

การจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ โดยใช้เทคนิคเหมืองข้อความ

Project Classification Based on Master Plan under Nation
Strategy Using Text Mining

กัญญารัตน์ พูลเพิ่ม^{1*} และ แกมกาญจน์ สมประเสริฐศรี²
Kanyarat Poolperm^{1*} And Gamgarn Somprasertsri²

สาขาสารสนเทศศาสตร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม^{1,2}
Information Science Faculty of Informatics, Mahasarakham University^{1,2}
E-Mail : nan1011285@gmail.com^{1*}, gamgarns@msu.ac.th²

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อความในการสร้างแบบจำลอง เพื่อจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ และวัดความถูกต้องของการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติโดยใช้แบบจำลองที่พัฒนาขึ้น โดยใช้ชุดข้อมูลโครงการจำนวน 1,150 โครงการจากระบบติดตามและประเมินผลแห่งชาติ (eMenscr) งานวิจัยนี้ใช้เทคนิคเหมืองข้อความเพื่อสร้างแบบจำลองในการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ โดยใช้เทคนิค 4 เทคนิค ได้แก่ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคป่าแบบสุ่ม เทคนิคนาอียเบย์ และเทคนิคการค้นหาเพื่อนบ้านใกล้สุด

ผลการวิจัย พบว่า เทคนิคป่าแบบสุ่มมีประสิทธิภาพสูงที่สุดในการจำแนกโครงการ โดยมีค่าความถูกต้อง 82.50% ความแม่นยำ 83.30% ความระลึก 82.50% และความถ่วงดุล 82.50% นอกจากนี้ยังพบว่า เทคนิคป่าแบบสุ่มมีความถูกต้องสูงที่สุดในการจำแนกโครงการใหม่จำนวน 22 โครงการ โดยมีค่าความถูกต้องเท่ากับ 90.91 %

คำสำคัญ: เหมืองข้อความ, การจำแนกข้อความ, ยุทธศาสตร์ชาติ, โครงการ

ABSTRACT

The objectives of this research were to compare the effectiveness of text mining technique used for developing a model to classify projects based on a master plan under a national strategy and measure the accuracy of a project classification model based on a master plan under a national strategy from the development model. The dataset was 1,150 projects collected from the National Monitoring and Evaluation System (eMenscr). In this research, the text mining technique was used to produce a project classification model based on a master plan under a national strategy which four techniques: Support Vector Machine, Random Forest, Naive Bayes, and k-Nearest Neighbors.

The results showed that Random Forest was the most effective technique, with accuracy at 82.50%, precision at 83.30%, recall at 82.50%, and F-measure at 82.50%. Moreover, Random Forest had the highest accuracy at 90.91% in classifying 22 new projects.

Keywords : Text mining, Text Classification, National Strategy, Project

บทนำ

รัฐธรรมนูญแห่งราชอาณาจักรไทย มาตรา 65 กำหนดให้รัฐพึงจัดให้มียุทธศาสตร์ชาติเป็นเป้าหมายการพัฒนาประเทศอย่างยั่งยืนตามหลักธรรมาภิบาล เพื่อใช้เป็นกรอบในการจัดทำแผนต่าง ๆ ดังนั้น แผนแม่บทภายใต้ยุทธศาสตร์ชาติ เป็นแผนแม่บทเพื่อบรรลุเป้าหมายตามที่กำหนดไว้ในยุทธศาสตร์ชาติ มีทั้งสิ้น 23 แผนแม่บท ซึ่งจะมีผลผูกพันต่อหน่วยงานของรัฐที่เกี่ยวข้องที่จะต้องปฏิบัติให้เป็นไปตามนั้น [1] ทั้งนี้ หน่วยงานภาครัฐที่ต้องจัดทำโครงการต่าง ๆ เพื่อขอรับงบประมาณจากภาครัฐ จึงมีความจำเป็นที่จะต้องกำหนดแผนแม่บทยุทธศาสตร์ชาติให้สอดคล้องกับโครง นอกจากนี้ในการเสนอโครงการทุกหน่วยงานจะต้องใช้ระบบติดตามและประเมินผลแห่งชาติ (Electronic Monitoring and Evaluation System of National Strategy and Country Reform : eMENSCR) ในการจัดเก็บและติดตามความก้าวหน้าของโครงการ

ภายใต้การดูแลการนำเข้าสู่ข้อมูลโครงการใน eMENSCR เป็นหน้าที่ของหน่วยงานยุทธศาสตร์และข้อมูลเพื่อการพัฒนาจังหวัด ซึ่งในขณะนำเข้าสู่ข้อมูลโครงการ พบปัญหา คือ การเชื่อมโยงโครงการกับแผนระดับต่าง ๆ โดยเฉพาะการเชื่อมโยงกับแผนแม่บทภายใต้ยุทธศาสตร์ชาติแต่ละด้าน เพราะส่วนของแผนแม่บทภายใต้ยุทธศาสตร์ชาติมีความละเอียดและซับซ้อน ทำให้หน่วยงานที่เสนอโครงการเลือกแผนแม่บทภายใต้ยุทธศาสตร์ชาติไม่สอดคล้องกับโครงการ ทำให้การอนุมัติโครงการเกิดความล่าช้า เนื่องจากสภาพนาเศรษฐกิจและสังคมแห่งชาติจะส่งคืนโครงการให้จังหวัดทำการปรับแก้ไขการเชื่อมโยงโครงการให้สอดคล้องกับแผนแม่บทภายใต้ยุทธศาสตร์ชาติที่ถูกต้อง ผู้วิจัยจึงเห็นว่า ถ้ามีการนำเทคโนโลยีมาช่วยในการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ ได้อย่างอัตโนมัติ จะช่วยแก้ปัญหาการเชื่อมโยงโครงการกับแผนแม่บทภายใต้ยุทธศาสตร์ชาติ ได้อย่างมีประสิทธิภาพ [2]

ปัจจุบันมีการนำเทคนิคเหมืองข้อความ (Text Mining) เข้ามาช่วยในการจำแนกเอกสารหรือข้อความแบบอัตโนมัติ เช่น งานวิจัยของ อังศุมล สุทธิภักดี [3] ศึกษาเกี่ยวกับการเปรียบเทียบประสิทธิภาพของการจำแนกหมวดหมู่ของข้อคิดเห็นในแบบสอบถามปลายเปิด โดยใช้เทคนิคนาอียเบย์และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน งานวิจัยของ สุภัตสร สมเจตนา [4] ศึกษาเกี่ยวกับการเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ตโฟนของบุตร โดยใช้เทคนิคเรปอร์ เทคนิคต้นไม้ตัดสินใจ ซี 4.5 เทคนิคนาอียเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคป่าแบบสุ่มและเทคนิคการค้นหาเพื่อนบ้านใกล้สุด งานวิจัยของ ประเดิม วงศ์กระโซ่ [5] ศึกษาเกี่ยวกับระบบจำแนกหมวดหมู่การแจ้งซ่อมบ้านออนไลน์โดยใช้เหมืองข้อความโดยใช้เทคนิคนาอียเบย์ เทคนิคต้นไม้ตัดสินใจและเทคนิคการค้นหาเพื่อนบ้านใกล้สุด ซึ่งพบว่าการใช้เทคนิคเหมืองข้อความสามารถนำมาใช้ในการจำแนกเอกสารหรือข้อความได้อย่างมีประสิทธิภาพ

ผู้วิจัยจึงสนใจพัฒนาและเปรียบเทียบประสิทธิภาพแบบจำลองการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ โดยงานวิจัยมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อความในการสร้างแบบจำลองจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ และวัดความถูกต้องของการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติโดยใช้แบบจำลองที่พัฒนาขึ้น การจำแนกโครงการใช้ชุดข้อมูลโครงการจำนวน 1,150 โครงการ จากระบบติดตามและประเมินผลแห่งชาติ (eMenscr) และใช้เทคนิค 4 เทคนิค ได้แก่ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคป่าแบบสุ่ม เทคนิคนาอียเบย์ และเทคนิคการค้นหาเพื่อนบ้านใกล้สุด โดยวัดค่าความแม่นยำ ค่าความระลึก ค่าความถูกต้อง และค่าความถ่วงดุล เพื่อให้ได้แบบจำลองสำหรับจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติที่เหมาะสม ที่สามารถนำไปพัฒนาเป็นระบบเชื่อมโยงโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติแบบอัตโนมัติต่อไป ซึ่งจะช่วยให้การปฏิบัติงานของเจ้าหน้าที่ในการนำโครงการเข้าในระบบติดตามและประเมินผลแห่งชาติสามารถตัดสินใจหรือเชื่อมโยงโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ ได้อย่างรวดเร็วและถูกต้อง

1. วัตถุประสงค์การวิจัย

1.1 เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อความในการสร้างแบบจำลองเพื่อจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ

1.2 เพื่อวัดความถูกต้องในการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติของแบบจำลองที่พัฒนาขึ้น

2. สมมติฐานการวิจัย

ประสิทธิภาพของแบบจำลองการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติที่พัฒนาขึ้นมีประสิทธิภาพมากกว่า 80%

3. เอกสารและงานวิจัยที่เกี่ยวข้อง

3.1 เอกสารที่เกี่ยวข้อง

ยุทธศาสตร์ชาติ หมายถึง กรอบการพัฒนาการบริหารประเทศระยะยาว เพื่อให้ประเทศสามารถบรรลุวิสัยทัศน์ที่ตั้งเป้าหมายเอาไว้ โดยเป็นเป้าหมายที่ทางภาครัฐ ภาคเอกชนและประชาชนร่วมมือกันกำหนดขึ้นเพื่อเป็นแนวทางในการพัฒนาประเทศ โดยปัจจุบันประเทศไทยใช้ ยุทธศาสตร์ชาติ (พ.ศ. 2561 - 2580) โดยมี 23 แผนแม่บทย่อย ประกอบด้วย 1) ความมั่นคง 2) การต่างประเทศ 3) การเกษตร 4) อุตสาหกรรมและบริการแห่งอนาคต 5) การท่องเที่ยว 6) พื้นที่และเมืองน่าอยู่อัจฉริยะ 7) โครงสร้างพื้นฐาน ระบบโลจิสติกส์ และดิจิทัล 8) ผู้ประกอบการและวิสาหกิจขนาดกลางและขนาดย่อมยุคใหม่ 9) เขตเศรษฐกิจพิเศษ 10) การปรับเปลี่ยนค่านิยมและวัฒนธรรม 11) ศักยภาพคนตลอดช่วงชีวิต 12) การพัฒนาการเรียนรู้ 13) การเสริมสร้างให้คนไทยมีสุขภาพที่ดี 14) ศักยภาพการกีฬา 15) พลังทางสังคม 16) เศรษฐกิจฐานราก 17) ความเสมอภาคและหลักประกันทางสังคม 18) การเติบโตอย่างยั่งยืน 19) การบริหารจัดการน้ำทั้งระบบ 20) การบริการประชาชนและประสิทธิภาพภาครัฐ 21) การต่อต้านการทุจริตและประพฤติมิชอบ 22) กฎหมายและกระบวนการยุติธรรม 23) การวิจัยและพัฒนานวัตกรรม เพื่อเป็นกรอบในการขับเคลื่อนภาครัฐและเอกชน ยุทธศาสตร์ชาติจึงมีความสำคัญอย่างยิ่งในการจัดทำแผนงานและโครงการ ในปัจจุบัน สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติจึงได้จัดทำระบบติดตามและประเมินผลแห่งชาติ (eMENSER) เพื่อเป็นเครื่องมือให้หน่วยงานของรัฐใช้รายงานผลการดำเนินการตามแผนระดับต่างๆ ของประเทศ ซึ่ง "หน่วยงานของรัฐทุกหน่วย" มีหน้าที่ดำเนินการเพื่อให้บรรลุเป้าหมายตามที่กำหนดไว้ในยุทธศาสตร์ชาติ [6]

การทำเหมืองข้อความ หมายถึง การค้นหาความรู้ใหม่หรือข้อเท็จจริงที่แฝงอยู่ในชุดข้อมูลเป็นเทคนิคเพื่อค้นหารูปแบบจากข้อมูลจำนวนมากโดยอัตโนมัติ [7] งานวิจัยนี้ใช้การจำแนกข้อความ (Text classification) ซึ่งเป็นงานด้านหนึ่งของการทำเหมืองข้อความมาช่วยในการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ โดยเทคนิคที่นำมาใช้ได้แก่ เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคการค้นหาเพื่อนบ้านใกล้สุด และเทคนิคป่าแบบสุ่ม

เทคนิคนาอิวเบย์ (Naive Bayes) ใช้คำนวณหาความน่าจะเป็นที่ข้อมูลน่าจะถูกจำแนกอยู่ในประเภทใด แต่จะลดความซับซ้อนลงโดยเพิ่มสมมติฐานที่ว่าคุณลักษณะต่าง ๆ ของข้อมูลจะไม่ขึ้นต่อกัน [8]

เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เป็นการใช้สมการเส้นตรงเพื่อแบ่งเขตข้อมูล 2 กลุ่มออกจากกัน ซึ่งหลักการของซัพพอร์ตเวกเตอร์แมชชีนคือการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นจำแนกประเภทข้อมูลในขั้นตอนการเรียนรู้และเลือกเส้นจำแนกประเภทข้อมูลที่เหมาะสมที่สุด [8]

เทคนิคการค้นหาเพื่อนบ้านใกล้สุด (k-Nearest Neighbors) การจำแนกข้อมูลด้วยวิธีการค้นหาเพื่อนบ้านใกล้สุดจะเป็นการเรียนรู้โดยการเปรียบเทียบกันระหว่างเรคคอร์ดของข้อมูลที่ต้องการจำแนก/ทำนายหมวดหมู่กับเรคคอร์ดทั้งหมดในชุดข้อมูลสอนที่มีลักษณะเหมือนกันหรือใกล้เคียงกันด้วยการพิจารณาข้อมูลแอตทริบิวต์ต่างๆ [8]

เทคนิคป่าแบบสุ่ม (Random Forest) แนวคิดของเทคนิคป่าแบบสุ่ม คือการสร้างโมเดลด้วยวิธีการสร้างต้นไม้ตัดสินใจ (Decision Tree) ขึ้นมาหลายๆ โมเดลโดยวิธีการสุ่มตัวแปร แล้วนำผลที่ได้แต่ละโมเดลมารวมกันพร้อมนับจำนวนผลที่มีจำนวนซ้ำกันมากที่สุด สกัดออกมาเป็นผลลัพธ์สุดท้าย [8]

สำหรับเทคนิคที่นำมาใช้ในการจำแนกข้อความจากงานวิจัยที่ผ่านมาวิเคราะห์ได้ว่าเทคนิคที่นิยมใช้ประกอบด้วย เทคนิคการค้นหาเพื่อนบ้านใกล้สุด เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคป่าแบบสุ่ม และเทคนิคต้นไม้ตัดสินใจ อย่างไรก็ตามเทคนิคป่าแบบสุ่มเป็นการรวมต้นไม้ตัดสินใจเข้าด้วยกัน ในงานวิจัยนี้ผู้วิจัยจึงเลือกใช้ เทคนิคการค้นหาเพื่อนบ้านใกล้สุด เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคป่าแบบสุ่ม ในการพัฒนาแบบจำลองการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์

3.2 งานวิจัยที่เกี่ยวข้อง

จากงานวิจัยที่เกี่ยวข้องเกี่ยวกับการจำแนกข้อความโดยใช้เทคนิคเมืองข้อความสามารถสรุปเทคนิคที่ใช้ในงานวิจัยได้ ดังต่อไปนี้

นุชนาฏ ปิ่นเมือง [9] ได้ศึกษาการจำแนกความคิดเห็นของคนไทยเกี่ยวกับสื่อออนไลน์โดยใช้การทำเหมืองข้อความ เปรียบเทียบเทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคการค้นหาเพื่อนบ้านใกล้สุด เทคนิคต้นไม้ตัดสินใจแบบ ซี 4.5 พบว่า เทคนิคนาอิวเบย์ ให้ผลดีที่สุดในการจำแนกความคิดเห็น โดยให้ค่าความถูกต้อง 93.88% ค่าความแม่นยำ 94.02% และค่าความระลึก 93.54%

สุภัสสรา สมเจตนา [4] ได้ศึกษาการเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร โดยใช้ 6 เทคนิค ได้แก่ เทคนิคริเบเปอร์ เทคนิคต้นไม้ตัดสินใจ แบบ ซี 4.5 เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคการค้นหาเพื่อนบ้านใกล้สุด และเทคนิคต้นไม้ป่าสุ่ม พบว่า เทคนิคป่าสุ่ม เป็นเทคนิคที่ดีที่สุดในการวิเคราะห์ความคิดเห็นสำหรับข้อมูลชุดนี้โดยให้ค่าความแม่นยำ 89.62% ค่าความระลึก 78.38% และค่าความถ่วงดุล 83.55%

จิรายุ ชัยมีบุญ [10] ได้ศึกษาการประยุกต์ใช้ปัญญาประดิษฐ์กับหุ่นยนต์สนทนาบนระบบไลน์ เพื่อช่วยในการบริหารงาน IT Support ยุค 4.0 โดยใช้เทคนิคที่ 3 เทคนิค ได้แก่ เทคนิคป่าแบบสุ่ม เทคนิคนาอิวเบย์ และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน พบว่าเทคนิคป่าแบบสุ่ม มีประสิทธิภาพในการจำแนกสูงที่สุด ค่าความถูกต้อง 81% ค่าความแม่นยำ 84% และค่าการวิเคราะห์เส้นโค้งคุณลักษณะสมบัติการทำงาน 96%

จากการวิเคราะห์เทคนิคที่ใช้ในงานวิจัยที่ผ่านมา พบว่าเทคนิคที่นิยมใช้ ประกอบด้วย เทคนิคการค้นหาเพื่อนบ้านใกล้สุด เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคป่าแบบสุ่ม และเทคนิคต้นไม้ตัดสินใจ อย่างไรก็ตามเทคนิคป่าแบบสุ่มเป็นการรวมต้นไม้ตัดสินใจเข้าด้วยกัน ผู้วิจัยจึงเลือกใช้ เทคนิคการค้นหาเพื่อนบ้านใกล้สุด เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคป่าแบบสุ่ม ในการพัฒนาแบบจำลองการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ

วิธีดำเนินการวิจัย

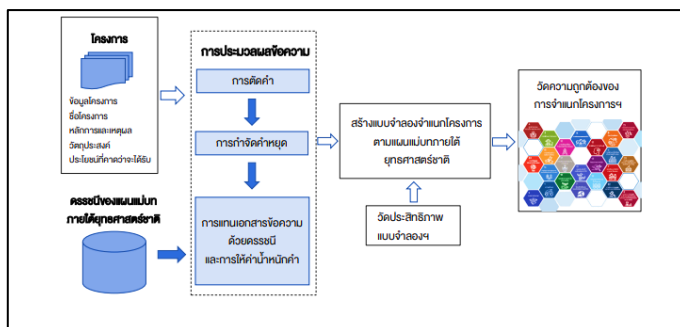
การวิจัยครั้งนี้เป็นการวิจัยและพัฒนาแบบจำลองการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ โดยใช้เทคนิคเมืองข้อความ มีรายละเอียดของการดำเนินการวิจัยดังนี้

1. เครื่องมือการวิจัย

1.1 โปรแกรม Google Colab ใช้ในการสร้างแบบจำลองการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ

1.2 เทคนิคที่ใช้ในการจำแนกข้อความ 4 เทคนิค ได้แก่ เทคนิคนาอิวเบย์ (Naive Bayes) เทคนิคการค้นหาเพื่อนบ้านใกล้สุด (k-Nearest Neighbors) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) และเทคนิคป่าแบบสุ่ม (Random Forest)

กระบวนการดำเนินการวิจัยประกอบด้วย 5 ขั้นตอน



ภาพที่ 1 ขั้นตอนในการดำเนินการวิจัย

จากภาพที่ 1 โดยเริ่มจากการรวบรวมข้อมูลโครงการ การประมวลผลข้อความ การสร้างแบบจำลอง วัดประสิทธิภาพแบบจำลอง และวัดความถูกต้องของการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ ซึ่งมีรายละเอียดดังนี้

2.1 การรวบรวมข้อมูลโครงการ

ในการรวบรวมข้อมูล ผู้วิจัยทำการเก็บรวบรวมข้อมูลจากระบบติดตามและประเมินผลแห่งชาติ (eMENSCR) จำนวน 1,150 โครงการ ซึ่งเป็นโครงการของปี พ.ศ. 2562-2564 ดังภาพที่ 2 โดยข้อมูลขอโครงการที่นำมาใช้ในการจำแนก ได้แก่ ชื่อโครงการ หลักการและเหตุผล วัตถุประสงค์ และประโยชน์ที่คาดว่าจะได้รับ

รายชื่อโครงการ / การดำเนินงาน						
แสดงเอกสาร		25	รายการ			
รหัสโครงการ	ชื่อโครงการ / การดำเนินงาน	บุคคลากรที่รับผิดชอบโครงการ	แผนภูมิรูปที่แสดงชื่อโครงการ	วันที่เปิดข้อมูลล่าสุด	สถานะ	จัดการโครงการ
บท1112-62-0030	งานผลิตสื่อวีดิทัศน์ โครงการส่งเสริมการเรียนรู้ด้านภาษาและวัฒนธรรม	นางสาว น.ก. 2562		1 ตุลาคม 2019 เวลา 9:42	จบแล้ว	
บท1120-61-0007	การพัฒนาระบบงานการติดตามและประเมินผลของบุคลากรด้านการปฏิบัติงาน (MENSOR)			13 กันยายน 2019 เวลา 11:33	กำลังพัฒนา	
บท1112-62-0024	โครงการการศึกษานานาชาติของบุคลากรด้านการปฏิบัติงาน			4 กันยายน 2019 เวลา 18:59	จบแล้ว	

ภาพที่ 2 ตัวอย่างโครงการในระบบ eMENSCR

2.2 การประมวลผลข้อความ

ขั้นตอนการประมวลผลข้อความที่เป็นกระบวนการในการจัดทำข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ข้อมูลได้ ซึ่งมีขั้นตอนดังนี้

2.2.1 การตัดคำภาษาไทย ผู้วิจัยใช้ฟังก์ชัน word_tokenize ของ PyThaiNLP ในการตัดคำตัวอย่างโครงการที่ผ่านขั้นตอนการตัดคำ

ตารางที่ 2 ตัวอย่างการตัดคำภาษาไทย

ข้อมูลโครงการ	ข้อมูลโครงการโดยการตัดคำ
เพิ่มประสิทธิภาพเชื่อมโยงข้อมูลสำนักงาน ปปง ศูนย์แลกเปลี่ยน ข้อมูลกระบวนการยุติธรรมกระบวนการยุติธรรมมีความจำเป็นต่อการ สร้างความสงบเรียบร้อยความยุติธรรมให้เกิดขึ้นแก่สังคม รัฐบาล จึงมีนโยบาย	เพิ่มประสิทธิภาพ, เชื่อมโยง, ข้อมูล, สำนักงาน, ปปง, ศูนย์, แลกเปลี่ยน, ข้อมูล, กระบวนการ, ยุติธรรม, กระบวนการยุติธรรม, มี ความจำเป็น, ต่อ, การ, สร้าง, ความสงบเรียบร้อย, ความยุติธรรม, ให้, เกิดขึ้น, แก่, สังคม, รัฐบาล, จึง, มี, นโยบาย

2.2.2 การกำจัดคำหยุด ผู้วิจัยใช้ฟังก์ชัน thai_stopwords ของ PyThaiNLP ตัวอย่างคำหยุดที่ใช้ในงานวิจัย

ตารางที่ 3 ตัวอย่างคำหยุดที่ใช้ในงานวิจัย

กล่าว	ต่างๆ	ระหว่าง	เดียว
กว่า	ต้อง	รับ	เดียวกัน
กัน	ถึง	ราย	เนื่องจาก
กับ	ถูก	ร่วม	เปิด
การ	ถ้า	ลง	เปิดเผย
ก็	ทั้ง	วัน	เป็น

ดังนั้นเมื่อทำการประมวลผลข้อความข้อมูลจากข้อมูลโครงการ จำนวนคำโดยเฉลี่ยของโครงการอยู่ที่ 350 คำ

2.2.3 การแทนเอกสารข้อความด้วยดรชนีและการให้ค่าน้ำหนักคำ มีรายละเอียดดังนี้

1) การแทนเอกสารข้อความด้วยดรชนี ผู้วิจัยนำข้อมูลโครงการที่ผ่านกระบวนการตัดคำและกำจัดคำหยุดแล้วไปหาความถี่ของคำ และเลือกคำที่มีความถี่มากกว่า 3 ครั้งขึ้นไป เพื่อนำมาสร้างดรชนี โดยให้นักวิเคราะห์นโยบายและแผนซึ่งเป็นผู้เชี่ยวชาญด้านการเชื่อมโยงแผนตรวจสอบคำดรชนีที่สอดคล้องกับแผนแม่บท 23 แผนย่อย ตัวอย่าง คำดรชนีที่สร้างขึ้นสำหรับแผนแม่บทภายใต้ยุทธศาสตร์ชาติ แผนย่อยความมั่นคง ซึ่งมีทั้งหมด 172 คำ

ตารางที่ 4 ตัวอย่างคำดรชนีของแผนแม่บทภายใต้ยุทธศาสตร์ชาติ ประเด็นความมั่นคง

แผนที่ 1 ความมั่นคง (172 คำ)		
ความมั่นคง	คดีอาญา	ผู้มีอำนาจ
คนต่างด้าว	คำประเวณี	พอกเงิน
ความยุติธรรม	ยาเสพติด	...

2) คำนำนั้ TF-IDF หลังจากแทนข้อมูลโครงการด้วยดรชนีแล้ว ผู้วิจัยสร้างเมตริกซ์ของดรชนี และใช้วิธีการให้น้ำหนักของคำในเอกสารแบบ TF-IDF ตัวอย่างเมตริกซ์ของดรชนีเอกสารที่ให้ค่าน้ำหนักคำด้วยค่า TF-IDF

ตารางที่ 5 ตัวอย่างเมตริกซ์ของกรณีเอกสาร

โครงการ	ยุทธธรรม	รัฐบาล	สารสนเทศ	พอกเงิน	อิเล็กทรอนิกส์	ที่ดิน	เกษตร	
ศูนย์แลกเปลี่ยนข้อมูลกระบวนการยุติธรรมกระบวนการยุติธรรมเป็นมีความจำเป็นต่อสร้างความสงบเรียบร้อยความยุติธรรมให้เกิดขึ้นแก่สังคม รัฐบาลจึงมีนโยบายให้มินำระบบเทคโนโลยีสารสนเทศมาใช้ในกระบวนการยุติธรรม	0.3	0.1	0.1	0	0	0	0	
แผนงานจัดทำมาตรฐานด้านตรวจพิสูจน์หลักฐานทางอิเล็กทรอนิกส์ของสำนักงานป้องกันปราบปรามพอกเงิน ส่วนตรวจพิสูจน์หลักฐานทางคอมพิวเตอร์	0	0	0	0.1	0.1	0	0	
นำร่องใช้แผนใช้ดินระดับตำบลไปปฏิบัตินำร่องโดยใช้ที่ดินระดับตำบลไปปฏิบัติในพื้นที่พัฒนาดินเป็นนำแผนใช้ดินรับปรับปรุงมาใช้ประโยชน์ในพื้นที่ของเกษตรกร	0	0	0	0	0	0.2	0.1	

2.3 การสร้างแบบจำลองการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ

การสร้างแบบจำลองการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ ผู้วิจัยสร้างแบบจำลองด้วยภาษา Python โดยใช้โปรแกรม Google Colaboratory (Colab) และใช้เทคนิคในการจำแนก 4 เทคนิค ได้แก่ เทคนิคการค้นหาเพื่อนบ้านใกล้สุด เทคนิคเคิร์ฟ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนและเทคนิคป่าแบบสุ่ม

2.4 การวัดประสิทธิภาพแบบจำลอง

ในการวัดประสิทธิภาพของแบบจำลองใช้วิธี 5-fold cross-validation จะทำการเลือกกลุ่มข้อมูลออกเป็น K ชุดเท่าๆ กัน จากนั้นจะทำการทดลองครั้งแรกด้วยข้อมูลชุดที่ 1 ซึ่งเป็นข้อมูลทดสอบและกำหนดให้ข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน และในการทดลองครั้งที่สองจะใช้ข้อมูลชุดที่ 2 เป็นชุดข้อมูลทดสอบและให้ข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน ทำจนกระทั่งข้อมูลทุกชุดข้อมูลได้ถูกนำมาเป็นชุดข้อมูลทดสอบทั้งหมด ซึ่งจำนวนในการทดสอบมีจำนวนเท่ากับ K ครั้ง โดยผลลัพธ์ที่ได้นั้นจะมากำหนดค่าเฉลี่ยความถูกต้องของการจำแนกข้อมูลในแต่ละรอบเนื่องจากข้อมูลโครงการที่มีจำนวน 1,150 โครงการ แยกตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ 23 แผน โดยแยกได้ผู้วิจัยจึงเลือกใช้ 5-fold cross-validation เหมาะสมกับชุดข้อมูล ในการวัดประสิทธิภาพของแบบจำลองใช้วิธี 5-fold cross-validation แสดงดังภาพที่ 2 และวัดประสิทธิภาพด้วยความแม่นยำ ค่าความระลึก และค่าความถ่วงดุลดังนี้

ค่าความแม่นยำ (Precision) คือ ค่าของการทำนายกลุ่มข้อมูลที่แบบจำลองพิจารณาจากจำนวนข้อมูลที่มีการทำนาย กลุ่มว่าเป็น “จริง” ทั้งหมดนั้น มีถูกต้องอย่างน้อยเพียงใด เมื่อเปรียบเทียบกับจำนวนข้อมูลทั้งหมดในกลุ่มที่เป็น “จริง” ดังสมการที่ 1

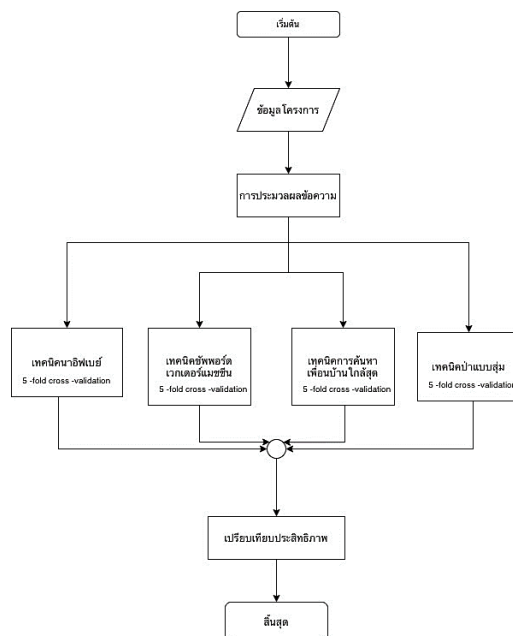
$$Precision = \frac{TP}{TP + FP} \quad (1)$$

ค่าความระลึก (Recall) คือ ค่าที่พิจารณาว่าข้อมูลที่มนุษย์หรือในความเป็นจริงบอกว่า “จริง” และเมื่อถูกพิจารณาด้วยแบบจำลองแล้ว แบบจำลองจะทำนายผล “จริง” ได้ถูกต้องและตรงกับคำตอบที่กระทำโดยมนุษย์หรือในความเป็นจริงอย่างน้อยเพียงใด ดังสมการที่ 2

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

ค่าความถ่วงดุล (F-measure) คือ ค่าวัดประสิทธิภาพโดยรวมของทั้งสองค่าความแม่นยำและค่าความระลึก พร้อมกันของแบบจำลอง โดยพิจารณาแยกทีละคลาส ดังสมการที่ 3

$$F - measure = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right) \quad (3)$$



ภาพที่ 3 ผังงานแสดงขั้นตอนการวัดประสิทธิภาพของแบบจำลอง

2.5 การวัดความถูกต้องของการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ

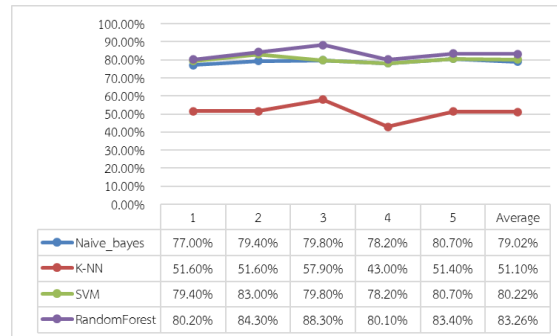
ในขั้นตอนนี้เป็นการวัดความถูกต้องของแบบจำลองในการจำแนกโครงการใหม่ ซึ่งมีรายละเอียด ดังนี้

- 1) รวบรวมโครงการใหม่ โดยใช้ข้อมูลโครงการจังหวัดมหาสารคาม ปี พ.ศ. 2565 จำนวน 22 โครงการ ที่ยังไม่มีมีการเชื่อมโยงแผนแม่บทภายใต้ยุทธศาสตร์ชาติ
- 2) ผู้วิจัยนำข้อมูลโครงการใหม่ จำนวน 22 โครงการ พร้อมแบบสอบถามให้ผู้เชี่ยวชาญ จำนวน 3 ท่าน ได้แก่ นักวิเคราะห์นโยบายและแผนชำนาญการ จำนวน 2 ท่าน และนักวิเคราะห์นโยบายและแผน จำนวน 1 ท่าน เพื่อทำการระบุแผนแม่บทฯ ของแต่ละโครงการ
- 3) นำแบบสอบถามจากผู้เชี่ยวชาญมาวิเคราะห์เพื่อหาแผนแม่บทของแต่ละโครงการ โดยพิจารณาจาก คำตอบของผู้เชี่ยวชาญทั้ง 3 ท่าน ซึ่งจะใช้คำตอบที่ตรงกันจากผู้เชี่ยวชาญอย่างน้อย 2 ท่าน กำหนดเป็นแผนแม่บท ของโครงการนั้น ๆ เพื่อใช้เป็นคำตอบในการเปรียบเทียบความถูกต้องกับผลลัพธ์ที่ได้จากแบบจำลอง
- 4) จำแนกโครงการใหม่ 22 โครงการ ด้วยแบบจำลองที่ผู้วิจัยพัฒนาขึ้น และเปรียบเทียบผลการจำแนก โครงการที่ได้จากแบบจำลองกับคำตอบจากผู้เชี่ยวชาญ

ผลการวิจัย

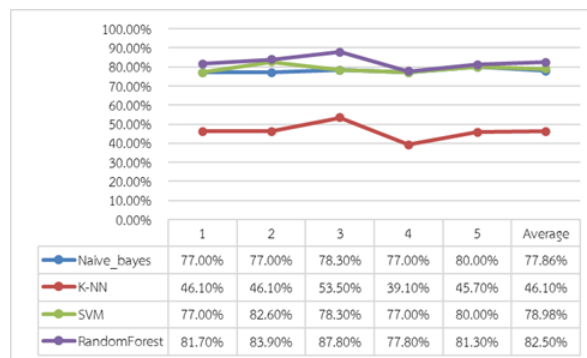
1. ผลการเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อความในการสร้างแบบจำลองเพื่อจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ

การวัดประสิทธิภาพแบบจำลองการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ โดยใช้เทคนิคนาอ์เบย์ เทคนิคการค้นหาเพื่อนบ้านใกล้สุด เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคป่าแบบสุ่ม



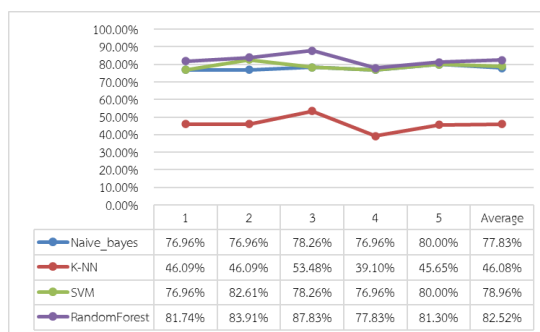
ภาพที่ 4 ค่าความแม่นยำ

จากภาพที่ 4 แสดงค่าความแม่นยำในการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติของเทคนิคนาอ์เบย์ เทคนิคการค้นหาเพื่อนบ้านใกล้สุด เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคป่าแบบสุ่มพบว่า เทคนิคป่าแบบสุ่มให้ค่าความแม่นยำเฉลี่ยสูงที่สุด เท่ากับ 83.26% รองลงมา คือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เท่ากับ 80.22% เทคนิคนาอ์เบย์ เท่ากับ 79.02% และเทคนิคการค้นหาเพื่อนบ้านใกล้สุดให้ค่าความแม่นยำเฉลี่ยน้อยที่สุด เท่ากับ 51.10%



ภาพที่ 5 ค่าความระลึก

จากภาพที่ 5 แสดงค่าความระลึกในการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติของเทคนิคนาอ์เบย์ เทคนิคการค้นหาเพื่อนบ้านใกล้สุด เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคป่าแบบสุ่มพบว่า เทคนิคป่าแบบสุ่มให้ค่าความระลึกเฉลี่ยสูงที่สุด เท่ากับ 82.06 % รองลงมา คือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เท่ากับ 78.98% เทคนิคนาอ์เบย์ เท่ากับ 77.86% และเทคนิคการค้นหาเพื่อนบ้านใกล้สุดให้ค่าความระลึกเฉลี่ยน้อยที่สุด เท่ากับ 46.10%

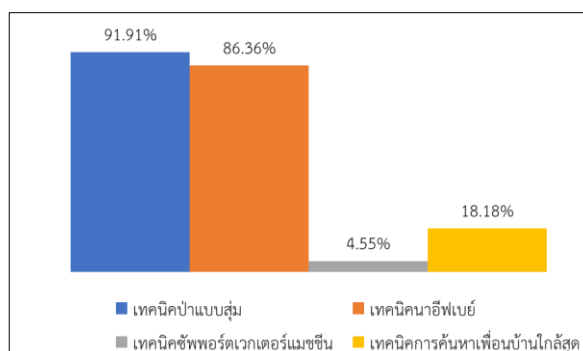


ภาพที่ 6 ค่าความถ่วงดุล

จากภาพที่ 6 แสดงค่าความถ่วงดุลการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติของเทคนิค Naïve Bayes, เทคนิคการค้นหาเพื่อนบ้านใกล้สุด, เทคนิคซ์พอร์ตเวกเตอร์แมชชีน และเทคนิคป่าแบบสุ่ม พบว่า เทคนิคป่าแบบสุ่มให้ค่าความถ่วงดุลเฉลี่ยสูงสุด เท่ากับ 82.52 % รองลงมา คือ เทคนิคซ์พอร์ตเวกเตอร์แมชชีน เท่ากับ 78.96% เทคนิค Naïve Bayes เท่ากับ 77.83% และเทคนิคการค้นหาเพื่อนบ้านใกล้สุดให้ค่าความถ่วงดุลเฉลี่ย น้อยที่สุด เท่ากับ 46.08%

2. ผลการวัดความถูกต้องในการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติของแบบจำลองที่พัฒนาขึ้น

การวัดความถูกต้องในการจำแนกโครงการใหม่ 22 โครงการ ด้วย 4 เทคนิค โดยค่าความถูกต้องของการ จำแนกคิดเป็นร้อยละ



ภาพที่ 7 ความถูกต้องของการจำแนกโครงการใหม่ 22 โครงการ

จากภาพที่ 7 เป็นผลความถูกต้องในการจำแนกโครงการใหม่ 22 โครงการของแบบจำลองที่พัฒนาขึ้น ทั้ง 4 เทคนิค พบว่า เทคนิคป่าแบบสุ่มสามารถจำแนกโครงการได้ถูกต้องจำนวน 20 โครงการ ซึ่งมีค่าความถูกต้อง สูงที่สุด คิดเป็นร้อยละ 90.91 รองลงมา คือ เทคนิค Naïve Bayes สามารถจำแนกโครงการได้ถูกต้องจำนวน 19 โครงการ คิดเป็นร้อยละ 86.36 เทคนิคการค้นหาเพื่อนบ้านใกล้สุดสามารถจำแนกโครงการได้ถูกต้องจำนวน 4 โครงการ คิดเป็นร้อยละ 18.18 และเทคนิคซ์พอร์ตเวกเตอร์แมชชีนสามารถจำแนกโครงการได้ถูกต้องจำนวน 1 โครงการ คิดเป็นร้อยละ 4.55 ตามลำดับ

อภิปรายผลการวิจัย

1. ผลการเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ พบว่า การใช้ข้อมูลของโครงการ ได้แก่ ชื่อโครงการ หลักการและเหตุผลของโครงการ วัตถุประสงค์ของโครงการ และประโยชน์ที่คาดว่าจะได้รับ โดยใช้เทคนิคป่าแบบสุ่มในการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติมีประสิทธิภาพสูงสุด โดยให้ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุลสูงกว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคนาอ็ฟเบย์ และเทคนิคการค้นหาเพื่อนบ้านใกล้สุด ซึ่งสอดคล้องกับงานวิจัยของ สุภัสสรา สมเจตนา [4] ได้ศึกษาการเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตรทำการทดลองและวัดประสิทธิภาพของแบบจำลอง ที่พบว่า เทคนิคป่าแบบสุ่มเป็นเทคนิคที่ดีที่สุดในการวิเคราะห์ความคิดเห็น

2. ผลการวัดความถูกต้องของการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติโดยใช้แบบจำลองที่พัฒนาขึ้น สำหรับโครงการใหม่ จำนวน 22 โครงการ ที่ยังไม่มีเชื่อมโยงแผนแม่บทภายใต้ยุทธศาสตร์ชาติ พบว่า เทคนิคป่าแบบสุ่ม ให้ค่าความถูกต้องสูงกว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคนาอ็ฟเบย์ และเทคนิคการค้นหาเพื่อนบ้านใกล้สุด สามารถจำแนกโครงการถูกต้องตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติ จำนวน 20 โครงการ สอดคล้องกับงานวิจัยของจิรายุ ชัยมีบุญ [10] ได้ศึกษาการประยุกต์ใช้ปัญญาประดิษฐ์กับหุ่นยนต์สนทนาบนระบบไลน์เพื่อช่วยในการบริหารงาน IT Support ยุค 4.0 พบว่าเทคนิคป่าแบบสุ่ม มีประสิทธิภาพในการจำแนกสูงที่สุด ทั้งนี้ อาจเป็นเพราะ เนื่องจากเทคนิคการสุ่มป่าไม่มีความได้เปรียบกว่าเทคนิคอื่นๆ ที่ได้นำมาทดสอบ เพราะเทคนิคนี้มีความสามารถในการทำนายผลได้อย่างแม่นยำ โดยที่ไม่มีปัญหาในเรื่องของ overfitting และเทคนิคป่าแบบสุ่มนำเสนอค่าประมาณสำหรับความสำคัญของตัวแปร ค่าที่หายไปจะถูกแทนที่ด้วยตัวแปรที่ปรากฏมากที่สุดในโหนด ในวิธีการจำแนกประเภทเทคนิคป่าแบบสุ่มให้ความแม่นยำสูงสุด และสามารถจำแนกกับข้อมูลขนาดใหญ่ที่มีข้อความ 1,000 ขึ้นไป สามารถปรับสมดุลชุดข้อมูลโดยอัตโนมัติ และวิธีการเทคนิคป่าแบบสุ่มยังจัดการกับตัวแปรได้อย่างรวดเร็ว ทำให้เหมาะสำหรับงานที่ซับซ้อน [11]

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

สามารถนำแบบจำลองการจำแนกโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติด้วยเทคนิคเหมืองข้อมูลไปพัฒนาระบบเชื่อมโยงโครงการตามแผนแม่บทภายใต้ยุทธศาสตร์ชาติแบบอัตโนมัติต่อไป

2. การศึกษาครั้งต่อไป

ในส่วนของการจำแนกโครงการอาจทำการศึกษาโดยใช้วิธีการให้ค่าน้ำหนักค่าในแบบอื่น ๆ เพื่อเพิ่มประสิทธิภาพในการจำแนก

เอกสารอ้างอิง

- [1] สำนักงานเลขาธิการของคณะกรรมการยุทธศาสตร์ชาติ. (2562). *ยุทธศาสตร์ชาติ พ.ศ. 2561 - 2580*. พิมพ์ครั้งที่ 2. กรุงเทพฯ : สำนักงานเลขาธิการของคณะกรรมการยุทธศาสตร์ชาติ สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ.
- [2] สำนักพัฒนาและส่งเสริมการบริหารราชการจังหวัด. (2552). *แผนกับการพัฒนาจังหวัด และกลุ่มจังหวัด*. กรุงเทพฯ : บทวิจารณ์พิมพ์ จำกัด.
- [3] อังศุมล สุทธิภักดี. (2554). *การเปรียบเทียบประสิทธิภาพของการจำแนกหมวดหมู่ของข้อความในแบบสอบถามปลายเปิด โดยวิธีเอนอ็ฟเบย์และซัพพอร์ตเวกเตอร์แมชชีน*. นนทบุรี: วิทยาลัยราชพฤกษ์.
- [4] สุภัสสรา สมเจตนา. (2564). *การเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร*. วารสารวิทยาศาสตร์และเทคโนโลยี, 23(1), 21-30.
- [5] ประเดิม วงศ์กระโซ่. (2560). *ระบบจำแนกหมวดหมู่การช้อปปิ้งออนไลน์ใช้เหมืองข้อมูล*. ปริญญาโทบริหารวิทยาศาสตร

มหาดบัณฑิต การจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

- [6] สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ. (2560). *ระบบติดตามและประเมินผลแห่งชาติ (eMENSCR)*. กรุงเทพฯ : สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ.
- [7] นงเยาว์ สอนจะโปะ. (2561). รูปแบบการจำแนกกลุ่มข้อความภาษาไทยแบบอัตโนมัติ โดยใช้การเรียนรู้ของเครื่อง (Machine Learning) ด้วยเทคนิค Unsupervised Learning ร่วมกับการประมวลผลภาษาธรรมชาติ (Natural Language Processing). *วารสารวิชาการศรีปทุม ชลบุรี*, 14(4), 95-106.
- [8] จันทิมา พลพินิจ. (2564). *การจำแนกเอกสารข้อความอัตโนมัติ*. มหาสารคาม: สำนักพิมพ์มหาวิทยาลัยราชภัฏมหาสารคาม.
- [9] นุชนาฏ ปิ่นเมือง. (2561). *การจำแนกความคิดเห็นของคนไทยเกี่ยวกับสื่อออนไลน์โดยใช้การทำเหมืองข้อความ*. ปรินิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม.
- [10] จิรายุ ชัยมีบุญ. (2564). *การประยุกต์ใช้ปัญญาประดิษฐ์กับหุ่นยนต์สนทนาบนระบบไลน์เพื่อช่วยในการบริหารงาน IT Support ยุค 4.0*. ปรินิพนธ์ปริญญา วิทยาศาสตรมหาบัณฑิต การจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- [11] Girma, R., Fürst, C., & Moges, A. (2022). Land use land cover change modeling by integrating artificial neural network with cellular Automata-Markov chain model in Gidabo river basin, main Ethiopian rift. *Environmental Challenges*, 6, 1-15. <https://doi.org/10.1016/j.envc.2021.100419>.