตามวัตถุประสงค์ที่กำหนดไว้ได้ ซึ่งมีรายละเอียดขั้นตอนการดำเนินงานวิจัยที่สำคัญ ดังนี้

5.2.1 การทำความสะอาดข้อมูล

(1) การลบคำเฉพาะ เช่น ชื่อยี่ห้อ รุ่น จำนวนหน้า รวมถึงการลบเครื่องหมายต่างๆที่เราไม่ต้องการ เช่น (!€@%#*&~) ในรายการคำของบประมาณ และรายละเอียดเหตุผลความจำเป็นจะช่วยลดมิติของข้อมูลที่นำไปสร้างแบบจำลองที่ 1 และแบบจำลองที่ 2

(2) การลบอักขระพิเศษ การใช้เครื่องมือในภาษาไพธอนซึ่งมีไลบรารีที่ชื่อว่า RegEx เข้ามาช่วยในการทำความสะอาดข้อมูล สามารถกรองคำจากข้อมูลคำของบประมาณและรายละเอียดเหตุผลความจำเป็นที่มี

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ทั้งตัวอักษรภาษาไทย ตัวเลข และภาษาอังกฤษได้ดีขึ้น ช่วยลดจำนวนของคำที่จะนำไปสร้างคลังคำศัพท์และกำหนดฟีเจอร์ชุดข้อมูลที่นำไปสร้างแบบจำลองและทดสอบแบบจำลอง

5.2.2 การตัดคำ

การเลือกใช้ไลบรารี PythaiNLP และกำหนดค่าพารามิเตอร์ newmm engine สำหรับการตัดคำในขั้นตอนแรก และใช้ไลบรารี thai_stopwords สำหรับการลบคำในภาษาไทยที่ไม่สำคัญออก ส่งผลให้การหาคำสำคัญสำหรับสร้างคลังคำศัพท์และจำนวนฟีเจอร์ของชุดข้อมูลที่สร้างแบบจำลอง และทดสอบแบบจำลองมีจำนวนลดลง ซึ่งส่งผลให้ประสิทธิภาพการทำนายผลของแบบจำลองสูงขึ้นตามไปด้วย

5.2.3 การสร้างแบบจำลอง และประเมินผลประสิทธิภาพแบบจำลอง

ในงานวิจัยฉบับนี้ เลือกใช้อัลกอริทึมประเภท Text Classification เพราะข้อมูลคลาสคำตอบของแบบจำลองที่ต้องการจัดอยู่ในประเภท Nominal หรือ categorical data คือ ข้อมูลที่จัดเป็นกลุ่ม เช่น แบบจำลองที่ 1 ในงานวิจัย คือ support และ network ทั้งนี้ ในการทดลองสร้างแบบจำลอง จำเป็นที่จะต้องกำหนดค่าพารามิเตอร์ให้แก่อัลกอริทึม ซึ่งผู้วิจัยได้ทำการทดลองสร้างแบบจำลองด้วยอัลกอริทึมแต่ละชนิดโดยใช้ค่าที่ไลบรารี sklearn กำหนดเป็นค่าเริ่มต้น (default) เปรียบเทียบกับค่าพารามิเตอร์ที่กำหนดขึ้นเอง พบว่า แบบจำลองสามารถทำนายผลถูกต้องมีค่าความถูกต้องและแม่นยำสูงขึ้นอยู่ในระดับ 80 % - 99 %

5.3 ข้อเสนอแนะ

สำหรับผลงานวิจัยฉบับนี้ ผู้วิจัยเห็นว่า ด้วยความก้าวหน้าของเทคโนโลยีในปัจจุบัน ซึ่งมีเทคนิคและเครื่องมือต่างๆ ที่ใช้สำหรับงานประมวลผลภาษาธรรมชาติ ดังนั้น เพื่อให้ผลงานวิจัยมีประสิทธิภาพที่สมบูรณ์ ครบถ้วนในการนำไปใช้ประโยชน์ต่อไป ผู้วิจัยมีข้อเสนอแนะเกี่ยวกับงานวิจัยนี้ 2 ประเด็น ดังนี้

5.3.1 การเตรียมข้อมูลคำของบประมาณ

เนื่องจากข้อมูลคำของบประมาณเป็นปัจจัยตั้งต้นของงานวิจัยนี้ ผู้วิจัยเห็นว่า เพื่อให้การสร้างแบบจำลองสามารถเรียนรู้ข้อมูลได้หลากหลายรูปแบบ จึงเห็นควรเพิ่มเติมปริมาณข้อมูลกลุ่มตัวอย่างที่ใช้สร้างแบบจำลองให้มากขึ้นกว่าเดิม และควรให้หน่วยงานที่เกี่ยวข้องกับการพิจารณางบประมาณปรับเปลี่ยนรูปแบบการจัดเก็บข้อมูลในรูปแบบเอกสารมาให้อยู่ในรูปแบบข้อมูลอิเล็กทรอนิกส์ เพื่อบูรณาการการใช้ประโยชน์ข้อมูลร่วมกันสำหรับพัฒนาต่อยอดระบบบริหารจัดการงบประมาณของสำนักงานศาลยุติธรรมให้มีประสิทธิภาพมากยิ่งขึ้น

5.3.2 การใช้เทคนิคและเครื่องมือการประมวลผลภาษาธรรมชาติ

ในปัจจุบันมีผู้พัฒนาเทคนิคและเครื่องมือที่ใช้สำหรับการประมวลภาษาธรรมชาติขึ้นมาใช้งานกันอย่างแพร่หลาย ซึ่งในงานวิจัยนี้ ผู้วิจัยเลือกใช้ PythaiNLP ซึ่งเป็นไลบรารีสำเร็จรูปในภาษาไพธอนเป็นเครื่องมือสำหรับการตัดคำ การสร้างคลังคำศัพท์ และการสร้างฟีเจอร์ เพื่อสร้างชุดข้อมูลเรียนรู้สำหรับการสร้างแบบจำลองที่ใช้จำแนกข้อมูลคำของบประมาณที่อยู่ในรูปแบบข้อความภาษาไทยเป็นส่วนใหญ่ ทั้งนี้ ในกระบวนการประมวลผลภาษาธรรมชาติด้วยไลบรารีข้างต้น ผู้วิจัยเลือกใช้โมดุล word_tokenize ในการตัดคำ และลดจำนวนคำภาษาไทยด้วย thai_stopword เป็นหลัก ซึ่งอาจไม่ใช่เครื่องมือและเทคนิคที่สามารถรองรับกับ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ความซับซ้อนของกลุ่มคำภาษาไทยในรายการคำของบประมาณในงานวิจัยนี้ รวมถึงการสร้างฟีเจอร์ของชุดข้อมูลเรียนรู้ ผู้วิจัยเลือกใช้วิธีการนับคำ ด้วยโมดุล Count Vectorizer เพียงเทคนิคเดียว แต่จากการศึกษาข้อมูลเพิ่มเติมยังมีเทคนิคการสร้างฟีเจอร์ด้วยการใช้ค่า TFIDF ดังนั้น เพื่อให้ผลงานวิจัยนี้สามารถนำไปใช้ให้เกิดประโยชน์สูงสุดกับการสร้างต้นแบบพัฒนาระบบงานงบประมาณด้วยการใช้เทคโนโลยีการประมวลผลภาษาธรรมชาติ จึงเห็นควรที่จะต้องทดลองใช้วิธีการตัดคำด้วยอัลกอริทึมอื่นๆ รวมถึงการสร้างฟีเจอร์ที่ซับซ้อนมากขึ้น เช่น word embedding เป็นต้น

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บรรณานุกรม

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บรรณานุกรม

[1] สำนักแผนงานและงบประมาณ สำนักงานศาลยุติธรรม. **ระบบบริหารงบประมาณศาลยุติธรรมและสำนักงานศาลยุติธรรม**, 2565.

[2] เฉลิมชัย พิเดช. **ระบบจำแนกตั๋วงานอัจฉริยะด้วยเทคนิควิเคราะห์ข้อความไทย – อังกฤษและการเรียนรู้ของเครื่อง**. สารนิพนธ์ปริญญามหาบัณฑิต คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ, 2562.

[3] วทัญญู ชูประจิตต์. **การจำแนกประเภทและวิธีการคัดเลือกคุณลักษณะสำหรับการสนับสนุนการวิเคราะห์ความรู้สึกต่อสถานที่ท่องเที่ยวไทย**. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยทักษิณ, 2563.

[4] ขวัญตา ศิลปะไพบูลย์พานิช. การใช้เทคโนโลยีการเรียนรู้ของเครื่องเพื่อทำนายผลการเรียนของนักเรียน. สารนิพนธ์ปริญญามหาบัณฑิต คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ, 2562.

[5] พรรณนาภรณ์ เกตุภู่พงษ์. **การประยุกต์ใช้การทำเหมืองข้อความเพื่อจำแนกประเภทโรคอาการ**. วิทยานิพนธ์ปริญญามหาบันฑิต คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2561.

[6] สันติ สุขเกษม. **การวิเคราะห์รูปแบบความคิดเห็นของข้อมูลโรงแรมโดยใช้เทคนิคการตัดคำแบบผสม**. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีไทย **–** ญี่ปุ่น, 2561.

[7] Hindawi Computational Intelligence and Neuroscience. Advances in Computational Intelligence Techniques for Next Generation Internet of Things, 2022. ค้นวันที่ 1 สิงหาคม 2565 จากhttps://www.hindawi.com/journals/cin/2022/1883698.

[8] Yih-Shan Sheu. Leveraging Text Mining and Analytical Technology to Enhance Financial Planning and Analysis, 2021. ค้นวันที่ 1 สิงหาคม 2565 จาก https://dr.lib.iastate.edu/server/api/core/bitstreams/c91f5f63-21ea-4340-b31d-b75e2ab7ea7d/content.

[9] Witchapong Daróntham. **ขั้นตอนการเตรียมข้อมูลประเภท** Text **ภาษาไทย แบบง่ายๆ โดยใช้** Python, 2018. ค้นวันที่ 1 สิงหาคม 2565 จาก https://medium.com.

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ภาคผนวก

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ภาคผนวก ก
เอกสารคำของบประมาณกิจกรรมจัดซื้อครุภัณฑ์คอมพิวเตอร์

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

แบบ ขงป 3 - 1 : สำหรับครุภัณฑ์แต่ละรายการ

แบบสรุปแผนความต้องการครุภัณฑ์ประจำปีงบประมาณ พ.ศ. .....

(สำหรับทุกหน่วยงาน)

◯ ครุภัณฑ์เครื่องถ่ายเอกสาร
◯ ครุภัณฑ์เครื่องเรือนฯ ข้าราชการตุลาการ
◯ ครุภัณฑ์คอมพิวเตอร์
◯ ครุภัณฑ์เครื่องปรับอากาศ
◯ ครุภัณฑ์เครื่องเรือนฯ ข้าราชการศาลยุติธรรม
◯ ครุภัณฑ์อื่นๆ

หน่วยงาน/ศาล..... **วันที่จัดทำข้อมูล.....**

ลำดับที่ (1)	รายการและคุณลักษณะครุภัณฑ์ (2)	หน่วยนับ (3)	จำนวน ของทดแทน (4/1)	จำนวน ของเพิ่ม/ของใหม่ (4/2)	จำนวน รวม (4/3)=(4/1)+(4/2)	ราคาต่อหน่วย (บาท) (5)	เป็นเงิน (บาท) (6)=(4/3)X(5)	ประเภทครุภัณฑ์ (จำเป็น/เสริม/สนับสนุน) (7)
1	(ระบุรายการครุภัณฑ์).....							
2	(ระบุรายการครุภัณฑ์).....							
3	(ระบุรายการครุภัณฑ์).....							
4	(ระบุรายการครุภัณฑ์).....							
5	(ระบุรายการครุภัณฑ์).....							
6	(ระบุรายการครุภัณฑ์).....							
7	(ระบุรายการครุภัณฑ์).....							
8	(ระบุรายการครุภัณฑ์).....							
9	(ระบุรายการครุภัณฑ์).....							
10	(ระบุรายการครุภัณฑ์).....							
11	(ระบุรายการครุภัณฑ์).....							
12	(ระบุรายการครุภัณฑ์).....							
				รวม		เป็นเงิน		

ลงชื่อ..... ผู้ตรวจสอบ
(.....)
ตำแหน่ง.....

ลงชื่อ..... ผู้จัดทำ
(.....)
ตำแหน่ง.....

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

คำอธิบายแบบ ขงป. 3 - 1

แบบสรุปแผนความต้องการครุภัณฑ์ประจำปีงบประมาณ พ.ศ. ...

(สำหรับทุกหน่วยงาน)

เป็นแบบสำหรับให้หน่วยงานนำข้อมูลจากแผนความต้องการครุภัณฑ์ในแบบ ขงป. 3 - 2 ที่หน่วยงานจัดทำมาจัดลำดับความสำคัญลงในแบบสรุปแผนความต้องการครุภัณฑ์ (แบบ ขงป. 3 - 1) ฉบับนี้ เพื่อให้สามารถเห็นภาพรวมความต้องการครุภัณฑ์แต่ละรายการของแต่ละหน่วยงานได้ชัดเจนยิ่งขึ้น โดยมีรายละเอียดคำชี้แจงแบบดังนี้

ช่อง (1) : เรียงลำดับความสำคัญของครุภัณฑ์

ช่อง (2) : ระบุรายการครุภัณฑ์ และคุณลักษณะตามมาตรฐานครุภัณฑ์ โดยเรียงลำดับความสำคัญ

ช่อง (3) : ระบุหน่วยนับของครุภัณฑ์

ช่อง (4/1) : ระบุจำนวนครุภัณฑ์ที่ต้องการขอทดแทน

ช่อง (4/2) : ระบุจำนวนครุภัณฑ์ที่ต้องการขอเพิ่ม/ขอใหม่

ช่อง (4/3) : ระบุจำนวนรวมของครุภัณฑ์ที่ต้องการขอ = ช่อง(4/1) + ช่อง(4/2)

ช่อง (5) : ระบุราคามาตรฐานของครุภัณฑ์ (ถ้ามี) ถ้าไม่มีให้ใช้ราคาตามท้องตลาด ณ ขณะนั้น ทั้งนี้ ต้องไม่ต่ำกว่า 5,000 บาท

ช่อง (6) : ระบุจำนวนเงินที่ขอ = ช่อง(4/3) X ช่อง(5)

ช่อง (7) : ระบุระดับความสำคัญต่อการปฏิบัติงานของครุภัณฑ์อย่างใดอย่างหนึ่ง ดังนี้

จำเป็น : เป็นครุภัณฑ์จำเป็นสำหรับการปฏิบัติงาน หากไม่ได้รับการจัดสรรจะไม่สามารถปฏิบัติงานได้

เสริม : ครุภัณฑ์เสริมการปฏิบัติงาน หากได้รับการจัดสรรจะทำให้การปฏิบัติงานมีประสิทธิภาพมากขึ้น

สนับสนุน : ครุภัณฑ์สนับสนุนการปฏิบัติ หากไม่ได้รับการจัดสรร ก็ยังสามารถปฏิบัติงานได้

(พร้อมแนบแบบ ขงป. 3 - 2 ของแต่ละรายการครุภัณฑ์ที่นำมาจัดลำดับความสำคัญมาด้วย)

หมายเหตุ *: ขอให้หน่วยงานระบุประเภทครุภัณฑ์และกรอกข้อมูลในแบบให้ถูกต้องครบถ้วน เพื่อประโยชน์ในการพิจารณาอนุมัติงบประมาณ*

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

แบบ ขจป. 3 - 2 : สำหรับครุภัณฑ์แต่ละรายการ

แบบรายละเอียดรายการครุภัณฑ์ตามแผนความต้องการครุภัณฑ์ประจำปีงบประมาณ พ.ศ. ....

แยกแต่ละรายการครุภัณฑ์

(สำหรับทุกหน่วยงาน)

หน่วยงาน/ศาล..... **วันที่จัดทำข้อมูล**.....

ที่	รายการและคุณลักษณะ	หน่วยนับ	จำนวน			ราคาต่อหน่วย**	เป็นเงิน	หมายเหตุ
			ทดแทน*	ของเพิ่ม/ของใหม่	รวม			
(1)	(2)	(3)	(4/1)	(4/2)	(4/3)=(4/1)+(4/2)	(บาท) (5)	(บาท) (6)=[(4/3)X(5)]	(7)
	◯ จำเป็น ◯ เสริม ◯ สนับสนุน							

ข้อมูลสำคัญประกอบการพิจารณา (โปรดระบุให้ครบถ้วน)

จำนวนครุภัณฑ์ของหน่วยงานที่มีอยู่ ณ ปัจจุบัน			จำนวนที่ควรมี (หน่วย) (d)	จำนวนที่ขาดอยู่ (หน่วย) (e)=(d)-(c)
ใช้การได้ (a)	ชำรุดเสียหายต้องทดแทน (b)	รวม (c)=(a)+(b)		

ข้อมูลประกอบกรณีขอทดแทนครุภัณฑ์เดิม

รายละเอียดครุภัณฑ์เดิมที่ต้องการทดแทน					
ลำดับที่	ประเภทเงิน	วิธีการได้มา	วันเดือนปีที่ได้รับครุภัณฑ์เดิม	ค่าสึกหรอสะสม (บาท)	ค่าซ่อมแซมสะสม (บาท)

*ที่มาข้อมูล : ทะเบียนคุมทรัพย์สิน

เอกสารแนบประกอบแผนความต้องการครุภัณฑ์

ใบเสนอราคา		รูปถ่าย		เอกสารแจ้งซ่อมแซม		ผังห้องหรืออัตราติดตั้ง		รายละเอียดการติดตั้ง		ทะเบียนคุมทรัพย์สิน	
มี	ไม่มี	มี	ไม่มี	มี	ไม่มี	มี	ไม่มี	มี	ไม่มี	มี	ไม่มี

ขอให้กรอกข้อมูลและ/หรือแนบเอกสารประกอบที่จำเป็นต่อการพิจารณาให้ครบถ้วนเพื่อประโยชน์ในการพิจารณาอนุมัติงบประมาณ

* เอกสารรายละเอียดการติดตั้งขอให้ระบุรายการค่าประกอบ ระยะทางเดินสายไฟหรือการเดินท่อ จุดเข้ารับ ลักษณะรายวิวแบบที่จะติดตั้ง หรือรายการค่าใช้จ่ายที่เกี่ยวเนื่องอื่นๆ ให้ครบถ้วน

เหตุผลความจำเป็น

- **วัตถุประสงค์และลักษณะการใช้งาน สถานที่ตั้งครุภัณฑ์ และพื้นที่ :** (ระบุวัตถุประสงค์และลักษณะการใช้งานที่ชัดเจน เป็นต้นว่า ใช้ประกอบการปฏิบัติงานเกี่ยวกับอะไร ครุภัณฑ์ตั้งอยู่ ณ ฝ่ายงานใด/กลุ่มงานใด การทำงานของครุภัณฑ์ครอบคลุมพื้นที่เท่าใด)

- **จำนวนผู้ใช้งานและปริมาณการใช้งาน :** *(ระบุจำนวนผู้ที่ใช้งานครุภัณฑ์/ปริมาณการใช้งาน/ความถี่ในการใช้งานแต่ละวัน)*

- **สภาพปัญหา :** *(ระบุสภาพปัญหาที่พบ ทำให้จำเป็นต้องขอเพิ่ม/ขอใหม่ หรือต้องได้รับการทดแทนของเดิม)*

- **อื่นๆ ประกอบ :** *(ระบุเหตุผลเพิ่มเติมที่แสดงให้เห็นถึงความจำเป็นที่ต้องได้รับการจัดสรรครุภัณฑ์รายการนี้ เช่น เพื่อรองรับการดำเนินการตามกฎระเบียบและข้อบังคับตามกฎหมายที่เกี่ยวข้อง สอดคล้องกับขอบเขตภารกิจของหน่วยงาน)*

ลงชื่อ..... ผู้ตรวจสอบ ลงชื่อ..... ผู้จัดทำ

หมายเหตุ : *กรณีขอทดแทน ขอให้ถ่ายรูปครุภัณฑ์เดิมและแนบทะเบียนคุมทรัพย์สินที่จะขอทดแทนมาด้วย เพื่อประกอบการพิจารณา

**ครุภัณฑ์นอกบัญชีราคามาตรฐานครุภัณฑ์ขอให้แนบใบเสนอราคา เพื่อประกอบการพิจารณา

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

แบบ ขงป. 3 - 2/1 : สำหรับครุภัณฑ์เครื่องเรือนฯ (ผ้าม่าน)

แบบรายละเอียดครุภัณฑ์ตามแผนความต้องการครุภัณฑ์ประจำปีงบประมาณ พ.ศ. ...

(ครุภัณฑ์เครื่องเรือนประจำบ้านพักฯ ข้าราชการฝ่ายตุลาการศาลยุติธรรม รายการผ้าม่าน)

(สำหรับทุกหน่วยงาน)

หน่วยงาน/ศาล.....

รายละเอียดข้อมูลพื้นฐาน (ขอให้ระบุรายละเอียดบ้านพัก โดยจำแนกต่อหลัง/ต่อหน่วย หากไม่ระบุให้ครบตามเงื่อนไข สำนักงานศาลยุติธรรมจะไม่พิจารณา)

1. บ้านพัก / อาคารชุด เลขที่..... ซอย.....ถนน.....ตำบล ..... อำเภอ.....จังหวัด.....

2. จำนวนช่องที่มี ได้แก่ ประตู.....ช่อง หน้าต่าง..... ช่อง

3. เคยได้รับจัดสรรครั้งล่าสุด เมื่อปี พ.ศ. ..... (หากระบุวัน/เดือน/ปี ได้ ขอให้ระบุมาด้วย ...../...../.....)

4. จำนวนช่องที่ต้องการติดตั้งผ้าม่าน ได้แก่ ประตู.....ช่อง หน้าต่าง.....ช่อง

ลำดับที่	รายการ	หน่วยนับ	จำนวนหน่วยที่ขอ (จำนวน)	ราคาต่อหน่วย (บาท)	วงเงินที่ขอ (บาท)	ขอเพื่อ	
						ทดแทน (จำนวน)	ใหม่ (จำนวน)
1	ผ้าม่านบ้านพักผู้พิพากษา/ข้าราชการ เลขที่.....	หลัง					
	ประตู						
	1) ขนาด .....ม.X .....ม. จำนวน.....ชุด						
	2) ขนาด .....ม.X .....ม. จำนวน.....ชุด						
	3) ขนาด .....ม.X .....ม. จำนวน.....ชุด						
	หน้าต่าง						
	1) ขนาด .....ม.X .....ม. จำนวน.....ชุด						
	2) ขนาด .....ม.X .....ม. จำนวน.....ชุด						
	3) ขนาด .....ม.X .....ม. จำนวน.....ชุด						

ขอรับรองว่า รายการผ้าม่านตามจำนวนหน่วยที่ขอติดตั้งข้างต้น มีความเหมาะสมแล้ว สำหรับหน่วยที่ยังเหลืออยู่ไม่ประสงค์ที่จะต้องติดตั้งเพิ่มเติมอีก ทั้งนี้ จนกว่าจะครบอายุการใช้งาน 5 ปีตามหลักเกณฑ์ที่กำหนด

ลงชื่อ..... ผู้จัดทำ
(.....)
ตำแหน่ง.....
โทร.....
วันที่จัดทำ.....

ลงชื่อ..... ผู้ตรวจสอบ
(.....)
ตำแหน่ง...ผู้อำนวยการ.....
โทร.....
วันที่จัดทำ.....

หมายเหตุ : ภายหลังจากกรอกข้อมูลข้างต้น และกรอกรายละเอียดในแบบ ขงป. 3 - 2 เรียบร้อยแล้ว ให้นำไปจัดลำดับความสำคัญลงในแบบ ขงป. 3 - 1 รวมกับครุภัณฑ์เครื่องเรือนประจำบ้านพักฯ ข้าราชการฝ่ายตุลาการศาลยุติธรรมรายการอื่นๆ ที่ศาลต้องการทั้งหมด พร้อมทั้งแนบใบเสนอราคาครุภัณฑ์ผ้าม่านแต่ละหลัง

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

คำอธิบายแบบ ขงป. 3 - 2

แบบแผนความต้องการครุภัณฑ์ประจำปีงบประมาณ พ.ศ. ...

แยกแต่ละรายการครุภัณฑ์

(สำหรับทุกหน่วยงาน)

เป็นแบบสำหรับให้หน่วยงานกรอกข้อมูลความต้องการครุภัณฑ์ในแต่ละรายการครุภัณฑ์ โดยมีรายละเอียดคำชี้แจงแบบฟอร์ม ดังนี้

ช่อง (1) : เรียงลำดับความสำคัญของครุภัณฑ์

ช่อง (2) : ระบุรายการครุภัณฑ์และคุณลักษณะตามมาตรฐานครุภัณฑ์ (ถ้ามี) โดยเรียงลำดับความสำคัญ
: ทำเครื่องหมายถูกลงในวงกลมใดวงกลมหนึ่งที่หน่วยงานพิจารณาแล้วเห็นว่าครุภัณฑ์ที่จัดทำแผนความต้องการครุภัณฑ์นั้น เป็นครุภัณฑ์ที่มีความสำคัญต่อการปฏิบัติงานในระดับใด

จำเป็น : เป็นครุภัณฑ์จำเป็นสำหรับการปฏิบัติงาน หากไม่ได้รับการจัดสรรจะไม่สามารถปฏิบัติงานได้

เสริม : ครุภัณฑ์เสริมการปฏิบัติงาน หากได้รับการจัดสรรจะทำให้การปฏิบัติงานมีประสิทธิภาพมากขึ้น

สนับสนุน : ครุภัณฑ์สนับสนุนการปฏิบัติงาน หากไม่ได้รับการจัดสรร ก็ยังสามารถปฏิบัติงานได้

ช่อง (3) : ระบุหน่วยนับของครุภัณฑ์ เช่น เครื่อง ชุด ตัว ตู้ เป็นต้น

ช่อง (4/1) : ระบุจำนวนครุภัณฑ์ที่ต้องการขอทดแทนครุภัณฑ์เดิม

ช่อง (4/2) : ระบุจำนวนครุภัณฑ์ที่ต้องการขอเพิ่ม/ขอใหม่ (กรณียังไม่มีครุภัณฑ์ไว้ใช้งาน หรือมีครุภัณฑ์สำหรับใช้งานไม่เพียงพอ)

ช่อง (4/3) : ระบุจำนวนรวมของครุภัณฑ์ที่ต้องการขอ = ช่อง(4/1) + ช่อง(4/2)

ช่อง (5) : ระบุราคามาตรฐานของครุภัณฑ์ (ถ้ามี) ถ้าไม่มีให้ใช้ราคาตามท้องตลาด ณ ขณะนั้น ทั้งนี้ ต้องไม่ต่ำกว่า 5,000 บาท

ช่อง (6) : ระบุจำนวนเงินที่ขอ = ช่อง(4/3) X ช่อง(5)

ช่อง (7) : ระบุหมายเหตุ

ระบุข้อมูลประกอบ ดังนี้

- ระบุจำนวนครุภัณฑ์ที่มีอยู่ ณ ปัจจุบัน (ใช้การได้/ชำรุดเสียหายต้องทดแทน)
- ระบุจำนวนครุภัณฑ์ที่หน่วยงานควรมีไว้ใช้งาน
- ระบุจำนวนครุภัณฑ์ที่ขาด (จำนวนครุภัณฑ์ที่ควรมี - จำนวนครุภัณฑ์ที่มีอยู่)

ระบุข้อมูลประกอบเพิ่มเติมกรณีขอทดแทนครุภัณฑ์เดิม ดังนี้

- ระบุลำดับที่ของครุภัณฑ์เดิมที่ต้องการขอทดแทน
- ระบุประเภทเงิน (งบประมาณ นอกงบประมาณ เงินบริจาค/เงินช่วยเหลือ อื่นๆ)
- ระบุวิธีการได้มา (ตกลงราคา สอบราคา ประกวดราคา วิธีพิเศษ รับบริจาค หรือประกาศเชิญชวนทั่วไป คัดเลือก เฉพาะเจาะจง) ตาม พ.ร.บ. การจัดซื้อจัดจ้างฯ พ.ศ. 2560)
- ระบุวันเดือนปีที่ได้รับครุภัณฑ์เดิม
- ระบุค่าเสื่อมราคาสะสม
- ระบุค่าซ่อมแซมสะสม

(หมายเหตุ : ข้อมูลประกอบเพิ่มเติมกรณีขอทดแทน มีที่มาจากทะเบียนคุมทรัพย์สินของหน่วยงาน)

ระบุเอกสารแนบประกอบแผนความต้องการครุภัณฑ์ ดังนี้

- ระบุใบเสนอราคา กรณีเป็นครุภัณฑ์นอกราคามาตรฐาน
- ระบุรูปถ่าย สำหรับกรณีต้องการขอทดแทน ให้แนบรูปถ่ายครุภัณฑ์ที่ใช้งานปัจจุบัน และกรณีครุภัณฑ์ที่มีความจำเป็นต้องนำขนาดพื้นที่หรือขนาดห้องมาใช้ประกอบการพิจารณา ให้แนบรูปถ่ายสถานที่หรือห้องมาด้วย
- ระบุเอกสารการแจ้งซ่อมแซมครุภัณฑ์ กรณีต้องการขอทดแทน
- ระบุผังห้องหรือผังการติดตั้ง สำหรับครุภัณฑ์ที่จำเป็นต้องคำนวนขนาดพื้นที่ห้องเพื่อกำหนดคุณลักษณะเฉพาะครุภัณฑ์ให้เหมาะสมกับการใช้งาน เช่น เครื่องปรับอากาศ เครื่องฟอกอากาศ เป็นต้น
- ระบุรายละเอียดการติดตั้ง สำหรับครุภัณฑ์ที่มีลักษณะเป็นชุดหรือระบบซึ่งต้องติดตั้งอุปกรณ์หลายชิ้นใช้งานร่วมกัน อาทิเช่น ระบบกล้องโทรทัศน์วงจรปิด ชุดเครื่องเสียงห้องประชุม ระบบสืบพยานเด็ก ชุดเครื่องเสียงตามสาย เป็นต้น โดยระบุอุปกรณ์ประกอบ ระยะการเดินสายไฟหรือการเดินท่อ จุดเด้ารับ ลักษณะบริเวณที่จะติดตั้ง แผนภาพการติดตั้ง (Diagram) หรือรายการค่าใช้จ่ายที่เกี่ยวเนื่องอื่นๆ ให้ครบถ้วน
- ระบุทะเบียนคุมทรัพย์สิน สำหรับกรณีต้องการขอทดแทน

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ระบุเหตุผลความจำเป็น ดังนี้

- วัตถุประสงค์และลักษณะการใช้งาน สถานที่ตั้ง และพื้นที่ เป็นต้นว่า ใช้ประกอบการปฏิบัติงานเกี่ยวกับอะไร ครุภัณฑ์ตั้งอยู่ ณ กลุ่มงานใด การทำงานของครุภัณฑ์ครอบคลุมพื้นที่เท่าใด ทั้งนี้ สำหรับครุภัณฑ์เครื่องปรับอากาศและเครื่องฟอกอากาศ เกี่ยวเนื่องกับการคำนวณพื้นที่ของห้องที่จะทำการติดตั้งเพื่อให้เกิดความเหมาะสม จึงขอให้ระบุขนาดพื้นที่ของห้องและจำนวนครุภัณฑ์ พร้อมทั้งคุณลักษณะที่ใช้ในห้องดังกล่าวมาด้วย (กว้าง x ยาว x สูง : ครุภัณฑ์ขนาด ... BTU/CFM จำนวน ... เครื่อง)
- ระบุจำนวนผู้ใช้งาน/ปริมาณการใช้งาน /ความถี่ในการใช้งานแต่ละวัน
- ระบุสภาพปัญหาที่พบ ที่เป็นสาเหตุให้ต้องขอรับการจัดสรรครุภัณฑ์
- ระบุเหตุผลอื่นๆ ประกอบ ที่แสดงให้เห็นถึงความจำเป็นที่ต้องได้รับการจัดสรรครุภัณฑ์รายการนี้

<u>หมายเหตุ</u> : 1. เมื่อกรอกข้อมูลลงในแบบ ขงป. 3 - 2 เรียบร้อยแล้ว ให้นำข้อมูลไปจัดลำดับความสำคัญลงในแบบสรุปแผนความต้องการครุภัณฑ์แยกตามประเภทครุภัณฑ์ที่กำหนดไว้ในแบบ ขงป. 3 - 1

2. กรณีครุภัณฑ์เครื่องเรือนฯ รายการผ้าม่านขอให้หน่วยงานจัดทำข้อมูลลงในแบบ ขงป. 3 - 2/1 และกรณีครุภัณฑ์คอมพิวเตอร์ขอให้หน่วยงานจัดทำข้อมูลลงในแบบ ขงป. 3 - 3 แนบมาพร้อมแบบ ขงป. 3 - 2 ด้วย

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

แบบ ขงป. ๓ - ๓ : รายละเอียดครุภัณฑ์

แบบรายละเอียดประกอบแผนความต้องการครุภัณฑ์คอมพิวเตอร์

๑. ชื่อส่วนราชการ.....

๒. ระบบหรืออุปกรณ์คอมพิวเตอร์ที่มีอยู่ในปัจจุบันของหน่วยงาน

๒.๑ อุปกรณ์คอมพิวเตอร์ที่มีอยู่ในปัจจุบัน (จำนวน.....เครื่อง/ชุด/ตัว)

งาน	เครื่องแม่ข่าย	เครื่องไมโครคอมพิวเตอร์	เครื่องพิมพ์		อุปกรณ์ต่อพ่วง	
			เลเซอร์	หัวเข็ม	Switch/Hub	อื่นๆ
๑.						
๒.						
๓.						
๔.						
๕.						
๖.						
๗.						
๘.						
๙.						
๑๐.						
รวม						

๒.๒ ระบบงานที่มีอยู่ในปัจจุบัน (ให้ระบุทุกระบบที่ใช้งาน) โดยประกอบด้วย

(๑) ชื่อระบบงาน : .....

ก. ลักษณะงาน : .....

.....

.....

.....

ข. จำนวนบุคลากรที่ใช้งานระบบ.....

ค. ปัญหาและอุปสรรคในการปฏิบัติงาน.....

.....

.....

.....

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

๒.๓ โครงรูปและการเชื่อมโยงระบบเครือข่ายคอมพิวเตอร์และอุปกรณ์
ให้แสดงรายละเอียดแผนภาพการเชื่อมโยงเครือข่ายคอมพิวเตอร์และอุปกรณ์ต่อพ่วง
(จัดทำเป็นเอกสารแนบเพิ่มเติม)

๓. ข้อมูลบุคลากร

กลุ่มงาน	จำนวนบุคลากร (คน)				
	ข้าราชการตุลาการ	ข้าราชการศาลยุติธรรม	ลูกจ้างประจำ	พนักงานข้าราชการ	ลูกจ้างชั่วคราว

๔. ระบบหรืออุปกรณ์คอมพิวเตอร์ที่มีอยู่ในปัจจุบันของหน่วยงาน

๔.๑ ปัญหาและอุปสรรคที่ขอรับการจัดสรรในครั้งนี้

.....

.....

.....

๔.๒ รายละเอียดระบบหรืออุปกรณ์คอมพิวเตอร์ที่ขออนุมัติ

.....

.....

.....

๔.๓ ปริมาณงานเดิมที่เกี่ยวข้องกับระบบหรืออุปกรณ์คอมพิวเตอร์ที่ขอรับการจัดสรรงบประมาณ

ชื่องาน.....

รายละเอียดของงาน.....

ปริมาณงาน.....

๔.๔ สถานที่ติดตั้งระบบและอุปกรณ์คอมพิวเตอร์ (ส่วน/กลุ่ม/งาน.....)

.....

.....

.....

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

แบบ ขจป. 3 - 4 : ภาคสรุปครุภัณฑ์แต่ละรายการ

แบบสรุปแผนความต้องการครุภัณฑ์ประจำปีงบประมาณ พ.ศ. ...

(สำหรับสำนักศาลยุติธรรมประจำภาค : สรุปความต้องการครุภัณฑ์ของหน่วยงานในสังกัดภาคฯ)

◯ ครุภัณฑ์เครื่องถ่ายเอกสาร ◯ ครุภัณฑ์เครื่องปรับอากาศ

◯ ครุภัณฑ์เครื่องเรือนฯ ข้าราชการตุลาการ ◯ ครุภัณฑ์เครื่องเรือนฯ ข้าราชการศาลยุติธรรม

◯ ครุภัณฑ์ครุภัณฑ์คอมพิวเตอร์ ◯ ครุภัณฑ์อื่นๆ

สำนักศาลยุติธรรมประจำภาค..... วันที่จัดทำข้อมูล.....

ลำดับที่ (1)	หน่วยงานในสังกัดภาค (2)	รายการและคุณลักษณะครุภัณฑ์ (3)	หน่วยนับ (4)	จำนวน ขอทดแทน (5/1)	จำนวน ขอเพิ่ม/ขอใหม่ (5/2)	จำนวน รวม (5/3)= (5/1)+(5/2)	ราคาต่อหน่วย (บาท) (6)	เป็นเงิน (บาท) (7)=(5/3)X(6)	ประเภทครุภัณฑ์ (จำเป็น/เสริม/สนับสนุน) (8)
1	ศาล.....	(ระบุรายการครุภัณฑ์).....							
2	ศาล.....	(ระบุรายการครุภัณฑ์).....							
3	ศาล.....	(ระบุรายการครุภัณฑ์).....							
4	ศาล.....	(ระบุรายการครุภัณฑ์).....							
5	ศาล.....	(ระบุรายการครุภัณฑ์).....							
6	ศาล.....	(ระบุรายการครุภัณฑ์).....							
7	ศาล.....	(ระบุรายการครุภัณฑ์).....							
8	ศาล.....	(ระบุรายการครุภัณฑ์).....							
9	ศาล.....	(ระบุรายการครุภัณฑ์).....							
10	ศาล.....	(ระบุรายการครุภัณฑ์).....							
11	ศาล.....	(ระบุรายการครุภัณฑ์).....							
12	ศาล.....	(ระบุรายการครุภัณฑ์).....							
					รวม		เป็นเงิน		

ลงชื่อ..... ผู้ตรวจสอบ
(.....)
ตำแหน่ง.....

ลงชื่อ..... ผู้จัดทำ
(.....)
ตำแหน่ง.....

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

คำอธิบายแบบ ขงป. 3 - 4

แบบสรุปแผนความต้องการครุภัณฑ์ประจำปีงบประมาณ พ.ศ. ...

(สำหรับสำนักศาลยุติธรรมประจำภาค : สรุปความต้องการครุภัณฑ์ของศาลในสังกัดภาคฯ)

เป็นแบบสำหรับให้ภาคจัดทำสรุปความต้องการครุภัณฑ์ของภาคและศาลในสังกัดฯ โดยภาครวบรวม วิเคราะห์ ตรวจสอบแผนความต้องการครุภัณฑ์แต่ละรายการของภาคและศาลในสังกัดฯ แล้วนำมาจัดลำดับความสำคัญลงในแบบสรุปแผนความต้องการฉบับนี้ เพื่อให้เห็นภาพรวมของความต้องการครุภัณฑ์แต่ละรายการของภาคและศาลในสังกัดฯ ได้ชัดเจนมากขึ้น โดยมีรายละเอียดคำชี้แจงดังนี้

ช่อง (1) : เรียงลำดับความสำคัญของครุภัณฑ์

ช่อง (2) : ระบุชื่อหน่วยงานในสังกัดภาค

ช่อง (3) : ระบุรายการครุภัณฑ์ และคุณลักษณะตามมาตรฐานครุภัณฑ์ (ถ้ามี) โดยเรียงลำดับความสำคัญ

ช่อง (4) : ระบุหน่วยนับของครุภัณฑ์

ช่อง (5/1) : ระบุจำนวนครุภัณฑ์ที่ต้องการขอทดแทน

ช่อง (5/2) : ระบุจำนวนครุภัณฑ์ที่ต้องการขอเพิ่ม/ขอใหม่

ช่อง (5/3) : ระบุจำนวนรวมของครุภัณฑ์ที่ต้องการขอ = ช่อง(5/1) + ช่อง(5/2)

ช่อง (6) : ระบุราคามาตรฐานของครุภัณฑ์ (ถ้ามี) ถ้าไม่มีให้ใช้ราคาตามท้องตลาด ณ ขณะนั้น ทั้งนี้ ราคาต่อหน่วยหรือต่อชุดต้องไม่ต่ำกว่า 5,000 บาท

ช่อง (7) : ระบุจำนวนเงินที่ขอ = ช่อง(5/3) X ช่อง(6)

ช่อง (8) : ระบุประเภทความสำคัญของครุภัณฑ์อย่างใดอย่างหนึ่ง ดังนี้

จำเป็น : เป็นครุภัณฑ์จำเป็นสำหรับการปฏิบัติงาน หากไม่ได้รับการจัดสรรจะไม่สามารถปฏิบัติงานได้

เสริม : ครุภัณฑ์เสริมการปฏิบัติงาน หากได้รับการจัดสรรจะทำให้การปฏิบัติงานมีประสิทธิภาพมากขึ้น

สนับสนุน : ครุภัณฑ์สนับสนุนการปฏิบัติงาน หากไม่ได้รับการจัดสรร ก็ยังสามารถปฏิบัติงานได้

(พร้อมแนบแบบ ขงป. 3 - 2 ที่เป็นรายละเอียดครุภัณฑ์แต่ละรายการ แบบ ขงป. 3 - 2/1 กรณีครุภัณฑ์เครื่องเรือนฯ รายการผ้าม่าน และ ขงป. 3 - 3 กรณีครุภัณฑ์คอมพิวเตอร์ ของภาคและศาลในสังกัดฯ ที่ภาคนำมาจัดลำดับความสำคัญมาด้วย)

หมายเหตุ *: ขอให้ภาคตรวจสอบแบบ ขงป. 3 - 2 แบบ ขงป. 3 - 2/1 และแบบ ขงป. 3 - 3 ที่ภาคและศาลในสังกัดฯ จัดทำให้ครบถ้วนเรียบร้อย ก่อนจัดส่งไปยังสำนักงานศาลยุติธรรม เพื่อประโยชน์ในการพิจารณาอนุมัติงบประมาณให้ต่อไป*

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

แบบ ขงป. 4

แบบคำของบประมาณกิจกรรมปรับปรุงซ่อมแซมอาคารศาล สถานที่และบ้านพัก ประจำปีงบประมาณ พ.ศ. ..... เพิ่มเติม

(ค่าใช้จ่ายด้านลงทุน)

หน่วยงาน / ศาล .....

ลำดับที่	รายการ	งบประมาณต่อรายการ (บาท)	เหตุผลความจำเป็น	เอกสารประกอบคำของบประมาณ							
				แบบรูป		ประมาณราคา		แผนผัง		รูปถ่าย	
				มี	ไม่มี	มี	ไม่มี	มี	ไม่มี	มี	ไม่มี
	รวมทั้งสิ้น										

หมายเหตุ : 1) ให้ตรวจสอบคำของบประมาณ ว่ารายการใดที่ได้รับจัดสรรและให้ตัดออก

2) ในการจัดทำคำขอตั้งงบประมาณประจำปีหรือระหว่างปี จะต้องแนบแบบรูปรายการและประมาณการราคา ของแต่ละรายการพร้อมคำขอตั้งงบประมาณทุกครั้ง หากรายการใดไม่มีเอกสารประกอบดังกล่าว จะส่งคืนเพื่อให้หน่วยงานจัดเตรียมเอกสารให้ครบถ้วนต่อไป

ลายมือชื่อ.....

(.....)

ผู้อำนวยการ.....

โทรศัพท์.....

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ประวัติผู้เขียน

ชื่อ – นามสกุล พรพรรษชล มาทัพ

ประวัติการศึกษา

พ.ศ. 2543 - วิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์) มหาวิทยาลัยศิลปากร

พ.ศ. 2552 - นิติศาสตรบัณฑิต มหาวิทยาลัยสุโขทัยธรรมาธิราช

ประสบการณ์ทำงาน

พ.ศ.2547 - ปัจจุบัน นักวิชาการคอมพิวเตอร์ สำนักงานศาลยุติธรรม