

# การพัฒนาตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่ง ด้วยเทคนิคการเรียนรู้ของเครื่อง

## Development of Models to Identify Emotions of Thai Country Song Lyrics

### Using Machine Learning Techniques

พิศาล สุขชี<sup>1\*</sup>, นฤมล สีดำ<sup>2</sup> และ รัตนากร คำเมือง<sup>3</sup>

Phisan Sookkhee<sup>1\*</sup>, Naruamon Seedam<sup>2</sup> and Rattanakon khammueang<sup>3</sup>

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยราชภัฏศรีสะเกษ<sup>1,2,3</sup>

phisan.s@sskru.ac.th<sup>\*</sup>, naromol.sida60@sskru.ac.th, rattanakon.kham60@sskru.ac.th

#### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่อง และ 2) ประเมินประสิทธิภาพของตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่อง ซึ่งผู้วิจัยทำการสร้างตัวแบบการจำแนกจาก 2 เทคนิค ประกอบด้วยซัพพอร์ตเวกเตอร์แมชชีน และการวิเคราะห์การถดถอยโลจิสติก โดยใช้ชุดข้อมูลเนื้อร้องเพลงลูกทุ่งจำนวน 1,500 เพลง แบ่งออกเป็นบทเพลงประเภท เศร้า, สนุกสนาน และให้กำลังใจ จำนวนประเภทละ 500 บทเพลงที่ไม่ซ้ำกัน และยังสามารถแบ่งข้อมูลเนื้อเพลงออกเป็นชุดย่อยเพื่อใช้สำหรับการสร้างตัวแบบเป็นจำนวน 8 ชุดประกอบด้วย 1) เนื้อเพลงเฉพาะท่อนนำ 2) เนื้อเพลงเฉพาะท่อนนำที่กำหนดคำหยุด 3) เนื้อเพลงเฉพาะท่อนเข้าเนื้อเรื่อง 4) เนื้อเพลงเฉพาะท่อนเข้าเนื้อเรื่องที่กำหนดคำหยุด 5) เนื้อเพลงเฉพาะท่อนหลัก 6) เนื้อเพลงเฉพาะท่อนหลักที่กำหนดคำหยุด 7) เนื้อเพลงทุกท่อนจากทั้งเพลง 8) เนื้อเพลงทุกท่อนจากเนื้อเพลงที่กำหนดคำหยุด โดยผู้วิจัยได้ใช้เทคนิคลดจนชั้นตอนวิธีในการจัดทำเหมืองข้อความ และการประมวลผลภาษาธรรมชาติในการจัดเตรียมข้อมูลอย่างเป็นระบบ และใช้วิธีการทางการเรียนรู้ของเครื่องเพื่อสร้างตัวแบบการจำแนก จากนั้นทำการทดสอบประเมินประสิทธิภาพความถูกต้องของตัวแบบด้วยวิธี K-fold Cross-validation

ผลของการศึกษาพบว่าในประเด็นด้าน 1) การพัฒนาตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่อง พบว่าการนำข้อมูลในชุดที่ 7 เนื้อเพลงทุกท่อนจากทั้งเพลง สามารถสร้างตัวแบบที่มีประสิทธิภาพในการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งได้ดีกว่าชุดข้อมูลอื่น และ 2) ประเมินประสิทธิภาพของตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่อง พบว่าตัวแบบที่ถูกสร้างด้วยเทคนิค ซัพพอร์ตเวกเตอร์แมชชีนมีความแม่นยำอยู่ที่ร้อยละ 73.73 และ การวิเคราะห์การถดถอยโลจิสติกมีความแม่นยำอยู่ที่ร้อยละ 71.87

**คำสำคัญ:** การเรียนรู้ของเครื่อง, การประมวลผลภาษาไทย, เหมืองข้อความ, การจำแนก, เพลงลูกทุ่ง

#### ABSTRACT

The research purpose the development of models to identify emotions of Thai country song lyrics using machine learning techniques. Two techniques were selected which are Support Vector Machine and Logistic Regression to construct the model. The data used in this study are the country song of 1,500 lyrics, divided into 500 unique consists of sad, fun, and encouraging genres. Each song is divided into 8 sections, consisting of 1) verse lyrics, 2) verse lyrics with stop word removal, 3) pre-chorus lyrics, 4) pre-chorus lyrics with stop word removal, 5) chorus lyrics, 6) chorus lyrics with stop word removal, 7) all

lyrics without stop word removal 8) all lyrics with stop word removal. We have used algorithms of text mining and natural language processing in systematic preparation data. K-fold Cross-validation method was selected to evaluate the accuracy.

The results of the study on the issue of the model development. Found that bringing information in the 7<sup>th</sup>) lyrics of every verse from the whole song without stop word removal. They were able to generate more efficient models to classify emotions of Thai country song lyrics than other datasets, and the results of assess the effectiveness of models. Found that the model that was created with Support Vector Machine has an accuracy of 73.73% and a logistic regression analysis has an accuracy of 71.87 %.

**Keywords:** Machine Learning, Thai Language Procession, Text Mining, Classification, Thai Country Song

## บทนำ

เพลงลูกทุ่งเป็นเพลงที่มีพัฒนาการจากเพลงพื้นบ้านและเพลงไทยเดิม มีวิธีการร้องและมีเอกลักษณ์เฉพาะตัว มีทำนองเพลงฟังแล้วเข้าใจง่าย มีลีลาเฉพาะ เน้นความสนุกสนาน ทำนองเพลงมาจากเพลงพื้นบ้าน เนื้อร้องของเพลงลูกทุ่งจะมีลักษณะเด่น การร้องจะใช้ลูกคอหรือจังหวะเร็วเสียงและมีการเอื้อนเสียงเป็นสำคัญ มากไปกว่านั้นการถ่ายทอดอารมณ์เพลงถือเป็นหัวใจสำคัญของการขับร้องเพลงลูกทุ่ง เนื้อร้องของเพลงลูกทุ่งจะมีลักษณะเด่น มีเนื้อหาเกี่ยวกับชีวิตจริงแล้วถ่ายทอดออกมาด้วยความซาบซึ้งกินใจ (สุนทรี ดวงทิพย์, 2558: 67) บทเพลงมักกล่าวถึงเรื่องราวในชีวิตผสมผสานกับจินตนาการของผู้ประพันธ์ถ่ายทอดความรู้สึกนึกคิด ตลอดจนสภาพสังคมในยุคนั้น ๆ สามารถสร้างอารมณ์สะท้อนใจ อารมณ์เศร้า รัก ผิดหวัง สนุกสนาน ตลกขบขัน หรือปลุกสำนึก ทั้งนี้ขึ้นอยู่กับชั้นเชิงการประพันธ์ ตลอดจนอารมณ์ร่วมรับรู้ของผู้ฟัง (พรทิพย์ ฉายก็ และ จันทนา แก้ววิเชียร, 2561: 87)

ในยุคปัจจุบันการพัฒนาก้าวกระโดดของเทคโนโลยีด้านการสื่อสารและอินเทอร์เน็ต ทำให้ข้อมูลต่าง ๆ ได้ถูกนำมาเผยแพร่ในรูปแบบสื่อมัลติมีเดียผ่านเว็บไซต์ ไม่ว่าจะเป็น หนังสือนิยาย รูปภาพ คลิปวิดีโอ ทำให้ผู้บริโภคสามารถเข้าถึงข้อมูลที่ตนเองต้องการได้รวดเร็ว ตลอดจนเทคโนโลยีที่ช่วยให้ผู้ใช้งานสามารถดำเนินงานต่าง ๆ ที่ตนเองต้องการได้ผ่านระบบออนไลน์ได้อย่างง่ายดายมากยิ่งขึ้น (พิศาล สุขซี และคณะ, 2562: 84) ด้วยเช่นนี้ทำให้ระบบอินเทอร์เน็ตเปรียบเสมือนคลังข้อมูลสารสนเทศด้านต่าง ๆ ขนาดใหญ่ที่เติบโตขึ้นเรื่อย ๆ เนื่องด้วยมีปริมาณข้อมูลที่เพิ่มมากขึ้นอย่างต่อเนื่องในทุก ๆ วัน ไม่เว้นแม้แต่เพลงลูกทุ่งเองก็เช่นกัน ปัจจุบันระบบอินเทอร์เน็ตก็เป็นช่องทางที่ใหญ่และสำคัญที่สุดในการเผยแพร่ผลงานเพลงลูกทุ่งไม่ว่าจะเป็นเพลงดั้งเดิมที่เคยมีอยู่แล้ว หรือบทเพลงที่ถูกประพันธ์ขึ้นมาใหม่ก็ตาม ซึ่งถูกเผยแพร่ในรูปแบบมิวสิควิดีโอเพลง และเนื้อเพลงชื่อของศิลปินผู้ขับร้อง ตลอดจนชื่อผู้ประพันธ์บทเพลงด้วยเช่นกัน

เมื่อพิจารณาว่าปัจจุบันเครือข่ายอินเทอร์เน็ตเป็นคลังข้อมูลที่มีขนาดใหญ่มหาศาล ปัญหาที่สำคัญอย่างยิ่งในยุคที่ข้อมูลมีความหลากหลายและเติบโตขึ้นอย่างต่อเนื่อง คือการไม่สามารถเชื่อมโยงข้อมูลที่หลากหลายเข้าด้วยกัน และไม่สามารถสกัดองค์ความรู้จากข้อมูลที่มีจำนวนมหาศาลเพื่อให้ได้องค์ความรู้ที่แท้จริงและตอบสนองต่อความต้องการในการใช้งานได้อย่างตรงประเด็น (พิศาล สุขซี และคณะ, 2562: 84) โดยการเติบโตของสารสนเทศในรูปแบบข้อความ (Textual Information) มีอัตราที่สูงขึ้น ทั้งจากเครือข่ายอินเทอร์เน็ตและเอกสารที่องค์กรจัดเก็บไว้ และแม้ว่าเทคโนโลยีการสืบค้น (Search Engine) จะมีขีดความสามารถสูงขึ้น โดยสามารถค้นหาคำสำคัญ (Keywords) และให้ผลลัพธ์อย่างรวดเร็ว แต่วิธีดังกล่าวก็ยังไม่สามารถสรุปความ ประมวลความหมาย หรือความสัมพันธ์ของคำได้อย่างตรงประเด็น ทำให้การสืบค้นเอกสารไม่ตรง กับความต้องการของผู้ใช้ (วรรณวิภา วงศ์วิไลสกุล, 2556: 158)

ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยต้องการนำเอาเนื้อร้องเพลงลูกทุ่งที่ถูกเผยแพร่อยู่บนระบบอินเทอร์เน็ตซึ่งถือว่าเป็นข้อมูลสารสนเทศประเภทข้อความที่กระจัดกระจายอย่างไม่เป็นระเบียบมาใช้ให้เกิดประโยชน์ ด้วยการรวบรวม จัดหมวดหมู่ และนำมาประมวลผลด้วยวิธีทางด้านการทำเหมืองข้อความ การประมวลผลภาษาธรรมชาติ และเทคนิควิธีการเรียนรู้ของเครื่อง เพื่อนำข้อมูลเนื้อร้องเพลงลูกทุ่งที่มีอยู่ในระบบอินเทอร์เน็ตมาสร้างเป็นตัวแบบจำลองสำหรับจำแนกอารมณ์ของบทเพลงลูกทุ่ง โดยคาดหวังว่าตัวแบบจำลองที่ได้นี้จะสามารถนำไปใช้ให้เกิดประโยชน์ได้อย่างหลากหลายทางเช่น ใช้สำหรับจัดกลุ่มหรือจัดหมวดหมู่ของเพลงจากเนื้อเพลงแบบอัตโนมัติ หรือใช้สำหรับเป็นตัวแบบในการช่วยให้ผู้ที่ต้องการประพันธ์บทเพลงลูกทุ่งใหม่ ๆ ใช้สำหรับวิเคราะห์เนื้อเพลงของตนเองว่าเนื้อเพลงลูกทุ่งที่ตนเองได้ทำการประพันธ์มานั้นมีเนื้อหาอยู่ในอารมณ์เพลงประเภทใด ตรงตามความตั้งใจในการใช้ถ้อยคำเรียบเรียงบทเพลงเพื่อให้ผู้รับฟังเกิดความรู้สึกอย่างกับผู้ประพันธ์นั้นต้องการจริง ๆ หรือไม่

## 1. วัตถุประสงค์การวิจัย

- 1.1 เพื่อพัฒนาตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่อง
- 1.2 เพื่อประเมินประสิทธิภาพของตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่อง

## 2. เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาวิจัยในครั้งนี้ผู้วิจัยได้ทำการศึกษาค้นคว้าความรู้จากเอกสารต่าง ๆ ตลอดจนได้มีการทบทวนวรรณกรรมที่เกี่ยวข้องกับการวิจัยในครั้งนี้ดังต่อไปนี้

การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) เป็นวิทยาการแขนงหนึ่งในหมวดหมู่ของเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) ซึ่งช่วยให้คอมพิวเตอร์สามารถเข้าใจ ตลอดจนตีความและใช้งานภาษาปกติที่มนุษย์ใช้สื่อสารได้ โดยเทคโนโลยี NLP นี้มีรากฐานจากวิทยาการหลากหลายสาขา โดยเฉพาะด้านวิทยาการคอมพิวเตอร์ (Computer Science) และภาษาศาสตร์เชิงคำนวณ (Computational Linguistics) เพื่อวัตถุประสงค์ในการปิดช่องว่างทางการสื่อสารระหว่างมนุษย์และระบบคอมพิวเตอร์ (SAS, 2020) ซึ่งจะช่วยให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติของมนุษย์เพื่อให้คอมพิวเตอร์สามารถเข้าใจภาษาเขียน รูปแบบความรู้ ที่คอมพิวเตอร์สามารถนำไปใช้งานได้ เช่น การสรุปบทความ การตอบโต้ทางแชตบอท ระบบตัดคำ เป็นต้น (กรมวุฒิ หนุน และคณะ, 2559)

การทำเหมืองข้อความ (Text Mining) เป็นเทคนิคการค้นหาคำหรือความรู้ใหม่จากข้อมูลประเภทข้อความที่มีปริมาณมากโดยอัตโนมัติ โดยการสกัดคำ คำนวณรูปแบบ และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อความเอกสาร เพื่อให้เกิดความหมายและสามารถนำมาใช้ประโยชน์ได้ โดยสถาปัตยกรรมระบบของเหมืองข้อความมีความคล้ายคลึงกับสถาปัตยกรรมระบบในการค้นหาคำรู้ในฐานข้อมูลประเภทข้อความ ซึ่งประกอบด้วย ขั้นตอนการเตรียมข้อมูล ขั้นตอนการค้นหาคำรู้ความสัมพันธ์ และขั้นตอนการนำเสนอผลลัพธ์ที่ชัดเจน ส่วนประโยชน์ของเหมืองข้อความนั้น คือการทำให้สารสนเทศในรูปแบบข้อความสามารถเข้าถึง วิเคราะห์ และประมวลผลภายในเวลาอันรวดเร็ว โดยผู้ใช้สามารถเข้าใจและใช้ประโยชน์จากเอกสารที่จัดเก็บไว้เป็นจำนวนมากได้อย่างมีประสิทธิภาพ แนวทางสำคัญในการประยุกต์ใช้เหมืองข้อความ คือ การประยุกต์ใช้ตามหน้าที่ และตามกลุ่มการใช้งาน และขณะนี้ได้มีการพัฒนาเหมืองข้อความขึ้นเป็นเหมืองข้อความแสดงความคิดเห็น ซึ่งผลงานเทคนิคของการสืบค้นข้อมูลเข้ากับการประมวลผลภาษาธรรมชาติ ทำให้สามารถสรุปความคิดเห็นที่หลากหลายบนเครือข่ายสังคมออนไลน์ให้เข้าใจมากยิ่งขึ้น (วรรณวิภา วงศ์วิไลสกุล, 2556)

การจำแนกประเภทข้อมูล (Classification) เป็นกระบวนการสร้างตัวแบบจำลองจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ โดยจะนำข้อมูลส่วนหนึ่งมาสอนให้ระบบเรียนรู้ (Training Data) เพื่อจำแนกข้อมูลออกเป็นกลุ่มที่กำหนดไว้ ผลลัพธ์ที่

ได้จากการเรียนรู้คือ ตัวแบบจำลองจำแนกประเภทข้อมูล (Classifier Model) และจะนำข้อมูลส่วนที่เหลือจากข้อมูลสอนระบบมาเป็นข้อมูลที่ใช้ทดสอบ (Testing Data) ซึ่งกลุ่มที่แท้จริงของข้อมูลที่ใช้ทดสอบนี้จะถูกนำมาเปรียบเทียบกับกลุ่มที่ทำมาได้จากโมเดลเพื่อทดสอบความถูกต้อง และปรับปรุงโมเดลจนกว่าจะได้ค่าความถูกต้องในระดับที่น่าพอใจ หลังจากนั้นหากมีข้อมูลใหม่เข้ามา จะนำข้อมูลมาผ่านตัวแบบจำลอง โดยตัวแบบจำลองจะสามารถจำแนกกลุ่มของข้อมูลที่เข้ามาใหม่ได้ (ศศิมา มณฑาสวรรณ, 2557: 10) โดยในการสร้างแบบจำลองจำแนกประเภทข้อมูล อาจใช้เทคนิคหรืออัลกอริทึม ต่าง ๆ เช่น ต้นไม้ตัดสินใจ (Decision Tree) เครือข่ายประสาทเทียม (Neural Networks) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) หรือ การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression) เป็นต้น

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) เป็นขั้นตอนวิธีในจำแนกกลุ่มข้อมูลโดยอาศัยระนาบการตัดสินใจที่เรียกว่า ระนาบเกิน หรือไฮเปอร์เพลน (Hyperplane) มาใช้ในการจำแนกกลุ่มข้อมูล ซึ่งจะใช้สมการเส้นตรงสำหรับแบ่งข้อมูลออกเป็น 2 กลุ่มแยกออกจากกัน ในการหาระนาบเกินที่เหมาะสมที่สุด จะทำการหาตำแหน่งของซัพพอร์ตเวกเตอร์ (Support Vector) เพื่อใช้เป็นตัวแทนของกลุ่มข้อมูลทั้งคู่ ในการพิจารณาเกณฑ์แบ่งกลุ่มโดยอาศัยหลักการคือ จะใช้ระนาบเกินที่เป็นระยะห่างที่สุดระหว่างข้อมูล 2 กลุ่ม ที่อยู่ใกล้กันมากที่สุดเพียงระนาบเดียวเท่านั้น ในทางทฤษฎีจะต้องไม่มีข้อมูลเกินเข้ามาในระหว่างขอบระนาบทั้งสอง จากนั้นจึงหาระนาบที่รักษาระยะห่างจากขอบมากที่สุด (Maximum Margin) และถือว่าระนาบดังกล่าวคือระนาบสำหรับการแบ่งกลุ่มที่สมมติที่สุด (นิตพร จันทร์กลิ่น, 2557: 8)

การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression) มีลักษณะคล้ายกับการวิเคราะห์การถดถอย (Regression Analysis) และการวิเคราะห์จำแนกประเภท (Discriminant Analysis) คือ ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระหรือตัวแปรพยากรณ์ และทำการสร้างสมการถดถอยขึ้นมาเพื่อทำนายหรือประมาณตัวแปรตาม (กัลยา วานิชย์บัญชา, 2551: 425 - 428) แต่การวิเคราะห์การถดถอยโลจิสติกเป็นเทคนิคสถิติที่ใช้พยากรณ์ความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจ เมื่อตัวแปรตามเป็นตัวแปรเชิงคุณภาพ ส่วนตัวแปรอิสระหรือตัวแปรพยากรณ์ (Predictor) เป็นตัวแปรที่สนใจหรือปัจจัยที่สามารถพยากรณ์การเกิดเหตุการณ์ที่สนใจ ซึ่งเป็นได้ทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพ สามารถมีได้มากกว่า 1 ตัว จากสมการ

$$\text{Logit}(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

เมื่อ  $Y$  คือความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจ

$\alpha$  คือส่วนตัดแกน  $Y$  หรือค่าของ  $Y$  เมื่อ  $X = 0$

$\beta_i$  คือสัมประสิทธิ์ของการถดถอยของตัวแปรอิสระ ตัวที่  $i$  โดยที่  $i = 1, \dots, k$

$X_i$  คือค่าของตัวแปรอิสระตัวที่  $i$  โดยที่  $i = 1, \dots, k$

$\epsilon$  คือค่าความคลาดเคลื่อนของการพยากรณ์

ถ้ามีตัวแปรอิสระ 1 ตัวเรียกว่าการวิเคราะห์การถดถอยโลจิสติกอย่างง่าย (Simple Logistic Regression) แต่ถ้ามีตัวแปรอิสระ 2 ตัวขึ้นไป เรียก การวิเคราะห์การถดถอยโลจิสติกเชิงพหุ (Multiple Logistic Regression) ส่วนถ้าพิจารณาตามจำนวนกลุ่มของตัวแปรตาม ถ้าตัวแปรที่มีค่าเพียง 2 ค่า หรือแบ่งข้อมูลออกเป็น 2 กลุ่มย่อย เช่น การศึกษาการเป็นโรคไข้หวัดใหญ่ เมื่อ  $Y =$  การเป็นโรคไข้หวัดใหญ่ จะได้

$$\text{จะได้ } Y = \begin{cases} 1 & \text{เป็นโรคไข้หวัดใหญ่} \\ 0 & \text{ไม่เป็นโรคไข้หวัดใหญ่} \end{cases}$$

เรียกว่า การวิเคราะห์การถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression) ถ้าตัวแปรตามมีค่ามากกว่า 2 ค่า หรือแบ่งข้อมูลออกได้มากกว่า 2 กลุ่มย่อย เช่น การศึกษาการดื่มแอลกอฮอล์ของผู้ป่วยโรคเมี่ยงดับ เมื่อ  $Y =$  การดื่มแอลกอฮอล์ จะได้

$$\text{จะได้ } Y = \begin{cases} 1 & \text{ดื่มเป็นประจำ} \\ 2 & \text{ดื่มนาน ๆ ครั้ง} \\ 3 & \text{ไม่ดื่มเลย} \end{cases}$$

เรียกว่าการวิเคราะห์การถดถอยลอจิสติกแบบหลายกลุ่ม (Multiple Logistic Regression) โดยงานวิจัยนี้ได้นำเทคนิคการวิเคราะห์การถดถอยลอจิสติกแบบหลายกลุ่มมาใช้ในการสร้างตัวแบบจำแนกอารมณ์จากเนื้อเพลง มีตัวแปรตามจำนวนทั้งสิ้น 3 ค่า ประกอบไปด้วย สนุก, เศร้า และให้กำลังใจ

การให้น้ำหนักคำ (Term Weighting) เป็นวิธีการให้น้ำหนักสำหรับคำที่มีความสำคัญหรือใช้เป็นตัวแทนของเอกสารที่ควรจะถูกจัดอยู่เป็นจำนวนมากในเนื้อหาของเอกสารเฉพาะฉบับนั้น และปรากฏอยู่น้อยในชุดของเอกสารที่เหลือทั้งหมด แต่ถ้าคำนั้นปรากฏเป็นจำนวนมากในทุก ๆ เอกสาร แสดงว่าคำดังกล่าวไม่สามารถเป็นตัวแทนของเอกสารใด ๆ ได้ ซึ่งคำเหล่านั้นเรียกว่าคำหยุด (Stop word) เช่น กับ, แก่, แต่ เป็นต้น โดยการศึกษาวิจัยในครั้งนี้คณะผู้วิจัยเลือกใช้วิธีการให้น้ำหนักคำสำคัญแบบ TF-IDF (Term Frequency – Inverse Document Frequency) ซึ่งเป็นวิธีที่นิยมใช้มากเพราะทำให้สามารถลดจำนวนคำที่ไม่มีความสำคัญลงให้คงเหลือแต่คำสำคัญที่สามารถเป็นตัวแทนที่ดีของเอกสารอย่างแท้จริง (Mahgoub, 2008; Chatree และคณะ, 2016)

วิธีการตรวจสอบไขว้ (K-fold Cross-validation) วิธีนี้เป็นวิธีที่นิยมในการทำงานวิจัยเพื่อใช้ในการทดสอบประสิทธิภาพของตัวแบบจำลองเนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัดประสิทธิภาพด้วยวิธี K-fold Cross-validation นี้ จะทำการแบ่งข้อมูลออกเป็นหลายส่วนซึ่งมักจะแสดงด้วยค่า K เช่น 10-fold cross-validation คือ การแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นจะทำการแยกข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล (Testing Data) ซึ่งการทำงานจะเป็นการทำซ้ำไปเช่นนี้จนครบจำนวนที่แบ่งไว้ โดยในงานวิจัยนี้ใช้ค่า K เท่ากับ 10

การศึกษาวิจัยตลอดจนการพัฒนาโปรแกรมประยุกต์ในปัจจุบันได้มีการนำเอาความรู้ทางด้านการประมวลผลภาษาธรรมชาติ กระบวนการทำเหมืองข้อความ ตลอดจนเทคนิคการเรียนรู้ของเครื่องมาประยุกต์ใช้อย่างหลากหลาย โดยงานวิจัยที่ผู้วิจัยได้เลือกมาสำหรับบททวนวรรณกรรมมีดังนี้

งานวิจัยประเภทการนำความรู้พื้นฐานด้านเหมืองข้อความมาประยุกต์ใช้งาน ประกอบด้วย Chatree J. และคณะ (2016) นำเสนอวิธีการใช้เหมืองข้อความเพื่อสกัดสารสนเทศจากข้อมูลสรพคุณสมุนไพรมต่าง ๆ ที่เผยแพร่อยู่บนเว็บไซต์ภาษาไทยและใช้การค้นหากฎความสัมพันธ์เพื่อวิเคราะห์ความสัมพันธ์ระหว่างอาการเจ็บป่วยและสมุนไพรมที่มีสรรพคุณในการรักษาอาการนั้น ๆ, พิศาล สุขชี และคณะ (2562) ได้ประยุกต์วิธีการพัฒนาระบบสนับสนุนการตัดสินใจ ด้วยการผสานเทคนิคของการสืบค้นข้อมูลเข้ากับการประมวลผลภาษาธรรมชาติ และเทคนิคเหมืองข้อความเพื่อการค้นหากฎความสัมพันธ์ของอาการเจ็บป่วยและสมุนไพรมที่มีสรรพคุณตอบสนองต่ออาการนั้น ๆ ด้วยการวิเคราะห์ข้อมูลจำนวนมากที่มีอยู่บนระบบอินเทอร์เน็ต และนำความสัมพันธ์ของคำที่เป็นอาการเจ็บป่วยและสมุนไพรมที่ปรากฏอยู่ด้วยกันมาสร้างเป็นกฎความสัมพันธ์และประยุกต์ออกมาในรูปแบบของระบบสนับสนุนการตัดสินใจ

งานวิจัยประเภทที่นำเอาเทคนิคการเรียนรู้ของเครื่องมาใช้สร้างตัวแบบจำแนกข้อความภาษาไทย ประกอบด้วย กานดา แผ้วพัฒนากุล และปราโมทย์ ลีอนาม (2556) พัฒนาตัวแบบประเภทความคิดเห็นบนเครือข่ายสังคมออนไลน์, วัชรวรรณ จิตต์สกุล และสุนันทา สดสี (2560) นำเสนอการใช้เทคนิค Random Forest ร่วมกับการวัดความสำคัญของคำโดยพิจารณาจากความใกล้ชิด เพื่อประยุกต์ในการจำแนกข้อความป้อนกลับของลูกค้าเมื่อซื้อสินค้าจากระบบพาณิชย์อิเล็กทรอนิกส์ โดยสามารถจำแนกความคิดเห็นออกเป็น ความคิดเห็นเชิงบวก เชิงลบ และเป็นกลาง, นพมาศ ปักเข็ม และคณะ (2560) ได้พัฒนาตัวแบบข้อความเพื่อการจัดเกทของภูมิปัญญาท้องถิ่นของไทย 10 ลักษณะโดยใช้ข้อมูลจากเว็บไซต์ทางภูมิปัญญาท้องถิ่นไทยโดยเทคนิคที่ผู้วิจัยได้นำมาทดสอบเพื่อการสร้างตัวแบบจำแนกประกอบด้วย ต้นไม้ตัดสินใจ (Decision Tree), ตัวแบบจำแนกข้อมูลเบย์อย่างง่าย (Naïve Bays Classification) และเทคนิคเพื่อนบ้านที่ใกล้ที่สุด K ตัว (K-Nearest

Neighbors-KNN) และนัยเพียร์ และพยุง มีสัจ (2556) ได้นำเสนอการวิจัยที่แสดงให้เห็นว่าการใช้เทคนิคซ์พอร์ดเวกเตอร์แมชชีนนั้นมีความเหมาะสมในการนำมาสร้างตัวแบบในการจำแนกและคัดกรองเพื่อการจำแนกข้อความ

## วิธีดำเนินการวิจัย

### 1. ขั้นตอนการดำเนินการวิจัย ประกอบด้วย 3 ขั้นตอน ดังนี้

#### 1.1 จัดเก็บและรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการรวบรวมข้อมูลตามขอบเขตการวิจัยที่ได้ตั้งไว้โดยพิจารณาแหล่งข้อมูลที่น่ามาใช้ตามหลักเกณฑ์การคัดกรองแหล่งข้อมูลดังต่อไปนี้ 1) เป็นข้อมูลประเภทข้อความเนื้อร้องเพลงของเพลงลูกทุ่ง 2) สามารถเข้าถึงข้อมูลแบบออนไลน์บนอินเทอร์เน็ตได้ 3) เป็นข้อมูลที่เผยแพร่บนเว็บไซต์ภาษาไทย 4) ลักษณะของการเผยแพร่ข้อมูลบนเว็บไซต์เป็นลักษณะให้ความบันเทิงแก่สาธารณะชน 5) มีการจัดหมวดหมู่อารมณ์เพลงเป็นระบบ จากเกณฑ์การคัดกรองแหล่งข้อมูลดังกล่าวกับการสืบค้นข้อมูลเกี่ยวกับเพลงลูกทุ่งผ่านระบบอินเทอร์เน็ตทำให้ได้แหล่งข้อมูลต้นฉบับที่ผู้จัดทำนำมาใช้ในการจัดทำโครงการทั้งสิ้นจำนวน 12 แหล่ง

#### 1.2 การเตรียมข้อมูลและสร้างตัวแบบ

ข้อมูลต้นฉบับที่จะถูกนำมาใช้ประมวลผลเพื่อสร้างตัวแบบสำหรับการจำแนกจะอยู่ในรูปแบบของข้อความ (Text) ซึ่งจะต้องผ่านกระบวนการในการจัดเตรียมข้อมูลจากข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) ให้เป็นข้อมูลที่มีโครงสร้าง (Structured Data) เพื่อให้เหมาะสมต่อการประมวลผลต่อไป โดยมีขั้นตอนในการจัดเตรียมดังนี้

ขั้นที่ 1 ประมวลข้อความภาษาไทยภายในเอกสาร โดยการตัดคำภาษาไทย (Thai Word Segmentation) โดยในการดำเนินงานวิจัยนี้ผู้วิจัยใช้การตัดคำด้วยไลบรารี PyThaiNLP เวอร์ชัน 2.2.3 ด้วยเทคนิคการตัดคำภาษาไทยแบบ Maximum Matching Algorithm

ขั้นที่ 2 กำจัดคำหยุด (Stop Words) เอกสารโดยเปรียบเทียบกับพจนานุกรมคำหยุด (Stop Words Dictionary) ของ PyThaiNLP ซึ่งมีจำนวน 1030 คำ เพื่อตัดคำที่ไม่มีความสำคัญในเอกสารและเพื่อลดขนาดของเอกสารให้มีขนาดเล็กลง

ขั้นที่ 3 แบ่งชุดข้อมูลออกเป็นทั้งหมด 8 ชุด ประกอบไปด้วย 1) เนื้อเพลงเฉพาะท่อนนำ 2) เนื้อเพลงเฉพาะท่อนนำที่กำหนดคำหยุดออกไป 3) เนื้อเพลงเฉพาะท่อนเข้าเนื้อเรื่อง 4) เนื้อเพลงเฉพาะท่อนเข้าเนื้อเรื่องที่กำหนดคำหยุดออกไป 5) เนื้อเพลงเฉพาะท่อนหลัก 6) เนื้อเพลงเฉพาะท่อนหลักที่กำหนดคำหยุดออกไป 7) เนื้อเพลงทุกท่อนจากทั้งเพลง 8) เนื้อเพลงทุกท่อนจากเนื้อเพลงที่กำหนดคำหยุดออกไป

ขั้นที่ 4 สร้างเวกเตอร์ของเอกสาร (Document Vector) ของข้อมูลแต่ละชุด โดยเวกเตอร์ของเอกสารนี้หมายถึงตารางที่มีแถวเป็นเอกสารทั้งหมดที่นำมาประมวลผลและมีคอลัมน์เป็นคำที่ปรากฏในเอกสารทุก ๆ เอกสารคือคำที่ปรากฏในพจนานุกรม ซึ่งภายในสร้างเวกเตอร์ของเอกสารประกอบด้วยค่าความถี่และคำที่พิจารณาที่พบในเอกสาร แล้วนำค่าความถี่นั้นมาแทนค่าทำด้วยการคำนวณหาค่า TF-IDF

ขั้นที่ 5 สร้างตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่ง ในการสร้างตัวแบบผู้วิจัยได้ใช้ Scikit-learn ซึ่งเป็นไลบรารีในภาษาไพธอนสำหรับการพัฒนาโปรแกรมโดยใช้การเรียนรู้ของเครื่องซึ่งได้รับความนิยมอย่างสูงในปัจจุบัน สำหรับการสร้างตัวแบบจากเทคนิคซ์พอร์ดเวกเตอร์แมชชีน ผู้วิจัยเลือกใช้ SVC จากไลบรารี sklearn.svm และกำหนดให้ใช้เคอร์เนลแบบเส้นตรง และสำหรับการวิเคราะห์การถดถอยลอจิสติก ผู้วิจัยใช้ LogisticRegression จากไลบรารีจาก sklearn.linear\_model.logistic

#### 1.3 การเปรียบเทียบประสิทธิภาพของตัวแบบ

ผลลัพธ์ของการเปรียบเทียบประสิทธิภาพการทำงานของตัวแบบที่ถูกสร้างขึ้นจากเทคนิคซ์พอร์ดเวกเตอร์แมชชีน (SVM) และการวิเคราะห์การถดถอยลอจิสติก (LR) จากชุดข้อมูลที่เตรียมไว้ทั้ง 8 ชุด โดยใช้วิธีการตรวจสอบไขว้ ใช้ค่า K มีค่า



เท่ากับ 10 ซึ่งจะให้ค่าผลลัพธ์ของการทำงานของตัวแบบเป็นค่าความถูกต้องเฉลี่ย (Average Accuracy) ซึ่งผู้วิจัยใช้เครื่องมือจากไลบรารี Scikit-learn คือ `model_selection.cross_val_score` สำหรับการประเมินประสิทธิภาพของตัวแบบจำแนก ซึ่งมีการกำหนดอาร์กิวเมนต์สำหรับการทำงานดังต่อไปนี้

```
1. kfold = model_selection.KFold(n_splits=10, random_state=seed)
2. cv_results = model_selection.cross_val_score(model,
3.                                             X_train_tfidf,
4.                                             Y_raw,
5.                                             cv=kfold,
6.                                             scoring='accuracy')
7.
8. print("%s:\t%.2f%\t(%.2f)" % (model_name, cv_results.mean()*100, cv_results.std()))
```

จากโค้ดโปรแกรมด้านบน `X_train_tfidf` คือ เวกเตอร์ของเอกสารจากชุดข้อมูลที่เรากำลังพิจารณาจากทั้ง 8 ชุดที่ได้ทำการเตรียมเอาไว้ซึ่งมีการกำหนดการให้น้ำหนักของคำสำคัญแบบ TF-IDF และ `Y_raw` คือคำตอบหรือคลาสของชุดข้อมูลที่เรากำลังพิจารณา และ `scoring` หรือวิธีการวัดประสิทธิภาพกำหนดให้วัดโดยใช้ค่าความถูกต้อง หรือ accuracy

## 2. เครื่องมือการวิจัย

ในกระบวนการพัฒนาและประเมินประสิทธิภาพของตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่องผู้วิจัยได้ใช้เครื่องมือต่าง ๆ ดังต่อไปนี้

- 2.1 ภาษาไพธอน เวอร์ชัน 3.7.6
- 2.2 ไลบรารี PythiaNLP เวอร์ชัน 2.2.3 สำหรับใช้ตัดคำภาษาไทย
- 2.3 ไลบรารี scikit-learn เวอร์ชัน 0.22.1 สำหรับพัฒนาตัวแบบด้วยเทคนิคการเรียนรู้ของเครื่อง

## 3. สถิติที่ใช้ในการประเมินประสิทธิภาพของตัวแบบการจำแนก

การวัดค่าความแม่นยำจากการจำแนก (Accuracy) ซึ่งทำได้จาก Confusion Matrix ซึ่งเมตริกนี้จะเป็นการประเมินผลลัพธ์การจำแนกเทียบกับประเภทข้อมูลนั้นจริง ๆ ที่หาได้โดยค่าความแม่นยำนี้สามารถคำนวณได้จากสมการที่ 1

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (1)$$

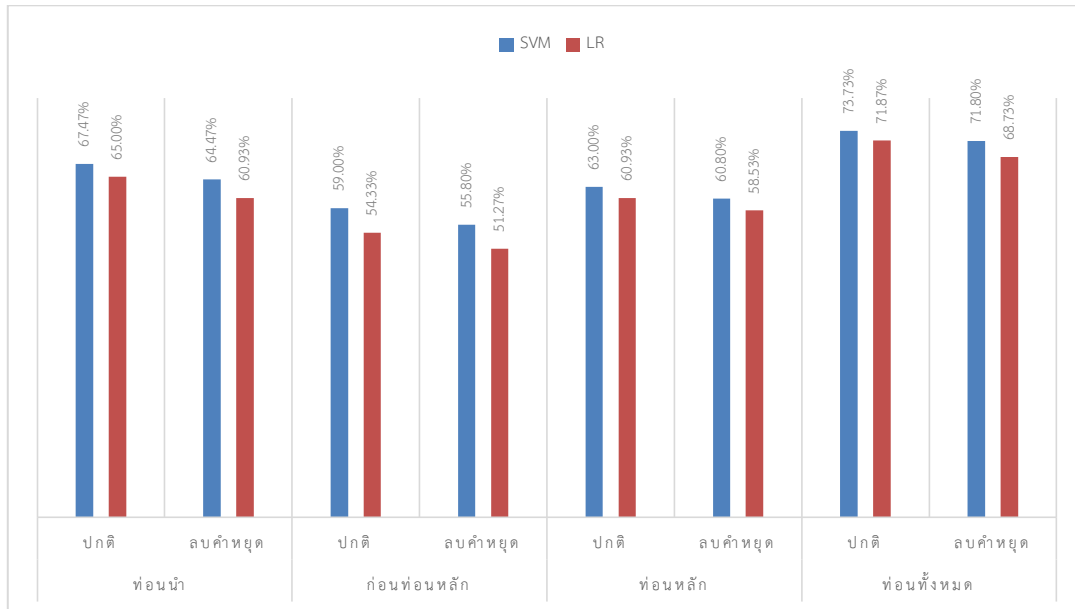
### ผลการวิจัย

#### 1. ผลการพัฒนาตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่อง

ผู้วิจัยทำการสร้างตัวแบบการจำแนกจาก 2 เทคนิค ประกอบด้วยซัพพอร์ตเวกเตอร์แมชชีน และการวิเคราะห์การถดถอยโลจิสติก โดยใช้ชุดข้อมูลเนื้อร้องเพลงลูกทุ่งจำนวน 1,500 เพลง แบ่งออกเป็นบทเพลงประเภท เศร้า, สนุกสนาน และให้กำลังใจ จำนวนประเภทละ 500 บทเพลงที่ไม่ซ้ำกัน และยังสามารถแบ่งข้อมูลเนื้อเพลงออกเป็นชุดย่อยเพื่อใช้สำหรับการสร้างตัวแบบเป็นจำนวน 8 ชุดประกอบด้วย 1) เนื้อเพลงเฉพาะท่อนนำ 2) เนื้อเพลงเฉพาะท่อนนำที่กำจัดคำหยุด 3) เนื้อเพลงเฉพาะท่อนเข้าเนื้อเรื่อง 4) เนื้อเพลงเฉพาะท่อนเข้าเนื้อเรื่องที่กำจัดคำหยุด 5) เนื้อเพลงเฉพาะท่อนหลัก 6) เนื้อเพลงเฉพาะท่อนหลักที่กำจัดคำหยุด 7) เนื้อเพลงทุกท่อนจากทั้งเพลง 8) เนื้อเพลงทุกท่อนจากเนื้อเพลงที่กำจัดคำหยุด โดยผู้วิจัยได้ใช้เทคนิคลดจนขั้นตอนวิธีในการจัดทำเหมืองข้อความ และการประมวลผลภาษาธรรมชาติในการจัดเตรียมข้อมูลอย่างเป็นระบบ และได้พบว่าการนำข้อมูลในชุดที่ 7) เนื้อเพลงทุกท่อนจากทั้งเพลง สามารถสร้างตัวแบบที่มีประสิทธิภาพในการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งได้ดีกว่าชุดข้อมูลอื่น

## 2. ผลการประเมินประสิทธิภาพของตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่อง

การทดลองเปรียบเทียบประสิทธิภาพด้านความถูกต้องในการจำแนกข้อมูลอารมณ์ของเนื้อร้องเพลงลูกทุ่งของตัวแบบที่สร้างจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน และการวิเคราะห์การถดถอยลอจิสติก โดยการใช้ชุดข้อมูลที่เตรียมไว้ทั้งหมด 8 ชุด ดังแสดงผลลัพธ์ของการเปรียบเทียบในภาพที่ 1 พบว่า ตัวแบบจำแนกที่สร้างโดยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน และใช้ชุดข้อมูลชุดที่ 7) คือเนื้อเพลงทุกท่อนจากทั้งเพลง ให้ค่าความถูกต้องสูงที่สุดอยู่ที่ร้อยละ 73.73 รองลงมาคือตัวแบบที่สร้างด้วยเทคนิคการวิเคราะห์การถดถอยลอจิสติก โดยใช้ชุดข้อมูลชุดที่ 7 เช่นกัน ให้ค่าความถูกต้องอยู่ที่ร้อยละ 71.87



ภาพที่ 1. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบ

### อภิปรายผลการวิจัย

1. การพัฒนาตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่องพบว่าปัจจัยที่ส่งผลให้ตัวแบบการจำแนกสามารถทำงานได้อย่างมีประสิทธิภาพคือจำนวนคำศัพท์ที่นำมาสร้างเป็นเวกเตอร์เอกสาร โดยในการสอนให้ตัวแบบได้เรียนรู้ นั้น ยังมีจำนวนคำศัพท์จำนวนมากจะยิ่งส่งผลต่อประสิทธิภาพความแม่นยำของการจำแนกของตัวแบบ ซึ่งจากผลการทดลองชี้ให้เห็นว่าการใช้เนื้อร้องเพลงทั้งเพลงโดยไม่มีการกำจัดคำหยุดมีความเหมาะสมมากที่สุดในการบ่งชี้ถึงอารมณ์ของเพลงทั้งเพลง โดยผู้วิจัยได้ใช้กระบวนการจัดเตรียมข้อมูล และการดำเนินการจัดทำเหมือนข้อความที่นำเสนอไว้ในงานวิจัยของ กานดา แฝ่วพัฒนากุล และปราโมทย์ ลีอนาม (2556) และงานวิจัยของพิศาล สุขชี และคณะ (2562)

2. ผลการประเมินประสิทธิภาพของตัวแบบการจำแนกอารมณ์จากเนื้อร้องเพลงลูกทุ่งด้วยเทคนิคการเรียนรู้ของเครื่อง พบว่าตัวแบบจำแนกที่พัฒนาด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพการจำแนกได้ดีที่สุด ซึ่งสอดคล้องกับงานวิจัยของวาทีนิ น้อยเพียร และพยุ่ง มีสัจ (2556) ที่บ่งชี้ว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพในการจำแนกเอกสารภาษาไทยได้เป็นอย่างดี



## ข้อเสนอแนะ

สำหรับแนวคิดในการพัฒนางานวิจัยต่อไปในอนาคต หรือการนำเอาตัวแบบการจำแนกที่ได้จากการวิจัยในครั้งนี้ไปใช้งานนั้น สามารถทำได้อย่างหลากหลายเช่น การนำไปพัฒนาระบบจำแนกหมวดหมู่เพลงลูกทุ่งโดยอัตโนมัติจากเนื้อร้อง หรือ การพัฒนาระบบช่วยแต่งเพลงลูกทุ่งตามรูปแบบอารมณ์เพลงของเนื้อร้องที่ผู้แต่งต้องการ เป็นต้น

## เอกสารอ้างอิง

- Chatree, J., Sookkhee, P., & Sookkhee, S. (2016). Text Mining Approach to Extract Information of Symptoms and Herbs from Data on Thai Web Sites. in *The International Conference on Applied Statistics 2016 (ICAS 2016)*. The International Conference on Applied Statistics 2016 (ICAS 2016), Phuket.
- SAS. (2020). การประมวลผลภาษาธรรมชาติ. สืบค้น 18 กันยายน 2563, จาก [https://www.sas.com/th\\_th/insights/analytics/what-is-natural-language-processing-nlp.html](https://www.sas.com/th_th/insights/analytics/what-is-natural-language-processing-nlp.html)
- กานดา แผ้วฉันทกุล, ปราโมทย์ ลือนาม. (2556). การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์. *วารสารการจัการสมัยใหม่*. 11(2), 11-20.
- กรมวุฒิ นงนุช, อนุชา ซาเฮา และ สุวุฒิ ต้มทอง. (2559). การวิเคราะห์บทความอัตโนมัติโดยใช้กระบวนการภาษาธรรมชาติ. ใน *การประชุมวิชาการระดับชาติมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ครั้งที่ 1*.
- กัลยา วานิชย์บัญชา. (2551). *การวิเคราะห์ข้อมูลหลายตัวแปร*. พิมพ์ครั้งที่ 3. กรุงเทพฯ:สำนักพิมพ์ธรรมสาร.
- นพมาศ ปักเข็ม, ชนิดา จันมณี, ศิวกร อุษย. (2560). การจำแนกประเภทภูมิปัญญาท้องถิ่นของไทยแบบอัตโนมัติด้วยวิธีการทางเหมืองข้อมูล. *วารสารมหาวิทยาลัยทักษิณ*. 20(ฉบับพิเศษ), 301-307.
- นิติพร จันทร์กลิ่น. (2557). *การปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนสำหรับการจำแนกข้อมูลภาพใบโอเมตริกซ์*. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. สาขาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีสุรนารี, นครราชสีมา.
- พรทีย์ ฉายกี, จันทนา แก้ววิเชียร. (2561). วิเคราะห์แนวคิดและศิลปะการใช้ภาษาในวรรณกรรมเพลงลูกทุ่งที่แต่งโดย สลา คุณวุฒิ. *วารสารราชภัฏเพชรบูรณ์สาร*, 20(1), 87-96.
- พิศาล สุขชี, เกษภา ชาติ และ สิริกัลยา สุขชี. (2562). การพัฒนาระบบสนับสนุนการตัดสินใจในการใช้สมุนไพรที่ตอบสนองต่ออาการเจ็บป่วยโดยใช้เทคนิคเหมืองข้อความ. *Veridian E-Journal Science and Technology Silpakorn University*, 6(2), 82-101.
- วาทีนี้ น้อยเพียร, พยุง มีสัจ. (2556). การเปรียบเทียบเทคนิคการคัดเลือกคุณลักษณะแบบการกรองและการควบรวมของการทำเหมืองข้อความเพื่อการจำแนกข้อความ. *วารสารวิชาการเทคโนโลยีอุตสาหกรรม*, 9(3), 118-129.
- วรรณวิภา วงศ์วิไลสกุล. (2556). เหมืองข้อความและการประยุกต์ใช้. *Panyapiwat Journal*, 4(ฉบับพิเศษ), 157-165.
- วัชรวิวัฒน์ จิตต์สกุล, สุนันทา สดสี. (2560). การยกระดับการจำแนกข้อความป้อนกลับของสินค้าโดยตัวจำแนกข้อความร่วมกับการวัดความสำคัญของคำ. *วารสาร มทร.อีสาน ฉบับวิทยาศาสตร์และเทคโนโลยี*, 10(3), 104-117.
- ศศิมา มณฑาสวรรณ. (2557). *การพัฒนาระบบค้นหารหัส ICD - 10 สำหรับงานเวชระเบียน*. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. สาขาวิชาสารสนเทศศาสตร์เพื่อการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร, นครปฐม
- สุนทรี ดวงทิพย์. (2558). การศึกษาวิเคราะห์วรรณศิลป์ในวรรณกรรมเพลงลูกทุ่งของหนูมีเตอร์. *สักทอง: วารสารมนุษยศาสตร์และสังคมศาสตร์ (สทมส.)*, 21(2), 67-8