โครงการ
การเรียนการสอนเพื่อเสริมประสบการณ์

ชื่อโครงการ	ระบบวิเคราะห์ความคิดเห็นต่อละครไทยบนทวิตเตอร์ Sentiment Analysis for Thai drama on Twitter	
ชื่อนิสิต	นายธนะสิทธิ์ เร่งสมบูรณ์สุข	5933629523
	นางสาวกวินธิดา สายยศ	5933602523
ภาควิชา	คณิตศาสตร์และวิทยาการคอมพิวเตอร์ สาขาวิชา วิทยาการคอมพิวเตอร์	
ปีการศึกษา	2562	

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ระบบวิเคราะห์ความคิดเห็นต่อละครไทยบนทวิตเตอร์

นายธนะสิทธิ์ เร่งสมบูรณ์สุข
นางสาวกวินธิดา สายยศ

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตร์บัณฑิต
สาขาวิชา วิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Sentiment Analysis for Thai drama on Twitter

Tanasit Rengsomboonsuk

Kawintida Saiyot

A Project Submitted in Partial Fulfillment of the Requirements

for the Degree of Bachelor of Science Program in Computer Science

Department of Mathematics and Computer Science

Faculty of Science

Chulalongkorn University

Academic Year 2019

Copyright of Chulalongkorn University

หัวข้อโครงงาน	ระบบวิเคราะห์ความคิดเห็นต่อละครไทยบนทวิตเตอร์
โดย	นายธนะสิทธิ์ เร่งสมบูรณ์สุข
	นางสาวกวินธิดา สายยศ
สาขาวิชา	วิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษาโครงงานหลัก	ผู้ช่วยศาสตราจารย์ ดร.ภควรรณ ปักษี

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับโครงงานฉบับนี้เป็นส่วนหนึ่ง ของการศึกษาตามหลักสูตรปริญญาบัณฑิต ในรายวิชา 2301499 โครงงานวิทยาศาสตร์ (Senior Project)

.....

(ศาสตราจารย์ ดร.กฤษณะ เนียมมณี) หัวหน้าภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะกรรมการสอบโครงงาน

.....

(ผู้ช่วยศาสตราจารย์ ดร.ภควรรณ ปักษี) อาจารย์ที่ปรึกษาโครงงานหลัก

.....

(รองศาสตราจารย์ ดร.จารุโลจน์ จงสถิตย์วัฒนา) กรรมการ

.....

(ผู้ช่วยศาสตราจารย์ ดร.กิติพร พลายมาศ) กรรมการ

ธนะสิทธิ์ เร่งสมบูรณ์สุข,กวินธิดา สายยศ : ระบบวิเคราะห์ความคิดเห็นต่อละครไทยบนทวิตเตอร์ (Sentiment Analysis for Thai drama on Twitter) อ.ที่ปรึกษาโครงงานหลัก: ผู้ช่วยศาสตราจารย์ ดร. ภควรรณ ปักษี, 84 หน้า.

เนื่องจากปัจจุบันผู้บริโภคส่วนใหญ่สนใจชมละคร และใช้สื่อออนไลน์ (Online media) เช่น ทวิตเตอร์ (Twitter) เป็นที่แลกเปลี่ยนแสดงความคิดเห็นเกี่ยวกับละคร ทำให้มีข้อความแสดงความคิดเห็นจำนวนมาก จึงต้องใช้เวลามากในการอ่านและทำความเข้าใจข้อความทั้งหมด ดังนั้นผู้พัฒนาจึงศึกษากลุ่มคำการจำแนกอารมณ์และความรู้สึกของข้อความที่เกี่ยวกับละครไทย และพัฒนาระบบวิเคราะห์ความคิดเห็นเกี่ยวกับละครไทยบนทวิตเตอร์ในรูปแบบของเว็บแอปพลิเคชัน (Web application) เป็นเครื่องมือที่ช่วยวิเคราะห์และสรุปผลความชอบที่มีต่อละครไทย โดยวิเคราะห์ข้อความแสดงความคิดเห็นออกเป็นชอบ ไม่ชอบ และกลาง ระบบนี้รวบรวมข้อความเกี่ยวกับละครที่เป็นภาษาไทย และไม่ครอบคลุมคำสแลง คำสะกดผิดหรือภาษาถิ่น ใช้การพัฒนาระบบด้วยภาษา Python และใช้ไลบรารี PyThaiNLP ช่วยในการตัดคำ และมีการเปรียบเทียบคำในข้อความกับคลังคำศัพท์สำหรับละครไทยที่สร้างขึ้น เพื่อสร้างเป็นเวกเตอร์ของข้อความสำหรับสร้างโมเดลการเรียนรู้ด้วยเครื่องแบบเทคนิคนาอีฟเบย์ หลังจากนั้นนำโมเดลที่ได้มาจำแนกข้อความแสดงความคิดเห็นเกี่ยวกับละครว่าชอบ หรือไม่ชอบละครเรื่องนั้น ๆ ในเรื่องใด ระบบที่พัฒนาขึ้นนี้คาดว่าจะเป็นเครื่องมือที่ช่วยในการตัดสินใจรับชมละครได้ง่ายขึ้น และเป็นประโยชน์ต่อผู้ผลิตละครสามารถนำไปวางแผนการผลิตละครในอนาคต

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์ ลายมือชื่อนิสิต ธนะสิทธิ์ เร่งสมบูรณ์สุข

ลายมือชื่อนิสิต กวินธิดา สายยศ

สาขาวิชา วิทยาการคอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาโครงงานหลัก วิไลวรรณ ปักษี

ปีการศึกษา 2562

5933629523, 5933602523 : MAJOR COMPUTER SCIENCE
KEYWORDS: SENTIMENT ANALYSIS / THAI DRAMA / NAIVE BAYES
TANASIT RENGSOMBOONSUK,KAWINTIDA SAIYOT: SENTIMENT ANALYSIS FOR THAI DRAMA ON TWITTER. ADVISOR: ASSIST. PROF. PAKAWAN PUGSEE, Ph.D., 84 pp.

Since most consumers are interested in watching dramas and using online media such as twitter to exchange opinions about dramas, a lot of comments are found and the consumers take more time of read and understand the overall of the other consumers views. Therefore, we have studied about word grouping, classification of emotions and feeling of the text about Thai dramas and then developed a system for analyzing opinions about Thai dramas on Twitter in the form of a web application, which is a tool that helps analyzing and summarizing preferences for Thai dramas. The process is analyzing the opinions expressed as likes, dislikes and middle level comments. This system collected the text about the dramas in Thai language and it does not cover slang, misspellings and dialects. It has been developed with Python language and text processing by the PyThaiNLP library. Each word will be looked up in the vocabulary library created for the Thai dramas. Then, the vectors of text are created for training a learning model using Naive Bay approach. After that, the model will classify the comments about the dramas whether most consumers like or not like the drama. The developed system is expected to be a tool that be able to make decision watching the dramas easier and this is beneficial to the dramas producers to facilitate planning the production of the dramas in the future.

Department : Mathematics and Computer Science Student's Signature Tanasit Rengsomboonsuk
Student's Signature Kawintida Saiyot
Field of Study : Computer Science Advisor's Signature Pakawan Pugsee
Academic Year : 2019

กิตติกรรมประกาศ

การจัดทำโครงงานระบบวิเคราะห์ความคิดเห็นต่อละครไทยบนทวิตเตอร์ สามารถลุล่วงไปได้ด้วยดี เนื่องจากได้รับความอนุเคราะห์และช่วยเหลือจากคณาจารย์และบุคลาการต่าง ๆ ดังนี้

ขอขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.ภควรรณ ปักษี อาจารย์ที่ปรึกษาโครงการ ที่คอยให้คำปรึกษา ข้อเสนอแนะทางวิชาการ อีกทั้งยังช่วยแนะนำ แก้ไข และชี้แนะตลอดการดำเนินการโครงงาน

ขอขอบพระคุณคณะกรรมการสอบ ได้แก่ ผู้ช่วยศาสตราจารย์ ดร.กิติพร พลายมาศ และรองศาสตราจารย์ ดร.จารุโลจน์ จงสถิตวัฒนา ที่ช่วยให้คำแนะนำ และข้อเสนอแนะ สำหรับพัฒนาโครงงานนี้ให้มีความถูกต้องและสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณ คุณพ่อและคุณแม่ที่คอยช่วยสนับสนุน ให้กำลังใจตลอดการทำโครงงาน

ขอขอบคุณเพื่อน ๆ ภาควิชาคณิตศาสตร์ สาขาวิชาวิทยาการคอมพิวเตอร์ ที่คอยช่วยเหลือและให้คำปรึกษาเกี่ยวกับโครงงาน

ท้ายที่สุดนี้ ขอขอบพระคุณทุกความกรุณาจากทุกท่านที่กล่าวมา รวมถึงบุคคลที่ไม่ได้กล่าวถึง ไว้ ณ ที่นี้อีกครั้งหนึ่ง สำหรับความช่วยเหลือและคำแนะนำต่าง ๆ ซึ่งทำให้โครงงานนี้ประสบสำเร็จ ลุล่วงไปด้วยดี

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า

บทที่ 1
บทนำ

1.1 ความเป็นมาและเหตุผลของโครงงาน

ปัจจุบันละครไทยเป็นที่นิยมของกลุ่มคนจำนวนมาก เนื่องจากปัจจุบันมีละครเพิ่มมากขึ้นทำ ให้ผู้คน สนใจดูละคร ซึ่งละครมีหลายรูปแบบ เช่น ละครตลก (comedy) ดราม่า (drama) และมีการแลกเปลี่ยนแสดงความคิดเห็นเกี่ยวกับละครผ่านทางสื่อออนไลน์ (online media) เช่น ทวิตเตอร์ (Twitter) ทำให้มีข้อมูลที่หลากหลาย ต้องใช้เวลานานในการค้นหาข้อมูลที่ต้องการ ซึ่งข้อมูลเหล่านี้สามารถนำไปวิเคราะห์ข้อความแสดงความคิดเห็น เพื่อใช้เป็นประโยชน์ในการตัดสินใจเลือกชมละคร ดังนั้นควรที่จะมีระบบวิเคราะห์ข้อความอัตโนมัติเพื่อลดปัญหาเหล่านี้ ตัวอย่างงานวิจัยเกี่ยวกับการวิเคราะห์ความคิดเห็นที่นำมาประยุกต์ใช้ เช่น งานวิจัย [1] มีการวิเคราะห์ความคิดเห็นต่อโรงแรมซึ่งใช้ข้อความรีวิวโรงแรมบนเว็บอโกด้าและบนทวิตเตอร์ โดยได้แบ่งประเภทของข้อความแสดงความคิดเห็นนี้เป็นข้อความเชิงบวก (positive) คือดี และข้อความเชิงลบ (negative) คือไม่ดี โดยใช้เทคนิค 4 วิธี คือ นาอีฟเบย์ (naïve bayes) ซับพอร์ตเวกเตอร์แมชชีนหรือเอสวีเอ็ม (support vector machine: SVM) เค-เนียร์เรสเนเบอร์ (k-nearest neighbor) และต้นไม้ตัดสินใจ (decision tree) ผลสรุปข้อมูลที่ได้จากการวิเคราะห์จากงานวิจัยนี้จะเห็นได้ว่า ค่าความถูกต้องของเทคนิคเอสวีเอ็มจะสูงที่สุด และรองลงมาคือนาอีฟเบย์ แต่จุดอ่อนของงานคือ ไม่มีการคำนวณหาค่าความแม่นยำของการจำแนกข้อความแต่ละด้าน อีกงานวิจัยหนึ่ง [2] เป็นงานวิจัยที่ศึกษาเกี่ยวกับการจำแนกความรู้สึกหรือความคิดเห็นเกี่ยวกับภาพยนตร์ โดยใช้ข้อมูลจากเว็บ https://www.imdb.com/ ซึ่งเป็นฐานข้อมูลรีวิวภาพยนตร์ (movie reviews database: IMDb) รวบรวมความคิดเห็นต่าง ๆ เกี่ยวกับภาพยนตร์ งานนี้มีการใช้เทคนิคต่าง ๆ ในการตัดคำและคิดค่าคะแนนเพื่อทำให้เทคนิคนาอีฟเบย์มีประสิทธิภาพในการจำแนกความรู้สึกที่รวดเร็วและถูกต้อง ดังนั้นจากการศึกษางานวิจัยด้านการวิเคราะห์ความคิดเห็น (opinion mining) และการวิเคราะห์อารมณ์และความรู้สึก ผู้พัฒนาพบว่า เทคนิคนาอีฟเบย์เป็นการเรียนรู้ด้วยเครื่องเทคนิคหนึ่งที่คนส่วนใหญ่นิยมใช้ในการวิเคราะห์ข้อความมากที่สุด จึงจะประยุกต์ใช้เทคนิคนาอีฟเบย์ในการวิเคราะห์ข้อความแสดงความคิดเห็นที่เป็นภาษาไทย

จากที่กล่าวมาข้างต้น ทางผู้พัฒนาจึงจะพัฒนาระบบวิเคราะห์อารมณ์และความรู้สึกต่อละครไทยบนทวิตเตอร์ โดยวิเคราะห์ข้อความแสดงความคิดเห็นออกเป็นข้อความที่ชอบ ไม่ชอบ หรือรู้สึกเป็นกลางต่อละครเรื่องนั้น ๆ และแสดงผลสรุปข้อมูลที่ได้จากการวิเคราะห์เป็นรูปกราฟแท่ง เพื่อให้ผู้บริโภคใช้เป็นเครื่องมือในการตัดสินใจเลือกชมละครที่สะดวกมากขึ้น และเป็นประโยชน์ต่อผู้ผลิตละครในการวางแผนการผลิตในอนาคต

1.2 วัตถุประสงค์ของโครงงาน

1. ศึกษาการจำแนกอารมณ์และความรู้สึกของข้อความภาษาไทยจากข้อความที่เกี่ยวข้องกับละครไทย
2. วิเคราะห์ความชอบที่มีต่อละครไทยและสรุปผลตอบรับจากข้อความที่อยู่บนทวิตเตอร์

1.3 ขอบเขตการวิจัย

1. โครงงานนี้ศึกษาเฉพาะข้อความบนทวิตเตอร์ที่เป็นภาษาไทย ไม่ครอบคลุมคำสแลง และไม่พิจารณาคำที่สะกดผิด
2. การเก็บตัวอย่างข้อความแสดงความคิดเห็นบนทวิตเตอร์จะใช้แฮชแท็ก (#) ชื่อละคร หรือ ชื่อตัวละคร (พระเอก/นางเอก) จากละครไทย 8 เรื่อง ในช่วงกันยายน – ธันวาคม ในปี พ.ศ. 2562 ได้แก่ รักฉุดใจนายฉุกเฉิน ลิขิตรักข้ามดวงดาว มธุรสโลกันต์ รองเท้านารี เพลิงรักเพลิงแค้น ฤกษ์สังหาร เขาวานให้หนูเป็นสายลับ และดาวหลงฟ้า โดยมีข้อความที่จะนำมาใช้วิเคราะห์ในโครงการนี้อย่างน้อย 10,000 ข้อความ
3. ผลลัพธ์ในการจำแนกข้อความแสดงความคิดเห็นจะถูกแบ่งออกเป็น ชอบ ไม่ชอบ และกลาง

1.4 ขั้นตอนการวิจัย

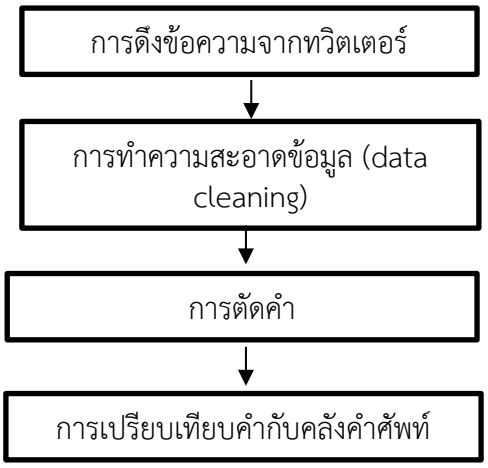
1. ศึกษาและค้นคว้าหาข้อมูลเกี่ยวกับการวิเคราะห์ความคิดเห็นของข้อความและการทำคลังศัพท์ (มีการศึกษากลุ่มคำด้านบวก (positive) ด้านลบ (negative) และเป็นกลาง (neutral) จาก [3] [4] [5])
2. กำหนดขอบเขตและขั้นตอนการดำเนินงาน
3. เก็บรวบรวมข้อความที่เกี่ยวกับละครไทยจากทวิตเตอร์ โดยใช้แฮชแท็ก (#) ชื่อละคร หรือชื่อตัวละครโดยใช้ไลบรารีที่ชื่อ Tweepy ซึ่งจะช่วยในการเชื่อมต่อไปยังทวิตเตอร์เอพีไอ
4. วิเคราะห์ข้อความแสดงความคิดเห็นและจำแนกประเภทของคำ โดยแบ่งข้อความออกเป็นประโยคและแบ่งประโยคออกเป็นคำ เพื่อใช้เป็นข้อมูลในการออกแบบระบบวิเคราะห์ข้อความแสดงความคิดเห็น
5. ออกแบบและพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทย ซึ่งจะมีการเปรียบเทียบผลลัพธ์ของการจำแนกข้อมูลระหว่างการใช้คลังคำศัพท์ (นับความถี่ของคำด้านบวกและด้านลบที่ปรากฏในประโยค) เพียงอย่างเดียวกับการใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ ซึ่งจะทดลองกับเซตของคุณลักษณะ (feature set) ที่แตกต่างกันเพื่อเลือกใช้โมเดลการจำแนกข้อมูลที่ดีที่สุดมาพัฒนาเป็นระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทย

ตัวอย่างเซตของคุณลักษณะที่คาดว่าจะใช้สำหรับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ เช่น

- แบบที่ 1 ใช้คำในประโยค
- แบบที่ 2 ใช้เฉพาะคำที่ตรงกับคำในคลังคำศัพท์
- แบบที่ 3 ใช้คำในประโยคและคำที่ตรงกับคำในคลังคำศัพท์

โครงสร้างของระบบจะแบ่งเป็น 2 ส่วน คือ Front-end และ Back-end ซึ่งในส่วนของ Front-end เป็นการแสดงผลลัพธ์การจำแนกประโยคข้อความผ่านทางหน้าเว็บแอปพลิเคชัน แบ่งออกเป็นข้อความ 3 กลุ่ม คือ ชอบ ไม่ชอบ และกลาง อีกทั้งสามารถแสดงผลข้อมูลในรูปแบบของกราฟแท่ง สำหรับการทำงานในส่วนของ Back-end จะมีการออกแบบวิธีการจำแนกข้อมูลหลาย ๆ วิธีการ ดังต่อไปนี้

1) การทำงานของระบบสำหรับการจำแนกข้อมูลโดยคลังคำศัพท์เพียงอย่างเดียว จะประกอบด้วย 4 มอดูล



มอดูล: การดึงข้อความจากทวิตเตอร์

ข้อมูลนำเข้า: แฮชแท็กชื่อละครหรือตัวละคร

กระบวนการ : ใช้ไลบรารี Tweepy ในการดึงข้อมูล

ข้อมูลที่ส่งออก : ประโยคข้อความจากทวิตเตอร์ที่ตรงกับแฮชแท็กที่ใช้ในการดึงข้อมูล

มอดูล: การทำความสะอาดข้อมูล
ข้อมูลนำเข้า: ประโยคข้อความจากทวิตเตอร์
กระบวนการ : ใช้วิธี manual ในการอ่านและลบข้อมูลที่ไม่เกี่ยวข้อง เช่น แฮชแท็ก ยูอาร์แอล ข้อความที่รีทวิต
ข้อมูลที่ส่งออก : ประโยคข้อความที่ทำความสะอาดแล้ว

มอดูล: การตัดคำ

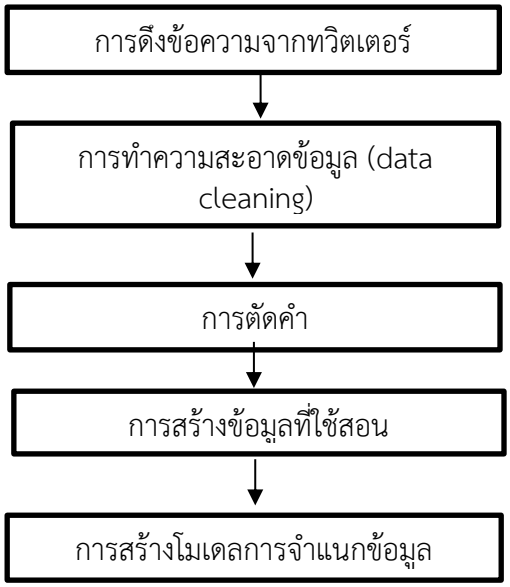
ข้อมูลนำเข้า: ประโยคข้อความที่ได้หลังจากการทำความสะอาด
กระบวนการ : ใช้ไลบรารี PyThaiNLP ด้วยเทคนิค Maximum Matching algorithm ในการตัดคำ
ข้อมูลที่ส่งออก : คำในประโยค

มอดูล: การเปรียบเทียบคำกับคลังคำศัพท์
ข้อมูลนำเข้า: คำในประโยค
กระบวนการ : ตรวจหาคำในประโยคที่ตรงกับคำในคลังคำศัพท์
ข้อมูลที่ส่งออก : คำในประโยคที่ตรงกับคำในคลังคำศัพท์

การจำแนกประโยคใช้การคำนวณผลรวมของจำนวนคำด้านบวกกับคำด้านลบที่ตรงกับคำในคลังคำศัพท์ คือ

- จำนวนคำด้านบวกเท่ากับจำนวนคำด้านลบจะได้เป็นประโยคในกลุ่มเป็นกลาง
- จำนวนคำด้านบวกมากกว่าจำนวนคำด้านลบจะได้เป็นประโยคในกลุ่มชอบ
- จำนวนคำด้านบวกน้อยกว่าจำนวนคำด้านลบจะได้เป็นประโยคในกลุ่มไม่ชอบ

2) การทำงานของระบบสำหรับการจำแนกข้อมูลโดยการใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ โดยไม่ใช้ข้อมูลจากคลังคำศัพท์ (เซตของคุณลักษณะแบบที่ 1) จะประกอบด้วย 5 มอดูล



มอดูลสามส่วนแรก คือ การดึงข้อความจากทวิตเตอร์ การทำความสะอาดข้อมูล และการตัดคำจะทำงานเหมือนที่กล่าวแล้วข้างต้น

มอดูล: การสร้างข้อมูลที่ใช้สอน
ข้อมูลนำเข้า: คำในประโยค
กระบวนการ : ใช้วิธี One-hot โดยให้คำที่ไม่ซ้ำกันแต่ละคำของข้อมูลทั้งหมดเป็นคุณลักษณะ ซึ่งถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0
ข้อมูลที่ส่งออก : เซตของคุณลักษณะแบบที่ 1 ที่จะนำไปเรียนรู้ด้วยเครื่อง

มอดูล : การสร้างโมเดลการจำแนกข้อมูล
ข้อมูลนำเข้า : เซตของคุณลักษณะแบบที่ 1 และกลุ่มของประโยคข้อความ
กระบวนการ : ใช้ไลบรารีนาอีฟเบย์ของเครื่องมือการเรียนรู้ด้วยเครื่อง
ข้อมูลที่ส่งออก : โมเดลการจำแนกข้อความออกเป็น 3 กลุ่ม ได้แก่ ชอบ ไม่ชอบ และกลาง

การจำแนกประโยคข้อความใช้การให้คำตอบจากโมเดลการจำแนกข้อมูล

3) การทำงานของระบบสำหรับการจำแนกข้อมูลโดยการใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ โดยใช้เฉพาะคำที่ตรงกับคำในคลังคำศัพท์ (เซตของคุณลักษณะแบบที่ 2) จะประกอบด้วย 6 มอดูล

มอดูลสี่ส่วนแรก คือ การดึงข้อความจากทวิตเตอร์ การทำความสะอาดข้อมูล การตัดคำ และการเปรียบเทียบคำกับคลังคำศัพท์ จะทำงานเหมือนที่กล่าวแล้วข้างต้น

มอดูล: การสร้างข้อมูลที่ใช้สอน
ข้อมูลนำเข้า: คำในประโยคที่ตรงกับคำในคลังคำศัพท์
กระบวนการ : ใช้วิธี One-hot โดยให้คำทั้งหมดในคลังคำศัพท์เป็นคุณลักษณะ ถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0
ข้อมูลที่ส่งออก : เซตของคุณลักษณะแบบที่ 2 ที่จะนำไปเรียนรู้ด้วยเครื่อง

มอดูล: การสร้างโมเดลการจำแนกข้อมูล
ข้อมูลนำเข้า: เซตของคุณลักษณะแบบที่ 2 และกลุ่มของประโยคข้อความ
กระบวนการ : ใช้ไลบรารีนาอีฟเบย์ของเครื่องมือการเรียนรู้ด้วยเครื่อง
ข้อมูลที่ส่งออก : โมเดลการจำแนกข้อความออกเป็น 3 กลุ่ม ได้แก่ ชอบ ไม่ชอบ และกลาง

การจำแนกประโยคข้อความใช้การให้คำตอบจากโมเดลการจำแนกข้อมูล

4) การทำงานของระบบสำหรับการจำแนกข้อมูลโดยการใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ โดยใช้คำในประโยคและคำที่ตรงกับคำในคลังคำศัพท์ (เซตของคุณลักษณะแบบที่ 3) จะประกอบด้วย 6 มอดูล ดังการทำงานของระบบในข้อ 3)

มอดูลสี่ส่วนแรก คือ การดึงข้อความจากทวิตเตอร์ การทำความสะอาดข้อมูล การตัดคำ และการเปรียบเทียบคำกับคลังคำศัพท์ จะทำงานเหมือนที่กล่าวแล้วข้างต้น

มอดูล: การสร้างข้อมูลที่ใช้สอน
ข้อมูลนำเข้า: คำในประโยคและคำในประโยคที่ตรงกับคำในคลังคำศัพท์
กระบวนการ : - คำในประโยคใช้วิธี One-hot โดยให้คำที่ไม่ซ้ำกันแต่ละคำของข้อมูลทั้งหมดเป็นคุณลักษณะ ซึ่งถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0
- คำในประโยคที่ตรงกับคำในคลังคำศัพท์ใช้วิธี One-hot โดยให้คำทั้งหมดในคลังคำศัพท์เป็นคุณลักษณะ ถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0
ข้อมูลที่ส่งออก : เซตของคุณลักษณะแบบที่ 3 ที่จะนำไปเรียนรู้ด้วยเครื่อง

มอดูล: การสร้างโมเดลการจำแนกข้อมูล
ข้อมูลนำเข้า: เซตของคุณลักษณะแบบที่ 3 และกลุ่มของประโยคข้อความ
กระบวนการ : ใช้ไลบรารีนาอีฟเบย์ของเครื่องมือการเรียนรู้ด้วยเครื่อง
ข้อมูลที่ส่งออก : โมเดลการจำแนกข้อความออกเป็น 3 กลุ่ม ได้แก่ ชอบ ไม่ชอบ และกลาง

การจำแนกประโยคข้อความใช้การให้คำตอบจากโมเดลการจำแนกข้อมูล

6. ตรวจสอบความถูกต้องของการจำแนกข้อมูล โดยนำผลลัพธ์ที่ได้มาตรวจสอบความถูกต้อง
7. จัดทำเอกสารรายงาน และคู่มือการใช้งานระบบ

ตารางเวลาการดำเนินงาน

การพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทยบททวิตเตอร์ เริ่มดำเนินงานตั้งแต่เดือนกันยายน 2562 ถึงเดือนเมษายน 2563 รวมระยะเวลา 8 เดือน โดยมีตารางเวลาการดำเนินงาน ดังนี้

ขั้นตอนดำเนินงาน	พ.ศ. 2562				พ.ศ. 2563			
	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.
1. ศึกษาและค้นคว้าหาข้อมูลเกี่ยวกับการวิเคราะห์ความคิดเห็นของข้อความ และการทำคลังศัพท์สำหรับการจำแนกกลุ่มคำของละครไทย								
2. กำหนดขอบเขตและขั้นตอนการดำเนินงาน								
3. เก็บรวบรวมข้อมูลที่เกี่ยวกับละครไทยจากทวิตเตอร์								
4. วิเคราะห์ข้อความแสดงความคิดเห็นและจำแนกประเภทของคำ								
5. ออกแบบและพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทย								
6. ตรวจสอบความถูกต้องของการจำแนกข้อมูล								
7.จัดทำเอกสารรายงาน และคู่มือการใช้งานระบบ								

1.5 ประโยชน์ที่ได้รับ

1. ประโยชน์ต่อผู้พัฒนา
   - ได้ศึกษาและเรียนรู้เทคนิคการวิเคราะห์ความคิดเห็นของข้อความจากทวิตเตอร์
   - ได้พัฒนาทักษะการวางแผนการดำเนินงาน

2. ประโยชน์ต่อผู้นำซอฟต์แวร์นี้ไปใช้งาน
   - เป็นเครื่องมือช่วยวิเคราะห์และสรุปผลความชอบที่มีต่อละครไทยและช่วยในการตัดสินใจการเลือกรับชม
   - เป็นประโยชน์ต่อผู้ผลิตละครสามารถนำข้อมูลนี้ไปวางแผนการผลิตละครในอนาคต

1.6 โครงสร้างของรายงาน

บทที่ 2 จะกล่าวถึงบทความและทฤษฎีที่เกี่ยวข้องกับโครงงาน

บทที่ 3 จะกล่าวถึงการรวบรวมข้อมูล และการวิเคราะห์ข้อมูล

บทที่ 4 จะกล่าวถึงการออกแบบและพัฒนาระบบ

บทที่ 5 จะกล่าวถึงผลการทดสอบระบบการวิเคราะห์ข้อมูลข้อความแสดงความคิดเห็นจากการออกแบบและพัฒนาระบบ ซึ่งจะทดสอบการจำแนกข้อความแสดงความรู้สึกที่เป็นบวก เป็นลบ และเป็นกลาง

บทที่ 6 จะกล่าวถึงข้อสรุป และข้อเสนอแนะทั้งหมดของโครงงานนี้

บทที่ 2
งานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงบทความและทฤษฎีที่เกี่ยวข้อง ซอฟต์แวร์ที่ใช้ในการพัฒนาวิธีการจำแนกข้อความแสดงความคิดเห็น ดังรายละเอียดต่อไปนี้

2.1 การวิเคราะห์อารมณ์และความรู้สึก (Sentiment Analysis)

การวิเคราะห์อารมณ์และความรู้สึก คือ การวิเคราะห์อารมณ์และความรู้สึกผ่านทางข้อความ โดยจะวิเคราะห์ข้อความแสดงความคิดเห็นเพื่อบ่งบอกความรู้สึกของผู้เขียน เช่น ความรู้สึกชอบ (positive) ความรู้สึกที่เป็นกลาง (neutral) และความรู้สึกไม่ชอบ (negative) ซึ่งการวิเคราะห์อารมณ์และความรู้สึกเป็นอีกสาขาหนึ่งของการประมวลผลทางภาษาธรรมชาติ (Natural Language Processing) โดยมีงานวิจัยที่เกี่ยวข้องดังต่อไปนี้

งานวิจัย [6] มีการวิเคราะห์ความคิดเห็นต่อผลิตภัณฑ์ดูแลผิวบนทวิตเตอร์ เพื่อวิเคราะห์ข้อความแสดงความคิดเห็นและจำแนกอารมณ์ความรู้สึกของข้อความแสดงความคิดเห็นต่อผลิตภัณฑ์ดูแลผิว โดยงานวิจัยนี้ใช้เทคนิคการทำเหมืองข้อมูล (Data mining process) เช่น การวิเคราะห์ความความคิดเห็น (Opinion analysis) การวิเคราะห์อารมณ์และความรู้สึก (Sentiment analysis) มาช่วยในการจำแนกข้อความแสดงความคิดเห็นและจำแนกอารมณ์ความรู้สึกด้านบวก (Positive) และด้านลบ (Negative) โดยใช้เทคนิคนาอีฟเบย์ และเทคนิคซับพอร์ตเวกเตอร์แมชชีน เพื่อนำข้อมูลไปประกอบการตัดสินใจสำหรับผู้ใช้ซึ่งเป็นผู้บริโภคที่มีความสนใจในผลิตภัณฑ์ดูแลผิว รวมถึงผู้ผลิตสามารถทราบถึงทัศนคติของผู้บริโภคที่มีต่อผลิตภัณฑ์ดูแลผิว

งานวิจัย [7] เป็นงานวิจัยที่เกี่ยวกับการวิเคราะห์ข้อความแสดงความคิดเห็นต่อโรงแรม สำหรับการจัดอันดับตามคุณลักษณะของโรงแรม ศึกษาคำที่ใช้บ่งบอกอารมณ์และความรู้สึกสำหรับข้อความแสดงความคิดเห็นต่อโรงแรม งานวิจัยนี้เก็บรวบรวมข้อมูลข้อความแสดงความคิดเห็นจาก www.agoda.com เพื่อวิเคราะห์ข้อความและจำแนกระดับความคิดเห็นด้านบวกกับด้านลบที่มีต่อโรงแรม ระบบจะวิเคราะห์ข้อความทั้งด้านวากยสัมพันธ์ (syntax) และด้านความหมาย (semantics) รวมถึงใช้การเรียนรู้ด้วยเครื่อง 4 วิธี เพื่อเลือกเทคนิคการเรียนรู้ด้วยเครื่องมี่เหมาะสมที่สุด ได้แก่ ซีโร่อาร์ (ZeroR) วันอาร์ (OneR) ต้นไม้การตัดสินใจ (Decision tree) แบบเจ48 (J48) และนาอีฟเบย์ (NaiveBayes) จากการทดสอบการจำแนกข้อความแสดงความคิดเห็นพบว่า แบบจำลองที่สร้างด้วยเทคนิคนาอีฟเบย์มีประสิทธิภาพดีที่สุด โดยให้ค่าความถูกต้อง ค่าความแม่นยำ และค่าการเรียกคืนที่มากกว่า 85% งานวิจัย

นี้สามารถช่วยให้ผู้ใช้งานเลือกโรงแรมที่จะเข้าพักตามความต้องการได้อย่างรวดเร็วและได้ข้อมูลที่น่าเชื่อถือมากขึ้น

จากที่กล่าวมาข้างต้น ผู้พัฒนาจึงเลือกเทคนิคนาอีฟเบย์มาใช้ในการพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทย โดยวิเคราะห์ความคิดเห็นออกเป็นข้อความที่ชอบ ไม่ชอบ หรือรู้สึกเป็นกลางต่อละครเรื่องนั้น ๆ

2.2 เครื่องมือการจำแนกนาอีฟเบย์ (Naïve Bayes Classifier)

เป็นโมเดลการจำแนกประเภทข้อมูลเป็นกลุ่มหรือคลาส (class) โดยใช้หลักความน่าจะเป็นตามทฤษฎีของเบย์ (Bayes's theorem) [8] มีสมมติฐานให้การเกิดเหตุการณ์ต่าง ๆ เป็นอิสระต่อกัน (independence) ซึ่งเป็นทฤษฎีที่พูดถึงความน่าจะเป็นในการเกิดสิ่งหนึ่งก็ต่อเมื่ออีกสิ่งที่ได้เกิดขึ้น โดยการจำแนกจะมีค่าความแม่นยำเพิ่มขึ้นเรื่อย ๆ หากมีข้อมูลที่นำมาเรียนรู้มากขึ้นและครบถ้วนทุกความน่าจะเป็น

จากทฤษฎีของเบย์ สามารถคำนวณความน่าจะเป็นของสมมติฐานต่าง ๆ โดยใช้สมการตามภาพที่ 2.1

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

ภาพที่ 2.1 สมการความน่าจะเป็นตามทฤษฎีของเบย์

โดยที่ D แทนข้อมูลที่นำมาใช้ในการคำนวณการแจกแจงความน่าจะเป็น 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 เมื่อทราบ D

P(D|h) คือ ความน่าจะเป็นของ D เมื่อทราบ h

2.3 ภาษาและเครื่องมือที่ใช้

ภาษาหลักที่ใช้สำหรับพัฒนาโครงงานนี้คือ ไพทอน (Python) เป็นภาษาสคริปต์ทำให้ใช้เวลาในการเขียนและคอมไพล์ไม่มาก ทั้งยังถูกออกแบบให้มีโครงสร้างที่ไม่ซับซ้อนสามารถอ่านทำความเข้าใจโปรแกรมและเขียนได้ง่าย

ภาษาเอชทีเอ็มแอลเป็นภาษาที่ใช้ในการเขียนเว็บแอปพลิเคชัน โดยใช้แท็ก (tag) ในการกำหนดการแสดงผล

ภาษาซีเอสเอสเป็นภาษาที่ใช้ตกแต่งเอกสารที่เป็นเอชทีเอ็มแอลให้มีหน้าตา สีสัน ตามที่ต้องการ และใช้จัดส่วน layout ของเว็บ

ในโครงงานนี้ได้เลือกใช้เครื่องมือสำหรับพัฒนาระบบ ได้แก่

1. Tweepy เป็นไลบรารีซึ่งช่วยในการเชื่อมต่อไปยังทวิตเตอร์ หรือที่เรียกกันว่าทวิตเตอร์เอพีไอ สำหรับการเก็บรวบรวมข้อมูลจากทวิตเตอร์ เช่น ชื่อผู้ใช้ทวิต วันที่ที่ทวิต ข้อความที่ทวิต ฯลฯ
2. PyThaiNLP เป็นไลบรารีของภาษาไพทอนเพื่อประมวลผลภาษาธรรมชาติ โดยโครงงานนี้ใช้วิธีการตัดคำหรือเอนจิน (engine) ของไลบรารีนี้ในการตัดคำสำหรับภาษาไทย คือ Maximum Matching algorithm ซึ่งเป็นการตัดคำให้ได้คำยาวที่สุดและได้คำในประโยคน้อยที่สุด
3. Pandas เป็นไลบรารีของภาษาไพทอน ที่มีความสามารถสำหรับจัดเตรียมข้อมูลให้พร้อมสำหรับการวิเคราะห์ โดยทำข้อมูลให้เป็นแถวเป็นแนวเพื่อสะดวกต่อการจัดการข้อมูล
4. Sklearn Naive Bayes เป็นไลบรารีการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ของ Sklearn ซึ่งเป็นเครื่องมือการจำแนกกลุ่มของข้อมูล พัฒนาโดย Devid Cournapeau
5. Flask web development คือ เว็บเฟรมเวิร์ก (web framework) สำหรับการพัฒนาเว็บแอปพลิเคชัน ซึ่ง Flask เป็นเฟรมเวิร์กขนาดเล็ก (micro-framework) ที่สะดวกและง่ายต่อการใช้งานสำหรับผู้พัฒนาเว็บ (web developer)

2.4 คลังคำศัพท์แสดงอารมณ์ความรู้สึก

คลังคำศัพท์แสดงอารมณ์ความรู้สึก [3] [4] [5] สร้างขึ้นโดย นายวรรณพงษ์ ภัททิยไพบูลย์ โดยเก็บรวบรวมข้อความที่ใช้สำหรับการวิเคราะห์ข้อความแสดงความคิดเห็นภาษาไทย โดยมีการใช้แบบฟอร์มผ่านทางกูเกิลฟอร์มเพื่อที่จะเก็บรวบรวมกลุ่มคำผ่านทางผู้คนที่สนใจในเรื่องนี้ มีการศึกษากลุ่มคำด้านบวก (positive) ด้านลบ (negative) และเป็นกลาง (neutral) มีคำในคลังคำศัพท์ทั้งหมด 1,063 คำ

บทที่ 3
การรวบรวมและวิเคราะห์ข้อมูล

ในบทนี้จะกล่าวถึงการรวบรวมข้อความแสดงความคิดเห็นเกี่ยวกับละครไทยจากทวิตเตอร์ คลังคำศัพท์ที่สร้างขึ้นสำหรับการวิเคราะห์ และการวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทยเพื่อจำแนกความรู้สึกที่เป็นบวก กลาง และลบ ดังรายละเอียดต่อไปนี้

3.1 การรวบรวมข้อมูล

ผู้พัฒนาได้เก็บรวมรวมข้อความแสดงความคิดเห็นต่อละครไทยผ่านทางทวิตเตอร์ ซึ่งเป็นเครือข่ายสังคมออนไลน์ที่นิยมในปัจจุบัน ผู้ใช้งานสามารถแสดงความคิดเห็นเกี่ยวกับละครไทยได้ง่าย จึงทำให้มีข้อความแสดงความคิดเห็นจำนวนมาก ผู้พัฒนาจึงเลือกทวิตเตอร์ในการรวบรวมข้อความแสดงความคิดเห็นเกี่ยวกับละครไทย

ลักษณะของข้อความแสดงความคิดเห็นที่ปรากฏบนทวิตเตอร์นั้น จะประกอบด้วยชื่อบัญชีผู้ใช้งานทวิตเตอร์ ข้อความแสดงความคิดเห็น และแฮชแท็ก (#) คือ สัญลักษณ์ที่ใส่ไว้หน้าคำที่ต้องการให้เป็นคำหลัก (Keywords) หรือหัวข้อที่ผู้ใช้สนใจในขณะนั้น โดยผู้ใช้งานสามารถส่งข้อความยาวไม่เกิน 280 ตัวอักษร จากที่ผู้พัฒนาศึกษาข้อความแสดงความคิดเห็นเกี่ยวกับละครไทยพบว่า ผู้ใช้งานทวิตเตอร์มักจะใช้ชื่อละครหรือชื่อตัวละครที่เป็นพระเอกหรือนางเอกเป็นแฮชแท็ก เพื่อแสดงความคิดเห็นเกี่ยวกับละครเรื่องนั้น ๆ

โครงงานนี้จึงเลือกเก็บข้อความแสดงความคิดเห็นเกี่ยวกับละครไทยจำนวน 8 เรื่อง ได้แก่ รักฉุด-ใจนายฉุกเฉิน ลิขิตรักข้ามดวงดาว มธุรสโลกันต์ รองเท้านารี เพลิงรักเพลิงแค้น ฤกษ์สังหาร เขาวานให้หนูเป็นสายลับ และดาวหลงฟ้า จากแฮชแท็กชื่อละคร ชื่อพระเอก และชื่อนางเอง โดยเก็บข้อมูลในช่วงเวลาเดือน กันยายน – ธันวาคม ปี พ.ศ. 2562 เก็บรวบรวมข้อความแสดงความคิดเห็นได้ทั้งหมดเป็นจำนวน 9,991 ข้อความ

3.2 การวิเคราะห์ข้อมูล

จากการเก็บรวมรวมข้อความแสดงความคิดเห็นเกี่ยวกับละครไทยบนทวิตเตอร์ทั้งหมด 8 เรื่อง ได้แก่ รักฉุดใจนายฉุกเฉิน ลิขิตรักข้ามดวงดาว มธุรสโลกันต์ รองเท้านารี เพลิงรักเพลิงแค้น ฤกษ์สังหาร เขาวานให้หนูเป็นสายลับ และดาวหลงฟ้า พบว่ามีข้อความด้านบวกคือชอบ ข้อความด้านลบคือไม่ชอบ

และข้อความที่เป็นกลางคือรู้สึกกลาง ๆ ต่อละคร ตัวอย่างรายละเอียดข้อมูลจำนวนข้อความชอบ ไม่ชอบ และข้อความที่เป็นกลาง ดังตารางที่ 3.1

ตารางที่ 3.1 รายละเอียดข้อมูลจำนวนข้อความชอบ ไม่ชอบ และข้อความที่เป็นกลาง

เรื่อง	ข้อความชอบ	ข้อความไม่ชอบ	ข้อความเป็นกลาง	รวม
รักฉุดใจนายฉุกเฉิน	531	192	626	1,349
ฤกษ์สังหาร	823	426	1,028	2,277
ลิขิตรักข้ามดวงดาว	540	251	548	1,339
มธุรสโลกันต์	348	64	118	530
รองเท้านารี	319	359	648	1,326
เพลิงรักเพลิงแค้น	191	427	151	769
เขาวานให้หนูเป็นสายลับ	698	125	510	1333
ดาวหลงฟ้า	291	534	243	1068
รวม	3,741	2,378	3,872	9,991

3.3 วิธีระบุคำตอบของข้อมูล

การจำแนกข้อความแสดงความคิดเห็นออกเป็นข้อความแสดงความรู้สึกชอบต่อละคร ข้อความแสดงความรู้สึกที่ไม่ชอบต่อละคร และข้อความแสดงความรู้สึกกลาง ๆ ต่อละคร ผู้พัฒนาตัดสินโดยพิจารณาข้อความแสดงความคิดเห็นตามรูปประโยคและความหมายของคำในประโยคนั้น ตามดุลยพินิจของผู้พัฒนา เพื่อจัดทำเฉลยคำตอบของประเภทข้อความไว้สำหรับการออกแบบระบบ ซึ่งสามารถแยกข้อความแสดงความคิดเห็นออกเป็น

- ตัวอย่างข้อความบวกที่แสดงว่าชอบ “ขอชื่นชมละครเขาวานให้หนูเป็นสายลับ สนุกมากจริง ๆ”
- ตัวอย่างข้อความลบที่แสดงว่าไม่ชอบ “ผิดหวังมากค่ะพี่ผา ฉากกระท่อมในตำนาน”
- ตัวอย่างข้อความที่เป็นกลาง ไม่แสดงความรู้สึก “หมอธันวา”หรือตัวอย่างข้อความที่เป็นกลาง แสดงความรู้สึกกลาง ๆ “ผมขอโอกาสอีกครั้ง”

3.4 การสร้างคลังคำศัพท์แสดงอารมณ์ความรู้สึก (Sentiment corpus)

คลังคำศัพท์แสดงอารมณ์ความรู้สึกที่สร้างขึ้นในโครงงานนี้ สร้างจากการนับความถี่ของคำที่พบในข้อความแสดงความคิดเห็นที่เก็บรวบรวมได้ทั้งหมด และเลือกคำจากการศึกษากลุ่มคำด้านบวก (positive) ด้านลบ (negative) และเป็นกลาง (neutral) จากการพบคำเหล่านี้ในประโยคที่เกี่ยวกับ

ข้อความแสดงความคิดเห็นต่อละคร โดยแบ่งออกเป็นคลังคำศัพท์กลุ่มคำด้านบวก และคลังคำศัพท์กลุ่มคำด้านลบ รวมทั้งหมด 272 คำ

- ตัวอย่างกลุ่มคำด้านบวก ดี โคตรสนุก สุดยอด น่ารัก ดีใจ ชอบจัง หยุดยิ้มไม่ได้
- ตัวอย่างกลุ่มคำด้านลบ แย่ ไม่ชอบ ไม่น่ารัก เสียความรู้สึก ทุเรศ เสื่อม เบื่อละคร

เมื่อเปรียบเทียบกับคลังคำศัพท์แสดงอารมณ์ความรู้สึก [3] [4] [5] ที่สร้างขึ้นโดยนายวรรณพงษ์ ภัททิยไพบูลย์ พบว่า จากคลังคำศัพท์แสดงอารมณ์ความรู้สึกที่สร้างขึ้นจำนวน 272 คำ มีคำที่เหมือนกันกับคลังคำศัพท์แสดงอารมณ์ความรู้สึก [3] [4] [5] จำนวน 81 คำ และคำที่แตกต่างจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกเดิม จำนวน 191 คำ เช่น ละมุน แซ่บ หมั่นเขี้ยว เสื่อม ทุเรศ

3.5 การสร้างคลังคำศัพท์ที่ใช้ในการวิเคราะห์ (Bag of words)

การสร้างคลังคำศัพท์ที่ใช้ในการวิเคราะห์ในโครงงานนี้แบ่งออกเป็น 3 แบบตามคุณลักษณะของข้อมูลที่ใช้ในการเรียนรู้ด้วยเครื่อง ได้แก่

- คำทุกคำที่พบในข้อความนำมาสร้างเป็นคลังคำศัพท์
- คำในข้อความที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึก นำมาสร้างเป็นคลังคำศัพท์
- คำที่ไม่ซ้ำกันทุกคำที่พบในข้อความและคำในข้อความที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึก นำมาสร้างเป็นคลังคำศัพท์

3.6 การทดลองเทคนิคการเรียนรู้ด้วยเครื่องในการจำแนกข้อความ

จากตัวอย่างข้อความแสดงความคิดเห็นทั้งหมด 1,999 ข้อความที่เก็บรวบรวมได้ในช่วงแรกของการพัฒนาโครงงาน ทดลองใช้การจำแนกอารมณ์และความรู้สึกของข้อความด้วยเทคนิคการเรียนรู้ด้วยเครื่องที่แตกต่างกัน โดยการใช้คำในประโยคทั้งหมดเป็นข้อมูลคลังคำศัพท์และใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์ (รายละเอียดจะอธิบายในบทที่ 4 หัวข้อ 4.3 และ 4.4 ตามลำดับ)

มีการแสดงผลลัพธ์ในรูปแบบของตารางคอนฟิวชัน (confusion matrix) ดังตารางที่ 3.2 ซึ่งคือตารางที่แสดงผลของการจำแนกข้อความแบ่งตามคลาส โดยตารางจะมีขนาด C x C โดย C คือคลาสที่ต้องการจำแนก ในโครงงานนี้จะแบ่งข้อความออกเป็น 3 คลาส คือ 1. คลาสข้อความชอบ 2. คลาสข้อความไม่ชอบ 3. คลาสข้อความเป็นกลาง

ตารางที่ 3.2 ตารางคอนฟิวชัน

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	**กลาง**	**ชอบ**
ข้อมูลจริง	**ไม่ชอบ**	ก	ข	ค
	กลาง	ง	จ	ฉ
	ชอบ	ช	ซ	ฌ

ความหมายของค่าต่าง ๆ ในตาราง มีดังนี้

ก คือ จำนวนของข้อมูลจริงที่เป็นข้อความไม่ชอบและแบบจำลองตอบว่าไม่ชอบ

ข คือ จำนวนของข้อมูลจริงที่เป็นข้อความไม่ชอบและแบบจำลองตอบว่ากลาง

ค คือ จำนวนของข้อมูลจริงที่เป็นข้อความไม่ชอบและแบบจำลองตอบว่าชอบ

ง คือ จำนวนของข้อมูลจริงที่เป็นข้อความกลางและแบบจำลองตอบว่าไม่ชอบ

จ คือ จำนวนของข้อมูลจริงที่เป็นข้อความกลางและแบบจำลองตอบว่ากลาง

ฉ คือ จำนวนของข้อมูลจริงที่เป็นข้อความกลางและแบบจำลองตอบว่าชอบ

ช คือ จำนวนของข้อมูลจริงที่เป็นข้อความชอบและแบบจำลองตอบว่าไม่ชอบ

ซ คือ จำนวนของข้อมูลจริงที่เป็นข้อความชอบและแบบจำลองตอบว่ากลาง

ฌ คือ จำนวนของข้อมูลจริงที่เป็นข้อความชอบและแบบจำลองตอบว่าชอบ

ในการประเมินผลประสิทธิภาพของเทคนิคการเรียนรู้ด้วยเครื่องใช้วัดจากค่าความถูกต้อง (accuracy) ค่าความแม่นยำ (precision) และค่าเรียกคืน (recall) โดยคำนวณจากสมการดังต่อไปนี้

ค่าความถูกต้อง คำนวณจาก (ก+จ+ฌ) / (ก+ข+ค+ง+จ+ฉ+ช+ซ+ฌ) * 100%

ค่าความแม่นยำข้อความไม่ชอบ คำนวณจาก ก / (ก+ง+ซ) * 100%

ค่าเรียกคืนข้อความไม่ชอบ คำนวณจาก ก / (ก+ข+ค) * 100%

ค่าความแม่นยำข้อความเป็นกลาง คำนวณจาก จ / (ข+จ+ซ) * 100%

ค่าเรียกคืนข้อความเป็นกลาง คำนวณจาก จ / (ง+จ+ฉ) * 100%

ค่าความแม่นยำข้อความชอบ คำนวณจาก ฌ / (ค+ฉ+ฌ) * 100%

ค่าเรียกคืนข้อความชอบ คำนวณจาก ฌ / (ช+ซ+ฌ) * 100%

ในโครงงานนี้ทดลองกับเทคนิคการเรียนรู้ด้วยเครื่อง 3 เทคนิค กับเซตของคุณลักษณะการเรียนรู้ด้วยเครื่อง 2 แบบ คือ การใช้คำในประโยคเป็นข้อมูลคลังคำศัพท์และการใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์ ได้แก่

- เทคนิคนาอีฟเบย์ ผลลัพธ์ในตารางที่ 3.3 -3.4 และตารางที่ 3.9-3.10
- เทคนิคการถดถอยเชิงเส้น ผลลัพธ์ในตารางที่ 3.5 -3.6 และตารางที่ 3.11-3.12
- เทคนิคเพอร์เซปตรอน ผลลัพธ์ในตารางที่ 3.7 -3.8 และตารางที่ 3.13-3.14

ตารางที่ 3.3 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเทคนิคนาอีฟเบย์โดยการใช้คำในประโยคเป็นข้อมูลคลังคำศัพท์

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	288	113	94
	กลาง	68	467	224
	ชอบ	43	134	568

ตารางที่ 3.4 ประสิทธิภาพการจำแนกข้อความด้วยเทคนิคนาอีฟเบย์โดยการใช้คำในประโยคเป็นข้อมูลคลังคำศัพท์

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	66%	72%	58%
กลาง		65%	62%
ชอบ		64%	76%

ตารางที่ 3.5 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเทคนิคการถดถอยเชิงเส้นโดยการใช้คำในประโยคเป็นข้อมูลคลังคำศัพท์

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	285	155	55
	กลาง	65	566	128
	ชอบ	49	165	531

ตารางที่ 3.6 ประสิทธิภาพการจำแนกข้อความด้วยเทคนิคการถดถอยเชิงเส้นโดยการใช้คำในประโยคเป็นข้อมูลคลังคำศัพท์

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	69%	71%	58%
กลาง		64%	75%
ชอบ		74%	71%

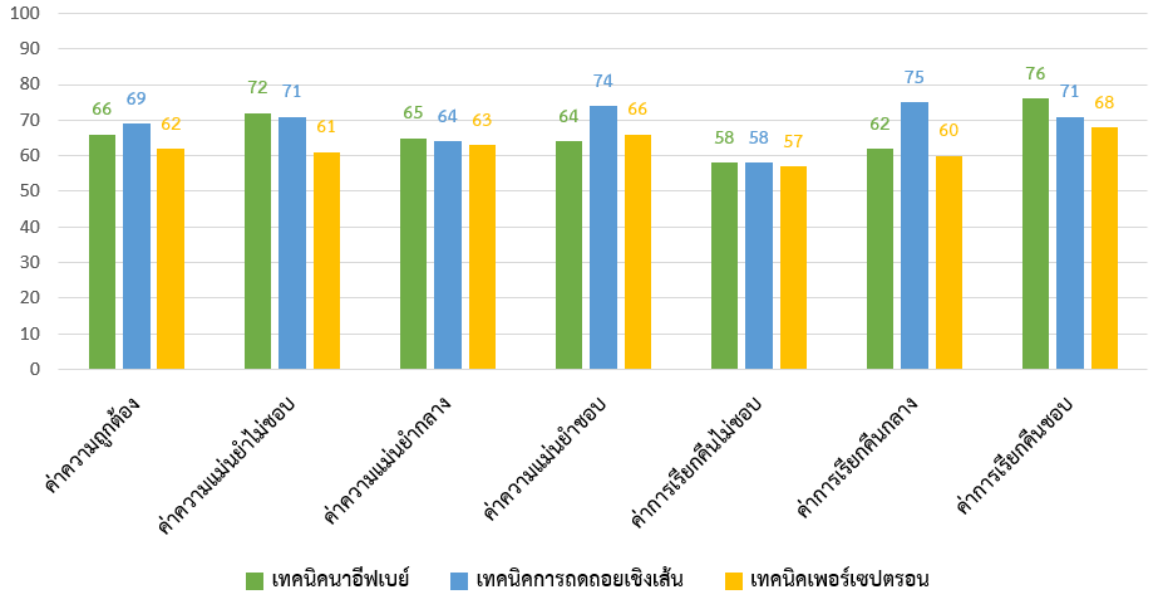
ตารางที่ 3.7 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเทคนิคเพอร์เซปตรอนโดยการใช้คำในประโยคเป็นข้อมูลคลังคำศัพท์

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	282	131	82
	กลาง	110	474	175
	ชอบ	70	165	510

ตารางที่ 3.8 ประสิทธิภาพการจำแนกข้อความด้วยเทคนิคเพอร์เซปตรอนโดยการใช้คำในประโยคเป็นข้อมูลคลังคำศัพท์

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	62%	61%	57%
กลาง		63%	60%
ชอบ		66%	68%

กราฟเปรียบเทียบประสิทธิภาพการจำแนกข้อความด้วยเทคนิคการเรียนรู้ด้วยเครื่องทั้ง 3 เทคนิค โดยการใช้คำในประโยคเป็นข้อมูลคลังคำศัพท์ แสดงดังภาพที่ 3.1



ภาพที่ 3.1 ผลลัพธ์การเปรียบเทียบประสิทธิภาพการจำแนกข้อความของเทคนิคการเรียนรู้ด้วยเครื่อง 3 เทคนิค เมื่อใช้คำในประโยคเป็นข้อมูลคลังคำศัพท์

จากภาพที่ 3.1 สรุปได้ว่า การจำแนกข้อความโดยการใช้ข้อมูลคลังคำศัพท์จากคำทั้งหมดด้วยเพอร์เซปตรอนมีประสิทธิภาพน้อยที่สุด ในขณะที่เทคนิคนาอีฟเบย์และการถดถอยเชิงเส้นมีประสิทธิภาพใกล้เคียงกัน

ตารางที่ 3.9 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเทคนิคนาอีฟเบย์โดยการใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	158	299	38
	กลาง	38	649	72
	ชอบ	35	283	427

ตารางที่ 3.10 ประสิทธิภาพการจำแนกข้อความด้วยเทคนิคนาอีฟเบย์โดยการใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	62%	68%	32%
กลาง		53%	86%
ชอบ		80%	57%

ตารางที่ 3.11 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเทคนิคถดถอยเชิงเส้นโดยการใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	183	253	59
	กลาง	61	559	99
	ชอบ	40	226	479

ตารางที่ 3.12 ประสิทธิภาพการจำแนกข้อความด้วยเทคนิคถดถอยเชิงเส้นโดยการใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	62%	64%	37%
กลาง		56%	79%
ชอบ		75%	64%

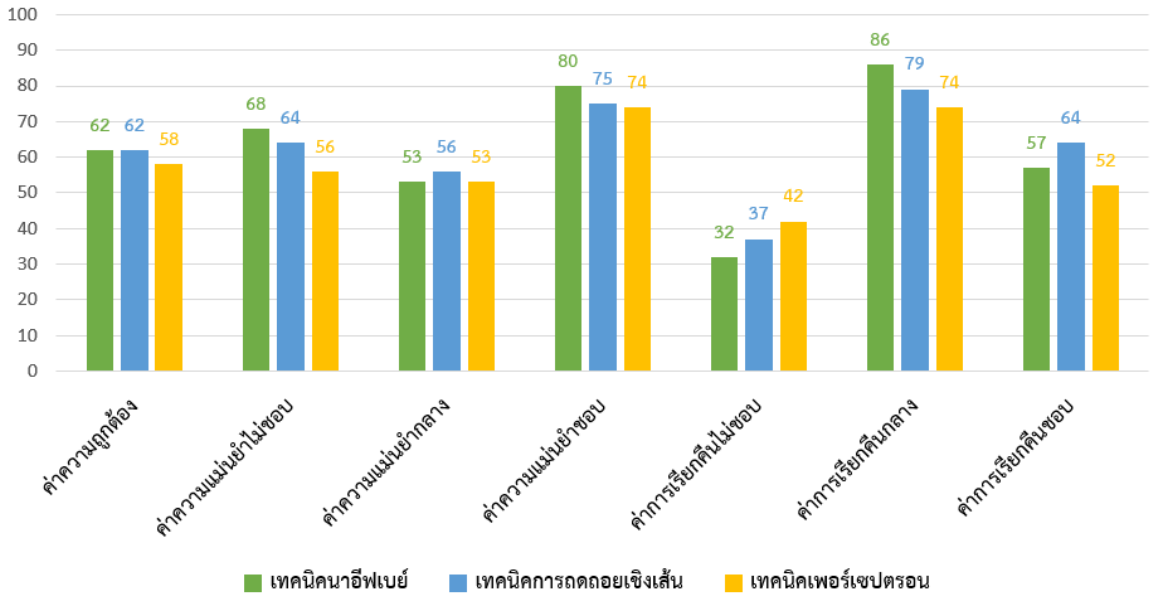
ตารางที่ 3.13 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเทคนิคเพอร์เซปตอนโดยการใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	208	240	47
	กลาง	90	581	88
	ชอบ	72	283	390

ตารางที่ 3.14 ประสิทธิภาพการจำแนกข้อความด้วยเทคนิคเพอร์เซปตอนโดยการใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	58%	56%	42%
กลาง		53%	74%
ชอบ		74%	52%

กราฟเปรียบเทียบประสิทธิภาพการจำแนกข้อความด้วยเทคนิคการเรียนรู้ด้วยเครื่องทั้ง 3 เทคนิค โดยการใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์ แสดงดังภาพที่ 3.2



ภาพที่ 3.2 ผลลัพธ์การเปรียบเทียบประสิทธิภาพการจำแนกข้อความของเทคนิคการเรียนรู้ด้วยเครื่อง 3 เทคนิค เมื่อใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์

จากภาพที่ 3.2 สรุปได้ว่า การจำแนกข้อความโดยการใช้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นข้อมูลคลังคำศัพท์ด้วยเพอร์เซปตรอนมีประสิทธิภาพน้อยที่สุด ในขณะที่เทคนิคนาอีฟเบย์และการถดถอยเชิงเส้นมีประสิทธิภาพใกล้เคียงกัน

โดยผลการประเมินประสิทธิภาพของระบบในการจำแนกข้อความไม่ชอบ เป็นกลาง และชอบ จากการจำแนกข้อความด้วยนาอีฟเบย์ เทคนิคการถดถอยเชิงเส้น และเทคนิคเพอร์เซปตรอน ทั้ง 3 เทคนิค จากการใช้ข้อมูลคลังคำศัพท์จากคำทั้งหมดและจากคำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึก พบว่า เทคนิคนาอีฟเบย์และเทคนิคการถดถอยเชิงเส้นมีค่าใกล้เคียงกัน บางค่าประสิทธิภาพเทคนิคนาอีฟเบย์มีประสิทธิภาพสูงกว่า ในขณะที่บางค่าเทคนิคการถดถอยเชิงเส้นดีกว่า จึงสรุปได้ว่าเทคนิคการถดถอยเชิงเส้นไม่ได้ดีไปกว่าเทคนิคนาอีฟเบย์ในการจำแนกข้อความเสมอไป ดังนั้นผู้พัฒนาจึงเลือกเทคนิคนาอีฟเบย์สำหรับการจำแนกข้อความแสดงความคิดเห็นในโครงงานนี้ ตามแนวทางที่กำหนดไว้เดิม

บทที่ 4
การออกแบบและพัฒนาระบบ

ในบทนี้จะกล่าวถึงการออกแบบระบบวิเคราะห์อารมณ์และความรู้สึกของข้อความแสดงความคิดเห็นเกี่ยวกับละครไทย ได้แก่ ภาพรวมของวิธีการวิเคราะห์ข้อความแสดงความคิดเห็นที่ออกแบบและการพัฒนาเว็บแอปพลิเคชัน

4.1 การออกแบบวิธีสร้างแบบจำลองเพื่อจำแนกข้อความแสดงความคิดเห็น

การออกแบบวิธีการสร้างแบบจำลองเพื่อจำแนกข้อความแสดงความคิดเห็นต่อละครไทย ใช้การเปรียบเทียบผลลัพธ์ของการจำแนกข้อมูลระหว่างแบบจำลองที่สร้างขึ้นจากคลังคำศัพท์แสดงอารมณ์ความรู้สึก การจำแนกข้อความด้วยเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ และการรวมข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกกับการทำงานของการจำแนกข้อความด้วยนาอีฟเบย์ โดยทดลองกับเซตของคุณลักษณะ (feature sets) ที่แตกต่างกัน เพื่อเลือกใช้โมเดลการจำแนกข้อมูลที่ดีที่สุดมาพัฒนาเป็นระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทย

แบบจำลองการวิเคราะห์และจำแนกข้อความแสดงความคิดเห็นในโครงการนี้ออกแบบมาทั้งหมด 3 รูปแบบ คือ

1. การใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกเพียงอย่างเดียว ใช้การนับความถี่ของคำด้านบวกและด้านลบที่ปรากฏในประโยคที่ตรงตามคลังคำศัพท์แสดงอารมณ์ความรู้สึก
2. การใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ โดยไม่ใช้ข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึก ซึ่งจะใช้เซตของคุณลักษณะที่สร้างจากคำในประโยคทุกคำ และแปลงเป็นเวกเตอร์คำของคลังคำศัพท์ทั้งหมดที่ใช้สอน (train) การเรียนรู้ด้วยเครื่อง
3. การใช้ข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ ซึ่งทดลองกับเซตของคุณลักษณะ 2 แบบ

    แบบที่ 1 ใช้เฉพาะคำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกสร้างเป็นเวกเตอร์ของคลังคำศัพท์ทั้งหมดที่ใช้สอนการเรียนรู้ด้วยเครื่อง

    แบบที่ 2 ใช้คำในประโยคทุกคำและคำที่ตรงกับคำในคลังคำศัพท์สร้างเป็นเวกเตอร์ของคลังคำศัพท์ทั้งหมดที่ใช้สอนการเรียนรู้ด้วยเครื่อง

4.2 การใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกเพียงอย่างเดียวในการจำแนกข้อความ

มอดูลที่ 1 ดึงข้อความจากทวิตเตอร์ ด้วยแฮชแท็กชื่อละครหรือตัวละคร และเก็บข้อมูลเป็น text file ดังภาพที่ 4.1

	A	B
1	created	text
2	######	#รักฉุดใจนายฉุกเฉิน หลบสปอยสุดชีวิต~~~
3	######	ใครเอาที่รักฉันมายืนยิ้มตรงนี้ ตกใจหมด นึกว่าผี □🤨🚑🏥#รักฉุดใจนายฉุกเฉิน #Myambulance https://t.co/H2Tpi3qP7g
4	######	เขียนดีสุด เหมือนแธิบายบท https://t.co/HK7YkEAeyW #รักฉุดใจนายฉุกเฉิน
5	######	โหมดจีบ " ผมขอบพี่ "โหมดแฟน " รักทานตะวัน "อุ๊ยเขิลลล□□❤🌷#ทีมน้องหมอ#รักฉุดใจนายฉุกเฉิน#MyAmbulance... https://t.co/lVZbyrraIB
6	######	ลมเย็นๆพัดมาแล้ว ฉันเพิ่งรู้สึก :)))#เวลาฟิง - Friday, 2545#รักฉุดใจนายฉุกเฉิน
7	######	" 15 ปี ยังไม่มีความหมาย นับประสาอะไรกับ 1 ปี 2 ปี 3ปี "😂😂#รักฉุดใจนายฉุกเฉิน#หมอเป้งทานตะวัน ❤😂
8	######	ละคร รักฉุดใจนายฉุกเฉินสอนให้รู้ว่า" 15 ปีที่ผ่านมาไม่มีความหมายอะไรเลย"😂😂 #ถ้ามันใช่มันใช่ไปนานแล้ว#รักฉุดใจนายฉุกเฉิน
9	######	หลายครั้งที่เราบอกเพื่อนที่อยู่ทีมเป้งว่า ให้ดูที่บทละคร อย่าดูที่ตัวนักแสดง ซันนี่เล่นก็จริง แต่บทมันไม่ใช่พระเอก... https://t.co/f3CoPUICkC
10	######	ใครที่ยังไม่มูฟออน พรุ่งนี้เรามากางกัน 20.10 กับ #รักฉุดใจนายฉุกเฉินEP17 555555 #รักฉุดใจนายฉุกเฉิน
11	######	นี่ค่าย concern กระแส anti นะเนี่ย ขนาดออกมา clarify message ที่ต้องการสื่อสารกับคนดูซีรีย์ ถ้า direction ที่วางมา... https://t.co/eJrEJTgHOx
12	######	Hi ผีหมอเต่า อุ้ย เป็นน้องหมอต่างหาก 👀 #รักฉุดใจนายฉุกเฉิน #HappyHalloween https://t.co/3sf5jABfU8
13	######	กระทู้นี้ตรงใจมาเลย เก็บตก #รักฉุดใจนายฉุกเฉิน - ความละเอียดของละครถึงความรักของทานตะวันกับฉลาม https://t.co/6TfXGj9qeN
14	######	ฝากเพลง Cover รักดิดไชเรน (Midnight Version) Ost.รักฉุดใจนายฉุกเฉิน ของฉันบน JOOX https://t.co/qc1qScAYIJ (#JOOX... https://t.co/iDnydOTo3Z
15	######	จริงด้วย พึ่งรู้เลย ว่าเด็กที่ไทเจอตอนจบ คือเด็กที่คุณพ่อเป็นโรคหัวใจ 😳 #รักฉุดใจนายฉุกเฉินตอนอวสาน... https://t.co/GOcl30H8w2
16	######	ไม่ไหวแล้วแม่น่ารักมาก 😍😍 #รักฉุดใจนายฉุกเฉิน#สกายวงศ์รวีสามีแห่งชาติ#mycloudy#ก้อนเมฆของสกาย#skywongravee @ Th... https://t.co/QDEIjs8cWs
17	######	ซึ่งต่างกับฉลามที่เข้ามาในช่วงเวลาที่ทตวต้องการใครสักคนจริงๆ คิดในแง่ของทตว เป็นเรา เราก็เลือกฉลาม. มันไม่ใช่แค่ตอน... https://t.co/j1Wzs9Bp7W
18	######	ทีมฉลามเพราะ เป้งบอกว่าอยากเป็นหมอเพราะทตว แต่พอเป็นหมอแล้วทตวเจ็บ เป้งกลับทิ้งทวตที่ต้องการเขามากที่สุดในตอนนั้น เ... https://t.co/tF4RlDQIXt
19	######	พุ่งนี้วันศุกแต่เราไม่สุข 🙁จะดูอะไรระนะ😒น้องหมอน่ารักดึกดักเวลาลงเล่นน้องหมอใจเย็นๆเวลาลงเล่นน่ารักๆ... https://t.co/tXq0J5boZe
20	######	สกาย-ปอนด์ ร่วมงาน THE MALL BANGKOK FOOD EXPO 2019 งานจัดระหว่างวันที่ 31 ตุลาคม – 3 พฤศจิกายน 2562 ณ เอ็มซีซี ฮอล... https://t.co/bkiVhod6B6
21	######	กูไม่ช้ำกับพระเอกเพราะเลือกถูก แต่กูช้ำกับเต่ามากๆ 😭😭😭#รักฉุดใจนายฉุกเฉินตอนอวสาน #รักฉุดใจนายฉุกเฉิน
22	######	So happy to meet both of you ❤□⬜🛩 #หมอฉลาม 🚁#หมอวาฬ #รักฉุดใจนายฉุกเฉิน#สกายวงศ์รวีสามีแห่งชาติ#mycloudy... https://t.co/ySis8PXm6a
23	######	นี่ไง คู่ใครคู่มันฉลาม ทานตะวัน เขาก็ถือของให้กันส่วน หมอเป้ง ก็ มองหน้า บะหมี่จบตามบทหรือเวอร์ตปท.ที่ออนยาวกว่า... https://t.co/pvNYtmqRhh
24	######	แล้วพรุ่งนี้จะดูอะไรอ่า #รักฉุดใจนายฉุกเฉิน

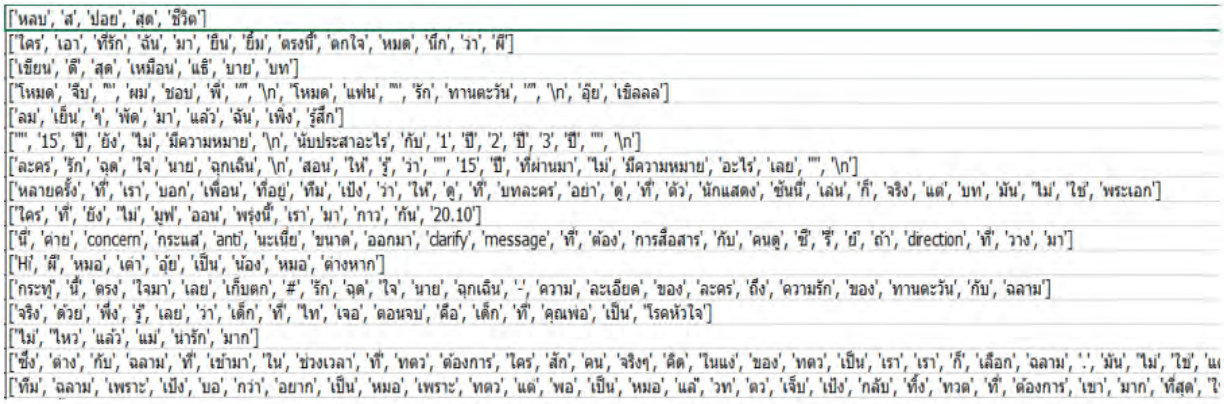
ภาพที่ 4.1 ข้อความจากทวิตเตอร์ที่ดึงออกมา

มอดูลที่ 2 ทำความสะอาดข้อมูล โดยลบข้อมูลที่ไม่เกี่ยวข้อง เช่น แฮชแท็ก ยูอาร์แอล ข้อความที่รีทวิต ดังภาพที่ 4.2

	text
2	หลบสปอยสุดชีวิต
3	ใครเอาที่รักฉันมายืนยิ้มตรงนี้ ตกใจหมด นึกว่าผี
4	เขียนดีสุด เหมือนแธิบายบท
5	โหมดจีบ " ผมขอบพี่ " โหมดแฟน " รักทานตะวัน " อุ๊ยเขิลลล
6	ลมเย็นๆพัดมาแล้ว ฉันเพิ่งรู้สึก
7	" 15 ปี ยังไม่มีความหมาย นับประสาอะไรกับ 1 ปี 2 ปี 3ปี "
8	ละคร รักฉุดใจนายฉุกเฉิน สอนให้รู้ว่า" 15 ปีที่ผ่านมาไม่มีความหมายอะไรเลย"
9	หลายครั้งที่เราบอกเพื่อนที่อยู่ทีมเป้งว่า ให้ดูที่บทละคร อย่าดูที่ตัวนักแสดง ซันนี่เล่นก็จริง แต่บทมันไม่ใช่พระเอก
10	ใครที่ยังไม่มูฟออน พรุ่งนี้เรามากางกัน 20.10
11	นี่ค่าย concern กระแส anti นะเนี่ย ขนาดออกมา clarify message ที่ต้องการสื่อสารกับคนดูซีรีย์ ถ้า direction ที่วางมา
12	Hi ผีหมอเต่า อุ้ย เป็นน้องหมอต่างหาก
13	กระทู้นี้ตรงใจมาเลย เก็บตก #รักฉุดใจนายฉุกเฉิน - ความละเอียดของละครถึงความรักของทานตะวันกับฉลาม
14	จริงด้วย พึ่งรู้เลย ว่าเด็กที่ไทเจอตอนจบ คือเด็กที่คุณพ่อเป็นโรคหัวใจ
15	ไม่ไหวแล้วแม่น่ารักมาก
16	ซึ่งต่างกับฉลามที่เข้ามาในช่วงเวลาที่ทตวต้องการใครสักคนจริงๆ คิดในแง่ของทตว เป็นเรา เราก็เลือกฉลาม. มันไม่ใช่แค่ตอน...
17	ทีมฉลามเพราะ เป้งบอกว่าอยากเป็นหมอเพราะทตว แต่พอเป็นหมอแล้วทตวเจ็บ เป้งกลับทิ้งทวตที่ต้องการเขามากที่สุดในตอนนั้น

ภาพที่ 4.2 ข้อความจากทวิตเตอร์ที่ผ่านการลบข้อมูลที่ไม่เกี่ยวข้องแล้ว

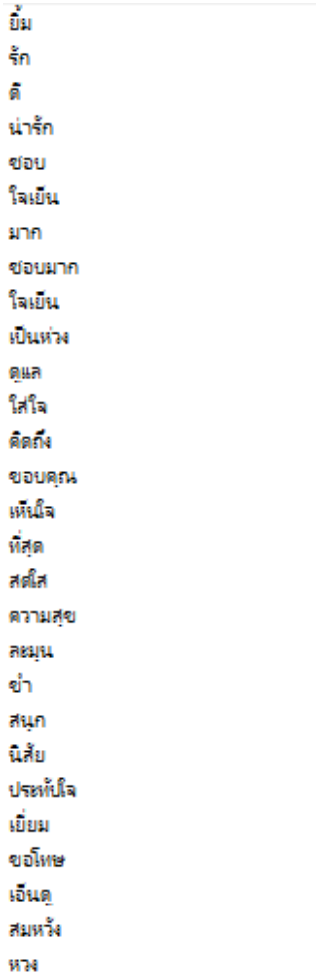
มอดูลที่ 3 การตัดคำ นำข้อมูลที่ได้จากการทำความสะอาดข้อมูลแล้วนำมาตัดประโยคออกเป็นคำ โดยใช้ไลบารี PyThaiNLP ด้วยเทคนิค Maximum Matching algorithm ในการตัดคำ ตัวอย่างดังภาพที่ 4.3



ภาพที่ 4.3 ข้อความที่ผ่านการตัดคำ

มอดูลที่ 4 การเปรียบเทียบคำกับคลังคำศัพท์แสดงอารมณ์ความรู้สึก โดยจะนำไปเปรียบเทียบกับคลังคำศัพท์แสดงอารมณ์ความรู้สึก 3 แบบ ได้แก่

1. คลังคำศัพท์แสดงอารมณ์ความรู้สึก ที่ผู้พัฒนาสร้างขึ้นเอง ในหัวข้อที่ 3.4



ภาพที่ 4.4 ตัวอย่างคลังคำศัพท์ที่ผู้พัฒนาสร้างขึ้นเอง

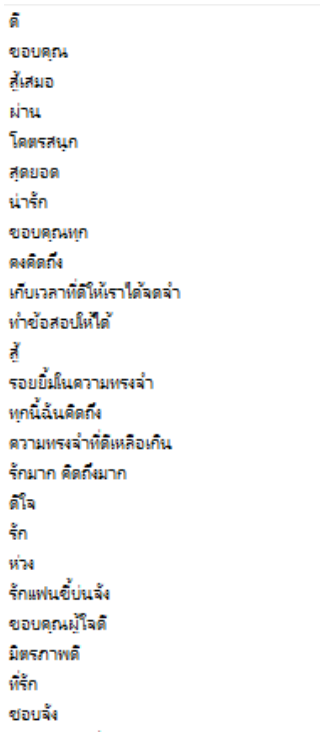
ตารางที่ 4.1 ตารางคอนฟิวชันผลลัพธ์การใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกที่ผู้พัฒนาสร้างขึ้นเอง ในการจำแนกข้อความ

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	767	1130	481
	กลาง	457	2721	694
	ชอบ	264	1263	2214

ตารางที่ 4.2 ประสิทธิภาพการใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกที่ผู้พัฒนาสร้างขึ้นเอง ในการจำแนกข้อความ

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	57%	52%	32%
กลาง		53%	70%
ชอบ		65%	59%

2. คลังคำศัพท์แสดงอารมณ์ความรู้สึก จาก [3] [4] [5]



ภาพที่ 4.5 ตัวอย่างคลังคำศัพท์แสดงอารมณ์ความรู้สึกจาก [3][4][5]

ตารางที่ 4.3 ตารางคอนฟิวชันผลลัพธ์การใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกจาก [3][4][5] ในการจำแนกข้อความ

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	601	1,539	238
	กลาง	335	3,059	478
	ชอบ	261	2,048	1,432

ตารางที่ 4.4 ประสิทธิภาพการใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกจาก [3][4][5] ในการจำแนกข้อความ

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	51%	50%	25%
กลาง		46%	79%
ชอบ		67%	38%

3. คลังคำศัพท์แสดงอารมณ์ความรู้สึกที่ผู้พัฒนาสร้างขึ้นเองรวมกับคลังคำศัพท์แสดงอารมณ์ความรู้สึก [3] [4] [5]

ยิ้ม
ฝัน
แรงใจ
ดวงใจลุกโชนความหวัง
แสงสว่างส่องใจ
การชำระจิตหรือใจให้ผ่องใสสิ่งสำคัญที่สุดที่พึงปรารถนาที่สุดของผู้ปัญญาทั้งหลาย
ทำจิตให้สงบจะพบทางออก
ความพร้อมอยู่ที่เรา
ลงมือทำอย่างต่อเนื่องเท่านั้นเอง
ไม่ถูกตัวเอง
ต้องเชื่อในศักยภาพของคุณและความพยายามที่
นิสิงล้ำค่า
ห่วงใย
สันติภาพสิ่งที่ดาดหวัง
หยิบยื่นโอกาส
สานต่อความฝันให้จริง
ไม่ต้องกังวลใจ
ทำได้ดีกว่านี้อีก
อยากให้อ่านกำลังใจขึ้น

ภาพที่ 4.6 คลังคำศัพท์แสดงอารมณ์ความรู้สึกที่ผู้พัฒนาสร้างขึ้นเองรวมกับคลังคำศัพท์แสดงอารมณ์ความรู้สึก [3] [4] [5]

ตารางที่ 4.5 ตารางคอนฟิวชันผลลัพธ์การใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกที่ผู้พัฒนาสร้างขึ้นเองรวมกับคลังคำศัพท์แสดงอารมณ์ความรู้สึก [3] [4] [5] ในการจำแนกข้อความ

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	958	971	449
	กลาง	568	2571	733
	ชอบ	322	1139	2280

ตารางที่ 4.6 ประสิทธิภาพการใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกที่ผู้พัฒนาสร้างขึ้นเองรวมกับคลังคำศัพท์แสดงอารมณ์ความรู้สึก [3] [4] [5] ในการจำแนกข้อความ

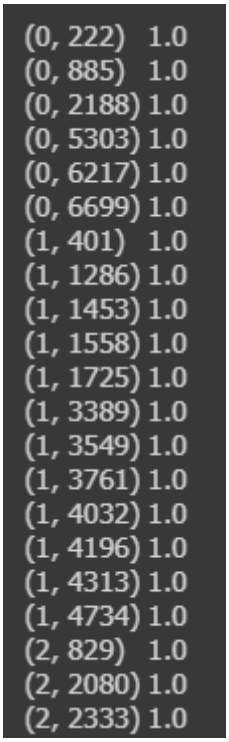
	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	58%	52%	40%
กลาง		55%	66%
ชอบ		66%	61%

จากประสิทธิภาพผลการจำแนกข้อความในตารางที่ 4.2, 4.4 และ 4.6 พบว่าประสทธิภาพการใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกที่ผู้พัฒนาสร้างขึ้นเองรวมกับคลังคำศัพท์แสดงความรู้สึก [3] [4] [5] มีค่าความถูกต้องอยู่ที่ 58% ซึ่งมากกว่าวิธีอื่น รองลงมาคือการใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกที่ผู้พัฒนาสร้างขึ้นเอง ในการจำแนกข้อความมีค่าความถูกต้องอยู่ที่ 57% และอันดับสุดท้ายคือการใช้คลังคำศัพท์แสดงอารมณ์ความรู้สึกจาก [3][4][5] ในการจำแนกข้อความ มีค่าความถูกต้องอยู่ที่ 51%

4.3 การใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ โดยไม่ใช้ข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึก (เซตของคุณลักษณะแบบที่ 1)

มอดูลที่ 1 - มอดูลที่ 3 ทำงานเหมือนในหัวข้อ 4.2

มอดูลที่ 4 การสร้างข้อมูลที่ใช้สอน ใช้วิธี One-hot ให้คำของข้อความทั้งหมดเป็นคุณลักษณะ โดยสร้างคลังคำศัพท์จากคำทั้งหมด และสร้างเวกเตอร์ของข้อความเทียบกับคำในคลังคำศัพท์ ถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0



ภาพที่ 4.7 การสร้างเวกเตอร์ของข้อความ

มอดูลที่ 5 การสร้างโมเดลการจำแนกข้อมูลโดยการใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ เพื่อจำแนกข้อความออกเป็น 3 กลุ่ม คือ ชอบ ไม่ชอบ และกลาง

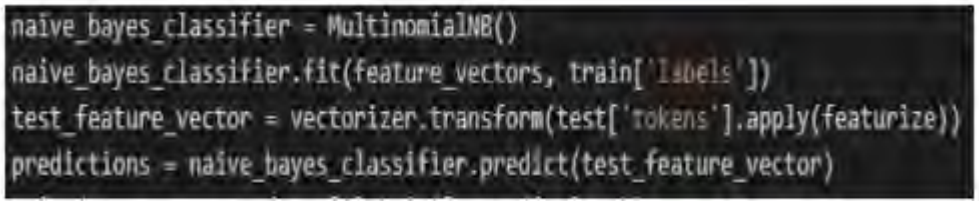
ภาพที่ 4.8 - ภาพที่ 4.10 แสดงโค้ดการทำงานในการจำแนกข้อความโดยการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ โดยไม่ใช้ข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึก

```
from pythainlp import word_tokenize
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction import DictVectorizer
from pythainlp.corpus import thai_stopwords
import pandas as pd
import numpy as np
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, classification_report
```

ภาพที่ 4.8 โค้ดการเรียกใช้ไลบรารี

```
vectorizer = DictVectorizer()
def featurize(token_list):
    features = {}
    for token in token_list:
        features[token] = 1
    return features
feature_vectors = vectorizer.fit_transform(train['tokens'].apply(featurize))
```

ภาพที่ 4.9 โค้ดการสร้างคุณลักษณะ



```
naive_bayes_classifier = MultinomialNB()
naive_bayes_classifier.fit(feature_vectors, train['labels'])
test_feature_vector = vectorizer.transform(test['tokens'].apply(featurize))
predictions = naive_bayes_classifier.predict(test_feature_vector)
```

ภาพที่ 4.10 โค้ดการสร้างโมเดล

ตารางที่ 4.7 - ตารางที่ 4.8 แสดงตัวอย่างผลลัพธ์การจำแนกข้อความโดยการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ โดยไม่ใช้ข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึก ซึ่งใช้การแบ่งข้อมูลเป็น -ข้อมูลสอน (training data) 80% และ -ข้อมูลทดสอบ (testing data) 20% ดังนั้นจากข้อมูลทั้งหมดจำนวน 9,991 ข้อความ จะเลือกข้อมูลทดสอบจำนวน 1,999 ข้อความ ประกอบด้วยข้อความแสดงความรู้สึกชอบต่อละคร 745 ข้อความ ข้อความแสดงความรู้สึกไม่ชอบต่อละคร 759 ข้อความ และข้อความแสดงความรู้สึกที่เป็นกลางต่อละคร 495 ข้อความ

ตารางที่ 4.7 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 1

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	288	113	94
	กลาง	68	467	224
	ชอบ	43	134	568

ตารางที่ 4.8 ประสิทธิภาพการจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 1

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	66%	72%	58%
กลาง		65%	62%
ชอบ		64%	76%

4.4 การใช้ข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ (เซตของคุณลักษณะแบบที่ 2)

มอดูลที่ 1 - มอดูลที่ 3 ทำงานเหมือนในหัวข้อ 4.3

มอดูลที่ 4 - มอดูลที่ 5 การสร้างข้อมูลที่ใช้สอน ใช้วิธี One-hot ให้ค่าของข้อความที่ตรงกับคำคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นคุณลักษณะ โดยสร้างคลังคำศัพท์จากคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึก และสร้างเวกเตอร์ของข้อความเทียบกับคำในคลังคำศัพท์ ถ้ามีคำอยู่ใน ข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0

```
(0, 34)    1.0
(1, 77)    1.0
(1, 207)   1.0
(2, 31)    1.0
(3, 3)     1.0
(4, 158)   1.0
(5, 166)   1.0
(6, 49)    1.0
(7, 59)    1.0
(8, 281)   1.0
(15, 158)  1.0
(17, 13)   1.0
(17, 158)  1.0
(18, 183)  1.0
(20, 34)   1.0
(22, 9)    1.0
```

ภาพที่ 4.11 การสร้างเวกเตอร์ของข้อความ

มอดูลที่ 6 ทำงานเหมือนในหัวข้อ 4.3 มอดูลที่ 5

ภาพที่ 4.12 - ภาพที่ 4.14 แสดงโค้ดการทำงานในการจำแนกข้อความโดยการใช้ข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์

```
from pythainlp import word_tokenize
from sklearn.feature_extraction import DictVectorizer
from pythainlp.corpus import thai_stopwords
import pandas as pd
import numpy as np
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, classification_report
from pandas import DataFrame
```

ภาพที่ 4.12 โค้ดการเรียกใช้ไลบรารี

```
vectorizer = DictVectorizer()
def featurize2my(token_list):
  with open("mixlex.txt", 'r',encoding = 'utf-8') as f:
    lexline = f.readlines()
    listlex = [e.strip() for e in lexline]
  features3 = {}
  for token in listlex:
    if token in token_list:
     features3[token]=1
  return features3
feature_vectors3 = vectorizer.fit_transform(train['tokens'].apply(featurize2my))
```

ภาพที่ 4.13 โค้ดการสร้างคุณลักษณะ

```
naive_bayes_classifier3 = MultinomialNB()
naive_bayes_classifier3.fit(feature_vectors3, train['labels'])
test_feature_vector3 = vectorizer.transform(test['tokens'].apply(featurize2my))
predictions3 = naive_bayes_classifier3.predict(test_feature_vector3)
```

ภาพที่ 4.14 โค้ดการสร้างโมเดล

ตารางที่ 4.9 - ตารางที่ 4.10 แสดงตัวอย่างผลลัพธ์การจำแนกข้อความโดยการใช้ข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ ซึ่งใช้ข้อมูลทดสอบข้อมูลเดียวกับในหัวข้อ 4.3 จำนวน 1,999 ข้อความ

ตารางที่ 4.9 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 2

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	158	299	38
	กลาง	38	649	72
	ชอบ	35	283	427

ตารางที่ 4.10. ประสิทธิภาพการจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 2

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	62%	68%	32%
กลาง		53%	86%
ชอบ		80%	57%

4.5 การใช้คำในข้อความและข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ (เซตของคุณลักษณะแบบที่ 3)

มอดูลที่ 1 – มอดูลที่ 3 ทำงานเหมือนในหัวข้อ 4.4

มอดูลที่ 4 - มอดูลที่ 5 การสร้างข้อมูลที่ใช้สอน

- คำในข้อความใช้วิธี One-hot ให้คำที่ไม่ซ้ำกันแต่ละคำของข้อความทั้งหมดเป็นคุณลักษณะ โดยสร้างคลังคำศัพท์จากคำที่ไม่ซ้ำกัน และสร้างเวกเตอร์ของข้อความเทียบกับคำในคลังคำศัพท์ ถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0

- คำในข้อความที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกใช้วิธี One-hot ให้คำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นคุณลักษณะ โดยสร้างคลังคำศัพท์จากคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึก และสร้างเวกเตอร์ของข้อความเทียบกับคำในคลังคำศัพท์ ถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0

```
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]
```

ภาพที่ 4.15 การสร้างเวกเตอร์ของข้อความ

มอดูลที่ 6 ทำงานเหมือนในหัวข้อ 4.3 มอดูลที่ 5

ภาพที่ 4.16 - ภาพที่ 4.17 แสดงโค้ดการทำงานในการจำแนกข้อความโดยการใช้คำในข้อความและข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์

```
vectorizer = DictVectorizer()
def featurize(token_list):
    features = {}
    for token in token_list:
      features[token] = 1
    return features
feature_vectors = vectorizer.fit_transform(train['tokens'].apply(featurize))
def featurize2my(token_list):
  with open("mixlex.txt", 'r',encoding = 'utf-8') as f:
    lexline = f.readlines()
    listlex = [e.strip() for e in lexline]
  features3 = {}
  for token in listlex:
    if token in token_list:
     features3[token]=1
  return features3
feature_vectors3 = vectorizer.fit_transform(train['tokens'].apply(featurize2my))
```

ภาพที่ 4.16 โค้ดการสร้างคุณลักษณะ

```
X_train = np.concatenate([feature_vectors.toarray(),feature_vectors3.toarray()],axis=1)
X_test = np.concatenate([test_feature_vector.toarray(),test_feature_vector3.toarray()],axis=1)

naive_bayes_classifiertokenlex = MultinomialNB()
naive_bayes_classifiertokenlex.fit(X_train, train['labels'])
test_feature_vectortokenlex = vectorizer.transform(test['tokens'].apply(featuretokenlex))
predictionstokenlex = naive_bayes_classifiertokenlex.predict(X_test)
print(accuracy_score(test['labels'], predictionstokenlex))
print (classification_report(test['labels'], predictionstokenlex))
```

ภาพที่ 4.17 โค้ดการสร้างโมเดล

ตารางที่ 4.11 - ตารางที่ 4.12 แสดงตัวอย่างผลลัพธ์การจำแนกข้อความโดยการใช้ใช้คำในข้อความและข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ ซึ่งใช้ข้อมูลทดสอบข้อมูลเดียวกับในหัวข้อ 4.3 จำนวน 1,999 ข้อความ

ตารางที่ 4.11 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 3

		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
ข้อมูลจริง	ไม่ชอบ	300	114	81
	กลาง	75	486	198
	ชอบ	42	136	567

ตารางที่ 4.12 ประสิทธิภาพการจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 3

	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
ไม่ชอบ	68%	72%	61%
กลาง		66%	64%
ชอบ		67%	76%

จากการเปรียบเทียบผลการจำแนกข้อความแสดงความคิดเห็นที่มีต่อละครไทยด้วยแบบจำลองที่สร้างขึ้น 4 รูปแบบตามหัวข้อที่ 4.2 – 4.5 สรุปผลได้ว่า แบบจำลองตามหัวข้อ 4.5 การใช้คำในประโยคและข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ มีค่าความถูกต้อง 68% ซึ่งมีค่ามากกว่าแบบจำลองอื่น จึงนำแบบจำลองที่ดีที่สุดไปสร้างเป็นเว็บแอปพลิเคชันการจำแนกข้อความแสดงความคิดเห็น

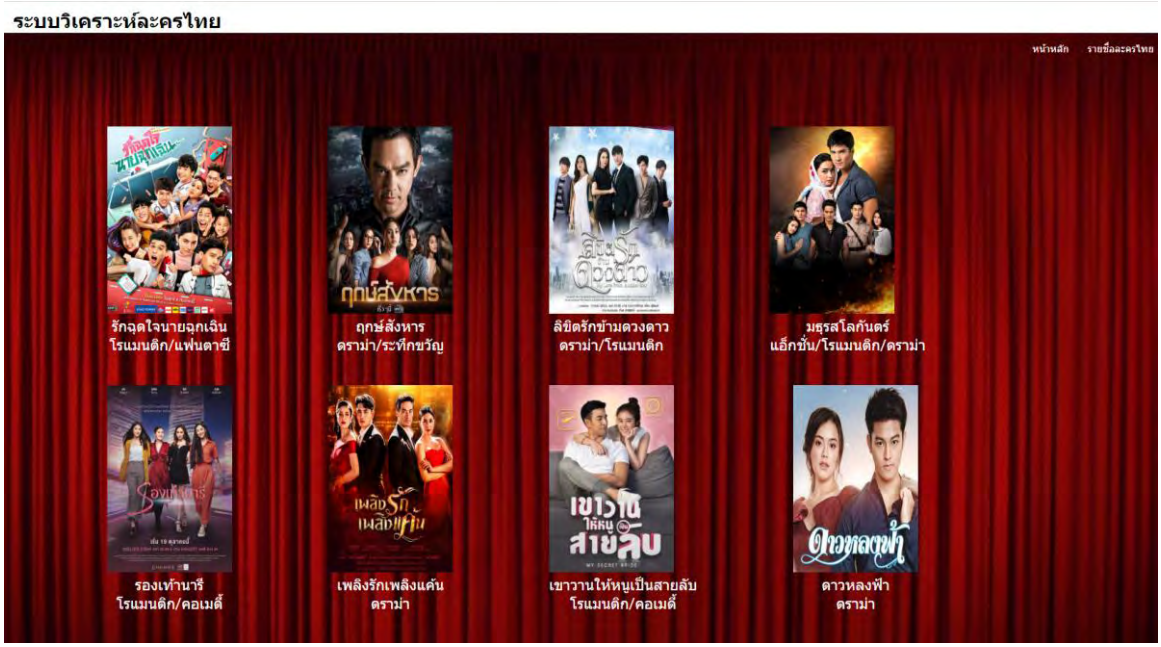
4.6 การออกแบบเว็บแอปพลิเคชัน

การออกแบบเว็บแอปพลิเคชันสำหรับระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทยบนทวิตเตอร์ ผู้พัฒนาได้เลือก Flask เป็น framework ใช้ภาษาไพทอน ภาษาเอชทีเอ็มแอล ภาษาซีเอสเอสในการพัฒนา

4.6.1 ส่วนต่อประสานกับผู้ใช้

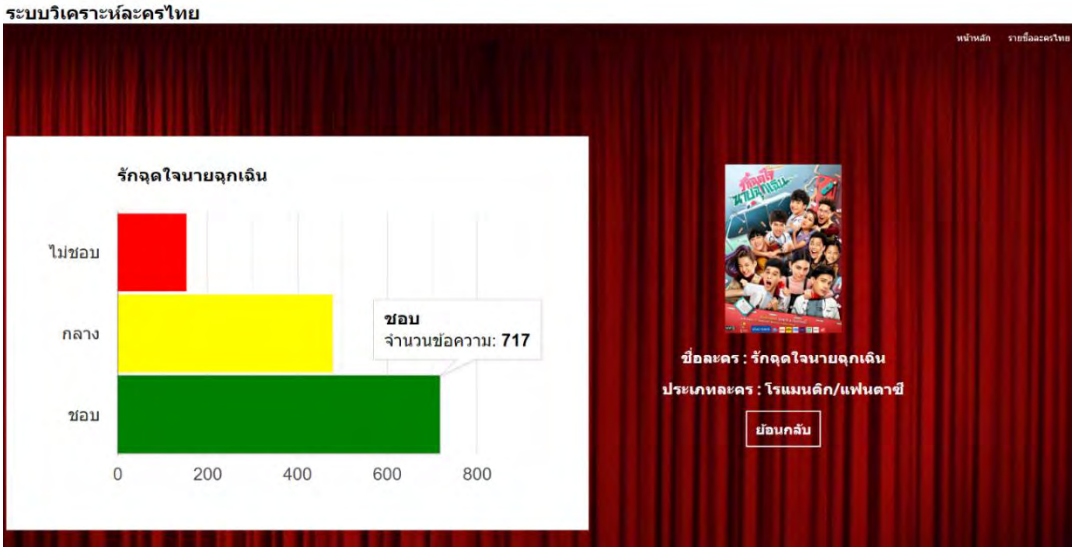
ภาพที่ 4.18 หน้าจอการทำงานเริ่มต้นของเว็บแอปพลิเคชัน

จากภาพที่ 4.18 เมื่อผู้ใช้เข้าสู่เว็บแอปพลิเคชัน กดเลือก “รายชื่อละครไทย” ผู้ใช้จะสามารถเปลี่ยนหน้าเพื่อดูรายละเอียดของละครแต่ละเรื่องในหน้าถัดไปดังภาพที่ 4.19



ภาพที่ 4.19 หน้าจอการทำงานของระบบ

จากภาพที่ 4.19 เมื่อผู้ใช้กดเลือกรูปละครที่สนใจ ผู้ใช้จะสามารถดูรายละเอียดผลการวิเคราะห์ของละครแต่ละเรื่องได้ มีการแสดงผลการวิเคราะห์ข้อความแสดงความรู้สึกของข้อความในรูปแบบของกราฟแท่ง โดยจำแนกข้อมูลเป็นชอบสีเขียว ไม่ชอบสีแดง และความรู้สึกที่เป็นกลางสีเหลือง พร้อมกับจำนวนข้อความที่ถูกจำแนกออกเป็นชอบ ไม่ชอบ และกลาง ดังภาพที่ 4.20



ภาพที่ 4.20 หน้าจอแสดงผลลัพธ์แบบกราฟแท่ง

บทที่ 5
การทดสอบระบบ

ในบทนี้จะกล่าวถึง การทดสอบโมเดลการวิเคราะห์ข้อความแสดงความคิดเห็นที่ออกแบบและพัฒนาในบทที่ 4 โดยเลือกโมเดลที่มีผลการทดสอบที่ดีที่สุดไปสร้างเป็นเว็บแอปพลิเคชัน โดยพิจารณาจากค่าความถูกต้อง ค่าความแม่นยำ และค่าการเรียกคืน รายละเอียดดังต่อไปนี้

5.1 บทนำ

ผู้พัฒนาได้เก็บรวมรวมข้อความแสดงความคิดเห็นต่อละครไทยผ่านทางทวิตเตอร์จำนวน 9,991 ข้อความ จากละครไทยทั้งหมด 8 เรื่อง ได้แก่ รักฉุดใจนายฉุกเฉิน ลิขิตรักข้ามดวงดาว มธุรสโลกันต์ รองเท้านารี เพลิงรักเพลิงแค้น ฤกษ์สังหาร เขาวานให้หนูเป็นสายลับ และดาวหลงฟ้า จากแฮชแท็กชื่อละคร ชื่อพระเอก และชื่อนางเอก โดยเก็บข้อมูลในช่วงเวลาเดือน กันยายน – ธันวาคม ปี พ.ศ. 2562 ซึ่งจำแนกกลุ่มเป็นข้อความแสดงความรู้สึกชอบต่อละคร 3,741 ข้อความ ข้อความแสดงความรู้สึกไม่ชอบต่อละคร 2,378 ข้อความ และข้อความแสดงความรู้สึกที่เป็นกลางต่อละคร 3,872 ข้อความ รวมทั้งหมด 9,991 ข้อความ

5.2 ผลการทดสอบระบบ

5.2.1 การทดสอบระบบด้วยแบบจำลองที่สร้างขึ้นในหัวข้อที่ 4.5 แยกตามแต่ละเรื่อง

ซึ่งเป็นการใช้คำในข้อความและข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ (เซตของคุณลักษณะแบบที่ 3) ทดสอบกับข้อความแสดงความคิดเห็นทั้งหมดที่เก็บรวบรวมได้จำนวน 9,991 ข้อความ แยกตามละครแต่ละเรื่อง คือ

รักฉุดใจนายฉุกเฉิน	จำนวน 1,349 ข้อความ
ลิขิตรักข้ามดวงดาว	จำนวน 1,339 ข้อความ
มธุรสโลกันต์	จำนวน 530 ข้อความ
รองเท้านารี	จำนวน 1,326 ข้อความ
เพลิงรักเพลิงแค้น	จำนวน 769 ข้อความ
ฤกษ์สังหาร	จำนวน 2,277 ข้อความ
เขาวานให้หนูเป็นสายลับ	จำนวน 1,333 ข้อความ
ดาวหลงฟ้า	จำนวน 1,068 ข้อความ

ตารางที่ 5.1 และ 5.2 แสดงผลลัพธ์การจำแนกข้อความจากละครทั้ง 8 เรื่อง

ตารางที่ 5.1 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความของละครทั้ง 8 เรื่อง

ข้อมูลจริง		ข้อมูลที่แบบจำลองตอบ		
		ไม่ชอบ	กลาง	ชอบ
รักฉุดใจนายฉุกเฉิน	ไม่ชอบ	80	44	68
	กลาง	53	355	218
	ชอบ	20	80	431
ลิขิตรักข้ามดวงดาว	ไม่ชอบ	87	106	58
	กลาง	23	409	116
	ชอบ	20	98	422
ฤกษ์สังหาร	ไม่ชอบ	190	176	60
	กลาง	108	719	201
	ชอบ	79	278	466
รองเท้านารี	ไม่ชอบ	171	118	70
	กลาง	65	432	151
	ชอบ	29	69	221
มธุรสโลกันต์	ไม่ชอบ	21	28	15
	กลาง	6	74	38
	ชอบ	20	92	236
เพลิงรักเพลิงแค้น	ไม่ชอบ	225	123	79
	กลาง	15	89	47
	ชอบ	11	38	142
ดาวหลงฟ้า	ไม่ชอบ	321	85	128
	กลาง	61	108	74
	ชอบ	22	33	236
เขาวานหนูเป็นสายลับ	ไม่ชอบ	55	48	22
	กลาง	33	339	138
	ชอบ	19	137	542

ตารางที่ 5.2 ประสิทธิภาพการจำแนกข้อความของละครทั้ง 8 เรื่อง

		ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
รักฉุดใจนายฉุกเฉิน	ไม่ชอบ	64%	52%	42%
	กลาง		74%	57%
	ชอบ		60%	81%
ลิขิตรักข้ามดวงดาว	ไม่ชอบ	69%	67%	35%
	กลาง		67%	75%
	ชอบ		71%	78%
ฤกษ์สังหาร	ไม่ชอบ	60%	50%	45%
	กลาง		61%	70%
	ชอบ		64%	57%
รองเท้านารี	ไม่ชอบ	62%	65%	48%
	กลาง		70%	67%
	ชอบ		50%	69%
มธุรสโลกันต์	ไม่ชอบ	62%	45%	33%
	กลาง		38%	62%
	ชอบ		82%	68%
เพลิงรักเพลิงแค้น	ไม่ชอบ	59%	90%	53%
	กลาง		36%	59%
	ชอบ		53%	74%
ดาวหลงฟ้า	ไม่ชอบ	62%	79%	60%
	กลาง		48%	44%
	ชอบ		54%	81%
เขาวานหนูเป็นสายลับ	ไม่ชอบ	70%	51%	44%
	กลาง		65%	66%
	ชอบ		77%	78%

จากผลการทดสอบการจำแนกข้อความของละครทั้ง 8 เรื่อง พบว่า ค่าความถูกต้องของเรื่องเขาวานหนูเป็นสายลับมีค่ามากที่สุด 70% อันดับที่ 2 คือลิขิตรักข้ามดวงดาว 69% และเรื่องอื่น ๆ มีค่าความถูกต้องอยู่ระหว่าง 59% - 64% ซึ่งใกล้เคียงกัน ในขณะที่ค่าความแม่นยำและค่าการเรียกคืนของข้อความ

ไม่ชอบ กลาง และชอบ สำหรับละครแต่ละเรื่องมีความแตกต่างกัน เช่น เรื่องเพลิงรักเพลิงแค้นมีค่าความแม่นยำของข้อความไม่ชอบสูงถึง 90% แต่เรื่องมธุรสโลกันต์มีค่าเพียง 45% เรื่องรักฉุดใจนายฉุกเฉินกับเรื่องดาวหลงฟ้ามีค่าการเรียกคืนของข้อความชอบมากถึง 81% แต่เรื่องฤกษ์สังหารมีค่าแค่ 57%

5.2.2 การทดสอบระบบด้วยการสร้างแบบจำลองตามในบทที่ 4

ซึ่งประกอบด้วยแบบจำลอง 3 แบบ คือ

1. การใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ โดยไม่ใช้ข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึก (เซตของคุณลักษณะแบบที่ 1) ในหัวข้อที่ 4.3 ผลลัพธ์แสดงในตารางที่ 5.3 และตารางที่ 5.4
2. การใช้ข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ (เซตของคุณลักษณะแบบที่ 2) ในหัวข้อที่ 4.4 ผลลัพธ์แสดงในตารางที่ 5.5 และตารางที่ 5.6
3. การใช้คำในข้อความและข้อมูลจากคลังคำศัพท์แสดงอารมณ์ความรู้สึกร่วมกับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ (เซตของคุณลักษณะแบบที่ 3) ในหัวข้อที่ 4.5 ผลลัพธ์แสดงในตารางที่ 5.7 และตารางที่ 5.8

โดยทดสอบกับข้อความแสดงความคิดเห็นทั้งหมดที่เก็บรวบรวมได้จำนวน 9,991 ข้อความ ด้วยเทคนิค 5-fold cross validation แบ่งข้อความทั้งหมดเป็น 5 ส่วน ส่วนละ 20% แต่ละส่วนใช้เป็นข้อมูลทดสอบกับแบบจำลองการจำแนกข้อความที่สร้างจากการสอนด้วยข้อมูลอีก 4 ส่วน (ผลลัพธ์การจำแนกข้อความแต่ละส่วนแสดงรายละเอียดในภาคผนวก ค) และทดลองแบ่งส่วนข้อมูลที่แตกต่างกันด้วยวิธีการสุ่มให้ครบ 5 แบบ

ตารางที่ 5.3 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 1

การสุ่มแบ่งข้อมูล							
แบบที่ 1	แบบจำลองตอบ			แบบที่ 2	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	1,371	537	410	ไม่ชอบ	1,349	600	429
กลาง	426	2,388	1,058	กลาง	409	2,403	1,060
ชอบ	201	680	2,860	ชอบ	201	675	2,865
แบบที่ 3	แบบจำลองตอบ			แบบที่ 4	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	1,362	594	422	ไม่ชอบ	1,363	593	422
กลาง	428	2,375	1,069	กลาง	427	2,378	1,067
ชอบ	196	695	2,850	ชอบ	215	665	2,861
แบบที่ 5	แบบจำลองตอบ						
	ไม่ชอบ	กลาง	ชอบ				
ไม่ชอบ	1,363	588	427				
กลาง	428	2,372	1,072				
ชอบ	200	696	2,845				

ตารางที่ 5.4 ประสิทธิภาพการจำแนกข้อความเซตของคุณลักษณะแบบที่ 1

		ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
การสุ่มแบ่งข้อมูลแบบที่ 1	ไม่ชอบ	66%	69%	58%
	กลาง		65%	62%
	ชอบ		66%	76%
การสุ่มแบ่งข้อมูลแบบที่ 2	ไม่ชอบ	67%	69%	57%
	กลาง		65%	62%
	ชอบ		66%	77%
การสุ่มแบ่งข้อมูลแบบที่ 3	ไม่ชอบ	66%	68%	57%
	กลาง		65%	61%
	ชอบ		66%	76%
การสุ่มแบ่งข้อมูลแบบที่ 4	ไม่ชอบ	66%	68%	57%
	กลาง		65%	61%
	ชอบ		66%	76%
การสุ่มแบ่งข้อมูลแบบที่ 5	ไม่ชอบ	66%	68%	57%
	กลาง		65%	61%
	ชอบ		65%	76%
ค่าเฉลี่ยของทั้ง 5 แบบ	ไม่ชอบ	66%	68%	57%
	กลาง		65%	61%
	ชอบ		66%	76%

ตารางที่ 5.5 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 2

การสุ่มแบ่งข้อมูล							
แบบที่ 1	แบบจำลองตอบ			แบบที่ 2	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	815	1,369	194	ไม่ชอบ	805	1,375	198
กลาง	195	3,294	383	กลาง	201	3,297	374
ชอบ	165	1,533	2,043	ชอบ	167	1,540	2,034
แบบที่ 3	แบบจำลองตอบ			แบบที่ 4	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	804	1,384	190	ไม่ชอบ	811	1,375	192
กลาง	203	3,290	379	กลาง	204	3,271	397
ชอบ	161	1,549	2,031	ชอบ	166	1,540	2,035
แบบที่ 5	แบบจำลองตอบ						
	ไม่ชอบ	กลาง	ชอบ				
ไม่ชอบ	824	1,375	179				
กลาง	216	3,275	381				
ชอบ	169	1,556	2,016				

ตารางที่ 5.6 ประสิทธิภาพการจำแนกข้อความเซตของคุณลักษณะแบบที่ 2

		ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
การสุ่มแบ่งข้อมูลแบบที่ 1	ไม่ชอบ	61%	69%	34%
	กลาง		53%	85%
	ชอบ		78%	55%
การสุ่มแบ่งข้อมูลแบบที่ 2	ไม่ชอบ	61%	69%	34%
	กลาง		53%	85%
	ชอบ		78%	54%
การสุ่มแบ่งข้อมูลแบบที่ 3	ไม่ชอบ	61%	69%	34%
	กลาง		53%	84%
	ชอบ		78%	54%
การสุ่มแบ่งข้อมูลแบบที่ 4	ไม่ชอบ	61%	69%	34%
	กลาง		53%	84%
	ชอบ		78%	54%
การสุ่มแบ่งข้อมูลแบบที่ 5	ไม่ชอบ	61%	68%	35%
	กลาง		53%	85%
	ชอบ		78%	54%
ค่าเฉลี่ยของทั้ง 5 แบบ	ไม่ชอบ	61%	69%	34%
	กลาง		53%	85%
	ชอบ		78%	54%

ตารางที่ 5.7 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 3

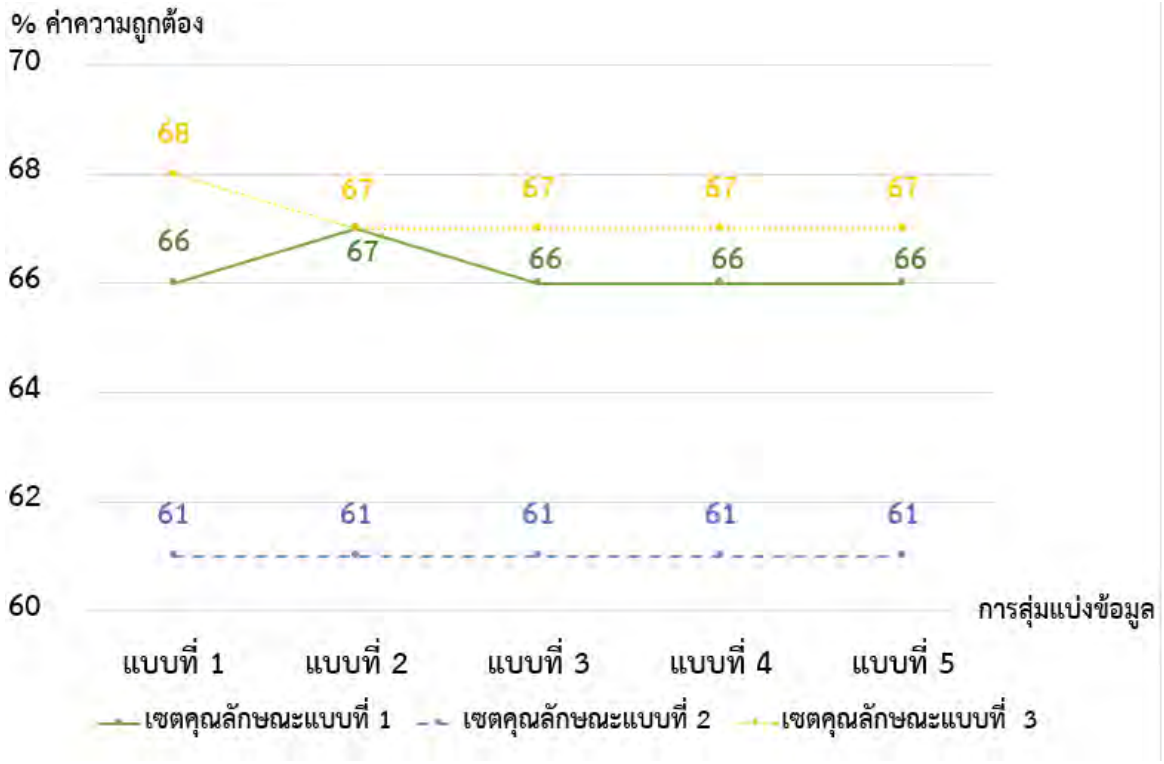
การสุ่มแบ่งข้อมูล							
แบบที่ 1	แบบจำลองตอบ			แบบที่ 2	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	1,444	594	340	ไม่ชอบ	1,419	600	359
กลาง	456	2,468	948	กลาง	439	2,474	959
ชอบ	223	664	2,854	ชอบ	224	674	2,843
แบบที่ 3	แบบจำลองตอบ			แบบที่ 4	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	1,427	606	345	ไม่ชอบ	1,428	587	363
กลาง	465	2,438	969	กลาง	447	2,468	957
ชอบ	219	705	2,817	ชอบ	220	696	2,825
แบบที่ 5	แบบจำลองตอบ						
	ไม่ชอบ	กลาง	ชอบ				
ไม่ชอบ	1,424	590	364				
กลาง	457	2,436	979				
ชอบ	220	688	2,833				

ตารางที่ 5.8 ประสิทธิภาพการจำแนกข้อความเซตของคุณลักษณะแบบที่ 3

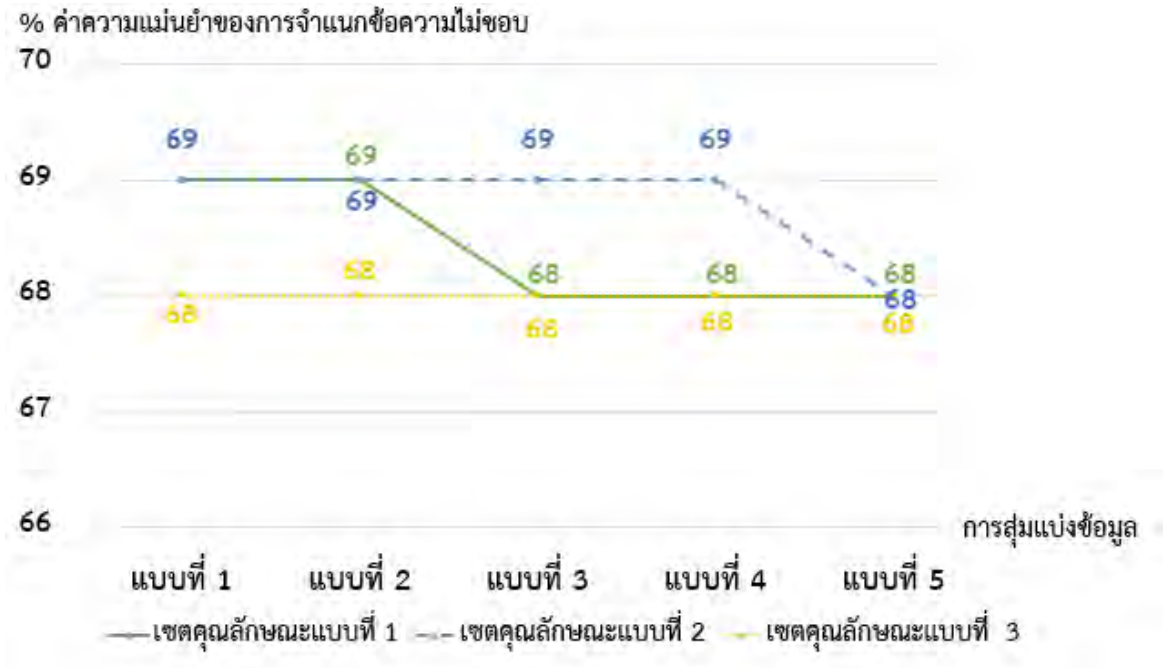
		ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าเรียกคืน
การสุ่มแบ่งข้อมูลแบบที่ 1	ไม่ชอบ	68%	68%	61%
	กลาง		66%	64%
	ชอบ		69%	76%
การสุ่มแบ่งข้อมูลแบบที่ 2	ไม่ชอบ	67%	68%	60%
	กลาง		66%	64%
	ชอบ		68%	76%
การสุ่มแบ่งข้อมูลแบบที่ 3	ไม่ชอบ	67%	68%	60%
	กลาง		65%	63%
	ชอบ		68%	75%
การสุ่มแบ่งข้อมูลแบบที่ 4	ไม่ชอบ	67%	68%	60%
	กลาง		66%	64%
	ชอบ		68%	76%
การสุ่มแบ่งข้อมูลแบบที่ 5	ไม่ชอบ	67%	68%	60%
	กลาง		66%	63%
	ชอบ		68%	76%
ค่าเฉลี่ยของทั้ง 5 แบบ	ไม่ชอบ	67%	68%	60%
	กลาง		66%	64%
	ชอบ		68%	76%

5.3 สรุปผลการทดลอง และการอภิปรายผล

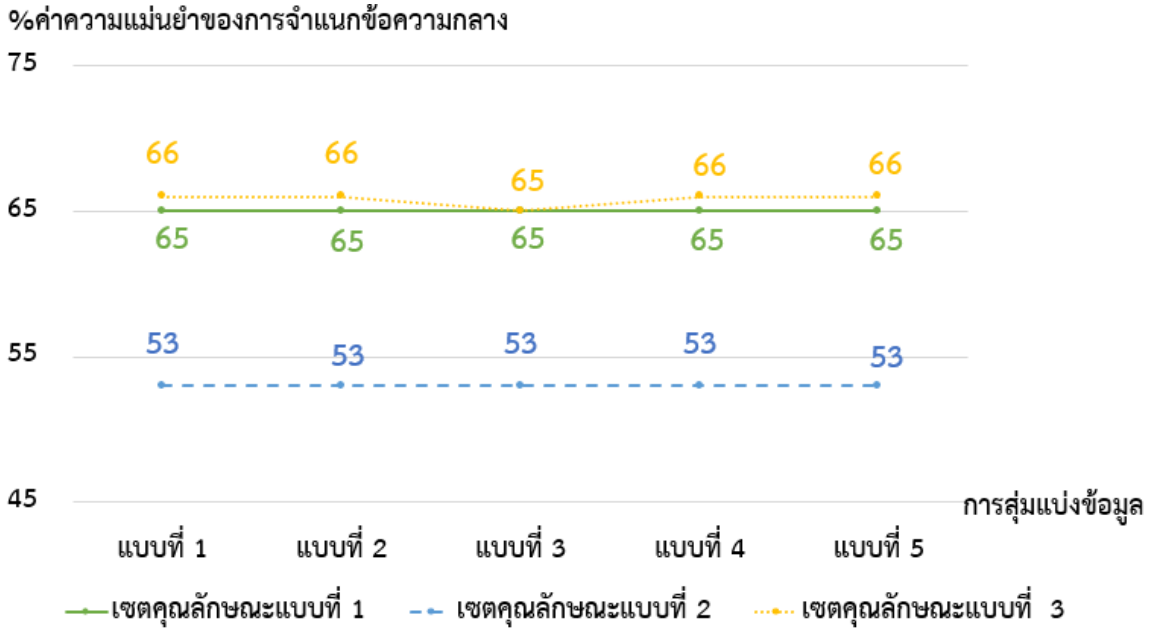
จากผลการทดสอบการจำแนกข้อความด้วยเซตของคุณลักษณะทั้ง 3 แบบ ภาพที่ 5.1 ถึง ภาพที่ 5.8 เป็นผลลัพธ์การเปรียบเทียบประสิทธิภาพของแบบจำลองทั้ง 3 แบบแยกตามการสุ่มข้อมูลแต่ละวิธี



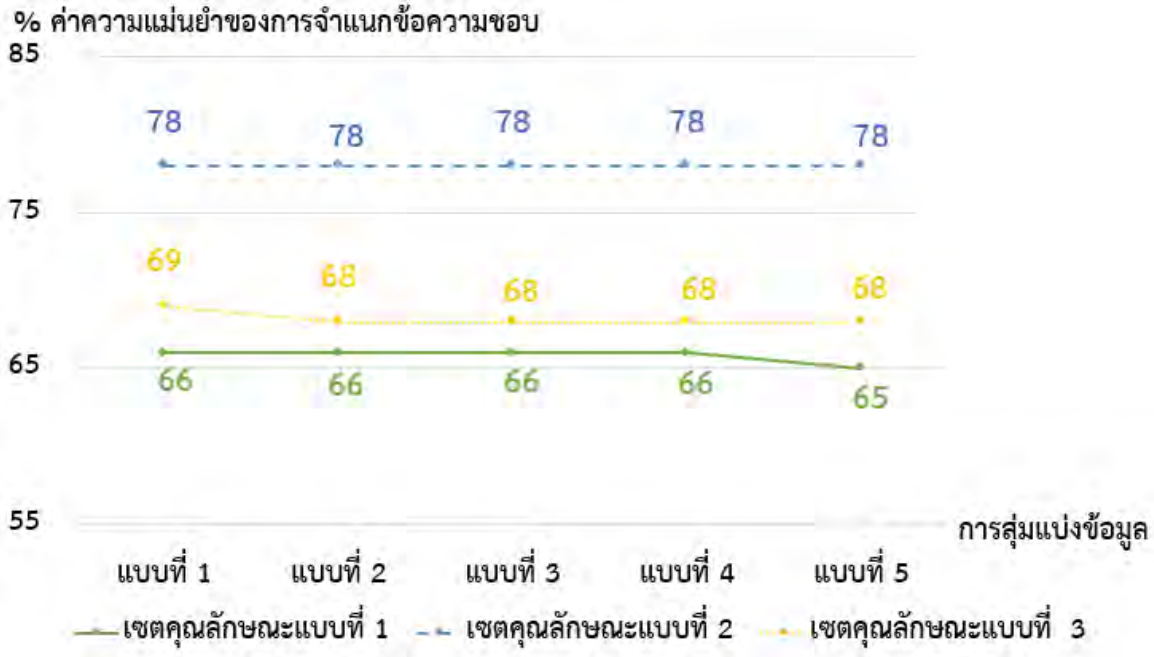
ภาพที่ 5.1 การเปรียบเทียบค่าความถูกต้องของการจำแนกข้อความ



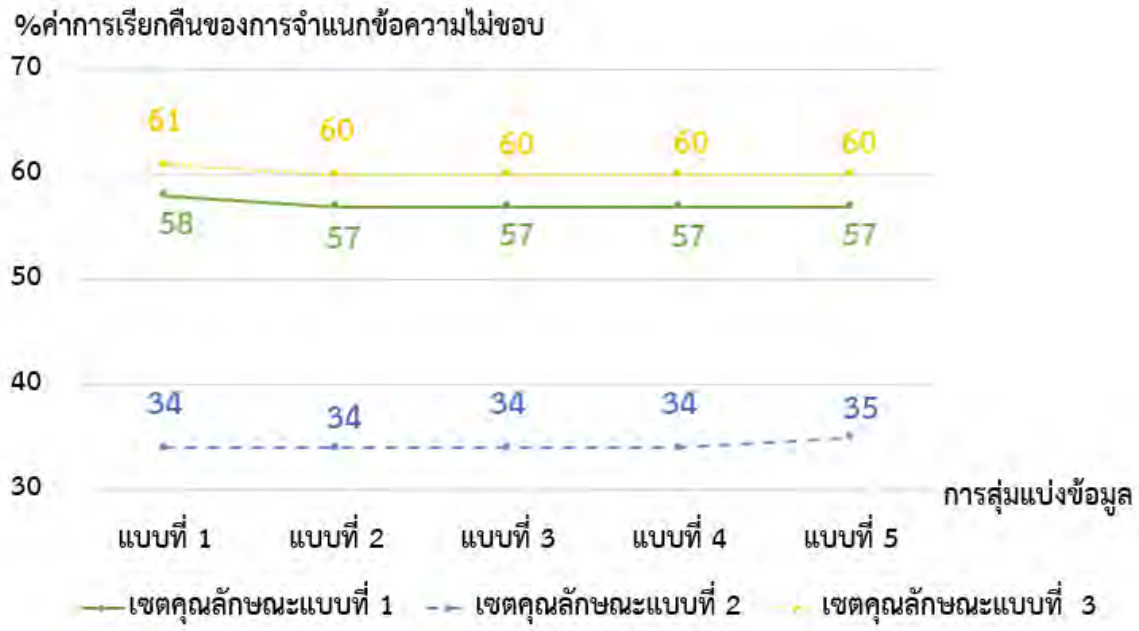
ภาพที่ 5.2 การเปรียบเทียบค่าความแม่นยำของการจำแนกข้อความไม่ชอบ



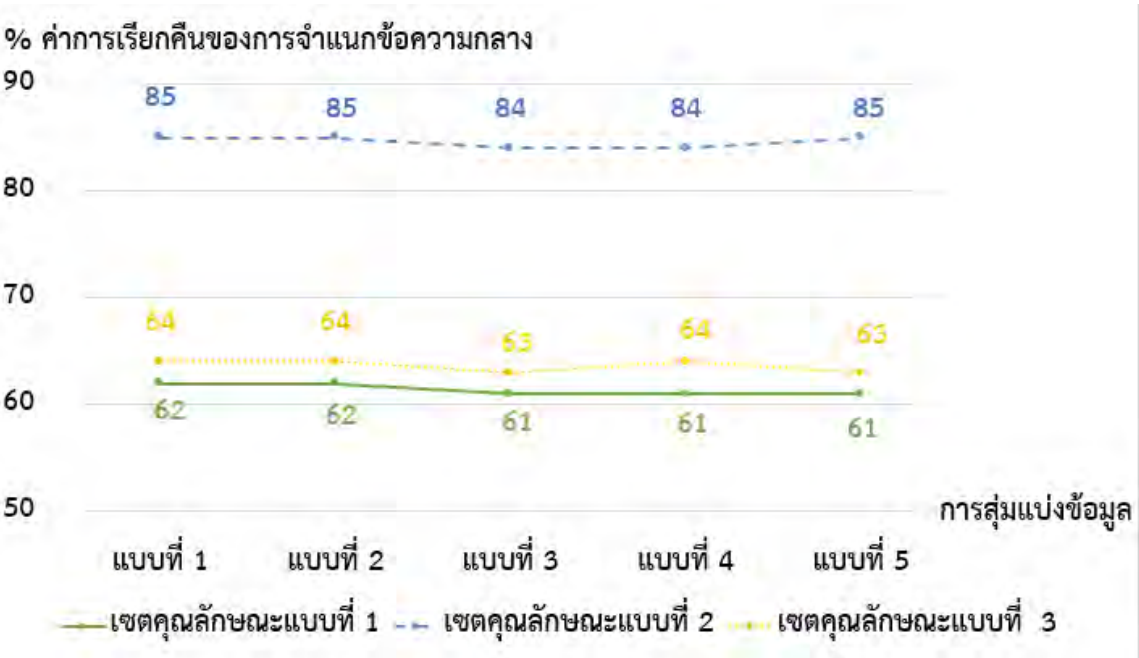
ภาพที่ 5.3 การเปรียบเทียบค่าความแม่นยำของการจำแนกข้อความกลาง



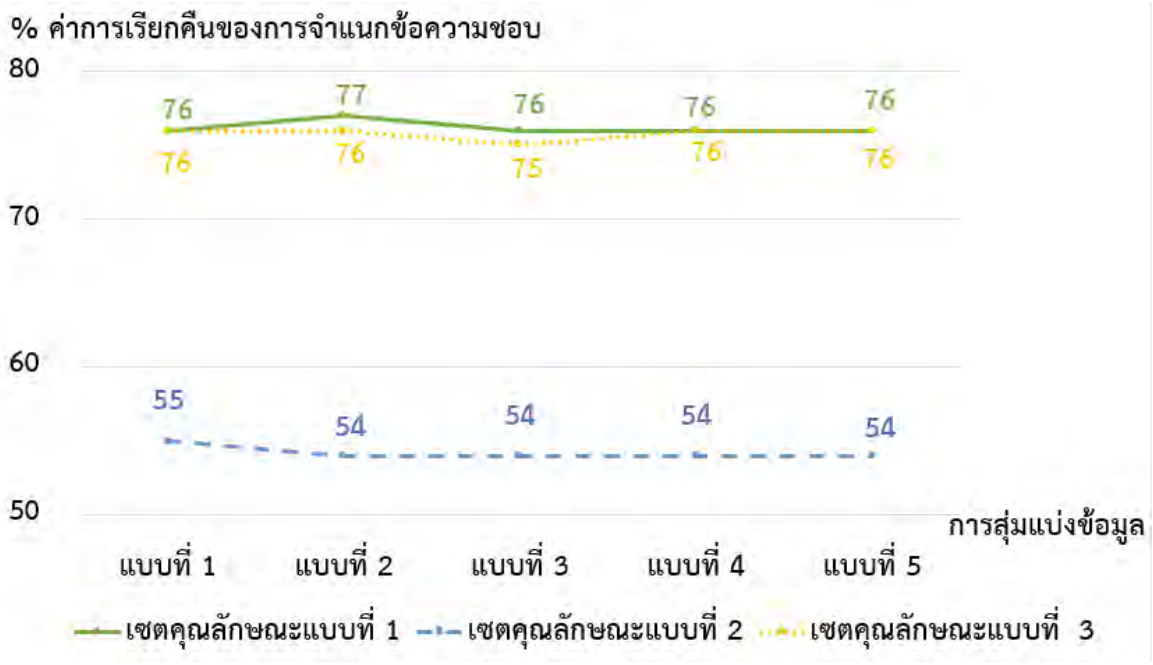
ภาพที่ 5.4 การเปรียบเทียบค่าความแม่นยำของการจำแนกข้อความชอบ



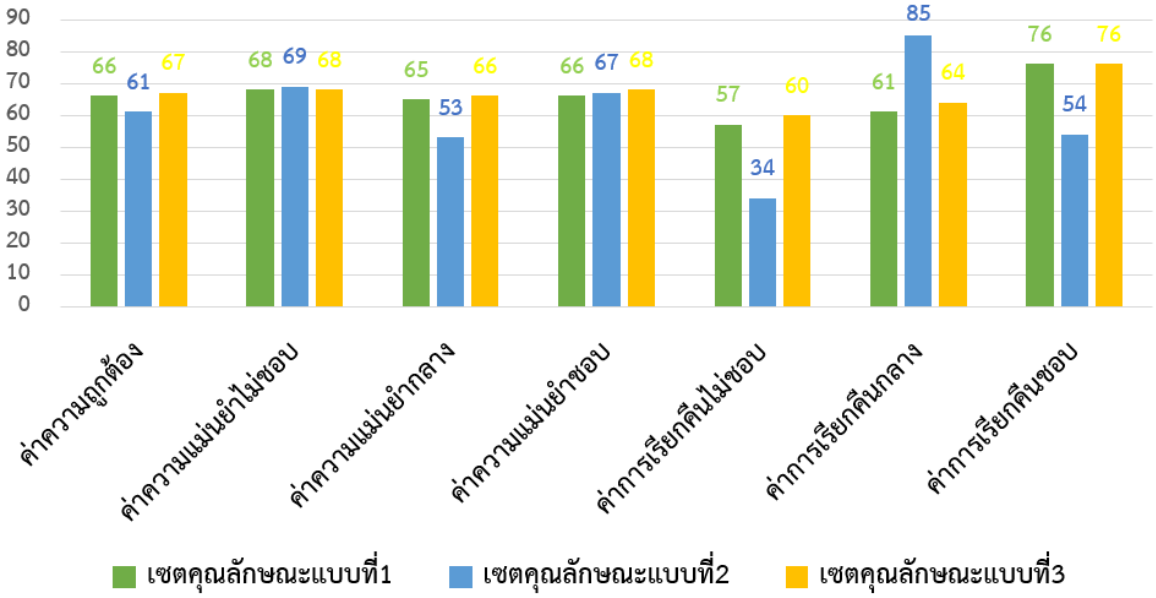
ภาพที่ 5.5 การเปรียบเทียบค่าการเรียกคืนของการจำแนกข้อความไม่ชอบ



ภาพที่ 5.6 การเปรียบเทียบค่าการเรียกคืนของการจำแนกข้อความกลาง



ภาพที่ 5.7 การเปรียบเทียบค่าการเรียกคืนของการจำแนกข้อความชอบ



ภาพที่ 5.8 การเปรียบเทียบค่าเฉลี่ยประสิทธิภาพของการจำแนกข้อความ

จากการทดสอบเซตคุณลักษณะแบบที่ 1 ในการจำแนกข้อความแสดงความคิดเห็นทั้ง 3 กลุ่ม พบว่า ผลการจำแนกข้อความที่แสดงอารมณ์ความรู้สึกให้ค่าเฉลี่ยความถูกต้อง ค่าเฉลี่ยความแม่นยำ และ

ค่าเฉลี่ยการเรียกคืนมากกว่า 57% เนื่องจากเซตคุณลักษณะแบบที่ 1 ผู้พัฒนาใช้คำที่ไม่ซ้ำในข้อความเป็นคลังคำศัพท์ ทำให้มีประสิทธิภาพปานกลางในการวิเคราะห์ข้อความแสดงความคิดเห็น

จากการทดสอบเซตคุณลักษณะแบบที่ 2 ในการจำแนกข้อความแสดงความคิดเห็นทั้ง 3 กลุ่ม พบว่า ผลการจำแนกข้อความที่แสดงความรู้สึกมีประสิทธิภาพที่น้อยลง เนื่องจากเซตคุณลักษณะแบบที่ 2 ผู้พัฒนาใช้คำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกเป็นคลังคำศัพท์ทำให้ความหลากหลายของคลังคำศัพท์มีจำนวนน้อยเมื่อเทียบกับเซตคุณลักษณะแบบที่ 1

จากการทดสอบเซตคุณลักษณะแบบที่ 3 ในการจำแนกข้อความแสดงความคิดเห็นทั้ง 3 กลุ่ม พบว่าผลการจำแนกข้อความที่แสดงความรู้สึกให้ค่าเฉลี่ยความถูกต้อง ค่าเฉลี่ยความแม่นยำ และค่าเฉลี่ยการเรียกคืนมากกว่า 60% ผลการจำแนกข้อความที่แสดงอารมณ์ความรู้สึกมีประสิทธิภาพที่สูงขึ้นเมื่อเทียบกับเซตคุณลักษณะแบบที่ 1 และเซตคุณลักษณะแบบที่ 2 เนื่องจากเซตคุณลักษณะแบบที่ 3 ผู้พัฒนาใช้คำที่ไม่ซ้ำในข้อความรวมกับคำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกสร้างเป็นคลังคำศัพท์ ทำให้ความหลากหลายของคลังคำศัพท์มีจำนวนมาก

จากผลการทดสอบการจำแนกข้อความด้วยเซตของคุณลักษณะทั้ง 3 แบบ สามารถสรุปได้ว่า การใช้คำที่ไม่ซ้ำในข้อความรวมกับคำที่ตรงกับคำในคลังคำศัพท์แสดงอารมณ์ความรู้สึกสร้างเป็นคลังคำศัพท์และแปลงเป็นเวกเตอร์ได้เซตคุณลักษณะแบบที่ 3 ให้ประสิทธิภาพการจำแนกข้อความที่ดีที่สุดโดยจากภาพที่ 5.8 เห็นได้ว่าค่าเฉลี่ยประสิทธิภาพของการจำแนกข้อความของเซตของคุณลักษณะแบบที่ 3 มีความแม่นยำอยู่ที่ประมาณ 67% ซึ่งมีค่าความแม่นยำมากกว่าแบบจำลองอื่น เนื่องจากข้อมูลที่แบบจำลองให้คำตอบการจำแนกข้อความมีความถูกต้องมากกว่าแบบจำลองอื่น แต่ยังมีประสิทธิภาพไม่เพียงพอเท่าที่ควรเนื่องจากยังไม่สามารถจำแนกข้อความด้านไม่ชอบได้ดีพอ (ค่าเรียกคืน 60%) เพราะปริมาณของข้อมูลในด้านลบน้อยทำให้ข้อมูลที่แบบจำลองตอบมีความคลาดเคลื่อน

บทที่ 6
ข้อสรุปและข้อเสนอแนะ

6.1 สรุปผล

ในโครงงานนี้เป็นการพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับละครไทยเพื่อ ใช้ประกอบการตัดสินใจเลือกชมละครได้สะดวกและรวดเร็ว วิเคราะห์ข้อความแสดงความคิดเห็น ออกเป็น ชอบ ไม่ชอบ และความคิดเห็นระดับกลาง โดยใช้ภาษาไพทอน ในการพัฒนาระบบและใช้ ไลบรารี PyThaiNLP ช่วยในการตัดคำ และสร้างโมเดลสำหรับการเรียนรู้ด้วยเครื่องโดยใช้เทคนิคนาอีฟเบย์ของไลบรารี Sklearn หลังจากนั้นนำโมเดลที่ได้มาจำแนกข้อความแสดงความคิดเห็นเกี่ยว ละครไทยว่ามีความรู้สึกชอบ ไม่ชอบหรือความรู้สึกที่เป็นกลางต่อละคร และใช้ Flask ในการพัฒนาเว็ปแอปพลิเคชัน ซึ่งระบบนี้จะช่วยให้ผู้ใช้งานนำไปใช้เป็นเครื่องมือช่วยวิเคราะห์และสรุปผลความชอบที่มีต่อละครไทยและช่วยตัดสินใจการเลือกชมละครได้ง่ายและรวดเร็ว รวมไปถึงผู้ผลิตละครสามารถนำข้อมูลเหล่านี้ปวางแผนการผลิตละครในอนาคตได้ โครงงานนี้ไม่สามารถวิเคราะห์คำที่สะกดผิดและคำสแลงได้ จากการจากการทดสอบการจำแนกข้อความแสดงความคิดเห็นพบว่า ค่าความถูกต้องและความแม่นยำประมาณ 70% มีประสิทธิภาพไม่เพียงพอเนื่องจากยังไม่สารมารถจำแนกข้อความด้านลบได้ดีพอ เพราะปริมาณข้อมูลด้านลบน้อยทำให้ข้อมูลที่แบบจำลองตอบมีความคลาดเคลื่อน

6.2 ผลที่ได้รับ

ผลที่ได้รับจากการพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทยบนทวิตเตอร์ จะแบ่งออกเป็น 2 ส่วน ได้แก่ ส่วนผู้พัฒนา และส่วนผู้ใช้งาน

ผู้พัฒนา

1. ผู้พัฒนาได้เรียนรู้และเข้าใจเทคนิคการวิเคราะห์ความคิดเห็นของข้อความจากทวิตเตอร์
2. ผู้พัฒนาได้เรียนรู้การเขียนโปรแกรมด้วยภาษาไพทอน
3. ผู้พัฒนาได้ฝึกทักษะการวางแผนการดำเนินการและแก้ไขปัญหาเฉพาะหน้า

ผู้ใช้งาน

1. เป็นเครื่องมือช่วยวิเคราะห์และสรุปผลความชอบที่มีต่อละครไทยและช่วยในการตัดสินใจการเลือกรับชม
2. เป็นประโยชน์ต่อผู้ผลิตละครสามารถนำข้อมูลนี้ไปวางแผนการผลิตละครในอนาคต

6.3 ปัญหาและอุปสรรค

1. เนื่องจากในปัจจุบันมีข้อความสะกดผิดและคำสแลงจำนวนมากจึงส่งผลให้มีตัวอย่าง ข้อมูลที่สามารถใช้งานได้ไม่เพียงพอตามเป้าหมายที่คาดหวัง

2. ผู้พัฒนาใช้เวลามากในการเก็บรวบรวมและทำผลเฉลยข้อความแสดงความคิดเห็น

6.4 วิธีการแก้ปัญหา

1. เก็บข้อมูลข้อความแสดงความคิดเห็นมากขึ้น
2. เริ่มรวบรวมและทำผลเฉลยให้เร็วขึ้น

เอกสารอ้างอิง

[1] รวิสุดา เทศเมือง และ นิเวศ จิระวิชิตชัย. การวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิว สินค้าออนไลน์โดยใช้ขั้นตอนวิธีซัพพอร์ตเวกเตอร์แมทชีน, วารสารวิศวกรรมศาสตร์ มหาวิทยาลัย สยาม, ปีที่ 18, ฉบับที่ 34, มกราคม-มิถุนายน ปี 2560, หน้า 1-12.

[2] Vivek Narayanan. Fast and accurate sentiment classification using an enhanced Naive Bayes model, Intelligent Data Engineering and Automated Learning (IDEAL 2013) Lecture Notes in Computer Science, Volume 8206, 2013, pp 194-201.

[3] นายวรรณพงษ์ ภัททิยไพบูลย์. “คลังข้อมูล sentiment analysis ภาษาไทย (คำด้านบวก)” [ออนไลน์] แหล่งที่มา : https://github.com/PyThaiNLP/lexicon-thai/blob/master/ข้อความ/pos.txt สืบค้นเมื่อ 9 กันยายน พ.ศ. 2562

[4] นายวรรณพงษ์ ภัททิยไพบูลย์. “คลังข้อมูล sentiment analysis ภาษาไทย (คำด้านลบ)” [ออนไลน์] แหล่งที่มา : https://github.com/PyThaiNLP/lexicon-thai/blob/master/ข้อความ/neg.txt สืบค้นเมื่อ 9 กันยายน พ.ศ. 2562

[5] นายวรรณพงษ์ ภัททิยไพบูลย์. “คลังข้อมูล sentiment analysis ภาษาไทย (คำเป็นกลาง)” [ออนไลน์] แหล่งที่มา : https://github.com/PyThaiNLP/lexicon-thai/blob/master/ข้อความ/neutral.txt สืบค้นเมื่อ 9 กันยายน พ.ศ. 2562

[6] วศิณี นุชศิริ และวันสิริ กิตติรุ่งเรือง การวิเคราะห์ความคิดเห็นต่อผลิตภัณฑ์ดูแลผิวบนทวิตเตอร์. โครงงานวิทยาศาสตร์ระดับปริญญาตรี สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์ และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์ มหาวิทยาลัย. ปี 2559.

[7] นิสาชล หยาง และวิลาสินี เอกพงศ์พันธุ์. การวิเคราะห์ข้อความแสดงความคิดเห็นต่อโรงแรมสำหรับการจัดอันดับตามคุณลักษณะของโรงแรม. โครงงานวิทยาศาสตร์ระดับปริญญาตรี สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์ และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์ มหาวิทยาลัย. ปี 2559

[8] นพมาศ ปักเข็ม, “การเรียนรู้แบบเบย์ (Bayesian Learning)”, ในการทำเหมืองข้อมูล, สาขาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยทักษิณ, ปี 2558.

ภาคผนวก

ภาคผนวก ก
แบบเสนอหัวข้อโครงงาน รายวิชา 2301399 Project Proposal
ปีการศึกษา 2562

ชื่อโครงงาน (ภาษาไทย)	ระบบวิเคราะห์ความคิดเห็นต่อละครไทยบนทวิตเตอร์
ชื่อโครงงาน (ภาษาอังกฤษ)	Sentiment Analysis for Thai Drama on Twitter
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.ภควรรณ ปักษี
ผู้ดำเนินการ	นายธนะสิทธิ์ เร่งสมบูรณ์สุข เลขประจำตัวนิสิต 5933629523
	นางสาวกวินธิดา สายยศ เลขประจำตัวนิสิต 5933602523
	สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
	คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

หลักการและเหตุผล

ปัจจุบันละครไทยเป็นที่นิยมของกลุ่มคนจำนวนมาก เนื่องจากปัจจุบันมีละครเพิ่มมากขึ้นทำให้ผู้คนสนใจดูละคร ซึ่งละครมีหลายรูปแบบ เช่นคอมมาดี้ (comedy) ดราม่า (drama) และมีการแลกเปลี่ยนแสดงความคิดเห็นเกี่ยวกับละครผ่านทางสื่อออนไลน์ (online media) เช่น ทวิตเตอร์ (Twitter) ทำให้มีข้อมูลที่หลากหลาย ต้องใช้เวลานานในการค้นหาข้อมูลที่ต้องการ ซึ่งข้อมูลเหล่านี้สามารถนำไปวิเคราะห์ข้อความแสดงความคิดเห็น เพื่อใช้เป็นประโยชน์ในการตัดสินใจเลือกชมละคร ดังนั้นควรที่จะมีระบบวิเคราะห์ข้อความอัตโนมัติเพื่อที่จะลดปัญหาเหล่านี้ได้

ตัวอย่างงานวิจัยเกี่ยวกับการวิเคราะห์ความคิดเห็นที่นำมาประยุกต์ใช้ เช่น งานวิจัย [1] มีการวิเคราะห์ความคิดเห็นต่อโรงแรมซึ่งใช้ข้อความรีวิวโรงแรมบนเว็บอโกด้าและบนทวิตเตอร์ โดยได้มีการแบ่งประเภทของข้อความการแสดงความคิดเห็นนี้ว่าเป็นข้อความที่เป็นเชิงบวก (positive) คือดี และข้อความที่เป็นเชิงลบ (negative) คือไม่ดี โดยใช้เทคนิค 4 วิธีคือ นาอีฟเบย์ (naïve bayes) ซับพอร์ตเวกเตอร์แมชชีนหรือเอสวีเอ็ม (support vector machine: SVM) เค-เนียร์เรสเนเบอร์ (k-nearest neighbor) และต้นไม้ตัดสินใจ (decision tree) ซึ่งผลลัพธ์จะแสดงผลสรุปข้อมูลที่ได้จากการวิเคราะห์ จากงานวิจัยนี้จะเห็นได้ว่าค่าความถูกต้องของเทคนิคเอสวีเอ็มจะสูงที่สุด และรองลงมาคือ นาอีฟเบย์ แต่จุดอ่อนของงานคือไม่มีการคำนวณหาค่าความแม่นยำของการจำแนกข้อความแต่ละด้าน อีกงานวิจัยหนึ่ง [2] เป็นงานวิจัยที่ศึกษาเกี่ยวกับการจำแนกความรู้สึกหรือความคิดเห็นเกี่ยวกับภาพยนตร์ โดยใช้ข้อมูลจากเว็บ

https://www.imdb.com/ ซึ่งเป็นฐานข้อมูลรีวิวภาพยนตร์ (movie reviews database: IMDb) ซึ่งจะรวบรวมความคิดเห็นต่าง ๆ เกี่ยวกับภาพยนตร์ งานนี้มีการใช้เทคนิคต่าง ๆ ในการตัดคำและคิดค่าคะแนนเพื่อทำให้เทคนิคนาอีฟเบย์มีประสิทธิภาพในการจำแนกความรู้สึกที่รวดเร็วและถูกต้อง ดังนั้นจากการศึกษางานวิจัยด้านการวิเคราะห์ความคิดเห็น (opinion mining) และการวิเคราะห์อารมณ์และความรู้สึก ผู้พัฒนาพบว่าเทคนิคนาอีฟเบย์เป็นการเรียนรู้ด้วยเครื่องเทคนิคหนึ่ง ที่คนส่วนใหญ่นิยมใช้ในการวิเคราะห์ข้อความมากที่สุด จึงจะประยุกต์ใช้เทคนิคนาอีฟเบย์ในการวิเคราะห์ข้อความแสดงความคิดเห็นที่เป็นภาษาไทย

จากที่กล่าวมาข้างต้นทางผู้พัฒนาจึงจะพัฒนาระบบวิเคราะห์อารมณ์และความรู้สึกต่อละครไทยบนทวิตเตอร์ โดยวิเคราะห์ความคิดเห็นออกเป็นข้อความที่ชอบ ไม่ชอบ หรือรู้สึกเป็นกลางต่อละครเรื่องนั้น ๆ และแสดงผลสรุปข้อมูลที่ได้จากการวิเคราะห์เป็นรูปกราฟแท่งเพื่อให้ผู้บริโภคใช้เป็นเครื่องมือในการตัดสินใจเลือกชมละครที่สะดวกมากขึ้น และเป็นประโยชน์ต่อผู้ผลิตละครในการวางแผนการผลิตในอนาคต

วัตถุประสงค์

1. ศึกษาการจำแนกอารมณ์และความรู้สึกของข้อความภาษาไทยจากข้อความที่เกี่ยวข้องกับละครไทย
2. วิเคราะห์ความชอบที่มีต่อละครไทยและสรุปผลตอบรับจากข้อความที่อยู่บนทวิตเตอร์

ขอบเขตของโครงงาน

1. โครงงานนี้ศึกษาเฉพาะข้อความบนทวิตเตอร์ที่เป็นภาษาไทยไม่ครอบคลุมคำสแลงและไม่พิจารณาคำที่สะกดผิด
2. การเก็บตัวอย่างข้อความแสดงความคิดเห็นบนทวิตเตอร์จะใช้แฮชแท็ก (#) ชื่อละครหรือชื่อตัวละคร (พระเอก/นางเอก) จากละครไทย 8 เรื่องในช่วงกันยายน – ธันวาคมในปี พ.ศ. 2562 ได้แก่ รักฉุดใจนายฉุกเฉิน ลิขิตรักข้ามดวงดาว มธุรสโลกันต์ รองเท้านารี เพลิงรักเพลิงแค้น ฤกษ์สังหาร เขาวานให้หนูเป็นสายลับ และดาวหลงฟ้าโดยมีข้อความที่จะนำมาใช้วิเคราะห์ในโครงการนี้อย่างน้อย 10,000 ข้อความ
3. ผลลัพธ์ในการจำแนกข้อความแสดงความคิดเห็นจะถูกแบ่งออกเป็นชอบ ไม่ชอบ และกลาง

วิธีการดำเนินงาน

1. ศึกษาและค้นคว้าหาข้อมูลเกี่ยวกับการวิเคราะห์ความคิดเห็นของข้อความและการทำคลังศัพท์ (มีการศึกษากลุ่มคำด้านบวก (positive) ด้านลบ (negative) และเป็นกลาง (neutral) จาก [3] [4] [5])
2. กำหนดขอบเขตและขั้นตอนการดำเนินงาน

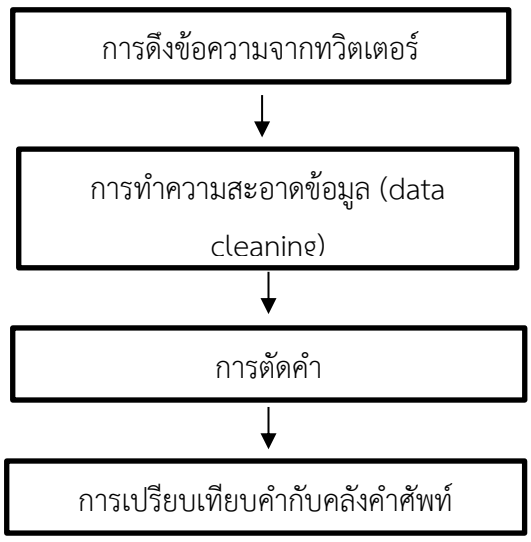
3. เก็บรวบรวมข้อความที่เกี่ยวกับละครไทยจากทวิตเตอร์โดยใช้แฮชแท็ก (#) ชื่อละครหรือชื่อตัวละครโดยใช้ไลบรารีที่ชื่อ Tweepy ซึ่งจะช่วยในการเชื่อมต่อไปยังทวิตเตอร์เอพีไอ
4. วิเคราะห์ข้อความแสดงความคิดเห็นและจำแนกประเภทของคำ โดยแบ่งข้อความออกเป็นประโยคและแบ่งประโยคออกเป็นคำเพื่อใช้เป็นข้อมูลในการออกแบบระบบวิเคราะห์ข้อความแสดงความคิดเห็น
5. ออกแบบและพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทยซึ่งจะมีการเปรียบเทียบผลลัพธ์ของการจำแนกข้อมูลระหว่างการใช้คลังคำศัพท์ (นับความถี่ของคำด้านบวกและด้านลบที่ปรากฏในประโยค) เพียงอย่างเดียวกับการใช้เทคนิคการเรียนรู้ด้วยนาอีฟเบย์ ซึ่งจะทดลองกับเซตของคุณลักษณะ (feature set) ที่แตกต่างกัน เพื่อเลือกใช้โมเดลการจำแนกข้อมูลที่ดีที่สุดมาพัฒนาเป็นระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทย

ตัวอย่างเซตของคุณลักษณะที่คาดว่าจะใช้สำหรับเทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟ-เบย์ เช่น

- แบบที่ 1 ใช้คำในประโยค
- แบบที่ 2 ใช้เฉพาะคำที่ตรงกับคำในคลังคำศัพท์
- แบบที่ 3 ใช้คำในประโยคและคำที่ตรงกับคำในคลังคำศัพท์

โครงสร้างของระบบจะแบ่งเป็น 2 ส่วน คือ Front-end และ Back-end ซึ่งในส่วนของ Front-end เป็นการแสดงผลลัพธ์การจำแนกประโยคข้อความผ่านทางหน้าเว็บแอปพลิเคชัน แบ่งออกเป็นข้อความ 3 กลุ่มคือ ชอบ ไม่ชอบ และกลางอีกทั้งสามารถแสดงผลข้อมูลในรูปแบบของกราฟแท่ง สำหรับการทำงานในส่วนของ Back-end จะมีการออกแบบวิธีการจำแนกข้อมูลหลายๆ วิธีการ ดังต่อไปนี้

1) การทำงานของระบบสำหรับการจำแนกข้อมูลโดยคลังคำศัพท์เพียงอย่างเดียว จะประกอบด้วย 4 มอดูล



มอดูล: การดึงข้อความจากทวิตเตอร์
ข้อมูลนำเข้า: แฮชแท็กชื่อละครหรือตัวละคร
กระบวนการ : ใช้ไลบรารี Tweepy ในการดึงข้อมูล
ข้อมูลที่ส่งออก : ประโยคข้อความจากทวิตเตอร์ที่ตรงกับแฮชแท็กที่ใช้ในการดึงข้อมูล

มอดูล: การทำความสะอาดข้อมูล
ข้อมูลนำเข้า: ประโยคข้อความจากทวิตเตอร์
กระบวนการ : ใช้วิธี manual ในการอ่านและลบข้อมูลที่ไม่เกี่ยวข้อง เช่น แฮชแท็ก ยูอาร์-แอล ข้อความที่รีทวิต
ข้อมูลที่ส่งออก : ประโยคข้อความที่ทำความสะอาดแล้ว

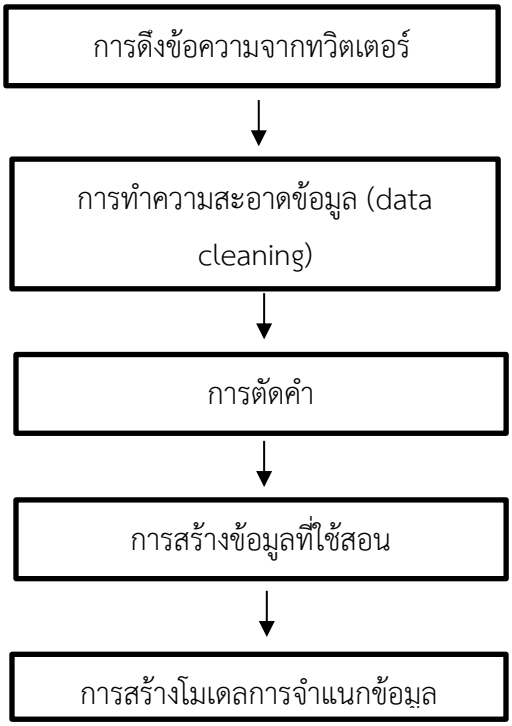
มอดูล: การตัดคำ
ข้อมูลนำเข้า: ประโยคข้อความที่ได้หลังจากการทำความสะอาด
กระบวนการ : ใช้ไลบรารี PyThaiNLP ด้วยเทคนิค Maximum Matching algorithm ในการตัดคำ
ข้อมูลที่ส่งออก : คำในประโยค

มอดูล: การเปรียบเทียบคำกับคลังคำศัพท์
ข้อมูลนำเข้า: คำในประโยค
กระบวนการ : ตรวจหาคำในประโยคที่ตรงกับคำในคลังคำศัพท์
ข้อมูลที่ส่งออก : คำในประโยคที่ตรงกับคำในคลังคำศัพท์

การจำแนกประโยคใช้การคำนวณผลรวมของจำนวนคำด้านบวกกับคำด้านลบที่ตรงกับคำในคลังคำศัพท์ คือ

- จำนวนคำด้านบวกเท่ากับจำนวนคำด้านลบจะได้เป็นประโยคในกลุ่มเป็นกลาง
- จำนวนคำด้านบวกมากกว่าจำนวนคำด้านลบจะได้เป็นประโยคในกลุ่มชอบ
- จำนวนคำด้านบวกน้อยกว่าจำนวนคำด้านลบจะได้เป็นประโยคในกลุ่มไม่ชอบ

2) การทำงานของระบบสำหรับการจำแนกข้อมูลโดยการใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์โดยไม่ใช้ข้อมูลจากคลังคำศัพท์ (เซตของคุณลักษณะแบบที่ 1) จะประกอบด้วย 5 มอดูล



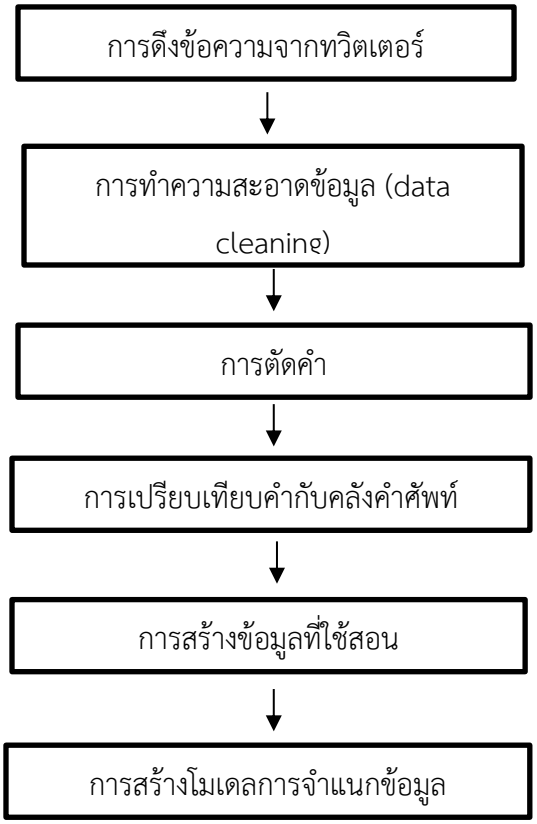
มอดูลสามส่วนแรก คือ การดึงข้อความจากทวิตเตอร์ การทำความสะอาดข้อมูล และการตัดคำจะทำงานเหมือนที่กล่าวแล้วข้างต้น

มอดูล: การสร้างข้อมูลที่ใช้สอน
ข้อมูลนำเข้า: คำในประโยค
กระบวนการ : ใช้วิธี One-hot โดยให้คำที่ไม่ซ้ำกันแต่ละคำของข้อมูลทั้งหมดเป็นคุณลักษณะ ซึ่งถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0
ข้อมูลที่ส่งออก : เซตของคุณลักษณะแบบที่ 1 ที่จะนำไปเรียนรู้ด้วยเครื่อง

มอดูล : การสร้างโมเดลการจำแนกข้อมูล
ข้อมูลนำเข้า : เซตของคุณลักษณะแบบที่ 1 และกลุ่มของประโยคข้อความ
กระบวนการ : ใช้ไลบรารีนาอีฟเบย์ของเครื่องมือการเรียนรู้ด้วยเครื่อง
ข้อมูลที่ส่งออก : โมเดลการจำแนกข้อความออกเป็น 3 กลุ่ม ได้แก่ ชอบ ไม่ชอบ และกลาง

การจำแนกประโยคข้อความใช้การให้คำตอบจากโมเดลการจำแนกข้อมูล

3) การทำงานของระบบสำหรับการจำแนกข้อมูลโดยการใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ โดยใช้เฉพาะคำที่ตรงกับคำในคลังคำศัพท์ (เซตของคุณลักษณะแบบที่ 2) จะประกอบด้วย 6 มอดูล



มอดูลสี่ส่วนแรก คือ การดึงข้อความจากทวิตเตอร์ การทำความสะอาดข้อมูล การตัดคำ และการเปรียบเทียบคำกับคลังคำศัพท์ จะทำงานเหมือนที่กล่าวแล้วข้างต้น

มอดูล: การสร้างข้อมูลที่ใช้สอน
ข้อมูลนำเข้า: คำในประโยคที่ตรงกับคำในคลังคำศัพท์
กระบวนการ : ใช้วิธี One-hot โดยให้คำทั้งหมดในคลังคำศัพท์เป็นคุณลักษณะ ถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0
ข้อมูลที่ส่งออก : เซตของคุณลักษณะแบบที่ 2 ที่จะนำไปเรียนรู้ด้วยเครื่อง

มอดูล: การสร้างโมเดลการจำแนกข้อมูล
ข้อมูลนำเข้า: เซตของคุณลักษณะแบบที่ 2 และกลุ่มของประโยคข้อความ
กระบวนการ : ใช้ไลบรารีนาอีฟเบย์ของเครื่องมือการเรียนรู้ด้วยเครื่อง
ข้อมูลที่ส่งออก : โมเดลการจำแนกข้อความออกเป็น 3 กลุ่ม ได้แก่ ชอบ ไม่ชอบ และกลาง

การจำแนกประโยคข้อความใช้การให้คำตอบจากโมเดลการจำแนกข้อมูล

4) การทำงานของระบบสำหรับการจำแนกข้อมูลโดยการใช้เทคนิคการเรียนรู้ด้วยเครื่องแบบนาอีฟเบย์ โดยใช้คำในประโยคและคำที่ตรงกับคำในคลังคำศัพท์ (เซตของคุณลักษณะแบบที่ 3) จะประกอบด้วย 6 มอดูล ดังการทำงานของระบบในข้อ 3)

มอดูลสี่ส่วนแรก คือ การดึงข้อความจากทวิตเตอร์ การทำความสะอาดข้อมูล การตัดคำ และการเปรียบเทียบคำกับคลังคำศัพท์ จะทำงานเหมือนที่กล่าวแล้วข้างต้น

มอดูล: การสร้างข้อมูลที่ใช้สอน
ข้อมูลนำเข้า: คำในประโยคและคำในประโยคที่ตรงกับคำในคลังคำศัพท์
กระบวนการ : - คำในประโยคใช้วิธี One-hot โดยให้คำที่ไม่ซ้ำกันแต่ละคำของข้อมูลทั้งหมดเป็นคุณลักษณะ ซึ่งถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0
- คำในประโยคที่ตรงกับคำในคลังคำศัพท์ใช้วิธี One-hot โดยให้คำทั้งหมดในคลังคำศัพท์เป็นคุณลักษณะ ถ้ามีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 1 แต่ถ้าไม่มีคำอยู่ในข้อความค่าคุณลักษณะจะเป็น 0
ข้อมูลที่ส่งออก : เซตของคุณลักษณะแบบที่ 3 ที่จะนำไปเรียนรู้ด้วยเครื่อง

มอดูล: การสร้างโมเดลการจำแนกข้อมูล
ข้อมูลนำเข้า: เซตของคุณลักษณะแบบที่ 3 และกลุ่มของประโยคข้อความ
กระบวนการ : ใช้ไลบรารีนาอีฟเบย์ของเครื่องมือการเรียนรู้ด้วยเครื่อง
ข้อมูลที่ส่งออก : โมเดลการจำแนกข้อความออกเป็น 3 กลุ่ม ได้แก่ ชอบ ไม่ชอบ และกลาง

การจำแนกประโยคข้อความใช้การให้คำตอบจากโมเดลการจำแนกข้อมูล

6. ตรวจสอบความถูกต้องของการจำแนกข้อมูล โดยนำผลลัพธ์ที่ได้มาตรวจสอบความถูกต้อง

7. จัดทำเอกสารรายงาน และคู่มือการใช้งานระบบ

ตารางเวลาการดำเนินงาน

การพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทยบททวิตเตอร์ เริ่มดำเนินงานตั้งแต่เดือนกันยายน 2562 ถึงเดือนมีนาคม 2563 รวมระยะเวลา 7 เดือน โดยมีตารางเวลาการดำเนินงานดังนี้

ขั้นตอนดำเนินงาน	2562				2563		
	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.
1. ศึกษาและค้นคว้าหาข้อมูลเกี่ยวกับการวิเคราะห์ความคิดเห็นของข้อความ และการทำคลังศัพท์สำหรับการจำแนกกลุ่มคำของละครไทย							
2. กำหนดขอบเขตและขั้นตอนการดำเนินงาน							
3. เก็บรวบรวมข้อมูลที่เกี่ยวกับละครไทยจากทวิตเตอร์							
4. วิเคราะห์ข้อความแสดงความคิดเห็นและจำแนกประเภทของคำ							
5. ออกแบบและพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็นต่อละครไทย							
6. ตรวจสอบความถูกต้องของการจำแนกข้อมูล							
7.จัดทำเอกสารรายงาน และคู่มือการใช้งานระบบ							

ประโยชน์ที่คาดว่าจะได้รับ

1. ประโยชน์ต่อผู้พัฒนา
   - ได้ศึกษาและเรียนรู้เทคนิคการวิเคราะห์ความคิดเห็นของข้อความจากทวิตเตอร์
   - ได้พัฒนาทักษะการวางแผนการดำเนินงาน

2. ประโยชน์ต่อผู้นำซอฟต์แวร์นี้ไปใช้งาน
   - เป็นเครื่องมือช่วยวิเคราะห์และสรุปผลความชอบที่มีต่อละครไทยและช่วยในการตัดสินใจการเลือกรับชม
   - เป็นประโยชน์ต่อผู้ผลิตละครสามารถนำข้อมูลนี้ไปวางแผนการผลิตละครในอนาคต

อุปกรณ์และเครื่องมือที่ใช้

1. ฮาร์ดแวร์
   - เครื่องคอมพิวเตอร์ ระบบปฏิบัติการ Window® แบบ 64 บิต
หน่วยประมวลผล Intel® core™ i5-6500
หน่วยความจำ DDR SDRAM 8 กิกะไบต์

2. ซอฟต์แวร์
   - โปรแกรมภาษา Python 3.7 ใช้พัฒนาระบบ
   - Tweepy เป็นไลบรารีซึ่งช่วยในการเชื่อมต่อไปยังทวิตเตอร์ หรือที่เรียกกันว่าทวิตเตอร์เอพีไอ สำหรับการเก็บรวบรวมข้อมูลจากทวิตเตอร์ เช่น ชื่อผู้ใช้ทวิต วันที่ที่ทวิต ข้อความที่ทวิต ฯลฯ
   - PyThaiNLP เป็นไลบรารีที่ไว้ใช้ในการตัดคำซึ่งจะรองรับเฉพาะ Python ที่เวอร์ชันสูงกว่า 3.4
3. อื่น ๆ
   - กระดาษเอสี่ (A4) สำหรับทำรายงาน
   - หมึกพิมพ์

งบประมาณ

1. สายเชื่อมต่อคอมพิวเตอร์กับช่องสัญญาณ VGA	ราคา 1,000 บาท
2. ฮาร์ดดิสก์พกพา (External Harddisk USB 3.0) ความจุ 2 TB	ราคา 2,000 บาท
3. Samsung SSD 970 Evo 500 GB	ราคา 4,400 บาท
4. เมาส์ Logitech G403 wired prodigy	ราคา 1,600 บาท
5. ค่ากระดาษ ค่าถ่ายเอกสาร และจัดทำรูปเล่มรายงาน	ราคา 1,000 บาท
รวม	10,000 บาท

หมายเหตุ : ทั้งนี้งบประมาณที่ตั้งไว้ขอถัวเฉลี่ยทุกรายการ

เอกสารอ้างอิง

1. รวิสุดา เทศเมือง และ นิเวศ จิระวิชิตชัย. การวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิวสินค้าออนไลน์โดยใช้ขั้นตอนวิธีซัพพอร์ตเวกเตอร์แมทชีน, วารสารวิศวกรรมศาสตร์ มหาวิทยาลัยสยาม, ปีที่ 18, ฉบับที่ 34, มกราคม-มิถุนายน ปี 2560, หน้า 1-12.
2. Vivek Narayanan. Fast and accurate sentiment classification using an enhanced Naive Bayes model, Intelligent Data Engineering and Automated Learning (IDEAL 2013) Lecture Notes in Computer Science, Volume 8206, 2013, pp 194-201.
3. นายวรรณพงษ์ ภัททิยไพบูลย์. “คลังข้อมูล sentiment analysis ภาษาไทย (คำด้านบวก)” [ออนไลน์] แหล่งที่มา : https://github.com/PyThaiNLP/lexicon-thai/blob/master/ข้อความ/pos.txt สืบค้นเมื่อ 9 กันยายน พ.ศ. 2562
4. นายวรรณพงษ์ ภัททิยไพบูลย์. “คลังข้อมูล sentiment analysis ภาษาไทย (คำด้านลบ)” [ออนไลน์] แหล่งที่มา : https://github.com/PyThaiNLP/lexicon-thai/blob/master/ข้อความ/neg.txt สืบค้นเมื่อ 9 กันยายน พ.ศ. 2562
5. นายวรรณพงษ์ ภัททิยไพบูลย์. “คลังข้อมูล sentiment analysis ภาษาไทย (คำเป็นกลาง)” [ออนไลน์] แหล่งที่มา : https://github.com/PyThaiNLP/lexicon-thai/blob/master/ข้อความ/neutral.txt สืบค้นเมื่อ 9 กันยายน พ.ศ. 2562
6. สถานีวิทยุโทรทัศน์ไทย. “รายชื่อละครที่ออกอากาศในช่วงกันยายน – ธันวาคม ปี พ.ศ. 2562”
7. หน่วยงานผู้พัฒนาทวิตเตอร์. “Twitter APIs” [ออนไลน์] แหล่งที่มา : https://developer.twitter.com/ apps สืบค้นเมื่อ 11 กันยายน พ.ศ. 2562
8. นพมาศ ปักเข็ม, "การเรียนรู้แบบเบย์ (Bayesian Learning)", ใน การทำเหมืองข้อมูล, สาขาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยทักษิณ, ปี 2558

ภาคผนวก ข
ตัวอย่างโค้ดที่ใช้ในการพัฒนาระบบ

1. ตัวอย่างโค้ดเมท็อด clean_n(x)

```
files = open("clean.txt","r")
  cleanline = files.readlines()
  listclean = [e.strip() for e in cleanline]
  list_tokenss = []

  for i in x:
      if i not in listclean:
        list_tokenss.append(i)
  return list_tokenss
```

2. ตัวอย่างโค้ดเมท็อด remove_ stopwords(x):

```
stopwords =list(thai_stopwords())


 def remove_stopwords(x):
   list_token = []
   print(stopwords)
   for i in x:
       if i not in stopwords:
         list_token.append(i)


  return list_token
```

3. ตัวอย่างโค้ดเมท็อด remove_ parentheses(x):

```
def remove_parentheses(x):
  list_tokens = []
  for i in x:
    if len(i) <=1:
      continue
    else:
      list_tokens.append(i)
  return list_tokens
```

ภาคผนวก ค

ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความอย่างละเอียด

ตารางที่ ค.1 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 1

การสุ่มแบ่งข้อมูลแบบที่ 1							
ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 1	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 2	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	261	125	90	ไม่ชอบ	282	102	92
กลาง	89	469	217	กลาง	84	478	230
ชอบ	34	144	570	ชอบ	38	126	566
ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 3	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 4	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	271	118	73	ไม่ชอบ	269	116	76
กลาง	82	497	224	กลาง	83	495	195
ชอบ	45	141	547	ชอบ	35	127	602
ข้อมูล	แบบจำลองตอบ						
ส่วนที่ 5	ไม่ชอบ	กลาง	ชอบ				
ไม่ชอบ	288	136	79				
กลาง	88	449	192				
ชอบ	49	142	575				
การสุ่มแบ่งข้อมูลแบบที่ 2							
ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 1	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 2	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	250	133	74	ไม่ชอบ	267	120	96
กลาง	98	473	189	กลาง	80	477	216
ชอบ	60	132	590	ชอบ	35	136	571
ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 3	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 4	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	258	123	92	ไม่ชอบ	288	112	76
กลาง	77	499	218	กลาง	75	491	206
ชอบ	43	125	563	ชอบ	29	140	581

ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 5	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	286	112	91
กลาง	79	463	231
ชอบ	34	142	560

การสุ่มแบ่งข้อมูลแบบที่ 3

ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 1	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 2	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	269	121	87	ไม่ชอบ	302	116	93
กลาง	92	463	212	กลาง	76	503	227
ชอบ	36	133	586	ชอบ	39	122	520
ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 3	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 4	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	264	115	63	ไม่ชอบ	239	129	85
กลาง	106	468	199	กลาง	86	474	208
ชอบ	41	161	581	ชอบ	37	144	596

ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 5	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	288	113	94
กลาง	68	467	223
ชอบ	43	135	567

การสุ่มแบ่งข้อมูลแบบที่ 4

ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 1	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 2	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	282	105	84	ไม่ชอบ	288	124	83
กลาง	82	492	185	กลาง	87	454	206
ชอบ	36	143	590	ชอบ	35	140	581
ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 3	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 4	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	279	118	92	ไม่ชอบ	243	115	83
กลาง	74	467	230	กลาง	92	488	221
ชอบ	52	120	566	ชอบ	51	130	575

ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 5	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	271	131	80
กลาง	92	477	225
ชอบ	41	132	549

การสุ่มแบ่งข้อมูลแบบที่ 5

ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 1	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 2	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	282	119	79	ไม่ชอบ	268	117	92
กลาง	79	474	201	กลาง	78	451	219
ชอบ	39	157	569	ชอบ	40	125	608
ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 3	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 4	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	281	115	91	ไม่ชอบ	288	121	73
กลาง	96	466	217	กลาง	85	476	217
ชอบ	42	142	548	ชอบ	45	135	558

ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 5	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	244	116	92
กลาง	90	505	218
ชอบ	34	137	562

ตารางที่ ค.2 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 2

การสุ่มแบ่งข้อมูลแบบที่ 1

ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 1	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 2	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	155	281	40	ไม่ชอบ	164	267	45
กลาง	46	670	59	กลาง	38	665	89
ชอบ	25	327	396	ชอบ	33	299	398

ข้อมูลส่วนที่ 3	แบบจำลองตอบ			ข้อมูลส่วนที่ 4	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	160	266	36	ไม่ชอบ	158	265	38
กลาง	37	691	75	กลาง	34	661	78
ชอบ	47	294	392	ชอบ	32	291	441

ข้อมูลส่วนที่ 5	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	178	290	35
กลาง	40	607	82
ชอบ	28	322	416

การสุ่มแบ่งข้อมูลแบบที่ 2

ข้อมูลส่วนที่ 1	แบบจำลองตอบ			ข้อมูลส่วนที่ 2	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	138	292	27	ไม่ชอบ	150	287	46
กลาง	36	651	73	กลาง	43	647	83
ชอบ	37	331	414	ชอบ	25	290	427

ข้อมูลส่วนที่ 3	แบบจำลองตอบ			ข้อมูลส่วนที่ 4	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	173	261	39	ไม่ชอบ	179	259	38
กลาง	53	659	82	กลาง	36	667	69
ชอบ	40	296	395	ชอบ	28	310	412

ข้อมูลส่วนที่ 5	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	165	276	48
กลาง	33	673	67
ชอบ	37	313	386

การสุ่มแบ่งข้อมูลแบบที่ 3

ข้อมูลส่วนที่ 1	แบบจำลองตอบ			ข้อมูลส่วนที่ 2	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	160	283	34	ไม่ชอบ	189	277	45
กลาง	36	649	82	กลาง	35	694	77
ชอบ	37	301	417	ชอบ	32	287	362

ข้อมูลส่วนที่ 3	แบบจำลองตอบ			ข้อมูลส่วนที่ 4	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	159	248	35	ไม่ชอบ	138	277	38
กลาง	42	650	81	กลาง	52	649	67
ชอบ	32	334	417	ชอบ	25	344	408

ข้อมูลส่วนที่ 5	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	158	299	38
กลาง	38	648	72
ชอบ	35	283	427

การสุ่มแบ่งข้อมูลแบบที่ 4

ข้อมูลส่วนที่ 1	แบบจำลองตอบ			ข้อมูลส่วนที่ 2	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	156	273	42	ไม่ชอบ	167	293	35
กลาง	47	635	77	กลาง	45	621	81
ชอบ	36	315	418	ชอบ	37	298	421

ข้อมูลส่วนที่ 3	แบบจำลองตอบ			ข้อมูลส่วนที่ 4	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	156	295	38	ไม่ชอบ	159	249	33
กลาง	35	657	79	กลาง	45	682	74
ชอบ	32	296	410	ชอบ	36	321	399

ข้อมูลส่วนที่ 5	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	173	265	44
กลาง	32	676	86
ชอบ	25	310	387

การสุ่มแบ่งข้อมูลแบบที่ 5

ข้อมูลส่วนที่ 1	แบบจำลองตอบ			ข้อมูลส่วนที่ 2	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	169	286	25	ไม่ชอบ	170	275	32
กลาง	38	642	74	กลาง	42	632	74
ชอบ	40	313	412	ชอบ	27	323	423

ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 3	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 4	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	171	278	38	ไม่ชอบ	166	280	36
กลาง	49	642	88	กลาง	35	675	68
ชอบ	37	310	385	ชอบ	43	312	383
ข้อมูล	แบบจำลองตอบ						
ส่วนที่ 5	ไม่ชอบ	กลาง	ชอบ				
ไม่ชอบ	148	256	48				
กลาง	52	684	77				
ชอบ	22	298	413				

ตารางที่ ค.3 ตารางคอนฟิวชันผลลัพธ์การจำแนกข้อความด้วยเซตของคุณลักษณะแบบที่ 3

การสุ่มแบ่งข้อมูลแบบที่ 1							
ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 1	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 2	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	274	130	72	ไม่ชอบ	299	96	81
กลาง	98	491	186	กลาง	87	497	208
ชอบ	33	142	573	ชอบ	53	131	546
ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 3	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 4	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	285	115	62	ไม่ชอบ	280	119	62
กลาง	82	519	202	กลาง	91	508	174
ชอบ	46	130	557	ชอบ	40	126	598
ข้อมูล	แบบจำลองตอบ						
ส่วนที่ 5	ไม่ชอบ	กลาง	ชอบ				
ไม่ชอบ	306	134	63				
กลาง	98	453	178				
ชอบ	51	135	580				
การสุ่มแบ่งข้อมูลแบบที่ 2							
ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 1	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 2	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	255	140	62	ไม่ชอบ	290	119	74
กลาง	100	481	179	กลาง	84	490	199
ชอบ	58	136	588	ชอบ	36	133	573

ข้อมูลส่วนที่ 3	แบบจำลองตอบ			ข้อมูลส่วนที่ 4	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	278	115	80	ไม่ชอบ	301	111	64
กลาง	88	504	202	กลาง	79	510	183
ชอบ	58	125	548	ชอบ	30	135	585

ข้อมูลส่วนที่ 5	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	295	115	79
กลาง	88	489	196
ชอบ	42	145	549

การสุ่มแบ่งข้อมูลแบบที่ 3

ข้อมูลส่วนที่ 1	แบบจำลองตอบ			ข้อมูลส่วนที่ 2	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	285	129	63	ไม่ชอบ	312	120	79
กลาง	102	477	188	กลาง	81	518	207
ชอบ	49	134	572	ชอบ	46	123	512

ข้อมูลส่วนที่ 3	แบบจำลองตอบ			ข้อมูลส่วนที่ 4	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	277	115	50	ไม่ชอบ	253	128	72
กลาง	113	472	188	กลาง	94	485	189
ชอบ	42	162	579	ชอบ	40	150	587

ข้อมูลส่วนที่ 5	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	300	114	81
กลาง	75	486	197
ชอบ	42	136	567

การสุ่มแบ่งข้อมูลแบบที่ 4

ข้อมูลส่วนที่ 1	แบบจำลองตอบ			ข้อมูลส่วนที่ 2	แบบจำลองตอบ		
	ไม่ชอบ	กลาง	ชอบ		ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	293	105	73	ไม่ชอบ	305	116	74
กลาง	96	493	170	กลาง	98	459	190
ชอบ	43	144	582	ชอบ	41	146	569

ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 3	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 4	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	285	127	77	ไม่ชอบ	261	110	70
กลาง	69	503	199	กลาง	93	512	196
ชอบ	51	127	560	ชอบ	45	135	576

ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 5	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	284	129	69
กลาง	91	501	202
ชอบ	40	144	538

การสุ่มแบ่งข้อมูลแบบที่ 5

ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 1	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 2	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	297	122	61	ไม่ชอบ	270	126	81
กลาง	80	480	194	กลาง	89	465	194
ชอบ	44	154	567	ชอบ	49	131	593

ข้อมูล	แบบจำลองตอบ			ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 3	ไม่ชอบ	กลาง	ชอบ	ส่วนที่ 4	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	301	108	78	ไม่ชอบ	292	127	63
กลาง	95	482	202	กลาง	89	496	193
ชอบ	48	134	550	ชอบ	46	136	556

ข้อมูล	แบบจำลองตอบ		
ส่วนที่ 5	ไม่ชอบ	กลาง	ชอบ
ไม่ชอบ	264	107	81
กลาง	104	513	196
ชอบ	33	133	567

ประวัติผู้เขียน

Mr Tanasit Rengsomboonsuk
นายธนะสิทธิ์ เร่งสมบูรณ์สุข ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ที่อยู่คอนโด ไอดีโอ รัชดา-ห้วยขวาง เลขที่ 62/329 ถ.ประชาราษฎร์บาเพ็ญ แขวงห้วยขวาง เขตห้วย ขวาง กรุงเทพมหานคร 10310
มือถือ 090-1015356
email sprite_love@hotmail.com

Miss Kawintida Saiyot
นางสาวกวินธิดา สายยศ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่อยู่92 หมู่3 ตำบลสระว่านพระยา อำเภอครบุรี จังหวัดนครราชสีมา 30250 มือถือ 090-257-7715
email kawintida5@gmail.com