

ระบบปัญญาประดิษฐ์สำหรับการจัดหมวดหมู่ข้อความและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐ

An artificial intelligence system for classifying public comments and suggestions toward government projects

อิทธิศักดิ์ ศรีดำ^{1*}
Idhisak Sridam^{1*}

Received: 9 July 2021 ; Revised: 24 January 2022 ; Accepted: 14 February 2022

บทคัดย่อ

ระบบรับฟังความคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการภาครัฐเป็นรูปแบบเว็บไซต์กระดานสนทนา รับฟังความคิดเห็นของประชาชน เป็นข้อความที่มีความหลากหลายไม่สามารถตีความหมายได้ และข้อความไม่ได้ถูกจำแนกหมวดหมู่ไว้อย่างชัดเจน รวมถึงขาดการสรุปผลสถิติข้อมูลที่เกี่ยวข้องกับความคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการภาครัฐเพื่อแสดงให้เห็นถึงการตอบกลับของภาครัฐให้ประชาชนได้รับทราบ บทความนี้มีจุดมุ่งหมายในการเสนอระบบการจัดหมวดหมู่ความคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐโดยวิธีปัญญาประดิษฐ์ ซึ่งเป็นการวิจัยเชิงทดลองด้วยการหาค่าความแม่นยำ ค่าความระลึก ค่าความถูกต้อง และค่าเอฟเมเชอร์ ผลการศึกษา พบว่า การประเมินผลแบบจำลองด้วยการวัดประสิทธิภาพการจำแนกหมวดหมู่ตามคุณลักษณะของข้อมูลทั้ง 3 ลักษณะ ได้แก่ หมวดหมู่ความคิดเห็นและข้อเสนอแนะ ประเภทโครงการ หน่วยงานหรือจังหวัดที่รับผิดชอบ การจำแนกคุณลักษณะด้านหมวดหมู่ข้อมูลความคิดเห็นและข้อเสนอแนะมีค่าความแม่นยำเท่ากับ 0.88 ค่าความระลึกเท่ากับ 0.86 ค่าความถูกต้องเท่ากับ 86.69% และค่าเอฟเมเชอร์เท่ากับ 0.87 ส่วนการจำแนกคุณลักษณะด้านประเภทโครงการ พบว่า ค่าความแม่นยำเท่ากับ 0.87 ค่าความระลึกเท่ากับ 0.85 ค่าความถูกต้องเท่ากับ 85.23% และค่าเอฟเมเชอร์เท่ากับ 0.85 การจำแนกคุณลักษณะด้านหน่วยงานหรือจังหวัดที่รับผิดชอบมีค่าความแม่นยำเท่ากับ 0.86 ค่าความระลึกเท่ากับ 0.83 ค่าความถูกต้องเท่ากับ 83.94% และค่าเอฟเมเชอร์เท่ากับ 0.84

คำสำคัญ: ระบบปัญญาประดิษฐ์ การจัดหมวดหมู่ ข้อความและข้อเสนอแนะ

Abstract

People participation systems for public comments and suggestions on government projects, using web-boards for public participation have a variety of content and the text data cannot be interpreted. The messages are not clearly categorized without summary of comments and suggestions data on government projects and cannot demonstrate public responses. This article aims to present a classification system for public comments towards government projects with artificial intelligence methods. This study used experimental research methodology with the model evaluation. The four indicators included precision value, recall value accuracy value and F-Measure value were measured. The study indicated that the model evaluation details are as follows: three data characteristics are categories of comments and suggestions indicated that precision = 0.88, recall = 0.86, accuracy = 86.69% and F-Measure = 0.87 ; project type indicated that precision = 0.87, recall = 0.85, accuracy = 85.23% and F-Measure = 0.85 ; and government agencies or provinces indicated that precision = 0.86, recall = 0.83, accuracy = 83.94% and F-Measure = 0.84.

Keywords: Artificial Intelligence System, Classification, Comments and Suggestions

¹ สาขาวิชาวิศวกรรมซอฟต์แวร์และระบบสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี สถาบันเทคโนโลยีปทุมวัน กรุงเทพฯ 10330

¹ Department of Software Engineering and Information System, Faculty of Science and Technology, Pathumwan Institute of Technology, Bangkok, Thailand 10330

* Corresponding author: Idhisak Sridam, Faculty of Science and Technology, Pathumwan Institute of Technology, Bangkok, Thailand 10330
E-mail: idhisak@pit.ac.th

บทนำ

การพัฒนาโครงการภาครัฐสำคัญต่อการพัฒนาประเทศในด้านการบริการประชาชนในรูปแบบสวัสดิการแห่งรัฐและการใช้ประโยชน์ร่วมกันของประชาชนในรูปแบบทรัพย์สินสาธารณะ โดยโครงการภาครัฐในแต่ละปีจะมีเป็นจำนวนมากและมีมูลค่าสูง ซึ่งตั้งแต่ปี พ.ศ. 2558-2561 ภาครัฐมีโครงการจำนวน 10,868,087 โครงการ มูลค่ารวม 3,478,323.21 ล้านบาท (Figure 1) นอกจากนี้ โครงการภาครัฐถือได้ว่าเป็นวิธีการ (Ways) ของยุทธศาสตร์ชาติ ที่ต้องสร้างความสมดุลของการใช้ทรัพยากร (Means) อย่างเหมาะสมและคุ้มค่าที่จะ

ทำให้บรรลุเป้าหมายของยุทธศาสตร์ชาติ (Ends) ในด้านต่างๆ และต้องให้ประชาชนเข้ามามีส่วนร่วม (People participation) ในกระบวนการจัดทำยุทธศาสตร์ชาติ รวมถึงต้องกำหนดให้มีการย้อนกลับ (Feedback) ของข้อมูล เพื่อปรับปรุงแก้ไขยุทธศาสตร์เมื่อพบกับปัญหาและอุปสรรคในระหว่างการขับเคลื่อนยุทธศาสตร์ชาติ (Colone, 2017) ซึ่งข้อมูลเกี่ยวกับข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการภาครัฐจัดอยู่ในส่วนของการให้ประชาชนเข้ามามีส่วนร่วมและการย้อนกลับของข้อมูล เพื่อปรับปรุงแก้ไขยุทธศาสตร์เมื่อพบกับปัญหาและอุปสรรคในระหว่างการขับเคลื่อนยุทธศาสตร์ชาติ

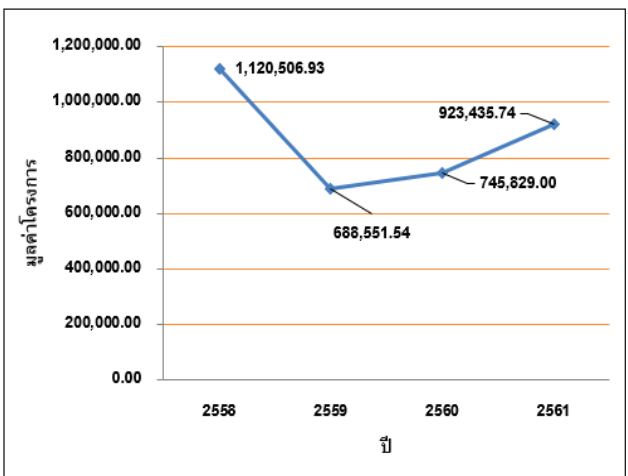
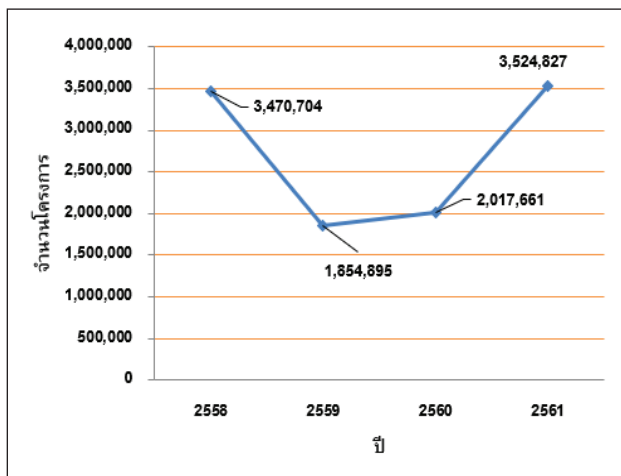


Figure 1 Government Projects Data (Thailand Government Spending, 2020)

ข้อมูลที่เกี่ยวข้องกับข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการภาครัฐมีความสำคัญต่อการพัฒนาด้านการวางแผนและการดำเนินงานของโครงการภาครัฐให้ได้ประโยชน์สูงสุดสามารถตอบสนองความจำเป็นของประชาชนและผลประโยชน์สูงสุดของชาติได้อย่างแท้จริง รวมถึงการนำข้อมูลที่ไต่ไปวิเคราะห์หาจุดปรับปรุงสำหรับการพัฒนาระบบรับฟังข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการภาครัฐให้มีประสิทธิภาพในอนาคต ปัจจุบันทางภาครัฐได้พัฒนาระบบรับข้อคิดเห็นและข้อเสนอแนะของประชาชนต่อโครงการภาครัฐ ในช่องทางออนไลน์ที่เว็บไซต์การรับฟังความคิดเห็นของประชาชน (Public communication) (Thailand Government Spending, 2020) ซึ่งประชาชนให้ความสนใจและได้แสดงข้อคิดเห็นและข้อเสนอแนะต่อโครงการภาครัฐผ่านทางเว็บไซต์การรับฟังความคิดเห็นของประชาชนระหว่างปี พ.ศ. 2549 ถึง พ.ศ. 2552 จำนวน 1,618 รายการ อย่างไรก็ตาม หลังจากปี พ.ศ. 2552 จนถึงปี พ.ศ. 2562 พบว่า ประชาชนเลิกให้ความสนใจในการแสดงข้อคิดเห็นและข้อเสนอแนะ ทำให้ภาครัฐขาดข้อมูลย้อนกลับในการพัฒนาการวางแผนและการดำเนินงานของโครงการภาครัฐ

ดังนั้น แสดงให้เห็นว่าระบบรับข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการภาครัฐไม่สามารถใช้ประโยชน์ได้เต็มประสิทธิภาพ และสาเหตุสำคัญที่ทำให้ประชาชนเลิกแสดงข้อคิดเห็นและข้อเสนอแนะต่อโครงการภาครัฐ คือหน่วยงานภาครัฐที่รับผิดชอบโครงการไม่ตอบกลับหรือแสดงสถานะให้ประชาชนเห็นว่าหน่วยงานภาครัฐที่รับผิดชอบโครงการได้รับฟังและนำข้อเสนอแนะไปใช้ประโยชน์ในอนาคต

การที่หน่วยงานภาครัฐที่รับผิดชอบโครงการไม่ตอบกลับหรือแสดงสถานะให้ประชาชนได้รับทราบ สาเหตุอาจมาจากจำนวนข้อมูลข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีจำนวนมากและไม่สามารถทราบว่ามีข้อมูลที่มีอยู่นั้นหน่วยงานภาครัฐใดเป็นผู้รับผิดชอบ ซึ่งข้อมูลเป็นข้อความที่มีความหลากหลายไม่สามารถตีความหมายได้ และข้อความไม่ได้ถูกจำแนกหมวดหมู่ไว้อย่างชัดเจน รวมถึงขาดการสรุปผลสถิติข้อมูลที่เกี่ยวข้องกับข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการภาครัฐเพื่อแสดงให้เห็นถึงการตอบกลับหรือแสดงสถานะของภาครัฐให้ประชาชนได้รับทราบ (Public Consultation, 2020)

ปัญญาประดิษฐ์ (Artificial intelligence) เป็นแนวทางการพัฒนาเทคโนโลยีในรูปแบบที่กำหนดให้ระบบงานคิดและตัดสินใจได้ใกล้เคียงกับมนุษย์ให้สามารถตอบสนอง การทำงานที่มากกว่าการเป็นระบบทั่วไป ประกอบด้วยลักษณะของระบบ 4 ลักษณะ ได้แก่ ความคิดที่เลียนแบบมนุษย์ การกระทำที่เหมือนมนุษย์ ความคิดอย่างมีเหตุผล และการกระทำแบบมีเหตุผล ข้อมูลข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐเป็นข้อมูลที่มีความหลากหลายไม่สามารถตีความหมายได้ เป็นข้อความที่ใช้ภาษาพูดทั่วไป และมีปริมาณข้อความ เป็นจำนวนมาก ดังนั้นการจัดหมวดหมู่ข้อมูลในลักษณะนี้แบบอัตโนมัติต้องอาศัยหลักการด้าน การเรียนรู้ของเครื่อง (Machine learning) ซึ่งเป็นส่วนหนึ่งของกรรมวิธีปัญญาประดิษฐ์ที่เน้นศึกษาและสร้างกรรมวิธี (Algorithm) ที่ทำให้ระบบเรียนรู้ได้จากข้อมูลสภาพแวดล้อมที่ถูกกำหนด เพื่อสร้างตัวแบบในการใช้หรือทำนายผลข้อมูลใหม่ซึ่งมีวัตถุประสงค์ในการเพิ่มประสิทธิภาพของระบบให้สูงขึ้น เมื่อระบบได้รับการเรียนรู้ในแต่ละครั้งแล้วระบบจะเก็บข้อมูลในลักษณะขององค์ความรู้ไว้ในฐานข้อมูลความรู้ในรูปแบบต่างๆ เช่น ฟังก์ชัน กฎ ข้อกำหนด เป็นต้น (Philip, 2019) ชนิดของการเรียนรู้ ประกอบด้วย 2 ชนิด ได้แก่ การเรียนรู้โดยการอนุมานแบบนिरนัย (Deductive) และการเรียนรู้โดยอนุมานแบบอุปนัย (Inductive) การเรียนรู้โดยการอนุมานแบบนिरนัย คือ การเรียนรู้โดยอาศัยความรู้ที่เป็นความจริงที่ได้รับการยอมรับเป็นสากล ซึ่งสามารถคาดการณ์เหตุการณ์ที่จะเกิดขึ้นได้อย่างแน่นอนตามรูปแบบของสภาพแวดล้อม ส่วนการเรียนรู้โดยอนุมานแบบอุปนัย คือ การเรียนรู้ สิ่งที่น่าสนใจ (Entity) หรือสถานการณ์ (Situation) โดยรู้ความจริงบางส่วนที่จะนำมาใช้อ้างอิงสำหรับการเรียนรู้และทำความเข้าใจเพื่อค้นหาความจริงในส่วนอื่นๆ ทั้งหมดจนเป็นความจริงที่ได้รับการยอมรับเป็นสากล โดยการเรียนรู้ของเครื่องเป็นการเรียนรู้โดยอนุมานแบบอุปนัย ประกอบด้วยกรรมวิธีเรียนรู้ 2 ลักษณะ ได้แก่ การเรียนรู้แบบมีผู้สอน (Supervised learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) การเรียนรู้แบบมีผู้สอน คือ การเรียนรู้ที่สามารถนำเสนอและจำแนกข้อมูลใดๆ ภายในชุดข้อมูลว่ามีผลลัพธ์ถูกหรือผิดได้ในลักษณะของการกำหนดเป้าหมายของคุณลักษณะของคำตอบ (Target attribute) นำไปใช้ในการจำแนกจัดหมวดหมู่ (Classification) และการทำนายหรือพยากรณ์ค่าของข้อมูลในลักษณะเชิงถดถอยต่อเนื่อง (Regression) ส่วนการเรียนรู้แบบไม่มีผู้สอน คือ การเรียนรู้ที่ไม่มีการกำหนดข้อมูลที่สนใจภายในชุดข้อมูลจึงไม่มีการจำแนกผลลัพธ์ว่าจะถูกผิดอย่างไร แต่จะเป็นการเรียนรู้ที่เกี่ยวข้องกับความสัมพันธ์ของข้อมูล ซึ่งนำไปใช้ในการจัดกลุ่มข้อมูลให้เป็นลักษณะกลุ่มๆ โดยไม่มีการกำหนดเป้าหมายของคุณลักษณะของคำตอบ (Non target attribute) ซึ่งนำไป ใช้

ในการรวบรวมข้อมูลให้เป็นกลุ่มข้อมูล (Clustering) (Philip, 2019 ; Alaa *et al.*, 2019)

การจัดหมวดหมู่ข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐเป็นการจำแนกข้อมูลในลักษณะของการจัดหมวดหมู่ในรูปแบบอัตโนมัติและไม่ซับซ้อนต้องอาศัยกรรมวิธีต้นไม้ตัดสินใจ (Decision tree) หรือการจำแนกแบบต้นไม้ (Classification) ซึ่งเป็นกรรมวิธีเรียนรู้แบบมีผู้สอนใช้สำหรับคาดคะเนหรือทำนายเหตุการณ์ที่จะเกิดขึ้นล่วงหน้าที่ได้จากการตัดสินใจในการจำแนกผลลัพธ์ด้วยการนำข้อมูลนำเข้ามาเป็นน้ำหนักในแต่ละโหนดของโครงสร้างข้อมูลต้นไม้ในลักษณะการแตกแขนงจากโหนดราก (Root) ถึงใบ (Leaf) และมีกิ่งก้าน (Branch) แตกออกไปตามเงื่อนไขหรือข้อมูลที่คาดคะเนไว้ การเกิดขึ้นของแต่ละเหตุการณ์ที่มีการเชื่อมโยงระหว่างสิ่งที่สนใจกับผลลัพธ์ที่คาดว่าจะเกิดขึ้นในสถานการณ์ต่างๆ ประกอบด้วย 3 องค์ประกอบ ได้แก่ ใบ กิ่งก้าน และราก ซึ่งใบ คือข้อมูลที่สนใจที่เป็นข้อมูลที่เกิดขึ้นโดยสภาพแวดล้อมตามสถานการณ์ต่างๆ หรือเป็นสิ่งที่กำหนดตามการคาดคะเนว่ามีโอกาสที่จะเกิดขึ้นตามสถานการณ์แวดล้อม ส่วน กิ่งก้าน คือ ข้อมูลที่เชื่อมโยงไปที่แตกออกมาจากโหนดต่างๆ เพื่อตัดสินใจว่าจะให้เกิดสถานการณ์ใดขึ้น ส่วนราก คือโหนดบนสุดที่จะส่งผลลัพธ์ไปสูใบ (ผลลัพธ์) ที่แตกต่างกัน (Philip, 2019) แสดงตัวอย่างกรรมวิธีต้นไม้ตัดสินใจ ดัง Figure 2

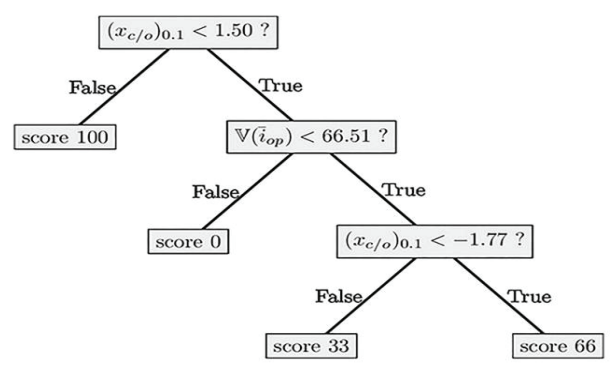


Figure 2 Example of Decision Tree Method (Audiffren *et al.*, 2016)

อย่างไรก็ตาม การใช้กรรมวิธีต้นไม้ตัดสินใจ ในการจำแนกข้อมูลในลักษณะของการจัดหมวดหมู่ในรูปแบบอัตโนมัติต้องอาศัยการค้นหาข้อความที่ใช้เป็นเป้าหมายของคุณลักษณะของคำตอบ ซึ่งข้อความดังกล่าวถูกซ่อนอยู่ในรูปประโยคต่างๆ ของข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐ การประมวลผลข้อความ (Text preprocessing) ถูกนำมาพิจารณาใช้ในการค้นหาข้อความที่ใช้เป็นเป้าหมายของคุณลักษณะของคำตอบของกรรมวิธีต้นไม้ตัดสินใจด้วยการสกัดข้อความเพื่อให้ได้

คุณลักษณะคำที่ต้องการ ประกอบด้วย 5 ขั้นตอน (Audiffren *et al.*, 2016) ได้แก่

1. การตัดคำ (Word extraction) เป็นขั้นตอนในการประมวลผลสำหรับการจำแนกหมวดหมู่ข้อความ สำหรับข้อความภาษาไทย มีจุดด้อยสำหรับการตัดคำเนื่องจากคำภาษาไทยมีลักษณะไม่มีเครื่องหมายวรรคตอนที่มีการแสดงการแบ่งคำอย่างชัดเจน ไม่มีช่องว่างคั่นที่แสดงขอบเขตของแต่ละคำเหมือนกับข้อความภาษาอังกฤษ ดังนั้น วิธีการตัดคำภาษาไทยวิธีหนึ่งที่สามารถใช้การตัดคำภาษาไทยได้คือ การตัดคำโดยใช้พจนานุกรม (Dictionary-based approach) ในการเปรียบเทียบสายอักขระระหว่างคำจากข้อความที่กำหนดกับคำที่ถูกจัดเก็บในพจนานุกรมร่วมกับการใช้เงื่อนไขในการตัดคำ

2. การกำจัดคำหยุด (Stop-word list removal) เป็นขั้นตอนนำคำในข้อความที่ไม่มีความสำคัญออกและยังคงความหมายของข้อความเหมือนเดิม

3. การหารากศัพท์ (Stemming) เป็นขั้นตอน การสร้างรูปแบบเดิมของคำหรือหาคำที่มีความหมายคล้ายกันให้อยู่เป็นกลุ่มคำที่มีความหมายเดียวกันโดยใช้คำใดคำหนึ่งเป็นตัวแทนของคำทั้งหมดในกลุ่มคำนั้น เพื่อเพิ่มประสิทธิภาพในการค้นคืนคำ

4. การสร้างดัชนีคำสำคัญ (Indexing) เป็นขั้นตอนการแปลงเอกสารที่เป็นภาษาธรรมชาติให้คอมพิวเตอร์สามารถเข้าใจและประมวลผลได้ การสร้างตัวแทนของกลุ่มข้อความโดยทั่วไปให้อยู่ในรูปแบบเวกเตอร์ (Word vector) คำนำหน้าคำซึ่งเป็นคำคุณลักษณะของกลุ่มข้อความ

5. การเลือกคุณลักษณะ (Feature selection) เป็นขั้นตอนเลือกคุณลักษณะที่จำเป็นต่อการนำข้อมูลไปใช้ประโยชน์เท่านั้น (Hacohen-Kerner *et al.*, 2020 ; Woo *et al.*, 2020) เนื่องจากกรรมวิธีต้นไม่ตัดสินใจ และกรรมวิธีอื่นของวิธีปัญญาประดิษฐ์เพื่อสร้างการจำแนกหมวดหมู่ข้อความ จะไม่สามารถรองรับการทำงานของจำนวนคุณลักษณะสูงๆ ได้ ซึ่งข้อคิดเห็นและข้อเสนอแนะของประชาชนจะมีแนวโน้มสูงขึ้นในอนาคต

จากความสำคัญและที่มาของปัญหาดังกล่าวข้างต้น การมีระบบการจัดหมวดหมู่ข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐในรูปแบบอัตโนมัติอาจจะเป็นแนวทางในการแก้ไขปัญหาดังกล่าวได้ การวิจัยนี้จึงเน้นศึกษาการจัดหมวดหมู่ข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐโดยวิธีปัญญาประดิษฐ์ เพื่อนำผลการศึกษาที่ได้ใช้เป็นข้อมูลอ้างอิงในการวิเคราะห์สำหรับการวางแผนและการดำเนินงานของโครงการภาครัฐให้ได้ประโยชน์สูงสุดสามารถตอบสนองความจำเป็นของประชาชนและผลประโยชน์สูงสุดของชาติ รวมถึงการนำข้อมูลที่ไปวิเคราะห์หาจุดปรับปรุงสำหรับการพัฒนาระบบรับฟังข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการภาครัฐให้มีประสิทธิภาพในอนาคต

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาระบบการจัดหมวดหมู่ข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐโดยวิธีปัญญาประดิษฐ์ในรูปแบบการเรียนรู้ของเครื่องด้วยกรรมวิธีต้นไม่ตัดสินใจ

วิธีดำเนินการวิจัย

การดำเนินงานวิจัยนี้ประกอบไปด้วย 4 ขั้นตอน ได้แก่ การเก็บรวบรวมข้อมูล การประมวลผลข้อความ การสร้างแบบจำลองการจำแนกหมวดหมู่ข้อความ และการประเมินผลแบบจำลอง โดยมีรายละเอียดดังนี้

การเก็บรวบรวมข้อมูล

การเก็บข้อมูลสำหรับงานวิจัยนี้ เป็นการเก็บรวบรวมข้อมูลจากระบบรับข้อคิดเห็นและข้อเสนอแนะของประชาชนต่อโครงการรัฐ (รูปแบบกระดานสนทนา) ในช่องทางออนไลน์ที่เว็บไซต์การรับฟังความคิดเห็นของประชาชน (Public communication) (Public Consultation, 2020) โดยเก็บข้อมูลข้อคิดเห็นและข้อเสนอแนะตั้งแต่วันที่ 1 สิงหาคม พ.ศ. 2549 ถึงเดือนเมษายน พ.ศ. 2552 จำนวน 1,618 รายการ โดยมีรายละเอียดและขอบเขตข้อมูล ดัง Table 1

Table 1 Details and scope of public comments and suggestions toward government projects

No	Filed	Description	Type
1	PostDate	วัน และเวลา	Date time
2	PostTitle	หัวข้อ	Varchar
3	PostDescription	รายละเอียด	Text
4	PostName	ผู้แสดงความคิดเห็น	Varchar

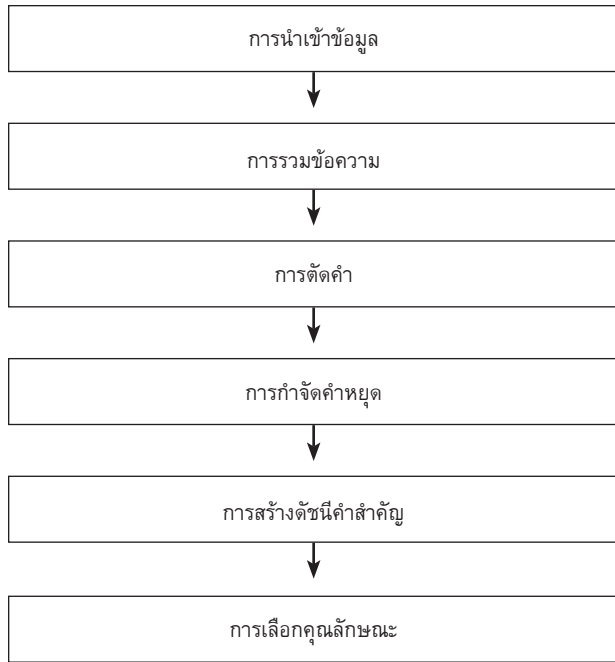


Figure 3 Text processing

การประมวลผลข้อความ

การประมวลผลข้อความเป็นการบวนการแปลงข้อความที่อยู่ในรูปแบบภาษาธรรมชาติที่ได้จากระบบรับข้อคิดเห็นและข้อเสนอแนะของประชาชนต่อโครงการรัฐให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถเรียนรู้ได้ด้วย การเขียนโปรแกรมคอมพิวเตอร์ ประกอบด้วย 6 ขั้นตอน ได้แก่ การนำเข้าข้อมูล การรวมข้อความ การตัดคำ การกำจัดคำหยุด การหารากศัพท์ การสร้างดัชนี คำสำคัญ และการเลือกคุณลักษณะ

1. การนำเข้าข้อมูล เป็นการรวบรวมข้อมูลจากระบบรับข้อคิดเห็นและข้อเสนอแนะของประชาชนต่อโครงการรัฐในช่องทางออนไลน์ที่เว็บไซต์การรับฟังความคิดเห็นของประชาชน³ มาแปลงไฟล์ข้อมูลที่มีนามสกุล.txt

2. การรวมข้อความ เป็นการรวมข้อความระหว่างหัวข้อ (หรือ Post title) และรายละเอียด (หรือ Post description) เป็นข้อความเดียวกัน เนื่องจากคำสำคัญจะถูกพบทั้งในหัวข้อและรายละเอียด ตัวอย่างดัง Table 2

Table 2 Example of text combine

Subject	Comments and suggestions	Text combine result
ทำไมเขตหนองแขมไม่มีการปรับปรุงทางเท้า	ทำไมเขตหนองแขมไม่มีการปรับปรุงทางเท้า ดิฉันเห็นเขตอื่นๆ ใน กทม. ก็เปลี่ยนพื้นทางเท้ากันแล้ว เวลาเดินบนทางเท้าของเขตหนองแขมแล้วหัวจะทิ่มคะ พื้นมันแหวกๆ โหวกๆ	ทำไมเขตหนองแขมไม่มีการปรับปรุงทางเท้าทำไมเขตหนองแขมไม่มีการปรับปรุงทางเท้า ดิฉันเห็นเขตอื่นๆ ในกทม. ก็เปลี่ยนพื้นทางเท้ากันแล้ว เวลาเดินบนทางเท้าของเขตหนองแขมแล้วหัวจะทิ่มคะ พื้นมันแหวกๆ โหวกๆ

3. การตัดคำ เป็นการสกัดคุณลักษณะด้วยการตัดคำเพื่อให้ได้คุณลักษณะจากกลุ่มข้อความในเอกสาร โดยใช้กลุ่มคลาสพจนานุกรมเล็กซ์ทู (LexTo) ซึ่งใช้เทคนิคการตัดคำแบบยาวที่สุด (Longest matching) สำหรับเปรียบเทียบสาย

อักขระระหว่างคำจากข้อความที่กำหนดกับคำที่ถูกจัดเก็บในพจนานุกรม (ไฟล์ lexitron.txt) ร่วมกับการใช้เงื่อนไขในการตัดคำ ตัวอย่างดัง Table 3

Table 3 Example of word segmentations

Comment and suggestion	Text combine result
ทำไมเขตหนองแขมไม่มีการปรับปรุงทางเท้าทำไมเขตหนองแขมไม่มีการปรับปรุงทางเท้า ดิฉันเห็นเขตอื่นๆ ใน กทม. ก็เปลี่ยนพื้นทางเท้ากันแล้ว เวลาเดินบนทางเท้าของเขตหนองแขมแล้วหัวจะทิ่มคะ พื้นมันแหวกๆ โหวกๆ	ทำไม เขต หนอง แขวง ไม่มี การปรับปรุง ทางเท้า ทำไม เขต หนอง แขวง ไม่มี การปรับปรุง ทางเท้า ดิฉัน เห็น เขต อื่นๆ ใน กทม. ก็ เปลี่ยน พื้น ทางเท้า กัน แล้ว เวลา เดิน บน ทางเท้า ของ เขต หนอง แขวง แล้ว หัว จะ ทิ่ม คะ พื้น มัน แหวก ๆ โหวก ๆ

4. การกำจัดคำหยุดเป็นการนำคำในข้อความ ที่ไม่มีความสำคัญ และเป็นคำที่ไม่มีความเกี่ยวข้องหรือไม่มีประโยชน์ในการจำแนกหมวดหมู่ออก แต่ยังคงความหมายของข้อความเหมือนเดิม เพื่อลดขนาดของดัชนีและลดขนาด

พื้นที่และเวลาในการประมวลผล ซึ่งใช้ประเภทของคำภาษาไทยมาใช้เป็นเกณฑ์หยุดคำ ได้แก่ คำบุพบท คำสันธาน คำสรรพนาม คำวิเศษณ์ และคำอุทาน ซึ่งตัวอย่างคำหยุดภาษาไทยดัง Table 4

Table 4 Example of thai stop words

Prepositions	Conjunctions	Pronouns	Adverbs	Interjections
บน	แต่	ผม	ไกล	เฮอ
ถึง	และ	ดิฉัน	ใกล้	แหม
จาก	จึง	เธอ	นี้	โ๊ะ
ด้วย	เพราะ	นาย	ง่าย	โถ
กับ	หรือ	คุณ	ยาก	เฮ้
ต่อ	พอ	กระผม	มาก	เอ๊ะ
สำหรับ	คือ	เขา	น้อย	หือ
เพื่อ	ทั้ง	มัน	ไหน	555

5. การสร้างดัชนีคำสำคัญ เป็นการสร้างตัวแทนของเอกสาร (Document representation) ในรูปแบบของเวกเตอร์ค่าน้ำหนักของคำ (Term weighting) โดยให้การหาค่าน้ำหนักรูปแบบคำเดียว (Single word) โดยค่าน้ำหนักของคำหาได้จากการเปรียบเทียบคำสำคัญในรายละเอียดข้อคิดเห็นและข้อเสนอแนะเทียบกับคำสำคัญในถ้อยคำแล้วนับจำนวนคำสำคัญที่พบในรายละเอียดข้อคิดเห็นและข้อเสนอแนะเพื่อนำความถี่มาคำนวณหาค่าน้ำหนักของคำหรือคุณลักษณะดังสมการ 1

6. การเลือกคุณลักษณะ เป็นการเลือกคุณลักษณะที่จำเป็นต่อการนำข้อมูลไปใช้ประโยชน์เท่านั้น เนื่องจากกรรมวิธีอื่นไม่ตัดสินใจหรือกรรมวิธีอื่นๆ ของวิธีปัญญาประดิษฐ์เพื่อสร้างการจำแนกหมวดหมู่ข้อความจะไม่สามารถรองรับการทำงานของจำนวนคุณลักษณะในปริมาณมากได้ ซึ่งข้อคิดเห็นและข้อเสนอแนะของประชาชนจะมีแนวโน้มสูงขึ้นในอนาคต โดยการวนทำซ้ำการสร้างดัชนีคำสำคัญ จำนวน 3 ครั้ง ตามลักษณะตัวแทนเอกสารในการจำแนกหมวดหมู่ 3 ลักษณะ ได้แก่ หมวดหมู่ข้อคิดเห็นและข้อเสนอแนะ ประเภทโครงการ หน่วยงานหรือจังหวัดที่รับผิดชอบ แสดงตัวอย่าง

รายละเอียดคุณลักษณะดัง Table 5 ส่วนหมวดหมู่ข้อคิดเห็นและข้อเสนอแนะ งานวิจัยนี้ได้กำหนดลักษณะตัวแทนเอกสารจากผลการวิเคราะห์ข้อมูลการสัมภาษณ์ผู้เชี่ยวชาญด้านภาษาไทยจำนวน 3 คน ประกอบด้วย 6 หมวดหมู่ ได้แก่ การประเมินผลกระทบสิ่งแวดล้อม การปฏิบัติตามระเบียบ/กฎหมาย การให้ประชาชนมีส่วนร่วม ความต้องการให้รัฐพัฒนาโครงการในพื้นที่ การตรวจสอบความโปร่งใสในการดำเนินงาน และการเร่งให้การดำเนินงานให้แล้วเสร็จ

$$W_{(f, d)} = TF_{(f, d)} \times \left(\log \frac{|D|}{|DF_{(f)}|} \right) \tag{1}$$

โดยที่	$W_{(f, d)}$	คือ	ค่าน้ำหนักของคุณลักษณะ (f) ที่ปรากฏในเอกสาร (d)
	$TF_{(f, d)}$	คือ	ความถี่ของคุณลักษณะ (f) ที่ปรากฏในเอกสาร (d)
	D	คือ	จำนวนเอกสารทั้งหมด
	$ DF_{(f)} $	คือ	จำนวนของเอกสารทั้งหมดที่มีคุณลักษณะ (f) ปรากฏอยู่

Table 5 Example of feature details

Example of suggestion	Feature	Example of document representation	Example of keywords in the bag of words
งานก่อสร้างถนนในตัวเทศบาลตำบลควรจะได้แล้วครับ ชาวบ้านลำบากมากในการเดินทางจากลูกหลานคนเมืองปทุมธานีครับ	หมวดหมู่ข้อคิดเห็นและข้อเสนอแนะ ประเภทโครงการ หน่วยงานหรือจังหวัดที่รับผิดชอบ	การเร่งให้การดำเนินงานให้แล้วเสร็จ *โครงการจ้างก่อสร้าง **องค์การบริหารส่วนตำบล หรือเทศบาลตำบล ***จังหวัดปทุมธานี	เมื่อไร, เสรี, ไม่เสร็จ, เวลา งานก่อสร้าง, ก่อสร้าง, สร้าง อบต., ตำบล, เทศบาลตำบล, ราชการประจำตำบล ปทุมธานี

* ลักษณะตัวแทนเอกสารตามคุณลักษณะด้านประเภทโครงการอ้างอิงมาจากข้อมูลประเภทโครงการ 5 ประเภท ได้แก่ โครงการจ้างก่อสร้าง โครงการจัดซื้อโครงการจ้างทำของ/เหมาบริการ โครงการเช่า และโครงการจ้างที่ปรึกษา (Thailand Government Spending, 2020)

** ลักษณะตัวแทนเอกสารตามคุณลักษณะด้านหน่วยงานหรือจังหวัดที่รับผิดชอบอ้างอิงมาจากรายชื่อหน่วยงานหลักของภาครัฐ จำนวน 15 หน่วยงาน (Thailand Government Spending, 2020)

*** ลักษณะตัวแทนเอกสารตามคุณลักษณะด้านจังหวัดที่รับผิดชอบอ้างอิงมาจากรายชื่อจังหวัด จำนวน 76 จังหวัด

การสร้างแบบจำลองการจำแนกหมวดหมู่ข้อความ

นำข้อมูลค่าน้ำหนักของคำสำคัญทั้ง 3 คุณลักษณะ ได้แก่ หมวดหมู่ข้อคิดเห็นและข้อเสนอแนะ ประเภทโครงการ หน่วยงานหรือจังหวัดที่รับผิดชอบ มาสร้างเป็นเอกสารนามสกุล .arff ซึ่งเป็นคุณลักษณะของข้อมูลทั้ง 3 ลักษณะ โดยใช้โปรแกรม Rapid Miner 9.7 โดยวิธีปัญญาประดิษฐ์ในรูปแบบการเรียนรู้ของเครื่องด้วยกรรมวิธีต้นไม้ตัดสินใจด้วยการประเมินผลแบบไขว้ 10 ชุด (10-fold cross validation) ต่อคุณลักษณะ โดยแบ่งข้อมูลเป็น 10 ชุด ซึ่งข้อมูลชุดที่ 1 เป็นชุดข้อมูลทดสอบ และอีก 9 ชุดที่เหลือเป็นชุดข้อมูลเรียนรู้ ทำวนจนครบ 10 ชุดข้อมูล

การประเมินผลแบบจำลอง

การประเมินผลแบบจำลองด้วยการวัดประสิทธิภาพการจำแนกหมวดหมู่ตามคุณลักษณะของข้อมูลทั้ง 3 ลักษณะ ได้แก่ หมวดหมู่ข้อคิดเห็นและข้อเสนอแนะ ประเภทโครงการ หน่วยงานหรือจังหวัดที่รับผิดชอบ โดยการหาค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าเอฟเมเชอร์ (F-measure) ดังสมการต่อไปนี้

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

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

โดยที่ TP คือ จำนวนข้อมูลที่ถูกต้องที่ถูกจำแนกออกมา
FP คือ จำนวนข้อมูลที่ไม่ถูกต้องที่ถูกจำแนกออกมา
TN คือ จำนวนข้อมูลที่ถูกต้องแต่ไม่ถูกจำแนกออกมา
FN คือ จำนวนข้อมูลที่ไม่ถูกต้องแต่ไม่ถูกจำแนกออกมา

ผลการวิจัย

ผลวิจัยจากการศึกษาระบบการจัดหมวดหมู่ข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐโดยวิธีปัญญาประดิษฐ์ด้วยการเรียนรู้ของเครื่องด้วยกรรมวิธีต้นไม้ตัดสินใจ (หรือ algorithm j48 (W-j48)) จากโปรแกรม Rapid Miner 9.7 แสดงได้ดัง Table 6

การจำแนกคุณลักษณะเกี่ยวกับหมวดหมู่ข้อมูลข้อคิดเห็นและข้อเสนอแนะ พบว่า ค่าความแม่นยำเท่ากับ 0.88 ค่าความระลึกเท่ากับ 0.86 ค่าความถูกต้องเท่ากับ 86.69% และค่าเอฟเมเชอร์เท่ากับ 0.87 ส่วนการจำแนกคุณลักษณะเกี่ยวกับประเภทโครงการ พบว่า ค่าความแม่นยำเท่ากับ 0.87 ค่าความระลึกเท่ากับ 0.85 ค่าความถูกต้องเท่ากับ 85.23% และค่าเอฟเมเชอร์เท่ากับ 0.85 และการจำแนกคุณลักษณะเกี่ยวกับหน่วยงานหรือจังหวัดที่รับผิดชอบ พบว่า ค่าความแม่นยำเท่ากับ 0.86 ค่าความระลึกเท่ากับ 0.83 ค่าความถูกต้องเท่ากับ 83.94% และค่าเอฟเมเชอร์เท่ากับ 0.84 โดยค่าต่างๆ ถ้ามีค่าสูงหมายถึงมีประสิทธิภาพสูง

Table 6 The result of a study of artificial intelligence for comments and suggestions classification system toward government projects using decision tree method

Features	Precision	Recall	Accuracy	F-measure
หมวดหมู่ข้อคิดเห็นและข้อเสนอแนะ	0.88	0.86	86.69%	0.87
ประเภทโครงการ	0.87	0.85	85.23%	0.85
หน่วยงานหรือจังหวัดที่รับผิดชอบ	0.86	0.83	83.94%	0.84

สรุปและอภิปรายผล

ผลวิจัยจากการศึกษา พบว่า ระบบปัญญาประดิษฐ์สำหรับการจัดหมวดหมู่ข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐโดยวิธีปัญญาประดิษฐ์ได้แก่ การเก็บรวบรวมข้อมูล การประมวลผลข้อความ การสร้างแบบจำลองการจำแนกหมวดหมู่ข้อความ และการประเมินผลแบบจำลอง เมื่อประเมินผลแบบจำลองด้วยการวัดประสิทธิภาพการจำแนกหมวดหมู่ตามคุณลักษณะของข้อมูลทั้ง 3 ลักษณะ ได้แก่ หมวดหมู่ข้อคิดเห็นและข้อเสนอแนะ ประเภทโครงการ หน่วยงานหรือจังหวัดที่รับผิดชอบ มีความเหมาะสมต่อ

การนำไปใช้งานในระดับสูง เนื่องจากการทดสอบการจำแนกคุณลักษณะเกี่ยวกับประเภทโครงการ พบว่า มีค่าความถูกต้อง ความแม่นยำ ความระลึก และค่าเอฟเมเชอร์อยู่ในระดับสูง โดยค่าความแม่นยำเท่ากับ 0.88 ค่าความระลึกเท่ากับ 0.86 ค่าความถูกต้องเท่ากับ 86.69% และค่าเอฟเมเชอร์เท่ากับ 0.87 ส่วนการจำแนกคุณลักษณะเกี่ยวกับประเภทโครงการ พบว่า ค่าความแม่นยำเท่ากับ 0.87 ค่าความระลึกเท่ากับ 0.85 ค่าความถูกต้องเท่ากับ 85.23% และค่าเอฟเมเชอร์เท่ากับ 0.85 และการจำแนกคุณลักษณะเกี่ยวกับหน่วยงานหรือจังหวัดที่รับผิดชอบ พบว่า ค่าความแม่นยำเท่ากับ 0.86

ค่าความระลึกเท่ากับ 0.83 ค่าความถูกต้องเท่ากับ 83.94% และค่าเอฟเมเชอร์เท่ากับ 0.84

ข้อเสนอแนะ

1. ควรนำผลวิจัยไปใช้อ้างอิงเพื่อปรับปรุงระบบรับข้อคิดเห็นและข้อเสนอแนะของประชาชนต่อโครงการรัฐในช่องทางออนไลน์ที่เว็บไซต์การรับฟังความคิดเห็นของประชาชน (Public communication) โดยการพัฒนาต่อยอดเป็นระบบจัดหมวดหมู่ข้อคิดเห็นและข้อเสนอแนะของประชาชนต่อโครงการรัฐด้วยการจำแนกเป็น 3 คุณลักษณะ ได้แก่ หมวดหมู่ข้อคิดเห็นและข้อเสนอแนะ ประเภทโครงการหน่วยงานหรือจังหวัดที่รับผิดชอบ

2. ในการทำวิจัยต่อไปควรพิจารณาเกี่ยวกับความจำเป็นในการหารากศัพท์ เพราะคำภาษาไทยบางคำสามารถสื่อความหมายได้ชัดเจนโดยศึกษาและพัฒนาคลังรากศัพท์ภาษาไทยเพื่อให้ระบบงานได้แม่นยำขึ้น

3. ในการทำวิจัยต่อไปควรพิจารณาเกี่ยวกับการนำวิธีปัญญาประดิษฐ์ในรูปแบบการเรียนรู้ของเครื่องด้วยการรวมวิธีอื่นๆ มาศึกษาเพิ่มเติม นอกจากกรรมวิธีต้นไม้ตัดสินใจ เช่น กรรมวิธีหาเพื่อนบ้านใกล้กันที่สุด (k-Nearest neighbor) กรรมวิธีสนับสนุนการเรียนรู้ของเครื่องแบบเวกเตอร์ (Support vector machine) กรรมวิธีแบบโครงข่ายใยประสาท (Neural networks) กรรมวิธีเชิงวิวัฒนาการ (Evolutionary computation) เป็นต้น

เอกสารอ้างอิง

- Alaa, A., Feras, H., Mohammad, K., Mahmoud, A. & Nahla, A. (2019). Dynamic detection of software defects using supervised learning techniques. *International Journal of Communication Networks and Information Security*, 11, 158-191.
- Audiffren, J., Bargiotas, I., Vayatis, N., Vidal, P., & Ricard, D.A. (2016). Non-linear scoring approach for evaluating balance: classification of elderly as fallers and non-fallers. *PLOS ONE*, 11, 1-12.
- Colone, S.S. (2017). *Strategic model and national strategy in the 21th century*, national defence-college of Thailand. Bangkok.
- Hacohen-Kerner, Y., Miller, D., & Yiga, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLOS ONE*, 15, 1-22.
- Philip, C.J. (2019). *Introduction to artificial intelligence* (7th ed.). Dover.
- Public Consultation. (2020). <http://www.publicconsultation.opm.go.th/Web/Index>
- Thailand Government Spending. (2020). from <https://govspending.data.go.th>
- Woo, H., Kim, J., & Lee, W. (2020). Validation of text data preprocessing using a neural network model. *Mathematical Problems in Engineering*, 10(11), 3681.