กลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์

EMOTION BEHAVIOR ANALYSIS MECHANISM BASED ON ONLINE SOCIAL MEDIA DATA

พิทวัส ชัยวุฒินันท์

โสพส เสียงไพโรจน์

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2563

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ปริญญานิพนธ์ปีการศึกษา 2563

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง กลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์

EMOTION BEHAVIOR ANALYSIS MECHANISM BASED ON ONLINE SOCIAL MEDIA DATA

ผู้จัดทำ

1. นายพิทวัส ชัยวุฒินันท์ รหัสนักศึกษา 60010714
2. นายโสพส เสียงไพโรจน์ รหัสนักศึกษา 60011121

_____อาจารย์ที่ปรึกษา

(รศ. ดร.ศักดิ์ชัย ทิพย์จักษุรัตน์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

กลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์

นายพิทวัส ชัยวุฒินันท์ 60010714

นายโสพส เสียงไพโรจน์ 60011121

รศ. ดร.ศักดิ์ชัย ทิพย์จักษุรัตน์ อาจารย์ที่ปรึกษา

ปีการศึกษา 2563

บทคัดย่อ

ปัจจุบันโรคซึมเศร้าเป็นปัญหาที่สำคัญปัญหาหนึ่งในสังคมและมีแนวโน้มที่จะเพิ่มขึ้น ดังนั้นถ้ามีกลไลการวิเคราะห์พฤติกรรมของบุคคลว่ามีแนวโน้มเป็นโรคซึมเศร้าน่าจะเป็นประโยชน์ และปัจจุบันรูปแบบการติดต่อระหว่างบุคคลหรือกลุ่มบุคคลจะใช้สื่อสังคมออนไลน์เป็นส่วนใหญ่ เราสามารถแสดงความคิดเห็นต่างๆ บนสื่อสังคมออนไลน์ได้อย่างอิสระ จากแนวคิดดังกล่าว จึงเป็นแรงบันดาลใจของงานวิจัยนี้ ในงานวิจัยนี้เราเสนอกลไลการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์ โดยเราจะเน้นสื่อสังคมออนไลน์ทวิตเตอร์เป็นหลัก โดยการประยุกต์ใช้การเรียนรู้ของเครื่อง (Machine Learning) และ การเรียนรู้เชิงลึก (Deep Learning) ในการวิเคราะห์ข้อความต่างๆ ที่บุคคลโพสต์ลงทวิตเตอร์ โดยกระบวนการทำงานจะใช้ข้อมูลทวีตที่ดึงมาจากทวิตเตอร์มาใช้เป็นชุดข้อมูลในขั้นตอนการเตรียมข้อมูลจะมีการทำการประมวลผลภาษาธรรมชาติ (Natural Language Processing) โดยจับคู่คำในข้อความกับ Subjectivity Dictionary ซึ่งจะมีคะแนนสำหรับคำที่บ่งบอกความรู้สึก คำนวณออกมาเป็นคะแนนแล้วจัดกลุ่มตามคะแนน แล้วนำข้อมูลส่วนหนึ่งนำเข้าไปฝึกฝนโมเดลแต่ละประเภทประกอบไปด้วย Naïve Bayes, ต้นไม้ตัดสินใจ, Random Forest และโครงข่ายประสาทแบบ Long short-term memory (LSTM) และข้อมูลอีกส่วนนำไปเป็นชุดทดสอบวัดประสิทธิภาพของโมเดล ซึ่งใช้วิธีวัดผล 2 รูปแบบได้แก่ เมทริกซ์ความสับสน (Confusion matrix) และ AUC (Area under ROC curve) โดยยึดถือตัววัดความถูกต้อง (Accuracy) เป็นหลัก ซึ่งจากการทดลองพบว่าโมเดล LSTM มีประสิทธิภาพมากที่สุด โดยมีความแม่นยำอยู่ที่ 92%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Emotion Behavior Analysis Mechanism based on Online Social Media Data

Mr. Pitawat Chaivutinun 60010714

Mr. Soros Seingpairod 60011121

Assoc.Prof. Dr.Sakchai Thipchaksurat Advisor

Academic Year 2020

ABSTRACT

Recently, major depression disorder is one of the major problems in society and it tends to increasing continuously. So, if there is a behavior analysis mechanism of the that person who is likely to have symptoms of depression disorder, it would be helpful. And nowadays, the most common form of contact between individuals or groups is a social media. We can express various opinions on social media freely. From that idea, this is the inspiration for this research. In this research, we propose a mechanism for analyzing emotional behavior based on social media data, and we focus on Twitter. By applying Machine Learning and Deep Learning to analyze various texts from a person who posts on Twitter. The progress will be used the tweet data extracted from Twitter to be used as a dataset. In data preparation, Natural Language Processing will be use to matching the words with the Subjectivity Dictionary, which will have a word score for indicating feeling. Calculated into scores and grouped by score. Part of the data will be trained by various models, each type of model comprising Naïve Bayes, Decision Tree, Random Forest and Long short-term memory (LSTM). And another part of the data was used to measure the model's performance. Then uses 2 types of measurement methods: confusion matrix and AUC (Area under ROC curve) by using accuracy as primary indicators. From the experiment, founding that the model LSTM was the most effective method with an accuracy of 92%.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

กิตติกรรมประกาศ

โครงงานเรื่องกลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์ (Emotion Behavior Analysis Mechanism based on Online Social Media Data) นั้นสำเร็จขึ้นได้โดยได้รับความช่วยเหลืออย่างดียิ่งจาก รศ. ดร.ศักดิ์ชัย ทิพย์จักษุรัตน์ อาจารย์ที่ปรึกษาโครงงาน ที่ได้ให้คำเสนอแนะ แนวคิดและให้ความรู้ในการจัดทำโครงงาน ตลอดจนการแก้ไขข้อบกพร่องต่างๆและช่วยดำเนินงานเอกสารมาโดยตลอด จนโครงงานนี้เสร็จสมบูรณ์ ผู้ศึกษาจึงขอกราบขอบพระคุณเป็นอย่างสูง

ท้ายสุดนี้คณะผู้จัดทำหวังเป็นอย่างยิ่งว่าโครงงานเรื่องกลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์ (Emotion Behavior Analysis Mechanism based on Online Social Media Data) จะเป็นประโยชน์ต่อการศึกษาค้นคว้าและเป็นประโยชน์ต่อผู้คนที่สนใจในเรื่องของการเรียนรู้ของเครื่อง การเรียนรู้เชิงลึก และโรคซึมเศร้า

พิทวัส ชัยวุฒินันท์

โสพส เสียงไพโรจน์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VI

สารบัญภาพ VII

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญ 1

1.2 วัตถุประสงค์การวิจัย 2

1.3 ขอบเขตการวิจัย 2

1.4 ข้อตกลงการวิจัย 3

1.5 ประโยชน์ที่คาดว่าจะได้รับ 3

1.6 แผนการดำเนินงาน 4

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง 7

2.1 แนวคิดที่สัมพันธ์กับเรื่องที่ทำการวิจัย 7

2.2 งานวิจัยที่เกี่ยวข้อง 15

บทที่ 3 การออกแบบ 20

3.1 ภาพรวมการออกแบบระบบ 20

3.2 การออกแบบโปรแกรมทำนายผล 21

3.3 การออกแบบเว็บแอปพลิเคชัน 30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use. Forbidden to modify the content, and cite the document when use.

บทที่ 4 ผลการทดลอง 44
4.1 แสดงภาพชุดข้อมูล 44
4.2 ผลการทดลองโมเดล 45
4.3 สรุปผลการทดลองโมเดล 50
4.4 การพิจารณาโมเดลที่เหมาะสม 50
4.5 ผลการทดลองทดสอบโมเดลกับชุดข้อมูล CLPsych 2015 51
4.6 เปรียบเทียบโครงงานกับงานวิจัยที่เกี่ยวข้อง 51
4.7 ผลการทดสอบเว็บแอปพลิเคชัน 53
บทที่ 5 สรุป 60
5.1 สรุปผลที่ได้จากการทดลอง 60
5.2 ปัญหาที่พบ 61
5.3 แนวทางในการพัฒนาต่อ 61

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญตาราง

ตาราง **หน้า**

1.1 แผนการดำเนินโครงงาน ภาคเรียนที่ 1/2563 5

1.2 แผนการดำเนินโครงงาน ภาคเรียนที่ 2/2563 6

2.1 เมทริกซ์ความสับสน (Confusion Matrix) 12

3.1 รูปแบบการแบ่งคลาสของข้อความทวีต 23

3.2 ความต้องการของเว็บแอปพลิเคชัน 31

4.1 สรุปค่าความแม่นยำของทุกโมเดล 50

4.2 สรุปค่า P, R และ F1 ของทุกโมเดล 50

4.3 ผลการทดลอง Classification of Depression on Social Media Using Text Mining 52

4.4 ผลการทดลอง Monitoring Tweets for Depression to Detect At-risk Users 52

4.5 ผลผลการทดลองการทำนายผลจากไทม์ไลน์ทวิตเตอร์ของผู้ใช้ 58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญรูป

รูป หน้า

2.1 ตัวอย่างการทำ Text Tokenization 8
2.2 ตัวอย่างการทำ Lemmatization 8
2.3 ตัวอย่างหาคะแนนเชิงขั้ว 9
2.4 สมการ Naive Bayes 10
2.5 ส่วนประกอบของต้นไม้ตัดสินใจ 10
2.6 ซัพพอร์ทเวกเตอร์แมชชีน 11
2.7 โครงข่ายประสาท 12
2.8 ROC curve 13
3.1 แผนภาพแสดงภาพรวมของการออกแบบระบบ 20
3.2 แผนภาพแสดงภาพรวมการสร้างโปรแกรมทำนายผล 21
3.4 แผนภาพแสดงขั้นตอนการทำความสะอาดข้อมูลตัวอักษร 21
3.5 ตัวอย่าง Subjectivity Dictionary 22
3.6 แผนภาพการแสดงขั้นตอนการหาขั้วของข้อความ 23
3.7 แผนภาพการแบ่งชุดข้อมูลทวีต 24
3.9 แผนผังแสดงการเทรนโมเดล 25
3.10 โครงสร้าง LSTM 26
3.11 โครงสร้าง LSTM แบบแสดง Input และ Output 26
3.12 แผนผังแสดงการวัดประสิทธิภาพของโมเดล 27
3.13 แผนผังแสดงการทำงานของโปรแกรมทำนายผลข้อความเดี่ยว 28
3.14 แผนผังแสดงการทำงานของโปรแกรมทำนายผลจากไทม์ไลน์ผู้ใช้ 28
3.15 แผนผังแสดงการหาอัตราส่วนข้อความที่เป็นขั้วลบเฉลี่ย 29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.16 แผนผังแสดงภาพรวมของการออกแบบเว็บแอปพลิเคชัน.....30
3.17 แผนผังแสดงการทำงานของเว็บแอปพลิเคชัน.....35
3.18 ส่วนติดต่อผู้ใช้หน้าหลัก.....36
3.19 ส่วนติดต่อผู้ใช้หน้าข้อมูลโรคซึมเศร้า.....37
3.20 ส่วนติดต่อผู้ใช้หน้าข้อมูลโรคซึมเศร้า (ก).....38
3.21 ส่วนติดต่อผู้ใช้หน้าข้อมูลโรคซึมเศร้า (ข).....38
3.22 ส่วนติดต่อผู้ใช้หน้าข้อมูลโรคซึมเศร้า (ค).....39
3.23 ส่วนติดต่อผู้ใช้หน้าแบบทดสอบ PHQ-9.....40
3.24 ส่วนติดต่อผู้ใช้หน้าเกี่ยวกับเรา.....41
3.25 แผนผังแสดงทำงานในส่วน Back-end.....42
3.26 แผนผังแสดงทำงานของ Python child process.....42
4.1 ตัวอย่างคำที่พบในข้อความทวีตที่มีความปกติ.....44
4.2 ตัวอย่างคำที่พบในข้อความทวีตที่มีความซึมเศร้า.....44
4.3 เมทริกซ์ความสับสน Multinomial Naïve Bayes.....45
4.4 เมทริกซ์ความสับสน Multinomial Naïve Bayes.....45
4.5 Multinomial Naïve Bayes AUC.....45
4.6 Multinomial Naïve Bayes AUC.....45
4.7 เมทริกซ์ความสับสน Decision Tree.....46
4.8 เมทริกซ์ความสับสน Decision Tree.....46
4.9 Decision Tree AUC.....46
4.10 Decision Tree AUC.....46
4.11 เมทริกซ์ความสับสน Random Forest.....47
4.12 เมทริกซ์ความสับสน Random Forest.....47
4.13 Random Forest AUC.....47

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.14 Random Forest AUC 47
4.15 เมทริกซ์ความสับสน SVM 48
4.16 เมทริกซ์ความสับสน SVM 48
4.17 SVM AUC 48
4.18 SVM AUC 48
4.19 เมทริกซ์ความสับสน LSTM 49
4.20 เมทริกซ์ความสับสน LSTM 49
4.21 LSTM AUC 49
4.22 LSTM AUC 49
4.23 การทดลองทดสอบโมเดลกับชุดข้อมูล CLPsych 2015.....51
4.24 การทดสอบเว็บหน้าทำนายโรคซึมเศร้าด้วยบัญชีทวิตเตอร์.....53
4.25 การทดสอบเว็บหน้าทำแบบทดสอบ PHQ-9.....54
4.26 การทดสอบเว็บหน้าข้อมูลโรคซึมเศร้า.....54
4.27 การทดสอบการเปลี่ยนภาษา.....55
4.28 การทดสอบใช้งานบนเดสก์ท็อป.....55
4.29 การทดสอบใช้งานบนแท็บเล็ต.....56
4.30 การทดสอบใช้งานบนสมาร์ทโฟน.....56
4.31 ตัวอย่าง POST Request ไปที่ส่วนต่อประสานโปรแกรม (ก).....59
4.32 Back-end Terminal เมื่อมี Request เข้ามา.....59
4.33 ตัวอย่าง POST Request ไปที่ส่วนต่อประสานโปรแกรม (ข).....59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญ

ในสังคมปัจจุบัน ประชากรกว่า 4.54 พันล้านคนหรือราวๆ 59 เปอร์เซ็นต์ทั่วโลกสามารถเข้าถึงอินเทอร์เน็ตได้และกว่า 3.8 พันล้านคนใช้งานสื่อสังคมออนไลน์ [1] ทำให้เกิดสื่อสังคมออนไลน์มากมายในปัจจุบัน ตัวอย่างเช่น เฟสบุ๊ค (Facebook) ทวิตเตอร์ (Twitter) อินสตาแกรม (Instagram) เป็นต้น กล่าวคือปัจจุบันมีประชากรกว่าครึ่งที่ใช้งานสื่อสังคมออนไลน์ และประชากรส่วนใหญ่เหล่านั้นล้วนได้รับอิทธิพลจากสื่อสังคมออนไลน์

ปัจจุบันสื่อสังคมออนไลน์ส่งผลกระทบต่อพฤติกรรมทางสังคมในหลายๆด้าน ไม่ว่าจะเป็นความเป็นส่วนตัว ปฏิสัมพันธ์ ภาพลักษณ์ หรือการแสดงออก [2] เนื่องจากสื่อสังคมออนไลน์ถือเป็นพื้นที่ในการแสดงออกที่ผู้คนทุกคนสามารถเข้าถึงได้ทำให้มีคนมากมายใช้สื่อสังคมออนไลน์เป็นพื้นที่ในการแสดงออกทัศนะ ความคิดเห็น อารมณ์ ความเป็นตัวตน โดยทั้งหมดเป็นการแสดงออกถึงตัวตนของคนนั้นๆ [3] ผู้คนส่วนใหญ่มีแนวโน้มที่จะแสดงออกถึงทัศนะได้ตรงกับความเป็นจริงมากกว่าเวลาปกติเมื่อคนๆนั้นแสดงความคิดเห็นบนสื่อสังคมออนไลน์ [4]

จากเหตุผลดังกล่าว จึงมีผู้ผลิตสินค้าหรือผู้ให้บริการต่างๆ ได้นำข้อมูลที่อยู่บนสื่อสังคมออนไลน์ มาใช้ในการวิเคราะห์เพื่อประโยชน์ทางการค้าหรือการให้บริการ เช่นการวิเคราะห์กลุ่มตลาดของสินค้า การวิเคราะห์แนวโน้มสินค้าที่ผู้บริโภคต้องการในอนาคต การวางแผนโฆษณาที่ตรงกลุ่มเป้าหมายของสินค้า หรือการวิเคราะห์ผลตอบรับต่อสินค้า/โฆษณาที่ปล่อยออกไป ทำให้เกิดเครื่องมือ เช่น Sprout Social [5], Awario [6], Snaplytics [7], Google Analytics [8] เป็นต้นที่ถูกใช้ในการวิเคราะห์สื่อสังคมออนไลน์ แต่ประเด็นที่คนส่วนมากมองข้ามไปคือสื่อสังคมออนไลน์ไม่ได้มีประโยชน์แค่ในด้านเศรษฐกิจเท่านั้น แต่สื่อสังคมออนไลน์ยังมีประโยชน์ในด้านการวิเคราะห์ถึงพฤติกรรมทางสังคมและสภาพจิตใจของคน เนื่องจากด้วยสื่อสังคมออนไลน์เป็นแหล่งข้อมูลที่คนจะแสดงออกถึงความคิดของตนได้อย่างตรงไปตรงมา ทำให้สื่อสังคมออนไลน์เป็นแหล่งข้อมูลที่ทรงคุณค่าแก่การวิเคราะห์สภาพจิตใจของมนุษย์ เพื่อสังเกตว่ามนุษย์ผู้นั้นมีความผิดปกติทางจิตโดยเฉพาะโรคซึมเศร้าหรือไม่

โรคซึมเศร้านั้นในปัจจุบันถือเป็นสาเหตุสำคัญสาเหตุหนึ่งที่ทำให้เกิดการฆ่าตัวตาย สถิติในปี 2562 สถิติการฆ่าตัวตายคือ 14.4 คน ต่อคนไทย 1 แสนคน ในจำนวนนี้มีผู้ป่วยโรคซึมเศร้าอยู่ถึง 7.8% โดยผู้ป่วยโรคซึมเศร้ามีอยู่ถึง 4% ของประชากรโลก และ 1.8% ของประชากรไทย [9] กล่าวคือคนไทยราวๆ 1,000,000 คนป่วยเป็นโรคซึมเศร้า และผู้ป่วยส่วนใหญ่เหล่านี้ไม่ได้รับการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

รักษาที่ถูกต้อง มีเพียง 4.9% [10] เท่านั้นที่ได้รับการส่งต่อไปรับการตรวจวินิจฉัยและรักษาจากโรงพยาบาลจิตเวช ซึ่งสาเหตุที่ไม่ยอมเข้ารับการตรวจโรคเกิดได้หลายปัจจัยตั้งแต่ ค่าใช้จ่าย ความลำบากในการเดินทาง จำนวนคลินิกจิตเวชที่ไม่เพียงพอ ไปถึงถึงความกังวลต่อสายตาของคนในสังคม

โครงงาน "กลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์ (Emotion Behavior Analysis Mechanism based on Online Social Media Data)" นี้ถูกจัดทำขึ้นเพื่อทำการศึกษาการใช้งานการเรียนรู้ของเครื่อง (Machine Learning) และการเรียนรู้เชิงลึก (Deep Learning) เพื่อวิเคราะห์ข้อความบนสื่อสังคมออนไลน์ เพื่อหาความผิดปกติทางจิตของผู้ใช้สื่อสังคมออนไลน์ กล่าวคือหากมีผู้ใช้ที่เป็นโรคซึมเศร้าที่มีโอกาสที่จะฆ่าตัวตายหรือมีแนวโน้มที่จะก่อเหตุก่อการร้ายเนื่องจากสภาวะจิตที่ไม่ปกติ เราจะพัฒนาระบบที่จะวิเคราะห์และค้นหาคนที่มีแนวโน้มเหล่านั้น โดยผลสำเร็จจะเป็นเว็บแอปพลิเคชันที่จะวิเคราะห์จากบัญชีทวิตเตอร์ที่ผู้ใช้ใส่ลงไป แล้วคาดคะเนแนวโน้มการป่วยเป็นโรคซึมเศร้าของบัญชีทวิตเตอร์นั้น

1.2 วัตถุประสงค์การวิจัย

1.) เพื่อศึกษาพฤติกรรมของบุคคลที่มีแนวโน้มเป็นโรคซึมเศร้า
2.) เพื่อพัฒนาตัวชี้วัดทั่วไปของการป่วยเป็นโรคซึมเศร้าที่มีประสิทธิภาพในการวิเคราะห์ข้อความ
3.) เพื่อศึกษาหลักการและทฤษฎีของการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึก
4.) ใช้วิธีการทางการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกเพื่อใช้ค้นหาบุคคลที่มีแนวโน้มจะเป็นโรคซึมเศร้าได้

1.3 ขอบเขตการวิจัย

ผู้วิจัยได้กำหนดขอบเขตการวิจัยและแบ่งออกเป็น 3 ด้านได้แก่ ขอบเขตของเนื้อหา ขอบเขตของแหล่งข้อมูล และขอบเขตของระยะเวลา โดยมีรายละเอียดดังนี้

1.3.1 ขอบเขตของเนื้อหา

พัฒนาโมเดลด้วยกระบวนการการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกประยุกต์กับการทำการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ภาษาอังกฤษ โดยจับคู่คำในข้อความที่นำมาจากทวิตเตอร์กับ Subjectivity Dictionary แล้วนำข้อมูลส่วนหนึ่งนำเข้าไปฝึกฝนกับโมเดลแต่ละประเภทประกอบไปด้วย Naïve Bayes, Decision Tree, Random Forest และ LSTM เพื่อสร้างโมเดลที่จะคาดคะเนโอกาสการป่วยเป็นโรคซึมเศร้าของเจ้าของข้อความ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เพื่อให้ง่ายต่อการใช้งานสำหรับผู้ใช้งาน เราได้พัฒนาเว็บแอปพลิเคชันโดยการนำโมเดลที่เราพัฒนาขึ้นเข้าระบบที่จะวิเคราะห์ความน่าจะเป็นที่จะเป็นโรคซึมเศร้าของบุคคลที่นำข้อความบนสื่อสังคมออนไลน์ (ทวิตเตอร์) มาวิเคราะห์โดยจะรองรับเฉพาะบัญชีทวิตเตอร์ที่ใช้งานภาษาอังกฤษเป็นหลัก และจะส่งผลลัพธ์ให้ผู้ใช้ทราบความเสี่ยงที่จะเป็นโรคซึมเศร้า

1.3.2 ขอบเขตของแหล่งข้อมูล

กลุ่มตัวอย่าง คือ ผู้ใช้งานทั่วไปที่ใช้สื่อสังคมออนไลน์ โดยอ้างอิงข้อมูลจากการทวีตจากบุคคลทั่วไปซึ่งสตรีมจากทวิตเตอร์ด้วยส่วนต่อประสานทวิตเตอร์(Twitter API)

เนื่องด้วยความแม่นยำของการวิเคราะห์ขึ้นอยู่กับปริมาณข้อมูลที่หาได้ งานวิจัยจะมีการวิเคราะห์ข้อความบนสื่อสังคมออนไลน์และจะมีข้อมูลการป่วยเป็นโรคซึมเศร้าของเจ้าของข้อความด้วยข้อกำหนดทางกฎหมายทางผู้วิจัยจึงต้องทำการขออนุญาตก่อนจึงจะนำข้อมูลมาใช้ได้ ส่งผลให้ข้อมูลที่สามารถนำมาวิเคราะห์ได้อย่างอิสระจึงมีจำนวนจำกัด ทำให้ข้อมูลที่นำมาวิเคราะห์จึงอาจไม่ได้มีปริมาณมากเท่าที่ควร

1.3.3 ขอบเขตของระยะเวลา

ระยะเวลาที่ใช้ในการวิจัยที่ถูกกำหนดไว้คือช่วงเดือนสิงหาคม 2563 จนถึงเดือนพฤศจิกายน 2563 สำหรับโครงงาน 1 และช่วงเดือนมกราคม 2563 จนถึงเดือนมีนาคม 2563 จะเป็นการดำเนินงานสำหรับโครงงาน 2

1.4 ข้อตกลงการวิจัย

ข้อมูลเก็บรวบรวมขากเจ้าของข้อความมาช่วงอายุและเพศที่แตกต่างกันไป โดยข้อมูลที่นำมาใช้ในงานวิจัยได้รับการอนุญาตจากเจ้าของข้อมูลแล้ว หรือมาจากแหล่งข้อมูลที่ปกปิดแหล่งที่มาอยู่แล้ว ข้อมูลจึงสามารถนำมาวิเคราะห์ได้อย่างอิสระตราบเท่าที่ไม่เป็นการเปิดเผยชื่อและอาการป่วยของเจ้าของข้อมูล โดยการเก็บรวบรวมข้อมูลเจ้าของข้อความมาช่วงอายุและเพศที่แตกต่างกันไป

เนื่องด้วยจริยธรรมทางการวิจัย ทางผู้วิจัยจึงต้องปิดบังชื่อของเจ้าของข้อความและอาการป่วยที่ไม่ได้รับการอนุญาติให้เปิดเผยชื่อเป็นความลับ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

สังคมทุกวันนี้มีผู้คนมากมายที่ต้องประสบกับปัญหาสุขภาพทางจิตและปัญหาความเครียดในชีวิตประจำวัน ซึ่งส่วนมากปัญหาเหล่านั้นไม่สามารถระบายไปให้ใครฟังได้ ทำให้ปัจจุบันมีคนมากมายที่ประสบกับปัญหาดังกล่าวจนก่อให้เกิดเป็นโรคซึมเศร้า และหลายๆคนที่เป็นโรคซึมเศร้าก็ไม่ได้มีโอกาสไปปรึกษากับจิตแพทย์ เนื่องด้วยทั้งสภาพสังคมและค่านิยมที่มองผู้ป่วยโรคซึมเศร้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เป็นคนไม่ปกติ ส่งผลให้ในสังคมปัจจุบันมีผู้ป่วยเป็นโรคซึมเศร้าที่ยังไม่รู้ตัวอีกเป็นจำนวนมาก หากผู้ป่วยยังไม่ได้รับการรักษาที่ถูกต้อง โรคซึมเศร้าอาจทวีความรุนแรงจนไม่สามารถใช้ชีวิตได้ตามปกติ

โครงงานนี้จึงมุ่งที่จะแก้ปัญหาดังกล่าว โดยการใช้การเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกในการวิเคราะห์ข้อความบนสื่อสังคมออนไลน์เพื่อหาความน่าจะเป็นที่จะเป็นโรคซึมเศร้า โดยผู้ใช้จะได้ผลลัพธ์ที่มีความแม่นยำและมีความเป็นส่วนตัวในการวิเคราะห์โดยไม่ต้องต้องกังวลต่อสายต่อของคนรอบข้าง และสามารถช่วยวิเคราะห์ให้กับคนใกล้ตัวตัวเพื่อระวังความเสี่ยงให้ได้ อย่างไรก็ตามหากผู้ถูกวิเคราะห์มีโอกาสเป็นโรคซึมเศร้าสูงควรจะไปปรึกษากับจิตแพทย์อย่างจริงจังต่อไป

1.6 แผนการดำเนินงาน

แผนการดำเนินงานจะถูกนำเสนอในรูปแบบตาราง โดยจะใช้สีเพื่อบอกระดับความสำคัญจากมากไปน้อยดังนี้ สีแดง > สีส้ม > สีเหลือง และใช้สีดำในการบอกกลุ่มของงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ID	Task Name	Aug 2020				Sep 2020				Oct 2020				Nov 2020	
		3	10	17	24	31	7	14	21	28	5	12	19	26	2	9
	วางแผน															
1	ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง															
2	เลือกภาษาและเครื่องมือที่ใช้พัฒนา															
3	ศึกษา Algorithms ที่ใช้ในการสร้าง Model															
	เตรียมข้อมูล															
4	ค้นหาและรวบรวมชุดข้อมูล															
5	คลีน แยกหมวดหมู่และจัดระเบียบชุดข้อมูล															
	พัฒนา model															
6	วางแผนการพัฒนาโมเดล															
7	ทบทวน Algorithms															
8	ทำ Normalize ข้อมูล															
9	Map คำศัพท์กับ Subjectivity Dictionary															
10	ใช้วิธี Lemmatization เพื่อแปลงคำให้อยู่ในรูป Basic form															
11	คำนวณ Score ของ Sentence															
12	นำโมเดลไปทดลองกับ Naïve Bayes															
13	นำโมเดลไปทดลองกับ Decision Tree															
14	นำโมเดลไปทดลองกับ LSTM															
15	ทดสอบและเปรียบเทียบประสิทธิภาพของโมเดล															
16	ปรับปรุงโมเดล															
17	**เขียนรายงานโครงงาน 1**															

ตาราง 1.1 แผนการดำเนินโครงงาน ภาคเรียนที่ 1/2563

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ID	Task Name	Jan 2021			Feb 2021				Mar 2021			
		17	24	31	7	14	21	28	7	14	21	28
1	รวบรวมชุดข้อมูล เพิ่มเติมสำหรับทดสอบ											
2	ใช้ข้อมูลจริงทดสอบและปรับปรุง model											
	พัฒนาเว็บแอปพลิเคชัน											
3	ศึกษาการพัฒนา เลือกภาษาและเครื่องมือที่เหมาะสม											
4	วางแผนการพัฒนา											
5	ออกแบบเว็บแอปพลิเคชัน											
6	พัฒนาและทดสอบเว็บแอปพลิเคชัน											
7	เชื่อมระบบกับโมเดลที่พัฒนา											
8	ทดสอบเว็บแอปพลิเคชันกับ Model											
9	**เตรียมรายงานโครงงาน 2 และปริญญานิพนธ์**											

ตาราง 1.2 แผนการดำเนินโครงงาน ภาคเรียนที่ 2/2563

การพัฒนาจะแบ่งเป็น 3 ส่วนใหญ่ๆ ได้แก่

1.) พัฒนาโมเดลคาดคะเนโอกาสเป็นโรคซึมเศร้า

2.) เว็บแอปพลิเคชันสำหรับผู้ใช้

3.) เชื่อมโมเดลกับเว็บแอปพลิเคชัน

ซึ่งในโครงงาน 1 จะเน้นไปที่การพัฒนาโมเดลคาดคะเนโอกาสเป็นโรคซึมเศร้า และจะพัฒนาเว็บแอปพลิเคชันสำหรับผู้ใช้และเชื่อมโมเดลกับเว็บแอปพลิเคชันในโครงงาน 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

ในบทนี้ ผู้วิจัยได้นำเสนองานวิจัยที่เกี่ยวกับการวิเคราะห์โรคซึมเศร้าด้วยการทำเหมืองข้อมูลผ่านกระบวนการทางการเรียนรู้ของเครื่อง (Machine Learning) และ การเรียนรู้เชิงลึก (Deep Learning), กระบวนการนำ การประมวลผลภาษาธรรมชาติ (Natural Language Processing) เพื่อใช้วิเคราะห์รูปแบบของภาษา, เทคนิคการทำ Dictionary เฉพาะทางสำหรับการวิเคราะห์สื่อสังคมออนไลน์ กำหนดหลักเกณฑ์และมาตรฐานการประเมินโรคซึมเศร้าแบบต่าง รวมไปถึงความแตกต่าง ข้อดี ข้อเสีย ของตัวชี้วัดทั่วไปของการป่วยเป็นโรคซึมเศร้า และความเชื่อมโยงกันระหว่างโรคซึมเศร้าและการใช้งานสื่อสังคมออนไลน์ เป็นต้น งานวิจัยได้มีการประยุกต์ใช้แนวคิดจากงานวิจัยเหล่านี้เพื่อทำการพัฒนาและปรับปรุง

2.1 แนวคิดที่สัมพันธ์กับเรื่องที่ทำการวิจัย

2.1.1 Typology of affective states

Scherer's (1984) [11] เป็นนักจิตวิทยาที่ทำการแบ่งอารมณ์ความรู้สึกที่มีผลต่อการแสดงความคิดเห็นของมนุษย์ (Typology of affective states) โดยแบ่ง 5 ปัจจัย ได้แก่ Emotion, Mood, Interpersonal stances, Attitudes, Personal traits รายละเอียดของปัจจัยทั้ง 5 สามารถอธิบายได้ดังนี้

- Emotion คือการตอบสนองทางร่างกายและระบบภายในของร่างกายที่มีต่อเหตุการณ์ต่างๆ รอบตัวที่อาจมีความสำคัญ เช่น โกรธมีความสุข แปลกใจ เศร้า อาย กลัว

- Mood คือสภาพทางอารมณ์อ่อนๆ มีผลในเวลาค่อนข้างนาน โดยส่วนใหญ่แล้วจะไม่รู้สาเหตุ เช่น ร่าเริง ซึมเศร้า อารมณ์เสีย ล่องลอย เบิกบาน ง่วง

- Interpersonal Stances คือท่าทางการแสดงออกทางอารมณ์เมื่อมีการปฏิสัมพันธ์กับผู้อื่น เช่น ห่างเหิน เย็นชา อบอุ่น รำคาญ จีบ

- Attitudes คือความเชื่อที่รับอิทธิพลมาจากอารมณ์ ความชอบส่วนตัว เช่นชอบ เกลียด ปรารถนา โหยหา

- Personality traits (ลักษณะทางบุคลิกภาพ) ลักษณะทางอารมณ์ที่ค่อนข้างเสถียรไม่เปลี่ยนแปลงง่าย และแนวโน้มเชิงพฤติกรรมที่เป็นลักษณะประจำของคนๆหนึ่ง เช่น เป็นมิตร ขี้อิจฉา ขี้หึง ซุ่มซ่าม บ้าพลัง

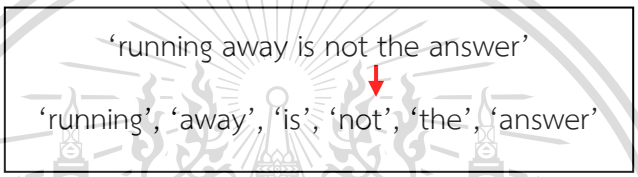
โดยในงานวิจัยนี้เราได้มุ่งเน้นในการหา Mood เกี่ยวกับอาการซึมเศร้าตาม Typology of affective states

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.1.2 การประมวลผลภาษาธรรมชาติ

การประมวลผลภาษาธรรมชาติ (Natural Language Processing) หรือที่รู้จักกันว่า NLP คือ กระบวนการที่ทำให้เครื่องคอมพิวเตอร์สามารถเข้าใจภาษาที่มนุษย์ใช้ในการสื่อสารหรือที่รู้จักกันในชื่อว่า “ภาษาธรรมชาติ” โดยมีวัตถุประสงค์เพื่อให้สามารถใช้ภาษาธรรมชาติในการปฏิสัมพันธ์ระหว่างมนุษย์กับคอมพิวเตอร์ได้ โดย NLP จะใช้กระบวนการในการย่อยประโยคที่มนุษย์ใช้ในการสื่อสารลงเป็นส่วนย่อย เพื่อให้ง่ายต่อการทำความเข้าใจของอุปกรณ์ แล้ววิเคราะห์ความสัมพันธ์ระหว่างคำหรือหน่วยย่อย ซึ่งทำให้คอมพิวเตอร์สามารถเข้าใจ และตีความภาษาธรรมชาติได้ กระบวนการทาง NLP ที่เราได้นำมาประยุกต์ใช้ ประกอบไปด้วย

- Text Tokenization คือ การตัดข้อความ ประโยค หรือย่อหน้า ให้แยกออกมาเป็นคำศัพท์ ซึ่งเป็นหน่วยย่อยสามารถเอาไปประมวลผลได้



รูป 2.1 ตัวอย่างการทำ Text Tokenization

- Lemmatization คือ การเปลี่ยนรูปของคำศัพท์ให้อยู่ในรูปพื้นฐาน (Basic Form) ซึ่งจะช่วยลดปริมาณคลังคำศัพท์และจำนวนฟีเจอร์ (Feature) ทำให้ชุดข้อมูลมีขนาดเล็กลง

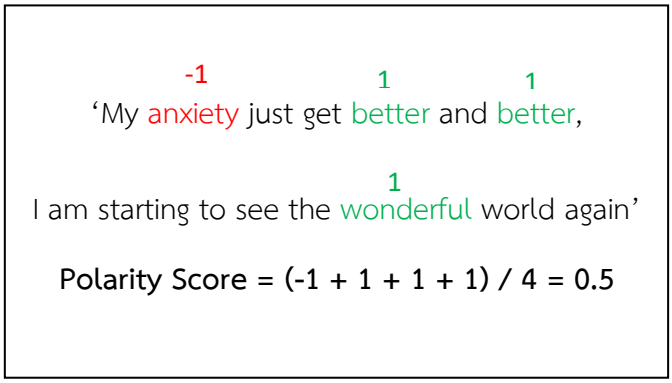
รูป 2.2 ตัวอย่างการทำ Lemmatization

2.1.3 Sentiment Analysis

Sentiment Analysis [12] คือ การวิเคราะห์ความรู้สึก วิเคราะห์อารมณ์จากข้อความ เพื่อบ่งบอกถึงความรู้สึกของผู้คนที่มีต่อบางสิ่งบางอย่างสามารถแบ่งออกเป็นขั้ว (Polarity) ได้เป็น 3 ขั้ว คือ Positive (ในทางที่ดี), Negative (ในทางที่ไม่ดี) และ Neutral (เป็นกลาง) โดยวิธีที่ผู้วิจัยได้ทำการใช้เพื่อคำนวณคะแนนของประโยค และใช้ในการแบ่งขั้วคือวิธีการใช้ Subjectivity Dictionary

- Subjectivity Dictionary คือ พจนานุกรมที่รวบรวมคำศัพท์ที่บ่งบอกความรู้สึกทางจิตใจพร้อมทั้งคะแนนของแต่ละคำศัพท์
- Calculate Polarity from Subjectivity Dictionary คือวิธีการคำนวณคะแนน เพื่อแบ่งขั้วของประโยคโดยใช้วิธีการจับคู่คำในประโยคกับคำที่มีอยู่ใน Subjectivity Dictionary แล้วหาผลรวมของคะแนนในประโยคแล้วหารด้วยจำนวนคำที่พบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้ทำ ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 2.3 ตัวอย่างหาคะแนนเชิงขั้ว

ซึ่งการจัดกลุ่มประโยคจะจัดจากคะแนนเชิงขั้ว (Polarity Score) ที่ได้คำนวณออกมาโดย ถ้ามากกว่า 0 จะอยู่ในกลุ่ม Positive ถ้าเท่ากับ 0 จะอยู่ในกลุ่ม Neutral และถ้าน้อยกว่า 0 จะอยู่ในกลุ่ม Negative โดยอ้างอิงการจัดกลุ่มจากคุณ Ashwanth Ramji [13] โดยช่วงการแบ่งกลุ่มที่กล่าวมาข้างต้นเป็นเพียงการแบ่งแบบพื้นฐาน ซึ่งสามารถแบ่งช่วงการแบ่งกลุ่มให้ละเอียดกว่านี้ได้โดยอาจอาศัยข้อมูลผู้ป่วยจริงมาใช้ปรับค่าได้ เป็นต้น

2.1.4 การทำเหมืองข้อมูล

คำนาย อภิปรัชญาสกุล (2557) [14] ได้ให้คำนิยามความหมายของเหมืองข้อมูลไว้ว่าคือ การหาคำตอบจากข้อมูลที่ซ่อนอยู่โดยใช้ซอฟต์แวร์ในการสร้างชุดคำถามขึ้นมาเพื่อหาคำตอบนั้น ตามรูปแบบที่กำหนดไว้ กฎเกณฑ์สำหรับการอ้างอิงในฐานข้อมูลขนาดใหญ่ และความสัมพันธ์ระหว่างข้อมูล โดยผลลัพธ์คือการคาดการณ์โอกาสและแนวโน้มของเหตุการณ์ที่จะเกิดขึ้นในอนาคตที่มีผลมาจากข้อมูลที่เก็บไว้ สร้างความได้เปรียบทางธุรกิจ ซึ่งสามารถนำการคาดการณ์ที่ได้รับจากข้อมูลไปใช้เป็นเครื่องมือในการสนับสนุนการตัดสินใจของผู้บริหาร นอกจากนั้นยังสามารถตอบคำถามในทางธุรกิจได้หลากหลาย และสามารถคาดการณ์โอกาสที่จะเกิดเหตุการณ์ในอนาคตได้อย่างแม่นยำ

2.1.4.1 Naive Bayes

เป็นเทคนิคการทำเหมืองข้อมูลแบบหนึ่งที่ไม่มีการหมุนวนที่ซับซ้อน โดยโมเดลจะอ้างอิงพื้นฐานของทฤษฎีของ Bayes และสมมติว่าการเกิดของเหตุการณ์ต่างๆเป็นอิสระต่อกัน เพื่อนำมาแยกประเภทข้อมูลผ่านหลักความน่าจะเป็น Naive Bayes [15] เป็นเทคนิคที่มีประโยชน์กับชุดข้อมูลขนาดใหญ่ ซึ่งจะทำให้การประมวลผลบนชุดข้อมูลขนาดใหญ่ทำงานได้อย่างมีประสิทธิภาพและรวดเร็ว หลักการคำนวณจะอ้างอิงจากทฤษฎีโดยสันนิษฐานว่าค่าที่เกิดจากตัวที่ใช้ทำนายเป็นอิสระต่อกัน โดยสามารถเขียนเป็นสมการได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

$$P(c|x) = \frac{P(x|c) \times P(c)}{P(x)} \quad \text{..... (1)}$$

รูป 2.4 สมการ Naive Bayes

P(c | x) คือค่าความน่าจะเป็นที่ข้อมูลที่มี attribute เป็น x จะมีคลาส c

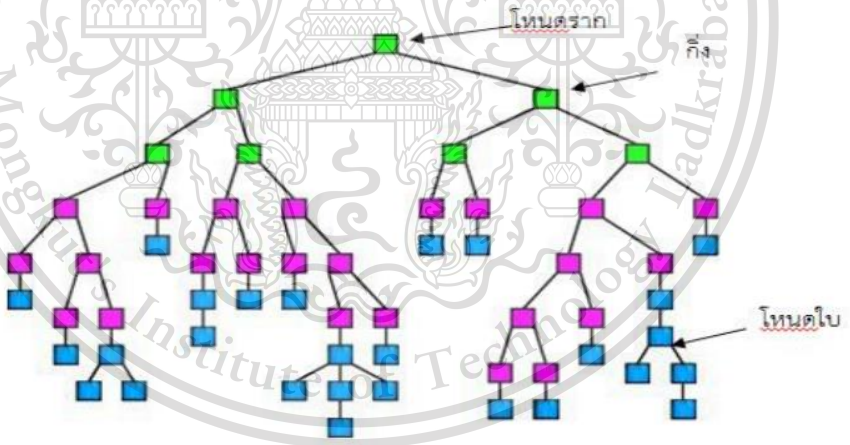
P(x | c) คือ ค่าความน่าจะเป็นที่ข้อมูลในชุดข้อมูลสอนที่มีคลาส c และมี attribute x โดยที่ $x = x_1 \cap x_2 \ldots \cap x_M$ โดยที่ M คือจำนวนของ attribute

P(c) คือ ค่าความน่าจะเป็นของคลาส C

P(x) คือ ค่าความน่าจะเป็นของ attribute x

2.1.4.2 ต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจ (Decision Tree) [16] เป็นเทคนิคทำเหมืองข้อมูลที่รู้จักกันดีมากที่สุด มีความสามารถในการจัดหมวดหมู่ข้อมูลเพื่อใช้สร้างต้นไม้ที่มีความสามารถในการพยากรณ์และตรวจสอบข้อมูล โดยใช้กฎในลักษณะ “ถ้า...แล้ว...” มาประกอบกันเพื่อสร้างต้นไม้ตัดสินใจ โดยจะเริ่มที่โหนดเริ่มต้นที่เรียกว่าโหนดราก หลังจากนั้นก็จะแตกกิ่งของ Attribute ไปเรื่อยๆ ดังแสดงในรูปที่ 2.5 จนแบ่งคลาสได้ชัดเจน ซึ่งต้นไม้ตัดสินใจจะประกอบด้วยองค์ประกอบดังนี้



รูป 2.5 ส่วนประกอบของต้นไม้ตัดสินใจ

- โหนด (Node) คือ คุณลักษณะ (Feature) ที่จะใช้ในการแบ่งข้อมูลว่าจะให้ไปในทิศทางใด โดยมีโหนดที่อยู่บนสุดเรียกว่าโหนดราก (root node)

- กิ่ง (Branch) คือ ตัวเชื่อมระหว่างโหนดที่แตกออกมา โดยจำนวนของกิ่งจะเท่ากับสมบัติของคุณสมบัติ (Feature)

- ใบ (Leaf) คือ กลุ่มของผลลัพธ์ ซึ่งโหนดที่อยู่ล่างสุดเรียกว่าโหนดใบ (Leaf Node) ซึ่งจะเป็นโหนดที่แสดงผลลัพธ์ของการกระทำตามเงื่อนไขที่เกิดขึ้น

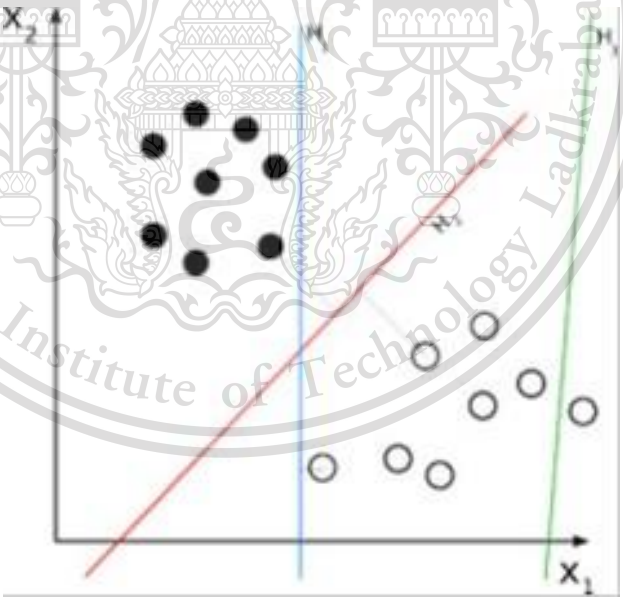
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.1.4.3 Random Forest

Random Forest [17] เป็นเทคนิคสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจหลายๆต้น โดยสุ่มข้อมูลจากชุดข้อมูลการเรียนรู้ (Training) และสุ่มเลือกฟีเจอร์ต่าง ๆ ออกมาเป็นหลายๆชุด แล้วนำคำตอบจากต้นไม้ตัดสินใจแต่ละต้นมาวิเคราะห์เพื่อหาค่าที่เหมาะสมที่สุด

2.1.4.4 ซัพพอร์ทเวกเตอร์แมชชีน

ซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine) หรือที่รู้จักกันว่า SVM [18] เป็นเทคนิคการทำเหมืองข้อมูลที่นิยมใช้แก้ปัญหาการจำแนกข้อมูล ซัพพอร์ทเวกเตอร์แมชชีนมักถูกใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูล โดยจะสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนโดยใช้หลักการของการหาสัมประสิทธิ์ของสมการดังแสดงในรูปที่ 2.6 จากนั้นจะให้ระบบเรียนรู้โดยให้ความสำคัญไปที่การเลือกเส้นที่จะแบ่งแยกแยะกลุ่มข้อมูลได้ดีที่สุด แนวความคิดของซัพพอร์ทเวกเตอร์แมชชีนเกิดจากการที่นำค่าของกลุ่มข้อมูลมาวางลงในฟีเจอร์สเปซ (Feature Space) จากนั้นจะสร้างเส้นแบ่ง (Hyperplane) ที่เป็นเส้นตรงขึ้นมาเพื่อแบ่งข้อมูลทั้งสองออกจากกัน ซึ่งซัพพอร์ทเวกเตอร์แมชชีนถูกออกแบบมาเพื่อใช้กับข้อมูลเชิงเส้นแต่ในปัจจุบันข้อมูลที่นำมาส่วนใหญ่มักเป็นข้อมูลแบบไม่เป็นเชิงเส้น ซึ่งปัญหานี้สามารถแก้ไขได้ด้วยการนำฟังก์ชัน Kernal มาใช้



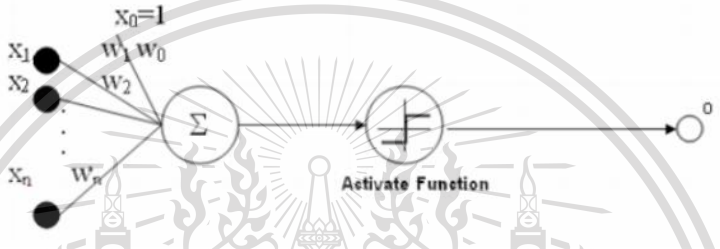
รูป 2.6 ซัพพอร์ทเวกเตอร์แมชชีน

2.1.4.5 โครงข่ายประสาท

โครงข่ายประสาท (Neural Network) [19] เป็นเทคนิคที่เลียนแบบการเรียนรู้ของสมองมนุษย์มีที่มาจากการพบว่าเมื่อมีการเรียนรู้สิ่งต่างๆในสมองของสิ่งมีชีวิต (การจดจำรูปหรือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

การแยกประเภทสิ่งต่างๆ) เซลล์ประสาท (นิวรอน) จะเกิดการกระตุ้นด้วยสัญญาณไฟฟ้าจะส่งต่อออกไปยังนิวรอนข้างเคียงต่อกันเป็นทอดๆ แม้นิวรอนจะมีรูปแบบการทำงานไม่ซับซ้อน แต่ด้วยการทำงานร่วมกันของนิวรอนหลายๆตัวทำให้แก้ปัญญาที่ซับซ้อนได้ โครงข่ายประสาทมีส่วนประกอบเป็นชั้นๆได้แก่ ระดับชั้นหนึ่งเป็นระดับชั้นนำเข้า (Input Layer) อีกระดับชั้นหนึ่งเป็นระดับชั้นส่งออก (Output Layer) และในระหว่าง 2 ชั้นจะมีระดับชั้นซ่อน (Hidden Layer) ซึ่งจะมีอย่างน้อย 1 ชั้น แต่ละชั้นจะเชื่อมต่อกันอย่างเป็นระบบโดยจะเชื่อมโยงอย่างเต็มที่กับระดับชั้นก่อนหน้าและระดับชั้นต่อไป ดังแสดงในรูปที่ 2.7 เส้นเชื่อม (Connection) ระหว่างนิวรอนจะมีค่าน้ำหนักกำกับซึ่งแสดงถึงน้ำหนัก/อิทธิพลของการเชื่อมนิวรอนต่ออีกนิวรอนหนึ่ง และเชื่อมโยงความสัมพันธ์ระหว่างตัวแปรนำเข้าและตัวแปรเป้าหมาย



รูป 2.7 โครงข่ายประสาท

2.1.5 การเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่ม (Evaluation)

2.1.5.1 เมทริกซ์ความสับสน

เมทริกซ์ความสับสน (Confusion Matrix) [20] เป็นรูปแบบของการสร้างตารางที่มีบทบาทเฉพาะในการเรียนรู้ของเครื่องและวิศวกรรมที่เกี่ยวข้อง ช่วยในการแสดงการทำนายและเรียกคืนในระบบที่ทราบค่าของข้อมูลการทดสอบ ดังแสดงเป็นตารางความผิดพลาดประเภทที่ 1 (Type I Error) ดังแสดงในตารางที่ 2.1

		ผลการจำแนก	
		คำตอบเป็นบวก	**คำตอบเป็นลบ**
ค่าที่แท้จริง	**คำตอบเป็นบวก**	**TP (True positive)**	**FN (False negative)**
	คำตอบเป็นลบ	**FP (False positive)**	**TN (True negative)**

ตาราง 2.1 เมทริกซ์ความสับสน (Confusion Matrix)

- True positive (TP) คือจำนวนข้อมูลที่จำแนกถูกว่าเป็นบวก
- True negative (TN) คือจำนวนข้อมูลที่จำแนกถูกว่าเป็นลบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

- False positive (FP) คือจำนวนข้อมูลที่จำแนกผิดว่าเป็นบวก ซึ่งค่าที่แท้จริงเป็นลบ
- False negative (FN) คือจำนวนข้อมูลที่จำแนกผิดว่าเป็นลบ ซึ่งค่าที่แท้จริงเป็นบวก

ซึ่งจะสามารถวัดประสิทธิภาพของโมเดลได้โดยค่า 3 ค่าที่จะคำนวณมาจากเมทริกซ์ความสับสนได้แก่ Precision, Recall,และ F1-Score

Precision คือการหาค่าทำนายผิด Positives หมายถึงทำนายผิดแต่ไม่ได้เป็นผลเสีย

$$\boldsymbol{Precision} = \frac{\boldsymbol{True\ positive}}{\boldsymbol{True\ positive + False\ positive}}$$

Recall คือการหาค่าทำนายผิด Negative แล้วเสียหาย หมายถึงทำนายผิดและมีผลเสีย

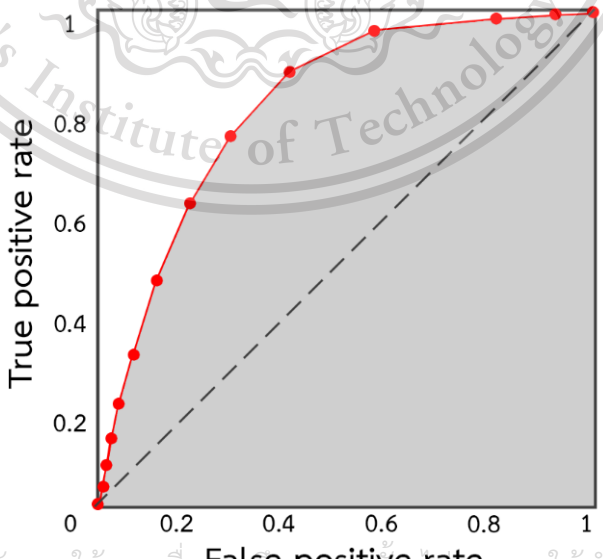
$$\boldsymbol{Recall} = \frac{\boldsymbol{True\ positive}}{\boldsymbol{True\ positive + True\ negative}}$$

F1-Score คือการหาค่าเฉลี่ย ระหว่าง Precision กับ Recall

$$\boldsymbol{F1 - Score} = \frac{\boldsymbol{2 \times Precision \times Recall}}{\boldsymbol{Precision + Recall}}$$

2.1.5.2 Area under the ROC curve (AUC)

คือค่าที่เกิดจากการเปรียบเทียบระหว่าง อัตราค่า True positive และค่า False positive ของแต่ละคลาสเป้าหมาย โดยถ้า ROC Curve มีค่าเข้าใกล้ 1 จะแสดงว่ามีประสิทธิภาพดี แสดงถึงค่า True Positive ที่มาก และ Area Under Curve (AUC) จะใช้แสดงค่าพื้นที่ใต้กราฟ ROC โดยถ้ามีค่าเข้าใกล้ 1 จะหมายถึงประสิทธิภาพที่สูงเช่นกัน



รูป 2.8 ROC curve

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.1.6 Python

Python เป็นหนึ่งในภาษาที่ใช่ในการเขียนโปรแกรม โดย Python สามารถใช้งานได้บนหลายแพลตฟอร์มไม่ว่าจะเป็น Windows 10, Windows 7, Windows XP, Unix, Linux ตลอดจนระบบ FreeBSD

ด้วยความที่ภาษา Python เป็นภาษาโอเพนซอร์ซ ทำให้เราสามารถพัฒนาโปรแกรมภาษา Python ได้ฟรีโดยไม่เสียค่าใช้จ่าย และด้วยความเป็นโอเพนซอร์ซของภาษา Python ทำให้มีผู้คนมากมายเข้ามาช่วยพัฒนาทำให้ปัจจุบันภาษา Python สามารถรองรับการทำงานได้อย่างหลากหลาย [21]

2.1.7 Node.js

Node ถูกสร้างขึ้นมาจากการที่ผู้พัฒนาอยากให้ JavaScript สามารถทำงานนอกเว็บบราวเซอร์ได้เพื่อเพิ่มความสามารถของเดิมมัน Node จึงถูกใช้เป็นรันไทม์สำหรับ JavaScript เพื่อให้ JavaScript สามารถทำงานได้ทุกที่เมื่อเราทำการติดตั้ง Node ลงไปในระบบแล้ว

อีกทั้ง Node.js ยังมีความเร็วของการประมวลผลที่สูง ทำให้เกิด Node.js แอปพลิเคชันขึ้นมาอย่างมากมาย ซึ่งส่วนหนึ่งก็ช่วยอำนวยความสะดวกในการพัฒนาเว็บไซต์ให้เป็นไปได้อย่างราบลื่น [22]

2.1.8 Express.js

ในการทำงานบนแพลตฟอร์มของ Node.js ตัว Express.js ซึ่งเป็นเซิร์ฟเวอร์ตัวหนึ่ง เป็นเว็บแอปพลิเคชันเฟรมเวิร์คที่ได้รับความนิยมเป็นอย่างมาก แต่ก่อนถ้าจะใช้เว็บแอปพลิเคชันเฟรมเวิร์ค ผู้พัฒนาจะต้องใช้งานหลายภาษา การเกิดของ Node.js กับ Express.js จึงช่วยลดความยุ่งยากในจุดนี้ เพียงสามารถใช้งาน Javascript ได้ก็สามารถเขียนได้ทั้งในฝั่งเซิร์ฟเวอร์และในฝั่งไคลเอนต์ และ Express.js ก็มีข้อดีในเรื่องความเร็วเช่นเดียวกับ Javascript อีกทั้งยังเป็นภาษาที่ง่ายต่อการเรียนรู้อีกด้วย [23]

2.1.9 React

React คือ JavaScript Library ที่ถูกพัฒนาโดย Jordan Walke วิศวกรซอฟต์แวร์ของ Facebook ซึ่งปัจจุบันได้รับความนิยมเป็นอยากมากในด้านของ front-end JavaScript library ในด้านของการพัฒนาเว็บไซต์ React ถูกใช้งานเป็นวงกว้างตั้งแต่บริษัทขนาดเล็กไปจนถึงบริษัทยักษ์ใหญ่ต่างๆ ตัวอย่างเช่น Netflix, Airbnb, Instagram, the New York Times เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

React คือ JavaScript Library ที่ถูกสร้างมาเพื่อการตอบสนองที่รวดเร็วของส่วนต่อประสานผู้ใช้ของเว็บไซต์หรือแอปพลิเคชันมือถือ React เป็น Open-Source, Component-Based, Front-End library ที่รับหน้าที่ในเฉพาะส่วนของ View ในสถาปัตยกรรม Model View Controller (MVC) [24]

2.1.10 TailwindCSS

Tailwind คือ CSS framework ดัดแปลงการออกแบบเฉพาะลงไปในเว็บไซต์เพื่อเปลี่ยนแปลงหน้าตาของเว็บไซต์ในคราวเดียว อีกทั้งยังรองรับการเลือกรูปแบบส่วนประกอบจากแคตตาล็อกของ CSS classes เพื่อให้สามารถเปลี่ยนแปลงหน้าตาของเว็บไซต์ทั้งหมดได้ในคราวเดียว [25]

2.1.11 Redux

Redux จะทำหน้าที่เก็บ State ทั้งหมดไว้ที่เดียวแบบ Global ทำให้เป็นตัวช่วยในการจัดการ State สำหรับหลากหลายเฟรมเวิร์ก เช่น Vue, Angular หรือ React ซึ่งเป็นแอปพลิเคชันที่พัฒนาด้วย JavaScript ส่งผลให้ Redux สร้างความสม่ำเสมอในการพัฒนาแอปพลิเคชัน, ทำให้แอปพลิเคชันทำงานในสภาพแวดล้อมที่แตกต่างกันได้ และช่วยให้แอปพลิเคชันง่ายต่อการทดสอบ [26]

2.1.12 Axios

Axios คือตัวกลางที่ใช้สำหรับเชื่อมต่อกับบริการ API สำหรับรับส่งข้อมูลแบบ RESTful API โดยในเรื่องของการรับส่งข้อมูล Axios จะทำหน้าที่เหมือนเป็นตัวกลางในการจัดการบริการต่างๆ นอกจากนี้ Axios ยังสามารถรองรับ Promises async/await ได้ รวมถึง Axios ยังรองรับได้ทั้งภาษา JavaScript และ TypeScript [27]

2.2 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวกับเรื่องการเรียนรู้ของเครื่อง, การเรียนรู้เชิงลึก และการประมวลผลภาษาธรรมชาติ เพื่อวิเคราะห์และคาดคะเนโรคซึมเศร้า ผู้วิจัยได้ค้นคว้าเอกสารงานวิจัยที่เกี่ยวข้องดังต่อไปนี้

V. Mishra and T. Garg (2018) [28] ได้ศึกษาเกี่ยวกับพฤติกรรมการใช้งานสื่อสังคมออนไลน์ของคนในยุคปัจจุบันซึ่งได้ผลกระทบต่อสุขภาพและจิตใจ งานวิจัยนี้มุ่งเน้นไปที่การหาวิธีวิเคราะห์ข้อความบนสื่อสังคมออนไลน์เพื่อตรวจหาอาการหดหู่ทางจิตใจ โดยงานวิจัยนี้จะคาดคะเนถึงโอกาสที่เจ้าของข้อความจะมีเป็นโรคซึมเศร้า โดยพิจารณาเทคนิคทางการเรียนรู้ของเครื่อง หลายๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

วิธีที่มีผู้วิจัยคนอื่นๆเคยใช้มาแล้ว เช่น ซัพพอร์ทเวกเตอร์แมชชีน, Single Gaussian classification, ต้นไม้ตัดสินใจ, Sparse L1-norm SVM (Linear kernel) หรือ Non sparse L2-norm SVM (Linear kernel) เป็นต้น เพื่อหาวิธีการที่มีประสิทธิภาพมากที่สุด จากนั้นใช้ Web-based tool ที่ชื่อว่า Social Mediated Patient Portal (SMPP) ทดสอบข้อมูลบนเฟซบุ๊กเพื่อคาดคะเนโอกาสเป็นโรคซึมเศร้าของแต่ละคนโดยเทียบกับคะแนน PHQ-9 โดยแบ่งระดับความรุนแรงตาม CESD-R ทางผู้วิจัยเลือกใช้การวิเคราะห์ความรู้สึก (Sentiment Analysis) ขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor Algorithm) และ ซัพพอร์ทเวกเตอร์แมชชีนในการวิเคราะห์เพื่อใช้คาดคะเนถึงโอกาสที่เจ้าของข้อความจะมีเป็นโรคซึมเศร้า และต่อยอดถึงการตรวจสอบตนเองของบุคคลทุกคนซึ่งอาจจะมีส่วนช่วยในการรักษาสภาวะทางจิตที่ดี

Munmun De Choudhury *et al*., (2013) [29] ได้ศึกษาเกี่ยวกับข้อมูลบนสื่อสังคมออนไลน์ (ทวิตเตอร์) ของคนที่เป็นโรคซึมเศร้าโดยแยกลักษณะข้อความ เช่น การทะเลาะ อารมณ์ ภาษา การพูดอัตตา และการพูดถึงโรคซึมเศร้า เป็นต้น แล้วใช้ Classifier ทางสถิติประยุกต์ใช้กับการสังเกตสัญญาณทางพฤติกรรมบนสื่อสังคมออนไลน์ เช่น ลดจำนวนการใช้สื่อสังคมออนไลน์ การเพิ่มขึ้นของเหตุการณ์ทางสุขภาพหรือการแสดงออกเกี่ยวกับศาสนาอย่างรุนแรง จากนั้นจึงนำไปเทียบกับ Ground truth ที่เก็บข้อมูลคะแนน BDI ของเจ้าของข้อมูล แล้วจึงใช้การเรียนรู้ของเครื่อง เพื่อเรียนรู้ลักษณะข้อมูลของคนที่ป่วยเป็นโรคซึมเศร้า โดยสุดท้ายผู้วิจัยเลือกใช้ซัพพอร์ทเวกเตอร์แมชชีนแบบเรเดียลเบสิสฟังก์ชันเป็น Classifier เพื่อใช้คาดคะเนสภาวะหดหู่ทางจิตใจของผู้ใช้ซึ่งอาจนำไปสู่โรคซึมเศร้า โดยวัดความแม่นยำได้อยู่ที่ 70%

Liu yi Lin, B.A *et al*., (2016) [30] ได้ศึกษาเกี่ยวกับความเชื่อมโยงกันระหว่างสื่อสังคมออนไลน์กับวัยรุ่นชาวอเมริกันซึ่งมีแนวโน้มป่วยเป็นโรคซึมเศร้า กลุ่มตัวอย่างคือคนอายุ 19-32 ปี จำนวน 1787 คน โดยสุ่มจากเบอร์โทรศัพท์และใช้ชุดคำถาม PROMIS Depression Scale แบบสั้นเพื่อวัดระดับของโรคซึมเศร้า แล้วจึงวิเคราะห์โดยใช้ Chi-square Test หาความเชื่อมโยงระหว่างตัวแปรกับ PROMIS Depression Score แล้ววิเคราะห์ผลด้วยโมเดลการถดถอยโลจิสติก ผู้วิจัยสรุปการวิจัยได้ว่าการเพิ่มขึ้นของการใช้งานสื่อสังคมออนไลน์มีความสัมพันธ์เชิงบวกกันกับภาวะโรคซึมเศร้า โดยปริมาณการใช้งานสื่อสังคมออนไลน์ใน 1 วัน มีผลโดยตรงกับโอกาสในการเป็นโรคซึมเศร้าในเชิงแปรผันตรง และ จำนวนวันที่ใช้สื่อสังคมออนไลน์ใน 1 สัปดาห์ มีผลโดยตรงกับโอกาสในการเป็นโรคซึมเศร้าในเชิงแปรผันตรง

Seung W. Choi *et al*., (2014) [31] ได้ศึกษาเกี่ยวกับความเชื่อมโยงตัวชี้วัดทั่วไปของการป่วยเป็นโรคซึมเศร้าทั้ง 3 ตัวได้แก่ BDI-II, CES-D, และ PHQ-9 กับ PROMIS Depression การวิจัยได้สรุปความแตกต่างของตัวชี้วัดทั่วไปของการป่วยเป็นโรคซึมเศร้าทั้ง 3 ตัวและความเหมาะสมในการใช้งานตามแต่โอกาส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ ดัดแปลง เผยแพร่ หรือนำไปใช้ประโยชน์ในเชิงการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โดยตัวชี้วัด CES-D ประกอบไปด้วย 20 คำถามที่จะเจาะจงไปที่ปัจจัยที่ทำให้เกิดอาการซึมเศร้าด้วยปริมาณคำถามที่มากแต่ได้ความละเอียดของเหตุการณ์และปัจจัยสูง จึงเหมาะที่จะใช้ในงานด้านการวิจัย PHQ-9 จะถูกออกแบบมาเพื่อใช้งานในเบื้องต้นจึงมีเพียง 9 คำถามเท่านั้น สนใจอาการทางกายภาพเป็นหลัก จึงเหมาะกับใช้งานในคลินิกเวชกรรม BDI-II ถูกต่อยอดมาจาก PHQ-9 แต่มีความละเอียดกว่าที่ 21 คำถาม แต่ BDI-II ถูกออกแบบมาเพื่อใช้งานส่วนบุคคล สุดท้าย PROMIS Depression ถูกออกแบบมาเพื่อใช้ทั้งด้านการวิจัยและการรักษา มีการเก็บรายละเอียดที่สูงประกอบด้วย 28 คำถาม และชุดคำถามถูกออกแบบมาเพื่อให้ง่ายต่อการเก็บสถิติ

M. Ghiassi, J. Skinner และ D. Zimbra (2013) [32] ได้ศึกษาการวิเคราะห์ความรู้สึกของชื่อแบรนด์จากข้อมูลในทวิตเตอร์โดยใช้วิธี N-grams ทำ Feature Reduction และพัฒนา Twitter-Specific Lexicon เพื่อใช้ในการวิเคราะห์ความรู้สึก ในงานวิจัยพิสูจน์ให้เห็นว่าการลดขนาด Lexicon ลงให้เหลือเพียงตัวสำคัญ การลดความซับซ้อนของโมเดล การจำกัดเอาเฉพาะคำที่มีการใช้งานบ่อยๆในทวิตเตอร์มีส่วนช่วยในการเพิ่มความแม่นยำของ Sentiment Classification และการกำหนดกลุ่มของความรู้สึกที่ Binary Classification (Positive และ Negative) จะได้ผลลัพธ์ดีกว่าแบ่งกลุ่มแบบ Fire-grained classification (Positive, Neutral และ Negative) มาก โดยผลลัพธ์งานวิจัย ผู้วิจัยได้พัฒนา Twitter-Specific Lexicon ซึ่งใช้งานกับ Twitter ได้ดีกว่า Lexicon ทั่วไป และพัฒนา DAN2 Machine Learning ซึ่งสร้าง Sentiment Classification ได้แม่นยำกว่าซัพพอร์ทเวกเตอร์แมชชีน

Reem ALRashdi และ Simon O'Keefe (2018) [33] ได้ทำการศึกษาการใช้งานการเรียนรู้เชิงลึกและการประมวลผลภาษาธรรมชาติเพื่อตรวจจับข้อความฉุกเฉิน (Crisis Response) บนทวิตเตอร์เพื่อใช้ในการตรวจจับสถานการณ์ฉุกเฉินของผู้ใช้งาน Twitter ที่ต้องการความช่วยเหลือ โดยได้ทำการทดลองบนโครงข่ายประสาท (CNN และ Bi-LSTM) ซึ่งทำ Word Embedding แบบ GloVe เทียบกับ Word Embedding คำเฉพาะสำหรับสถานะการณ์ฉุกเฉิน (Crisis Embedding) ซึ่งพัฒนามาจาก Skip-gram model ของ Word2Vec เพื่อพัฒนาประสิทธิภาพของโมเดลการแยกประเภททวีต

โดยผลงานวิจัยระบุว่าการทำ Word Embedding ของคำศัพท์ทั่วไป GloVe ด้วยกันกับ Bi-LSTM จะได้ผลลัพธ์ที่มีความแม่นยำอยู่ที่ 62.04% F1-Score ซึ่งสูงกว่าแบบอื่น ซึ่งคาดคะเนเหตุผลว่า Bi-LSTM จะ สามารถจับข้อมูลได้ดีกว่า CNN และถึงแม้ Crisis Embedding จะใช้ข้อมูลน้อยกว่า GloVe Embedding ในการเข้าใจบริบทของประโยชน์ แต่ไม่สามารถทำงานกับ Bi-LSTM ได้ดีอย่างที่คาดจึงได้ผลลัพธ์ที่ด้อยกว่า GloVe Embedding

Ahmed Husseini Orabi *et al*., (2018) [34] ได้ทำการวิจัยเพื่อหาความน่าจะเป็นของผู้ใช้ทวิตเตอร์ที่มีแนวโน้มจะเป็นโรคซึมเศร้า โดยจะมีวิธีการคือนำข้อมูลการทวีตจาก CLPsych2015 Shared Task ที่มีข้อมูลของทั้งคนที่ป่วยเป็นโรคทางจิต (PTSD และโรคซึมเศร้า) และคนปกติซึ่งผ่านการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Pre-Processing มาแล้ว มาทำ Word Embedding ทั้งแบบ Random Trainable, Skip-gram, CBOW และ Word Embedding optimized ที่ผ่านการปรับแต่งมาเฉพาะงานวิจัยนี้ จากนั้นก็ใช้การเรียนรู้เชิงลึกเรื่อง CNN (CNNWithMax, MultiChannelPoolingCNN, และ MultiChannelCNN) และ RNN (BiLSTM) มาใช้ในการวิเคราะห์แนวโน้ม โดยในงานวิจัยจะใช้หลายวิธีการเพื่อเปรียบเทียบผลลัพธ์ ซึ่งวัดผลโดยเทียบทั้ง Accuracy, Area-under-the-curve (AUC), Precision, Recall และ F1-measure ซึ่งผลลัพธ์ที่ดีที่สุดคือโมเดล CNNWithMax โดยทำ Word Embedding แบบ Optimized

Akshi Kumar *et al*., (2019) [35] ได้ทำการวิจัยความเกี่ยวเนื่องกันระหว่างสภาพจิตใจและสื่อสังคมออนไลน์ งานวิจัยทำการคาดคะเนภาวะซึมเศร้าจากข้อมูลทวิตเตอร์แบบเรียลไทม์โดยวิเคราะห์จาก 5 ปัจจัยหลักซึ่งดึงข้อมูลต่างๆ มาจากทวิตเตอร์ของผู้ใช้ได้แก่ คำ (Word), เวลา (Timing), ความถี่ (Frequency), ความรู้สึก (Sentiment), และความผิดแผก (Contrast) ซึ่งผลลัพธ์จะได้ขั้วของข้อความ (Polarity) ซึ่งจะแบ่งเป็นขั้วบวก (1) และขั้วลบ (0) จากนั้นจะนำปัจจัยหลัก 5 อย่างนี้มาใช้เป็นคุณลักษณะในการฝึกฝนโมเดล

ในงานวิจัยเราใช้โมเดลในการทำนาย 3 ตัวได้แก่ Naïve bayes, Gradient boosting, และ Random forest ในการวิเคราะห์ข้อมูลทวิตเตอร์ในช่วงหนึ่งเดือนของผู้ใช้ตามคุณลักษณะที่กล่าวมาข้างต้น และใช้ Ensemble voting classifier ในการลงคะแนนความเห็นส่วนใหญ่เพื่อทำนายว่าเจ้าของข้อความมีอาการของโรคซึมเศร้าหรือไม่ ซึ่งผลงานวิจัยจากการทดสอบข้อมูลทวิตเตอร์ของผู้ใช้จำนวน 100 คน ผลลัพธ์ได้ความแม่นยำอยู่ที่ 85.09%

David John Hughes *et al*., (2012) [36] ได้กล่าวถูกสื่อสังคมออนไลน์ 2 ตัวคือเฟสบุ๊คและทวิตเตอร์ซึ่งเป็นเครื่องที่มือยอดนิยมในการปฏิสัมพันธ์และแลกเปลี่ยนข้อมูลในปัจจุบัน งานวิจัยได้ศึกษาและเปรียบเทียบถึงบุคลิก ลักษณะนิสัย อายุ เพศ ของผู้ใช้สื่อสังคมออนไลน์ทั้ง 2 ตัว ซึ่งผลลัพธ์แสดงให้เห็นว่าบุคลิกภาพส่วนบุคคลมีอิทธิพลต่อการเรื่องใช้สื่อสังคมออนไลน์ และยังมีอิทธิพลไปถึงรวมไปถึงลักษณะการใช้งาน การเข้าสังคมและการและเปลี่ยนข่าวสาร ซึ่งประเด็นสำคัญของงานวิจัยนี้ที่นำมาอ้างถึงคือ “ถึงแม้เฟซบุ๊กจะเป็นสื่อสังคมออนไลน์ที่ใหญ่ที่สุด พร้อมไปด้วยสิ่งอำนวยความสะดวกออนไลน์มากมาย แต่เมื่อเทียบกับทวิตเตอร์ พวกเขานำเสนอบริการที่ไม่เหมือนกัน ทวิตเตอร์จะให้ความสำคัญไปที่การแสดงความคิดเห็นและการแลกเปลี่ยนข้อมูล (Kwak, Changhyun & Moon, 2010) มากกว่าเฟซบุ๊กที่สนใจที่การปฏิสัมพันธ์ทางสังคม (Huberman, Romero & Wu, 2009) ”

Angela Leis, PsyM *et al*., (2019) [37] มีเป้าหมายเพื่อศึกษาการระบุคุณสมบัติทางภาษาของทวิตในทวิตเตอร์ของภาษาสเปน และระบุรูปแบบพฤติกรรมของผู้ใช้ทวิตเตอร์คนๆนั้น เพื่อระบุถึงสัญญาณการป่วยเป็นโรคซึมเศร้า ในงานวิจัยมี 2 ขั้นตอนหลักได้แก่ ขั้นแรกคือเลือกผู้ใช้ทวิตเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

และทวิตแบบสุ่มตามเงื่อนไขแล้วจัดกลุ่มเป็นชุดข้อมูล โดยชุดข้อมูลแรกจะเป็นกลุ่มของผู้ใช้ที่ป่วยเป็นโรคซึมเศร้าชุดข้อมูลที่สองจะเป็นทวิตที่มีการบ่งชี้ว่าเป็นของคนที่ป่วยเป็นโรคซึมเศร้า และชุดข้อมูลสุดท้ายจะเป็นชุดข้อมูลควบคุมซึ่งสุ่มมาจะทวิตทั้งหมด ขั้นตอนที่ 2 จะเป็นการเปรียบเทียบและวิเคราะห์ผลจากชุดข้อมูลทั้งสามที่เลือกมา โดยผลลัพธ์ที่ผู้วิจัยสนใจคือในผู้ป่วยโรคซึมเศร้าจะมีข้อความที่เป็นขั้วลบเฉลี่ยอยู่ที่ 34% ของข้อความทั้งหมด และในทวิตที่ประเมินว่าเป็น Depressive ทวิตจะมีข้อความที่เป็นขั้วลบเฉลี่ยอยู่ที่ 46%

Nikie Jo Elauria Deocampo (2019) [38] เป็นโครงงานที่ต้องการหาอาการเริ่มต้นของโรคซึมเศร้าผ่านการทำเหมืองข้อความ (Text mining) โดยในโครงงานใช้ส่วนต่อประสานทวิตเตอร์ (Twitter API) เพื่อดึงข้อมูลการทวิตของผู้ใช้ทวิตเตอร์ทั่วไป โครงงานได้ทำการ Preprocess โดยการ Tokenize เป็นการตัดคำแล้วไปจับกับ Sentiment Dictionary เพื่อหา Polarity ของคำ จากนั้นนำ Polarity ทั้งหมดไปคำนวณในระดับประโยค (ทวิต) หลัง Preprocess ทวิตแล้วเสร็จต่อมาจึงนำข้อมูลที่ Preprocess แล้วไปเรียนรู้ในโมเดลต่างๆ เช่น Naive Bayes, ต้นไม้ตัดสินใจ (Decision Tree), Random Forest, ซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine), และวิธีการเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbour) เป็นต้น งานจะวัดผลลัพธ์ด้วยความแม่นยำ (Accuracy) ผ่านการทดสอบผ่านชุดที่ใช้ในการฝึกสอน (Validate Accuracy) ซึ่งผลลัพธ์ของโครงงานคือได้ความแม่นยำสูงสุดของโมเดล Decision Tree อยู่ที่ 98.55668748040587%

Zunaira Jamil *et al*., (2017) [39] เป็นงานวิจัยที่ใช้ข้อมูลบนสื่อสังคมออนไลน์โดยเฉพาะในทวิตเตอร์เพื่อหาผู้ใช้ที่มีความเสี่ยงในโรคซึมเศร้า โดยใช้ชุดข้อมูลของ BellLetsTalk 2015 ที่มีข้อมูล 156,612 ทวิตจากผู้ใช้ 25,362 คน งานวิจัยฝึกฝนโมเดลในระดับผู้ใช้และระดับทวิตเพื่อหาทำนายอาการของโรคซึมเศร้า โดยวัดผลด้วย Precision และค่าความครบถ้วน (Recall)

งานวิจัยเริ่มด้วยการทำความสะอาดชุดข้อมูล นำข้อมูลที่ไม่จำเป็นออก ใช้คำศัพท์เฉพาะเพื่อหาทวิตที่บ่งบอกถึงอาการของโรคซึมเศร้า จากนั้นก็เตรียม Label ของข้อมูลโดยแบ่งเป็นในระดับทวิตและในระดับผู้ใช้ งานวิจัยของเราจะสนใจผลการทดลองในระดับทวิต โดยในระดับทวิตงานวิจัยจะนำชุดข้อมูลที่แบ่งเป็น 7 ฟีเจอร์ไปเทรนในโมเดล SVM เพื่อหาความน่าเป็นที่ข้อความจะมีอาการของโรคซึมเศร้า ซึ่งผลลัพธ์ในระดับทวิตได้ความแม่นยำอยู่ที่ 61.02% Precision 12.37% และค่าความครบถ้วน 80.20%

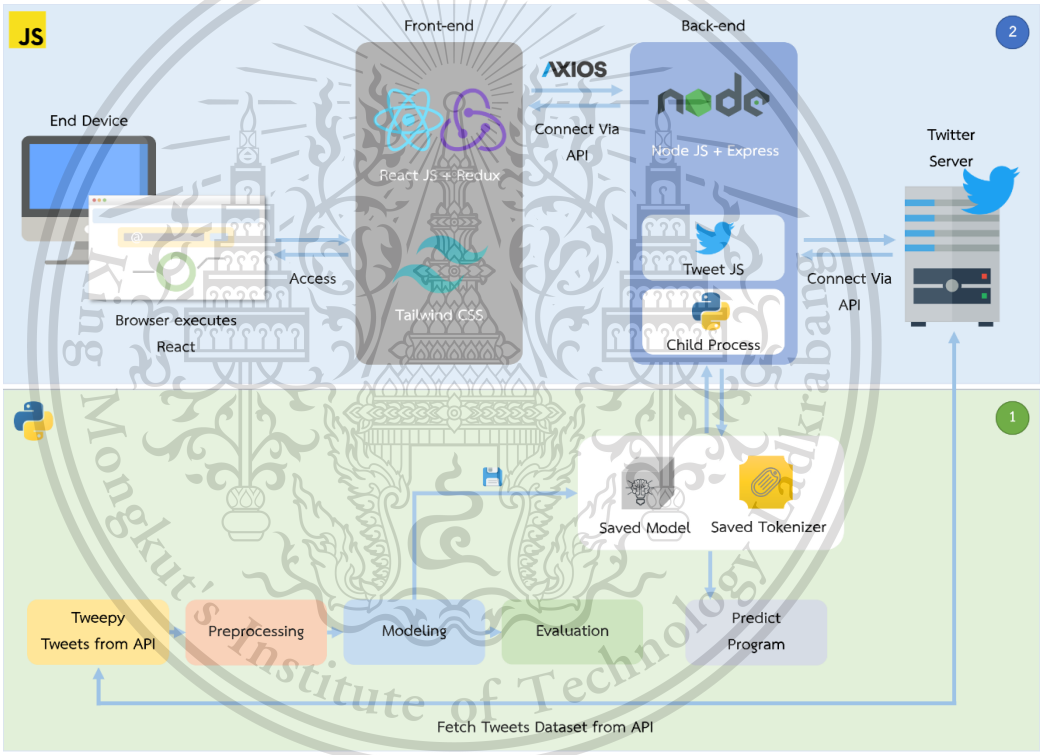
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 3

การออกแบบ

ในบทนี้ ผู้วิจัยได้นำเสนอการออกแบบโครงสร้างโดยรวมของระบบ ซึ่งประกอบไปด้วย 2 ส่วนหลัก ได้แก่ การออกแบบโปรแกรมทำนายผล เพื่อใช้ทำนายผลลัพธ์จากข้อความบนไทม์ไลน์ผู้ใช้ และการออกแบบเว็บแอปพลิเคชัน เพื่อให้ผู้ใช้สามารถใช้งานโปรแกรมทำนายผลผ่านอินเตอร์เฟซที่ผู้ใช้งานสามารถเข้าใจและเข้าถึงได้ง่าย

3.1 ภาพรวมการออกแบบระบบ



รูป 3.1 แผนภาพแสดงภาพรวมของการออกแบบระบบ

จากรูปที่ 3.1 แสดงให้เห็นถึงภาพรวมของการออกแบบระบบซึ่งถูกแบ่งออกเป็น 2 ส่วนหลัก ดังต่อไปนี้

1.) การออกแบบโปรแกรมทำนายผล – ส่วนนี้จะกล่าวถึงการสร้างโปรแกรมทำนายผล เริ่มตั้งแต่การเก็บรวบรวมชุดข้อมูลจนไปถึงได้สำเร็จออกมาเป็นโมเดล

2.) การออกแบบเว็บแอปพลิเคชัน – ส่วนนี้จะกล่าวถึงการออกแบบเว็บแอปพลิเคชันสำหรับผู้ใช้งานทั่วไปได้ใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.2 การออกแบบโปรแกรมทำนายผล

3.2.1 การออกแบบโดยรวมของโปรแกรมทำนายผล

รูป 3.2 แผนภาพแสดงภาพรวมการสร้างโปรแกรมทำนายผล

จากรูปที่ 3.2 การสร้างโปรแกรมทำนายผลประกอบไปด้วย 5 ส่วนหลัก ดังต่อไปนี้

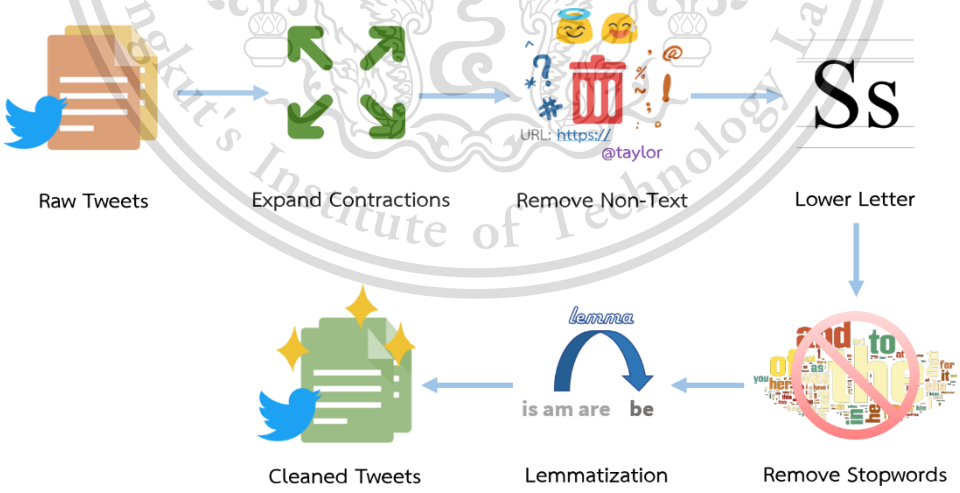
- Tweets from API คือ ส่วนการเก็บชุดข้อมูลจากทวิตเตอร์
- Preprocessing คือ ส่วนการเตรียมข้อมูล
- Modeling คือ ส่วนการสอนโมเดล
- Evaluation คือ ส่วนการวัดประสิทธิภาพของโมเดล
- Predict Program คือ ส่วนโปรแกรมทำนายผล

3.2.2 ส่วนการเก็บข้อมูลจากทวิตเตอร์

สำหรับการเก็บข้อมูลทวีตจากทวิตเตอร์ทำโดยใช้การสตรีมผ่านส่วนเชื่อมต่อประสานโปรแกรมทวิตเตอร์ ซึ่งมีการคัดกรองข้อความที่มีส่วนเกี่ยวข้องเกี่ยวกับประเด็นที่ได้ทำการศึกษา ได้แก่ Depression, Anxiety, Suicide, Stress และ Sad เป็นต้น

3.2.3 ส่วนการเตรียมข้อมูล

3.2.3.1 การทำความสะอาดข้อมูลตัวอักษร (Text Cleaning)



รูป 3.4 แผนภาพแสดงขั้นตอนการทำความสะอาดข้อมูลตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จากรูปที่ 3.4 การทำความสะอาดข้อมูลประเภทตัวอักษรประกอบไปด้วยขั้นตอนดังต่อไปนี้

1.) Expand Contractions คือ การขยายคำย่อ หรือตัวย่อ เช่น “I’m a cat person” เมื่อขยายคำย่อจะได้ “I am a cat person” ซึ่งทำให้สามารถตีความข้อความได้ง่ายขึ้น และเป็นการเพิ่ม Stopwords ทำให้ชุดข้อมูลมีขนาดเล็กลงเมื่อมีการกำจัด Stopwords ออกไป

2.) Remove Non-Text คือ การกำจัดสิ่งที่ไม่ใช่ตัวอักษร ไม่เกี่ยวข้องกับคำศัพท์ ซึ่งประกอบไปด้วย อีโมจิ วรรคตอน ตัวเลข เมนชั่น และ URLs เป็นต้น

3.) Lower Letter คือ การทำให้ตัวอักษรภาษาอังกฤษทุกตัวเป็นตัวพิมพ์เล็กเพื่อให้เป็นรูปแบบมาตรฐานเดียวกัน

4.) Remove Stopwords คือ การกำจัดคำที่ไม่ช่วยในการสื่อความหมายแต่พบได้ทั่วไปในประโยค เช่น a, an, the, of, quiet, unless, just, to เป็นต้น ซึ่งจะช่วยทำให้ชุดข้อมูลมีขนาดเล็กลง

5.) Lemmatization คือ การเปลี่ยนรูปคำภาษาอังกฤษให้อยู่ในรูปพื้นฐาน (Basic form) ของคำนั้นๆ เช่น “caring” เมื่อถูกเลมมาจะได้ “care” ซึ่งจะช่วยให้ Train Feature มีขนาดที่ไม่ใหญ่จนเกินไปจากคำเดียวกันแต่อยู่ในหลายรูปแบบ และทำให้สามารถจับคู่กับคำใน Subjectivity Dictionary ได้ดีขึ้น

3.2.3.2 การคำนวณหาขั้วของข้อความ (Calculate Polarity)

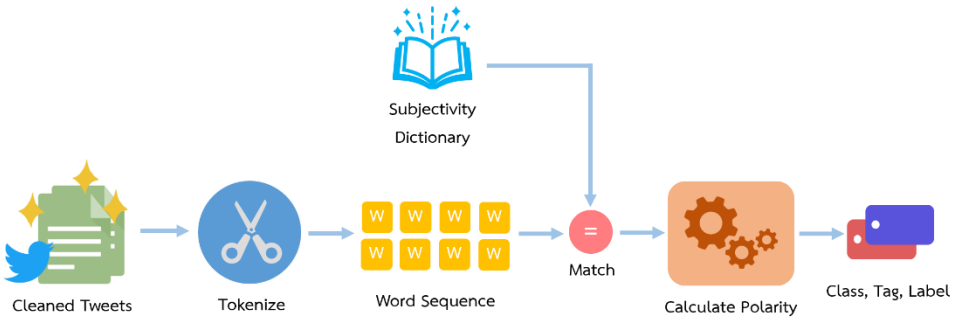
ในการคำนวณหาขั้วของข้อความนั้นได้มีการใช้ Subjectivity Dictionary เข้ามาช่วย ซึ่ง Subjectivity Dictionary คือ แหล่งรวมคำศัพท์ที่บ่งบอกอารมณ์ความรู้สึกทางจิตใจ โดยจะมีคะแนนของแต่ละคำเป็น -1 0 และ 1 ตามความรู้สึกของคำนั้นๆ ดังแสดงในรูปที่ 3.5

adversarial	adj	n	negative
adversary	noun	y	negative
adverse	adj	n	negative
adversity	noun	n	negative
advice	adj	n	positive
advice	noun	n	positive
advisable	noun	n	positive
advocate	verb	y	positive
advocacy	noun	n	positive
affable	adj	n	positive
affability	noun	n	positive

รูป 3.5 ตัวอย่าง Subjectivity Dictionary

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โดยขั้นตอนการคำนวณหาขั้วของข้อความประกอบไปด้วยขั้นตอนดังแสดงในรูปที่ 3.6



รูป 3.6 แผนภาพการแสดงขั้นตอนการหาขั้วของข้อความ

1.) Tokenize คือ การตัดคำจากประโยคให้อยู่ในรูปอนุกรมของคำ (Word Sequence) ซึ่งเป็นหน่วยย่อยของประโยคแทน เช่น “nothing can hurt us” เมื่อทำการตัดคำจะได้เป็น [“nothing”, “can”, “hurt”, “us”]

2.) Match word with Subjectivity Dictionary คือ การนำคำศัพท์แต่ละคำในอนุกรมของคำไปค้นหาใน Subjectivity Dictionary แล้วดูว่าคำนั้นมีคะแนนเท่าไร โดยคำที่มีขั้วเป็นบวกจะมีคะแนนเท่ากับ 1 คำที่มีขั้วเป็นกลางจะมีคะแนนเท่ากับ 0 และคำที่มีขั้วเป็นลบจะมีคะแนนเท่ากับ -1

3.) Calculate Polarity คือ การคำนวณหาขั้วของข้อความโดยหาได้จาก

$$\text{คะแนน} = \frac{\text{ผลรวมของคะแนนของคำในข้อความที่พบใน Subjectivity Dictionary}}{\text{จำนวนคำที่พบใน Subjectivity Dictionary}}$$

*โดยคะแนนของแต่ละคำหากพบมีการนำหน้าด้วย “ไม่” ซึ่งจะทำให้ความหมายเกิดการกลับขั้ว ต้องทำการคูณคะแนนของคำนั้นด้วย -1 ก่อนนำไปรวม

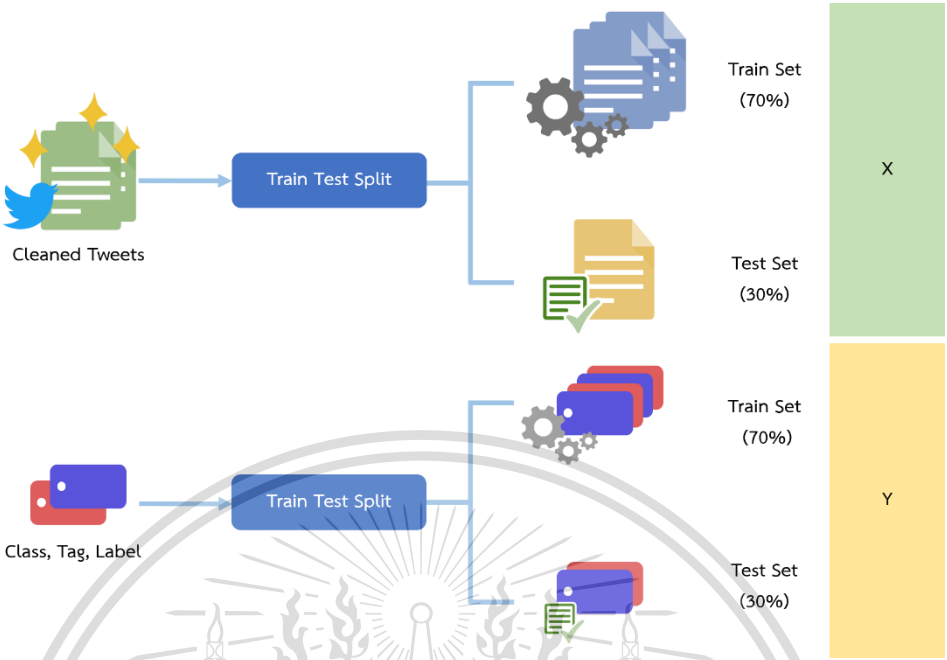
ซึ่งคะแนนที่ได้จะนำมาใช้ในการแบ่งคลาสของประโยคดังต่อไปนี้

เงื่อนไข	คลาส	ความหมาย
คะแนน >= 0	1	ข้อความทวีตที่มีความปกติ
คะแนน < 0	-1	ข้อความทวีตที่มีความซึมเศร้า

ตาราง 3.1 รูปแบบการแบ่งคลาสของข้อความทวีต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.2.3.3 การแบ่งชุดข้อมูล

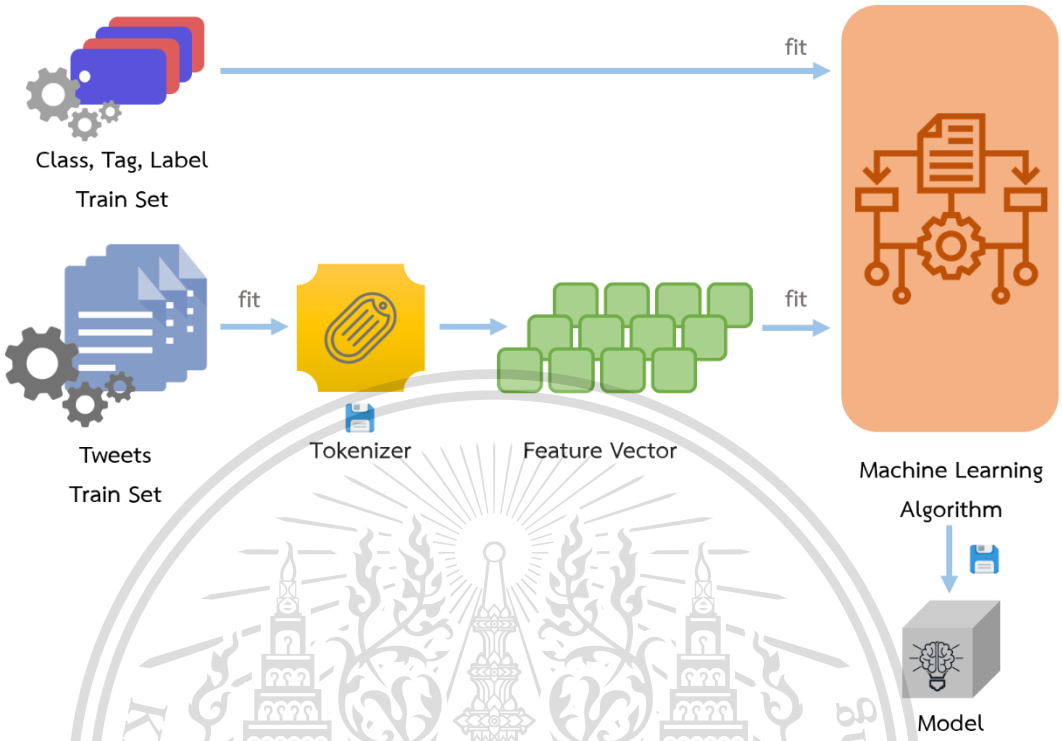


รูป 3.7 แผนภาพการแบ่งชุดข้อมูลทวีต

จากรูปที่ 3.7 ชุดข้อมูลจะถูกแบ่งออกเป็น 2 ส่วน โดยส่วนแรกใช้สำหรับเทรนโมเดล 70% และส่วนที่สองจะใช้สำหรับทดสอบประสิทธิภาพของโมเดลอีก 30%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.2.4 ส่วนการสอนโมเดล



รูป 3.9 แผนผังแสดงการเทรนโมเดล

จากรูปที่ 3.9 แสดงขั้นตอนการสอนโมเดล โดยมีขั้นตอนดังต่อไปนี้

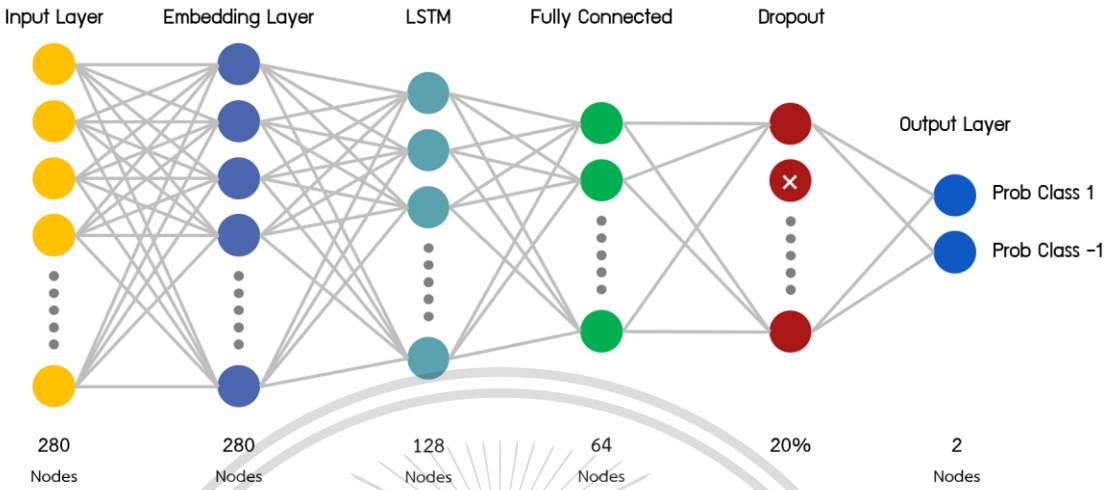
1.) นำชุดข้อมูล Tweets Train Set ไปทำการ fit กับ Tokenizer โดยตัว Tokenizer จะทำการตัดและแปลงคำศัพท์ให้อยู่ในรูปแบบของตัวเลขให้คอมพิวเตอร์สามารถใช้ประมวลผลได้และจัดให้อยู่ในรูปของ Feature Vector ซึ่งตัว Tokenizer จะทำการบันทึกการจับคู่คำศัพท์ไปเป็นตัวเลขเก็บไว้เป็นคลังคำศัพท์ (Dictionary) โดยเราจะทำการบันทึกตัว Tokenizer เพื่อนำไปใช้ในโปรแกรมทำนายผลต่อไป

2.) นำ Feature Vector และข้อมูลคลาสคำตอบ Train Set ไปทำการ fit กับโมเดลตามอัลกอริทึมต่างๆ ซึ่งในโครงงานนี้มี Algorithm ที่เลือกใช้ 5 รูปแบบ ได้แก่

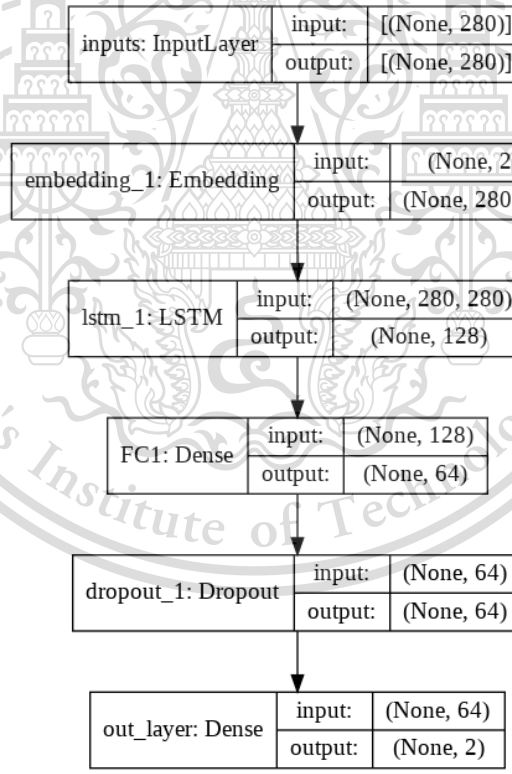
- Multinomial Naïve Bayes
- Decision Tree
- Random Forest (Depth = 100)
- Support Vector Machine (RBF Kernel)
- LSTM (RNN)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ซึ่งสำหรับ LSTM ที่เป็นโครงข่ายประสาทเทียม มีโครงสร้างแต่ละชั้น ดังแสดงในรูปที่ 3.10 และรูปที่ 3.11



รูป 3.10 โครงสร้าง LSTM

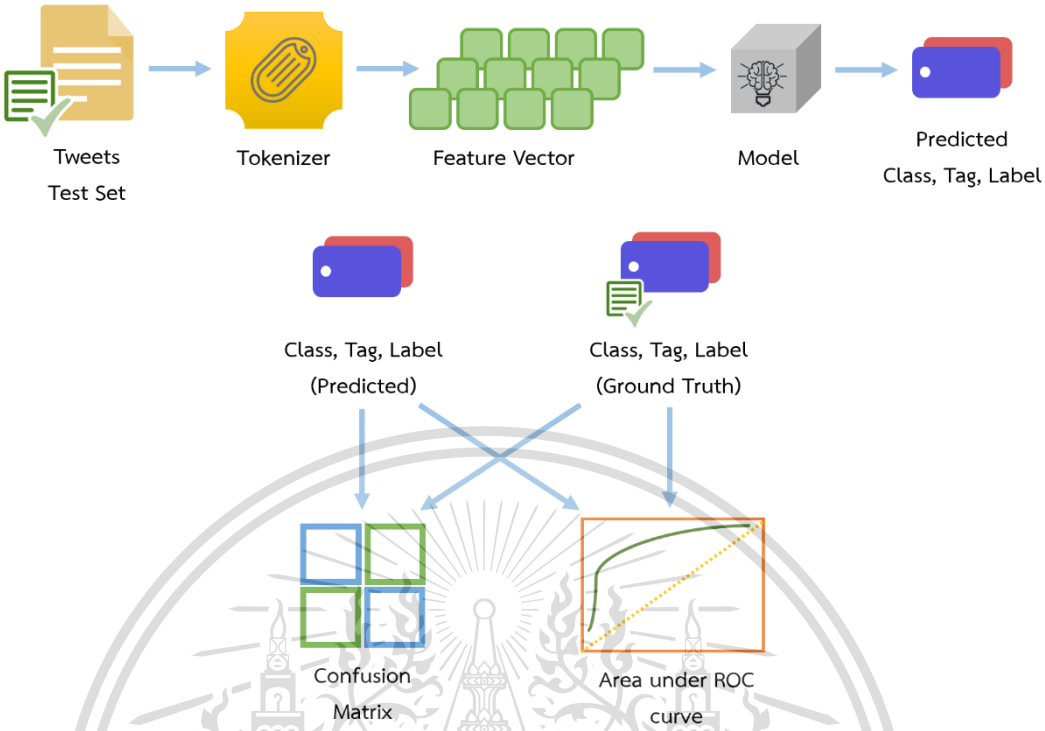


รูป 3.11 โครงสร้าง LSTM แบบแสดง Input และ Output

เมื่อทำการสอนโมเดลเสร็จสิ้น เราจะทำการบันทึกตัวโมเดลเพื่อนำไปใช้ในโปรแกรมทำนายผลต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.2.5 การวัดประสิทธิภาพของโมเดล



รูป 3.12 แผนผังแสดงการวัดประสิทธิภาพของโมเดล

จากรูปที่ 3.12 แสดงขั้นตอนการวัดประสิทธิภาพของโมเดล โดยได้ใช้วิธีการวัดประสิทธิภาพ 2 รูปแบบ ดังต่อไปนี้

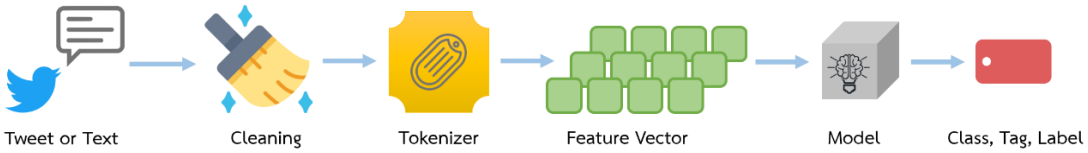
1.) เมทริกซ์ความสับสน (Confusion Matrix) เป็นการวัดประสิทธิภาพโดยใช้ค่า Precision Recall และ F1-Score ซึ่งหาได้จากเมทริกซ์ความสับสนและหา Accuracy จากเส้นทแยงมุม

2.) Area Under ROC Curve เป็นการวัดประสิทธิภาพโดยใช้ค่า AUC ซึ่งหาได้จากพื้นที่ใต้โค้ง ROC

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.2.6 โปรแกรมทำนายผล

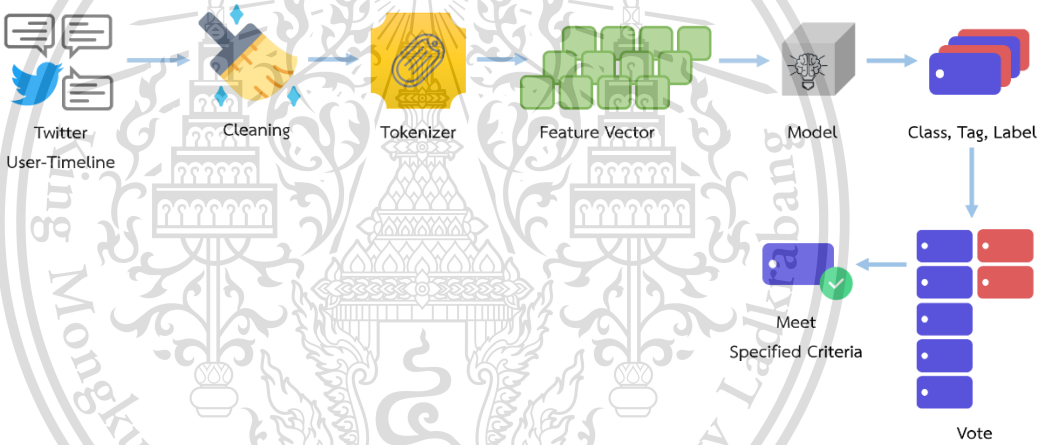
3.2.6.1 โปรแกรมทำนายผลจากข้อความเดี่ยว



รูป 3.13 แผนผังแสดงการทำงานของโปรแกรมทำนายผลข้อความเดี่ยว

จากรูปที่ 3.13 โปรแกรมจะรับข้อความเข้ามา โดยข้อความจะถูกทำความสะอาดและถูก Tokenize เพื่อแปลงข้อความให้กลายเป็น Feature Vector แล้วส่งเข้าโมเดลเพื่อทำนายคลาสของข้อความ

3.2.6.2 โปรแกรมทำนายผลจากไทม์ไลน์ผู้ใช้



รูป 3.14 แผนผังแสดงการทำงานของโปรแกรมทำนายผลจากไทม์ไลน์ผู้ใช้

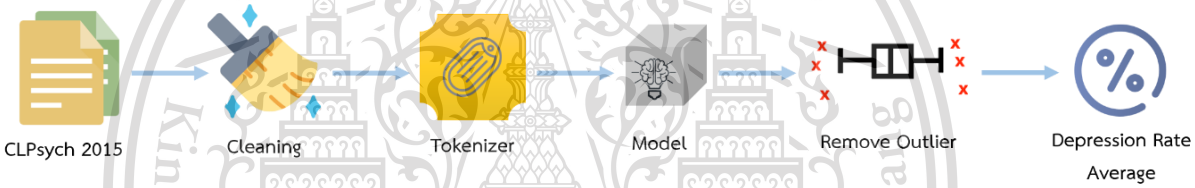
จากรูปที่ 3.14 โปรแกรมจะรับข้อความซึ่งถูกดึงมาจากไทม์ไลน์ผู้ใช้จากทวิตเตอร์ โดยข้อความจะถูกทำความสะอาด และถูก Tokenize เพื่อแปลงข้อความให้กลายเป็น Feature Vector แล้วส่งเข้าโมเดลเพื่อทำนายคลาสของข้อความแต่ละข้อความในไทม์ไลน์ของผู้ใช้ จากนั้นจะทำการโหวตเพื่อหาข้อสรุปว่าผู้ใช้บัญชีนี้มีภาวะโรคซึมเศร้าหรือไม่ตามเกณฑ์ที่กำหนด โดยเกณฑ์จะคิดดังนี้

นำชุดข้อมูล CLPsych 2015 Shared Task มาเป็นตัวช่วยในการกำหนดเกณฑ์ในการตัดสินว่าบัญชีทวิตเตอร์นี้มีภาวะของโรคซึมเศร้าหรือไม่ โดยชุดข้อมูล CLPsych 2015 Shared Task จะประกอบด้วยรายชื่อเจ้าของบัญชีทวิตเตอร์ ทวิต (ข้อความ) ของเจ้าบัญชีนั้น และอาการป่วยของโรคซึมเศร้า โดยเราได้ทำการรวบรวมจากคลินิกจิตเวชของต่างประเทศโดยมีปริมาณข้อมูลค่อนข้างมากและมีความน่าเชื่อถือของข้อมูลสูง โดยข้อมูลจะมี Label บอกข้อมูลต่างๆ ได้แก่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Depression คือผู้ป่วยโรคซึมเศร้า PTSD คือผู้ป่วยด้วยอาการ PTSD และ Control คือไม่มีอาการป่วย ในที่นี้เราจะไม่สนใจอาการของโรค PTSD

เราจะนำโมเดลทำนายโรคซึมเศร้าที่สร้างขึ้นไปทำนายทุกๆทวีต(ข้อความ) ของแต่ละบัญชีทวิตเตอร์ของชุดข้อมูลซึ่งกรองเอาเฉพาะบัญชีที่เจ้าของเป็นโรคซึมเศร้า เพื่อหาอาการของโรคซึมเศร้า (Depression) ในแต่และข้อความ จากนั้นก็คำนวณหาเปอร์เซ็นต์ของข้อความที่มีอาการของโรคซึมเศร้า (Depression) เมื่อเทียบกับข้อความทั้งหมดของเจ้าของบัญชีนั้นๆ ทำแบบนี้จนครบทุกบัญชีทวิตเตอร์ที่มีในชุดข้อมูล CLPsych 2015 Shared Task ต่อมาก็จัดการนำค่าผิดปกติ (Outlier) ออกจากข้อมูล โดยกำหนดให้ค่าที่น้อยกว่า Q3+1.5*IQR และค่าที่มากกว่า Q1-1.5*IQR คือค่าผิดปกติ จากนั้นจะคำนวณเพื่อให้ได้ค่าเฉลี่ยของอัตราส่วนข้อความที่มีอาการของโรคซึมเศร้า (Depression) ของบัญชีทวิตเตอร์ทุกบัญชีที่ 30.89% ซึ่งเมื่อเทียบกับงานวิจัยของ Angela Leis, PsyM et al., (2019) จะเห็นได้ว่าค่าเฉลี่ยมีค่าที่ใกล้เคียงกับอัตราส่วนข้อความที่เป็นขั้วลบเฉลี่ยในงานวิจัยมาก



รูป 3.15 แผนผังแสดงการหาอัตราส่วนข้อความมีอาการของโรคซึมเศร้าเฉลี่ย

ในงานวิจัยนี้เราจะใช้ค่า Quartile ของอัตราส่วนข้อความที่มีอาการของโรคซึมเศร้าของบัญชีผู้ใช้ในชุดข้อมูล CLPsych 2015 Shared Task ในการประเมินระดับของอาการโรคซึมเศร้า ซึ่งจากการคำนวณจะ ได้ค่า Quartile ดังนี้ Q1 = 28.14 / Q2 = 31 / Q3 = 33.75

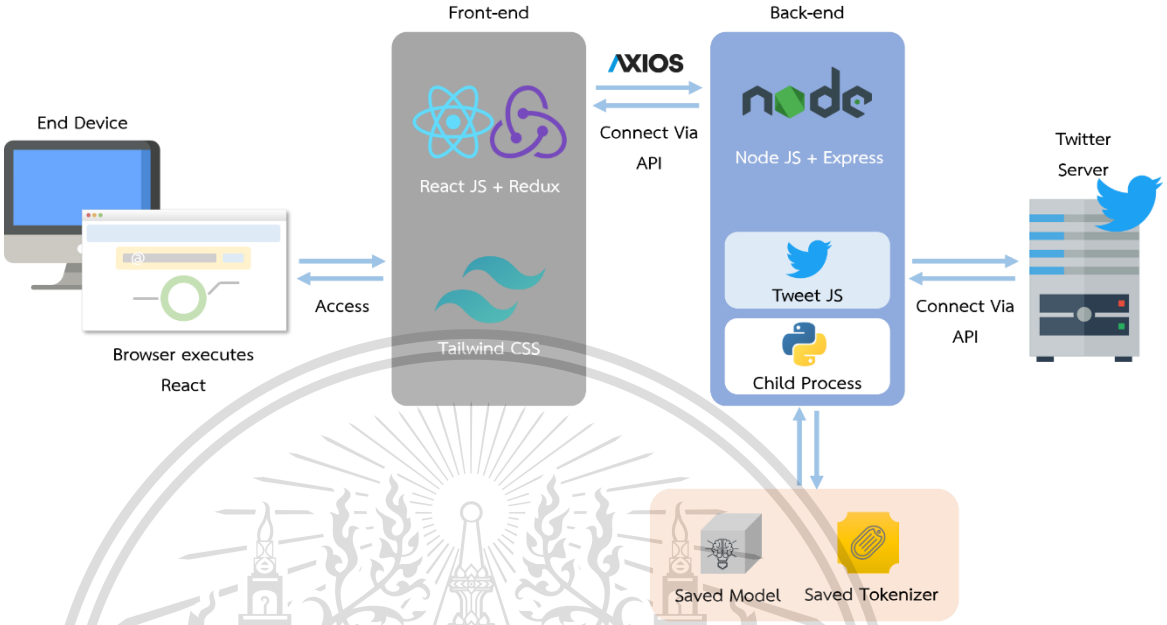
จากค่า Quartile ทำให้กำหนดใช้เกณฑ์ประเมิน (Criteria) ดังนี้

- หากอัตราส่วนข้อความที่มีอาการของโรคซึมเศร้าของผู้ใช้มีค่าน้อยกว่า Q1 จะประเมินว่าไม่มีอาการของโรคซึมเศร้า
- หากอัตราส่วนข้อความที่มีอาการของโรคซึมเศร้าของผู้ใช้มีค่ามากกว่า Q1 แต่น้อยกว่า Q2 จะประเมินว่ามีแนวโน้มอาการของโรคซึมเศร้าเล็กน้อย
- หากอัตราส่วนข้อความที่มีอาการของโรคซึมเศร้าของผู้ใช้มีค่ามากกว่า Q2 แต่น้อยกว่า Q3 จะประเมินว่ามีแนวโน้มอาการของโรคซึมเศร้าปานกลาง
- หากอัตราส่วนข้อความที่มีอาการของโรคซึมเศร้าของผู้ใช้มีค่ามากกว่า Q3 จะประเมินว่ามีแนวโน้มอาการของโรคซึมเศร้ารุนแรง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.3 การออกแบบเว็บแอปพลิเคชัน

3.3.1 การออกแบบโดยรวมของเว็บแอปพลิเคชัน



รูป 3.16 แผนผังแสดงภาพรวมของการออกแบบเว็บแอปพลิเคชัน

จากรูปที่ 3.16 แสดงภาพรวมการออกแบบเว็บแอปพลิเคชันซึ่ง ประกอบไปด้วยส่วนหลัก 3 ส่วน ดังนี้

1.) End Device คือ อุปกรณ์ที่ผู้ใช้งานใช้เข้าถึงเว็บไซต์ เช่น คอมพิวเตอร์ แท็บเล็ต สมาร์ทโฟน เป็นต้น

2.) Front-end คือ ส่วนหน้าบ้าน เป็นส่วนติดต่อผู้ใช้งาน แสดงอินเตอร์เฟซของเว็บไซต์ให้ผู้ใช้งานสามารถใช้งานได้

3.) Back-end คือ ส่วนหลังบ้าน เป็นส่วนที่ให้บริการส่วนต่อประสานโปรแกรมต่างๆ และเป็นส่วนที่เชื่อมต่อกับโมเดล และเซิร์ฟเวอร์ทวิตเตอร์

3.3.2 ผู้ใช้งาน (User)

ผู้ใช้งาน คือบุคคลทั่วไป ซึ่งมีความต้องการขั้นพื้นฐาน ดังต่อไปนี้

1.) ผู้ใช้ที่มีบัญชีใช้งานทวิตเตอร์เป็นภาษาอังกฤษสามารถตรวจภาวะโรคซึมเศร้าของตนเองได้ผ่านบัญชีทวิตเตอร์

2.) ผู้ใช้ทำแบบทดสอบโรคซึมเศร้าได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.) ผู้ใช้สามารถหาความรู้เกี่ยวกับโรคซึมเศร้าได้

4.) ผู้ใช้งานสามารถอ่านด้วยภาษาอังกฤษ หรือภาษาไทยได้

5.) ผู้ใช้สามารถใช้งานได้บนโทรศัพท์มือถือหรือแท็บเล็ต

3.3.3 ความต้องการของเว็บแอปพลิเคชัน (Requirement)

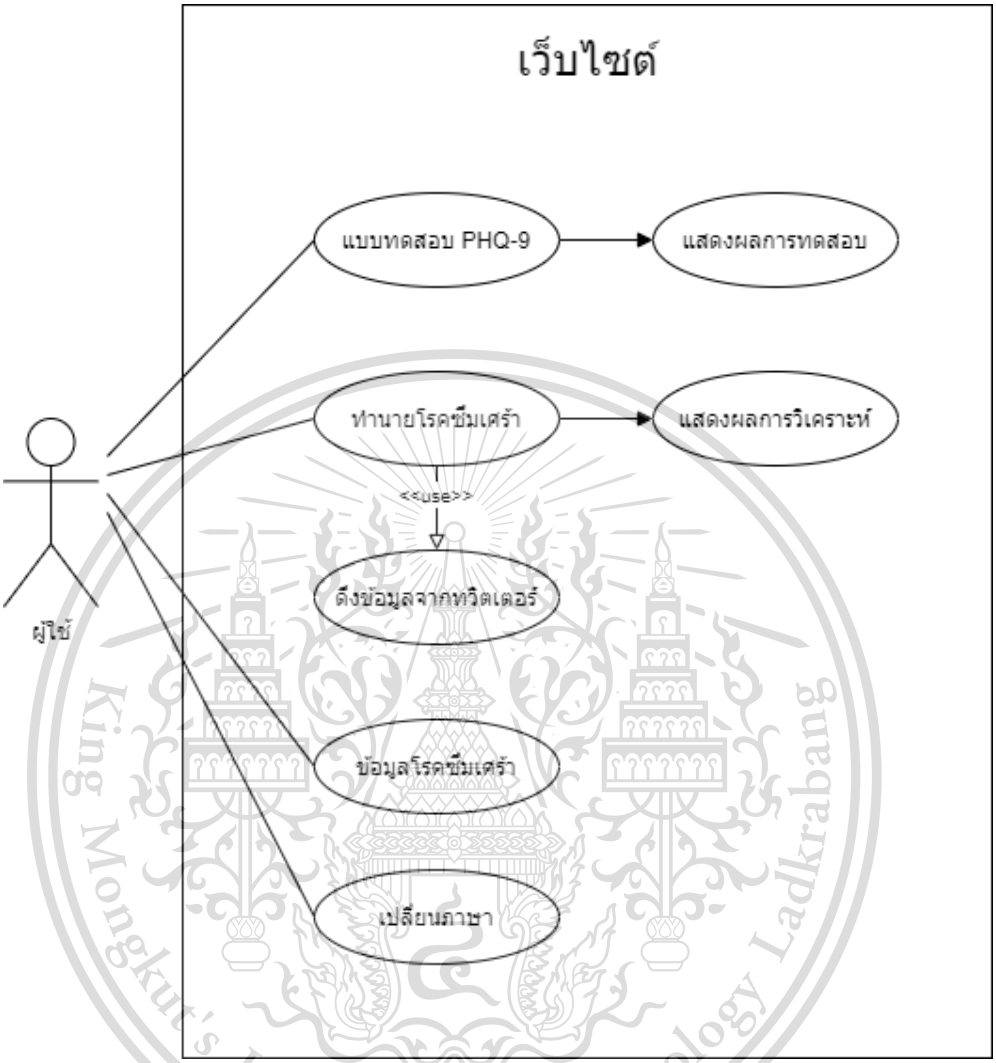
ID	Details	Type	Priority
R01	ผู้ใช้สามารถตรวจภาวะโรคซึมเศร้าของตนเองได้	• Product • Functional	MustHave
R02	ผู้ใช้สามารถตรวจภาวะโรคซึมเศร้าของของผู้อื่นได้	• Product • Functional	MustHave
R03	ผู้ใช้สามารถทำแบบทดสอบโรคซึมเศร้าได้	• Product • Functional	MustHave
R04	ผู้ใช้สามารถดูผลการทดสอบโรคซึมเศร้าได้	• Product • Functional	MustHave
R05	ผู้ใช้สามารถดูความรู้เกี่ยวกับโรคซึมเศร้าได้	• Product • Functional	MustHave
R06	ระบบสามารถรองรับทั้งภาษาไทยและภาษาอังกฤษ	• Usability • Non-Functional	ShouldHave
R07	เว็บไซต์ต้องรองรับ Responsive	• User Interface • Non-Functional	MustHave
R08	เว็บไซต์ต้องแสดงผลบนมือถือได้ครบถ้วน	• Compliance to Standards • Non-Functional	ShouldHave
R09	ระบบต้องเก็บข้อมูลผู้ใช้เป็นความลับ	• Security • Non-Functional	ShouldHave

ตาราง 3.2 ความต้องการของเว็บแอปพลิเคชัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.3.4 Use Case

3.3.4.1 Use Case Diagram



3.3.4.2 Use Case Detail

Use Case: แบบทดสอบ PHQ-9
Use Case ID: 1
Actor: ผู้ใช้
Preconditions: -
Flow of events: 1. ผู้ใช้งานกดไปที่หน้าแบบทดสอบ 2. ผู้ใช้ตอบคำถามตามความจริง 3. ผู้ใช้กดส่งผลการทดสอบ
Postconditions: แสดงผลการทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Use Case: แสดงผลการทดสอบ
Use Case ID: 2
Actor: ผู้ใช้
Preconditions: ทำแบบทดสอบ PHQ-9
Flow of events: 1. ระบบวิเคราะห์ข้อมูลจากแบบทดสอบของผู้ใช้ 2. แสดงผลการทดสอบ
Postconditions: -

Use Case: ทำนายโรคซึมเศร้า
Use Case ID: 3
Actor: ผู้ใช้
Preconditions: -
Flow of events: 1. ผู้ใช้กดเลือกทำนายด้วยบัญชีทวิตเตอร์ 2. ผู้ใช้ใส่ชื่อบัญชีทวิตเตอร์ 3. ผู้ใช้กดปุ่มทำนายผล
Postconditions: ดึงข้อมูลจากทวิตเตอร์, แสดงผลการวิเคราะห์

Use Case: ดึงข้อมูลจากทวิตเตอร์
Use Case ID: 4
Actor: ผู้ใช้
Preconditions: เริ่มทำนายโรคซึมเศร้า
Flow of events: 1. ระบบดึงข้อมูลไทม์ไลน์ในทวิตเตอร์จากชื่อบัญชีทวิตเตอร์
Postconditions: -

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Use Case: แสดงผลการวิเคราะห์
Use Case ID: 5
Actor: ผู้ใช้
Preconditions: ทำนายโรคซึมเศร้า
Flow of events: 1. ระบบวิเคราะห์ข้อมูลไทม์ไลน์ในทวิตเตอร์จากชื่อบัญชีทวิตเตอร์ 2. แสดงผลการวิเคราะห์
Postconditions: -

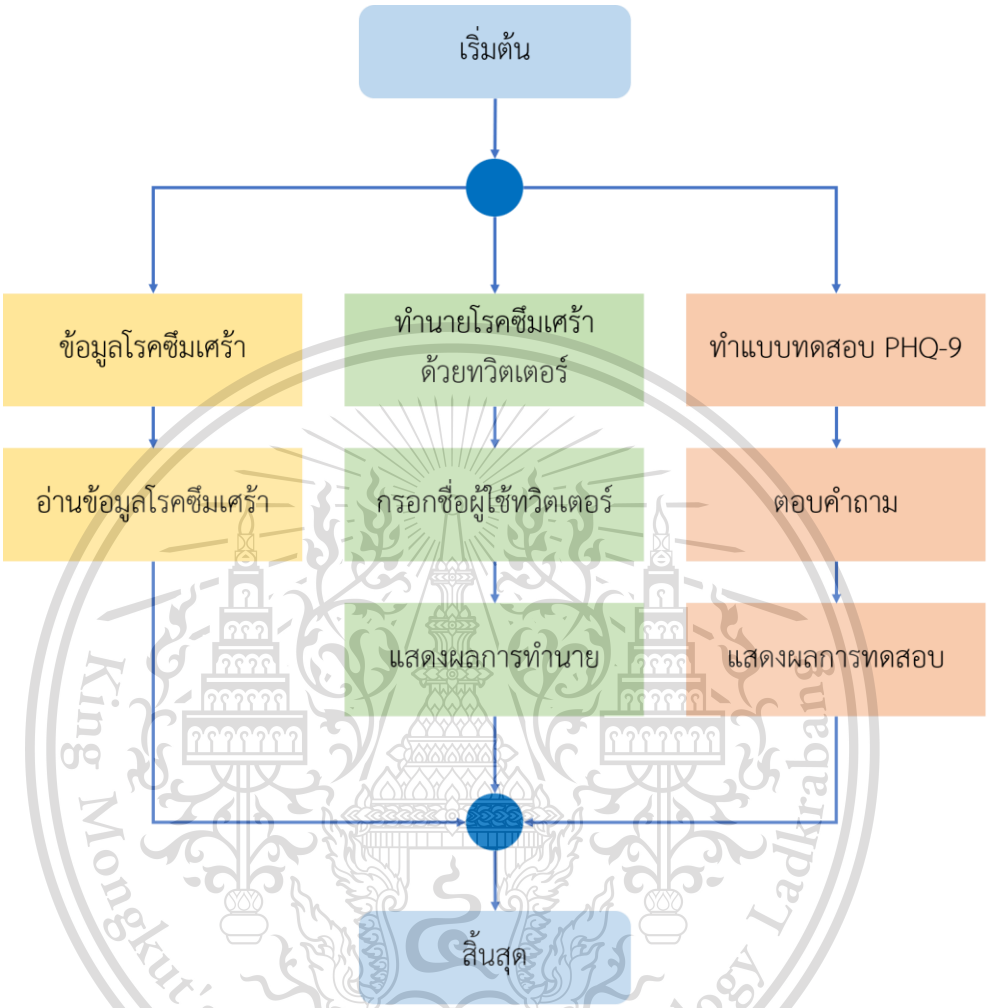
Use Case: ข้อมูลโรคซึมเศร้า
Use Case ID: 6
Actor: ผู้ใช้
Preconditions: -
Flow of events: 1. ผู้ใช้กดเลือกข้อมูลโรคซึมเศร้า 2. แสดงข้อมูลโรคซึมเศร้าที่น่าสนใจ
Postconditions: -

Use Case: เปลี่ยนภาษา
Use Case ID: 7
Actor: ผู้ใช้
Preconditions: -
Flow of events: 1. ผู้ใช้กดป๊อปอัพเปลี่ยนภาษา 2. ผู้ใช้เลือกภาษาที่ต้องการ
Postconditions: -

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.3.5 การพัฒนาส่วน Front-end

3.3.5.1 แผนผังการทำงานของเว็บแอปพลิเคชัน



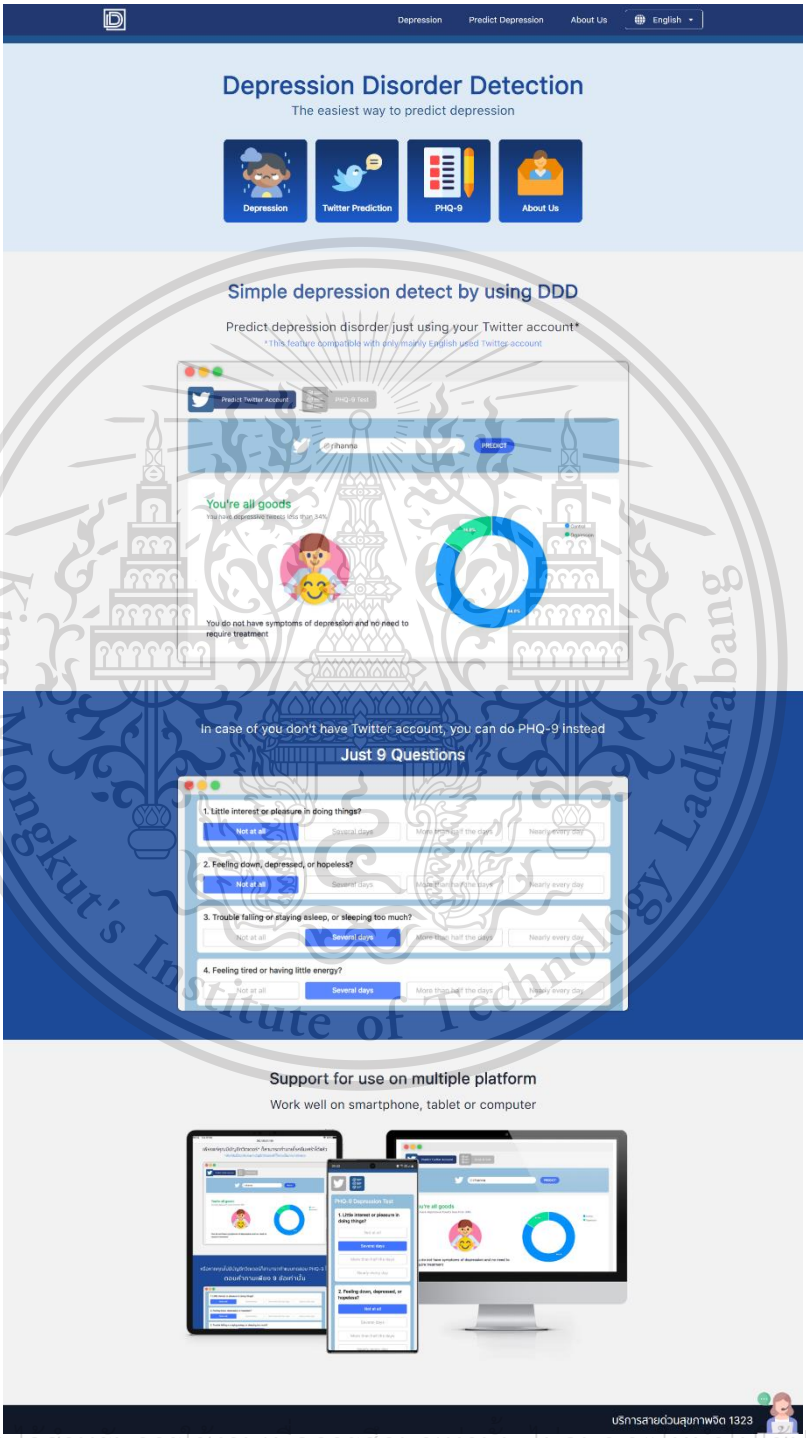
รูป 3.17 แผนผังแสดงการทำงานของเว็บแอปพลิเคชัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.3.5.2 การออกแบบส่วนติดต่อผู้ใช้ของเว็บแอปพลิเคชัน

ในการออกแบบส่วนติดต่อผู้ใช้ (User Interface) สามารถแบ่งออกได้ตามหน้าของเว็บแอปพลิเคชัน ดังต่อไปนี้

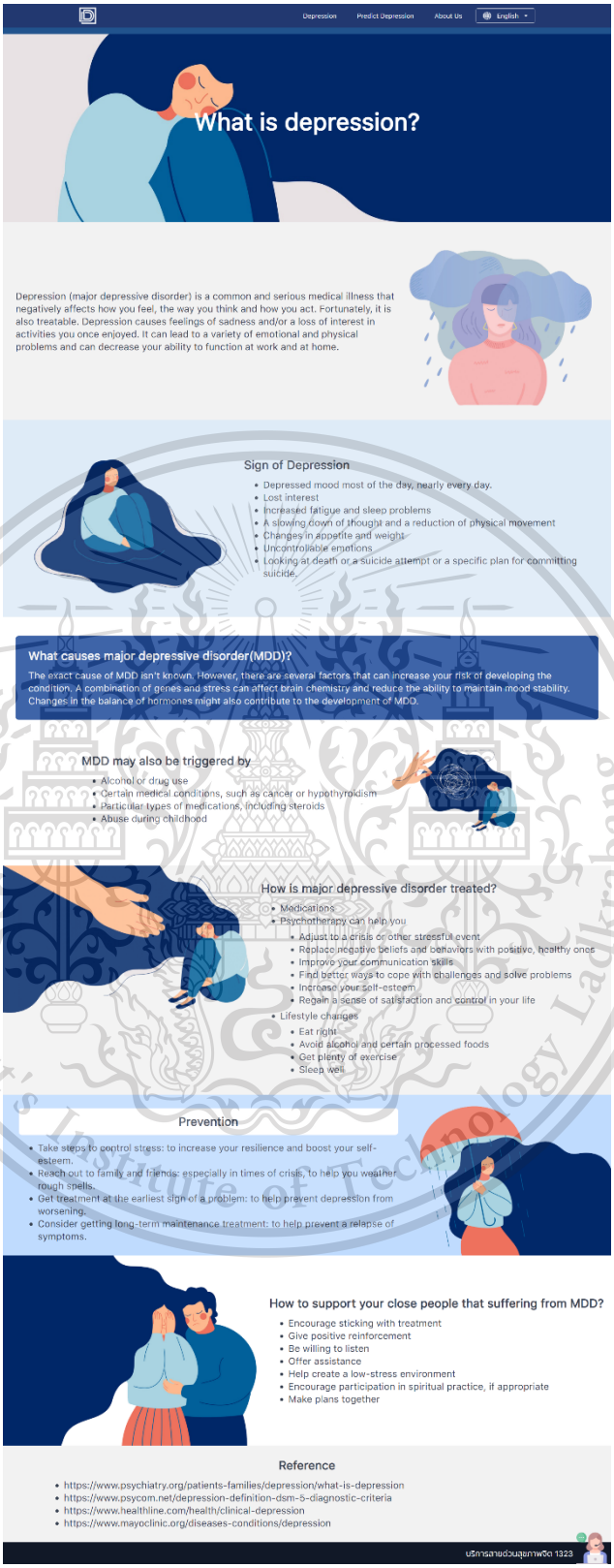
1.) หน้าหลัก (Homepage)



รูป 3.18 ส่วนติดต่อผู้ใช้หน้าหลัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

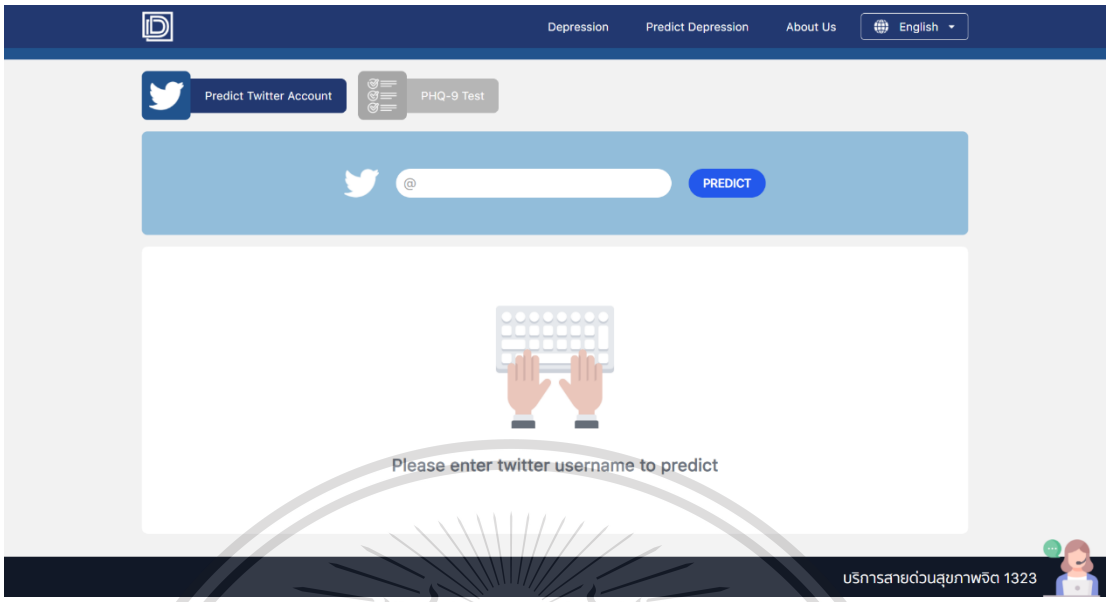
2.) หน้าแสดงข้อมูลโรคซึมเศร้า



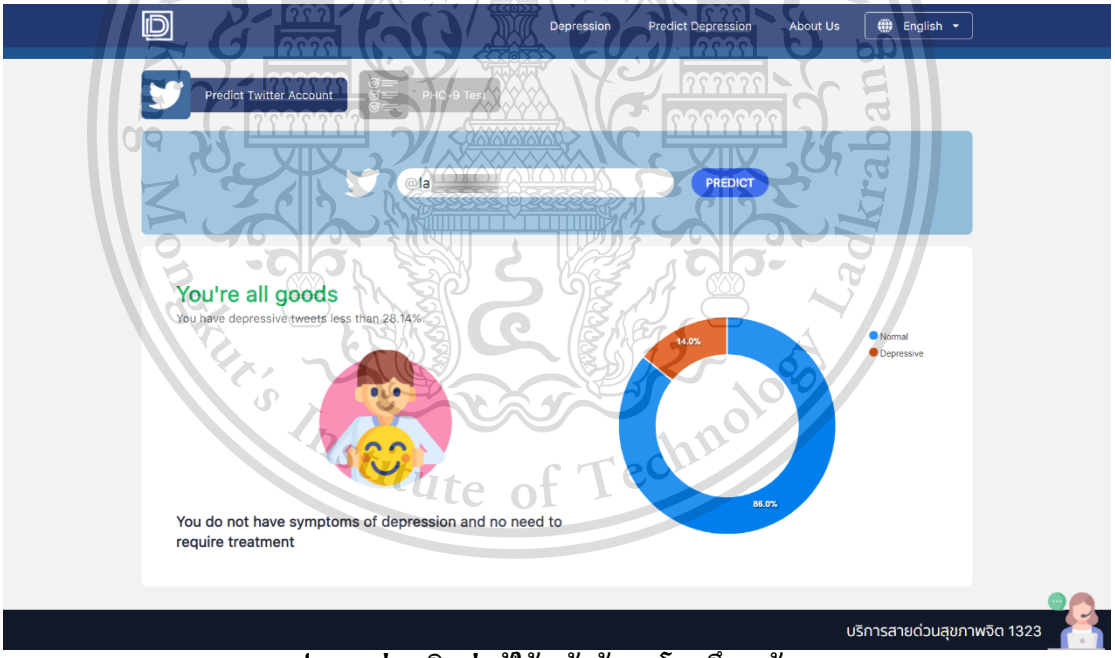
รูป 3.19 ส่วนติดต่อผู้ใช้หน้าข้อมูลโรคซึมเศร้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.) หน้าทำนายโรคซึมเศร้าด้วยบัญชีทวิตเตอร์

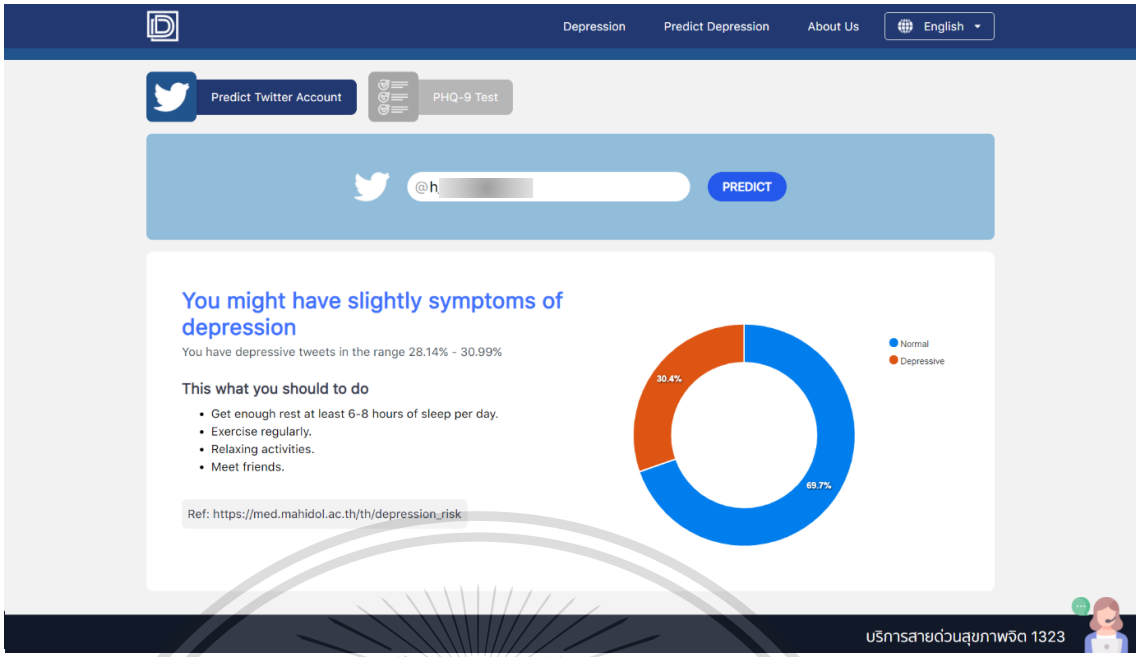


รูป 3.20 ส่วนติดต่อผู้ใช้หน้าข้อมูลโรคซึมเศร้า (ก)



รูป 3.21 ส่วนติดต่อผู้ใช้หน้าข้อมูลโรคซึมเศร้า (ข)

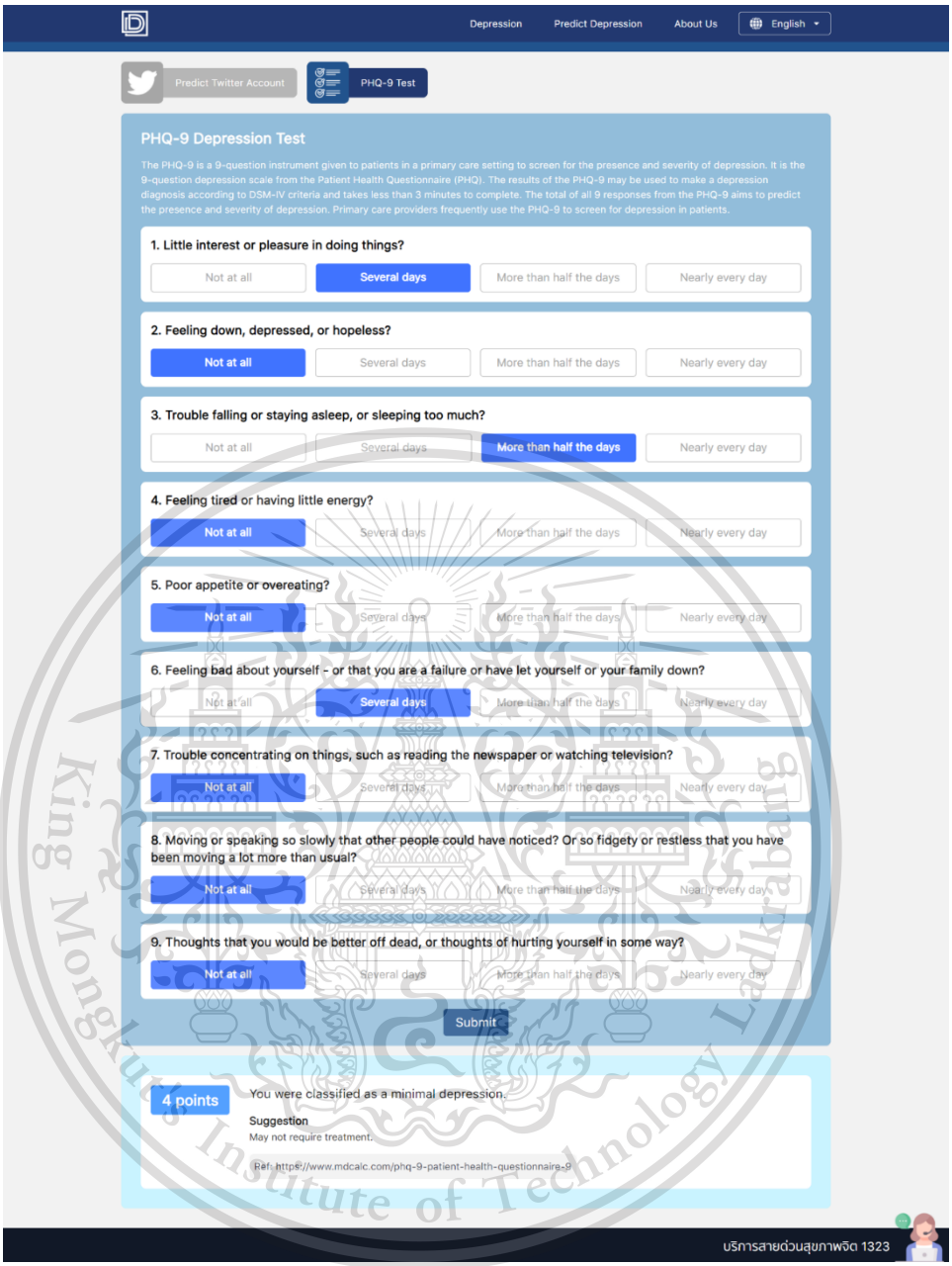
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 3.22 ส่วนติดต่อผู้ใช้หน้าข้อมูลโรคซึมเศร้า (ค)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

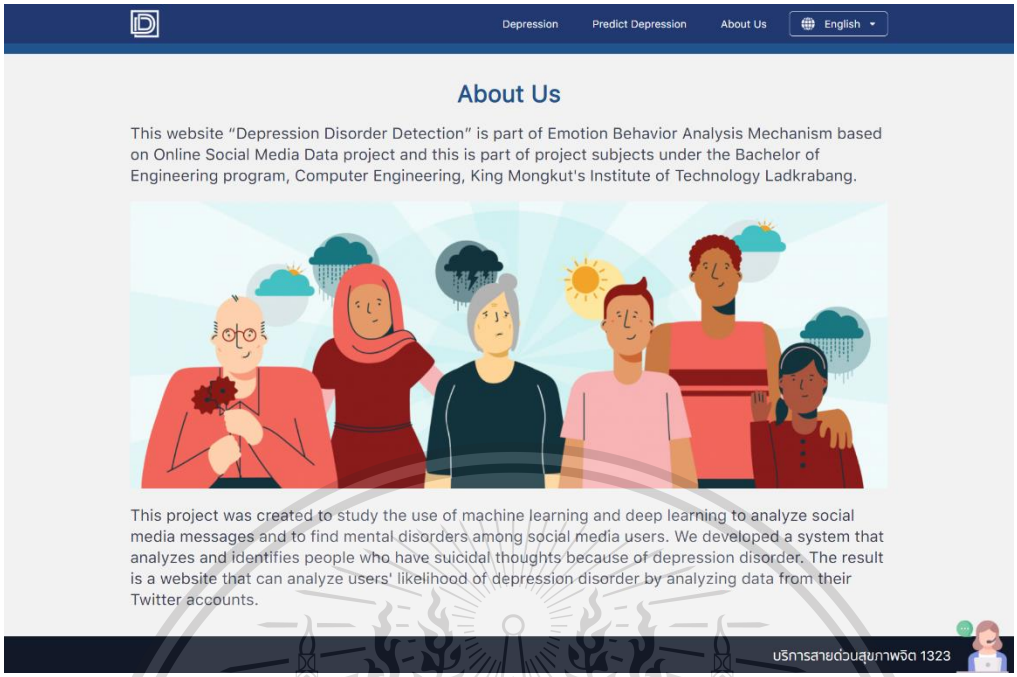
4.) หน้าแบบทดสอบ PHQ-9



รูป 3.23 ส่วนติดต่อผู้ใช้หน้าแบบทดสอบ PHQ-9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

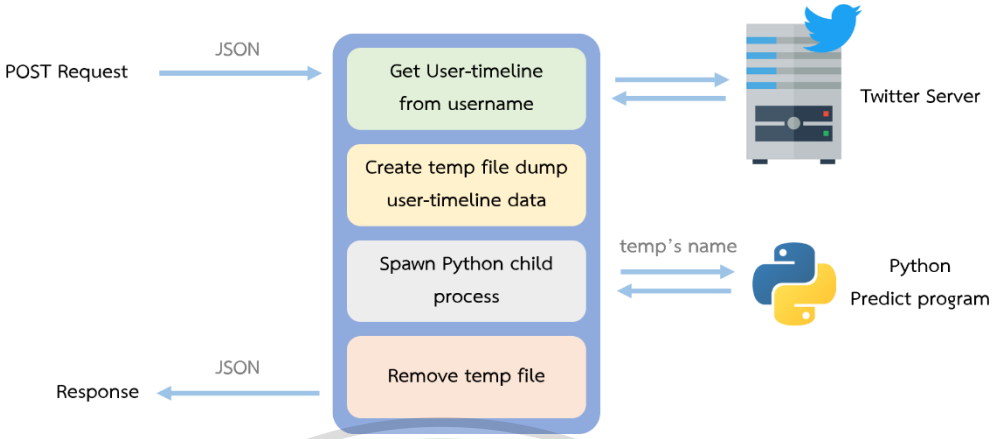
5.) หน้าเกี่ยวกับเรา



รูป 3.24 ส่วนติดต่อผู้ใช้หน้าเกี่ยวกับเรา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.3.6 การพัฒนาส่วน Back-end

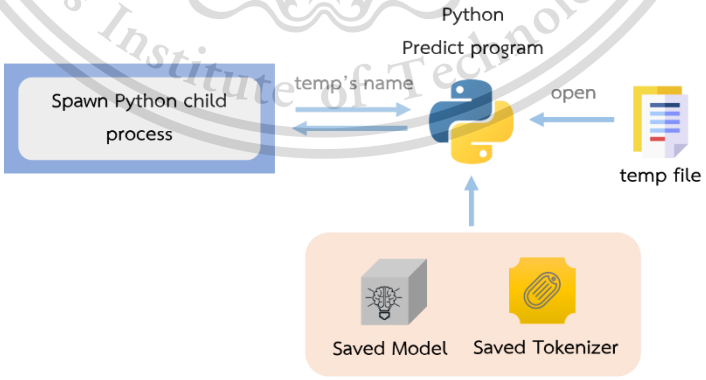


รูป 3.25 แผนผังแสดงทำงานในส่วน Back-end

จากรูปที่ 3.25 ฝั่ง Back-end จะมีหน้าที่ในการจัดหาส่วนต่อประสานโปรแกรมเพื่อรองรับ Request จากภายนอก เมื่อได้รับ Request มาแล้วจะทำการ Extract JSON นำชื่อผู้ใช้ไปดึงข้อมูลไทม์ไลน์จากทวิตเตอร์เซิร์ฟเวอร์ แล้วทำการสร้างไฟล์ชั่วคราวเก็บเอาไว้ จากนั้นจะทำการ Spawn child process ซึ่งรันโปรแกรม Python ที่ใช้เรียกโมเดลเพื่อทำนายผล เมื่อได้ผลลัพธ์ออกมาแล้ว จะทำการลบไฟล์ชั่วคราวแล้วส่ง Response ผลลัพธ์กลับไปในรูปแบบ JSON

3.3.7 การเชื่อมต่อระบบเข้ากับโมเดล

จากรูปที่ 3.25 เมื่อมีการ Spawn child process จะมีการส่ง Argument ซึ่งเป็นชื่อของไฟล์ช่วยคราวที่ถูกสร้างขึ้นเพื่อเก็บข้อมูลไทม์ไลน์ไปด้วย เพื่อให้โปรแกรม Python ใช้ในการอ่านข้อมูลไทม์ไลน์ และเรียกโมเดลเพื่อทำนายผล ดังแสดงในรูปที่ 3.26



รูป 3.26 แผนผังแสดงทำงานของ Python child process

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.3.8 ส่วนต่อประสานโปรแกรม (API Documentation)

ฝั่ง Back-end มีการจัดหาส่วนต่อประสานโปรแกรมไว้ ดังต่อไปนี้

1.) การทำนายด้วยบัญชีทวิตเตอร์

<table>
<tr><td>Path</td><td colspan="2">https://[Server_Address]/api/predict-twitter</td></tr>
<tr><td>Method</td><td colspan="2">POST</td></tr>
<tr><td>Header</td><td colspan="2">Content-Type: application/json</td></tr>
<tr><td>Request Body (JSON)</td><td colspan="2"><pre>{
    "username": "twitter_username"
}</pre></td></tr>
<tr><td>Response Body (JSON)</td><td colspan="2"><pre>{
    "status": 0,
    "message": "",
    "data": {}
}</pre></td></tr>
<tr><td>Status</td><td>Code</td><td>Message</td></tr>
<tr><td></td><td>200</td><td>OK Success</td></tr>
<tr><td></td><td>34</td><td>Error! Username does not exist</td></tr>
<tr><td></td><td>203</td><td>Error! Timeline is not enough to predict</td></tr>
<tr><td></td><td>301</td><td>Error! Mismatch JSON property</td></tr>
<tr><td></td><td>500</td><td>Unknown error occurs!</td></tr>
</table>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 4
ผลการทดลอง

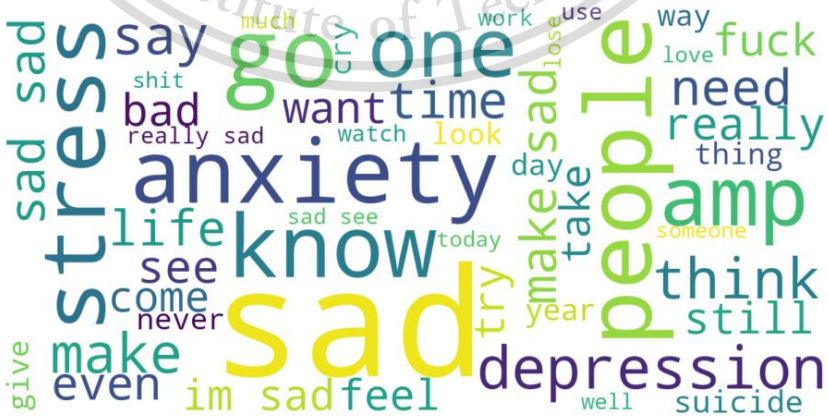
ในบทนี้ ผู้วิจัยได้นำเสนอแสดงภาพของชุดข้อมูลว่ามีคำใดที่ปรากฏบ่อยๆในกลุ่มข้อมูลประเภทต่างๆ และผลการทดลองจากการสอนโมเดลด้วยอัลกอริทึมรูปแบบต่างๆในรูปแบบของเมทริกซ์ความสับสน และกราฟ ROC พร้อมค่าพื้นที่ใต้กราฟ (AUC) แล้วเปรียบเทียบผลลัพธ์ระหว่างอัลกอริทึม และสรุปอัลกอริทึมที่ดีที่สุด รวมถึงทดสอบการใช้งานเว็บแอปพลิเคชัน

4.1 แสดงภาพชุดข้อมูล

4.1.1 ตัวอย่างคำที่พบในข้อความทวีตที่มีความปกติ

รูป 4.1 ตัวอย่างคำที่พบในข้อความทวีตที่มีความปกติ

4.1.2 ตัวอย่างคำที่พบในข้อความทวีตที่มีความซึมเศร้า



รูป 4.2 ตัวอย่างคำที่พบในข้อความทวีตที่มีความซึมเศร้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

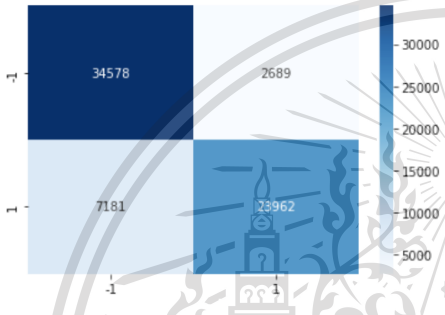
4.2 ผลการทดลองโมเดล

4.2.1 ผลการทดลอง Multinomial Naïve Bayes

- **ชุดตรวจสอบ**

--- Validate Set ---

	precision	recall	f1-score	support
-1	0.83	0.93	0.88	37267
1	0.90	0.77	0.83	31143
accuracy			0.86	68410
macro avg	0.86	0.85	0.85	68410
weighted avg	0.86	0.86	0.85	68410

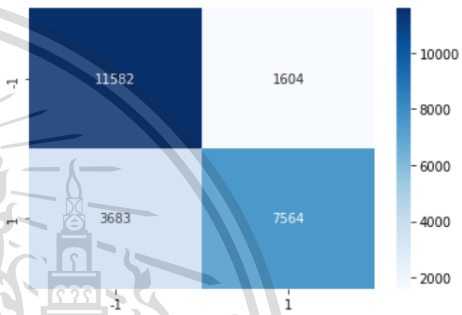


รูป 4.3 เมทริกซ์ความสับสน Multinomial Naïve Bayes ชุดตรวจสอบ

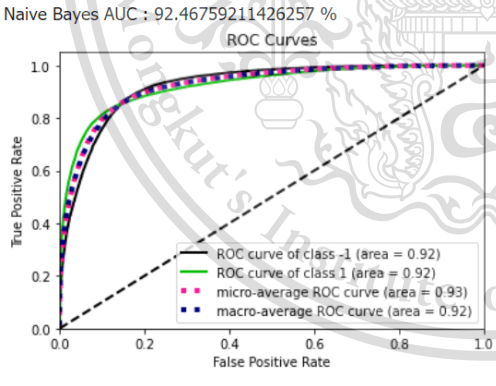
- **ชุดทดสอบ**

--- Test Set ---

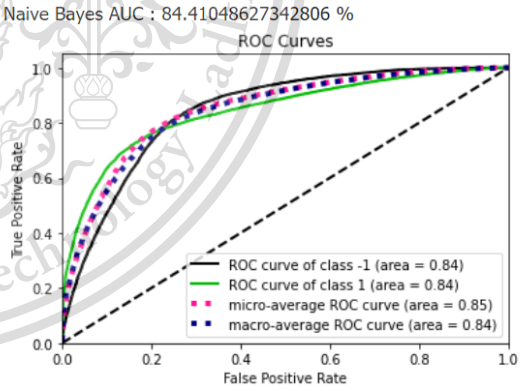
	precision	recall	f1-score	support
-1	0.76	0.88	0.81	13186
1	0.83	0.67	0.74	11247
accuracy			0.78	24433
macro avg	0.79	0.78	0.78	24433
weighted avg	0.79	0.78	0.78	24433



รูป 4.4 เมทริกซ์ความสับสน Multinomial Naïve Bayes ชุดทดสอบ



รูป 4.5 Multinomial Naïve Bayes AUC ชุดตรวจสอบ



รูป 4.6 Multinomial Naïve Bayes AUC ชุดทดสอบ

จากผลการทดลอง Multinomial Naïve Bayes พบว่าสำหรับข้อมูลชุดตรวจสอบโมเดลมีค่าความแม่นยำอยู่ที่ 86% และสำหรับข้อมูลชุดทดสอบโมเดลมีค่าความแม่นยำอยู่ที่ 78% และจากข้อมูลทั้ง 2 ชุดโมเดลมีความผิดพลาดในการทำนายคลาส 1 มากกว่าคลาส -1 ค่อนข้างมาก

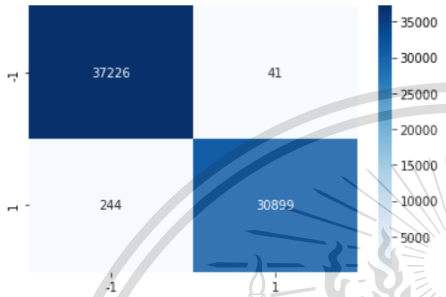
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.2.2 ผลการทดลองต้นไม้ตัดสินใจ

- **ชุดตรวจสอบ**

--- Validate Set ---

	precision	recall	f1-score	support
-1	0.99	1.00	1.00	37267
1	1.00	0.99	1.00	31143
accuracy			1.00	68410
macro avg	1.00	1.00	1.00	68410
weighted avg	1.00	1.00	1.00	68410

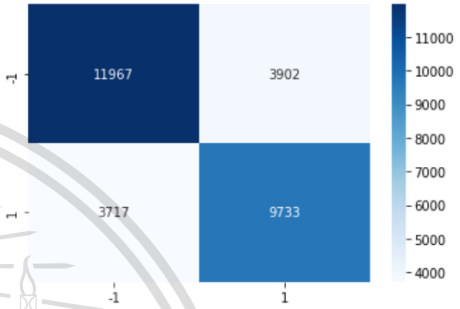


รูป 4.7 เมทริกซ์ความสับสน Decision Tree ชุดตรวจสอบ

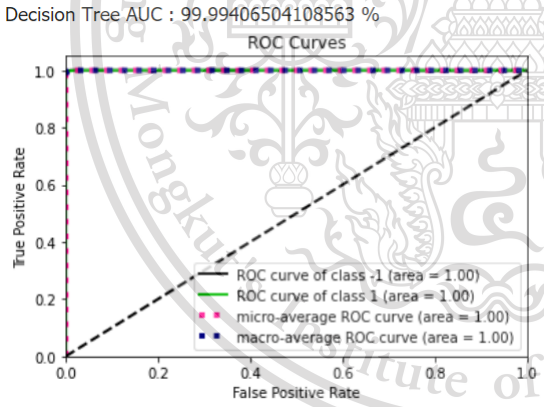
- **ชุดทดสอบ**

--- Test Set ---

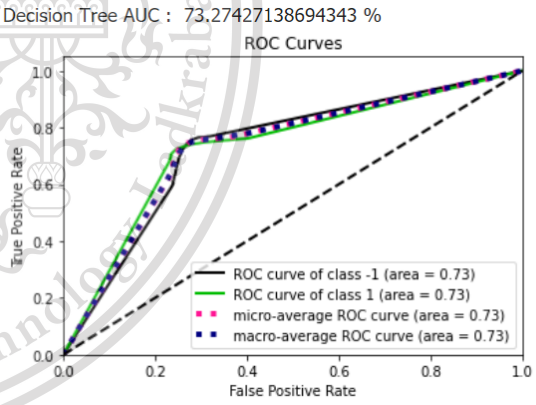
	precision	recall	f1-score	support
-1	0.76	0.75	0.76	15869
1	0.71	0.72	0.72	13450
accuracy			0.74	29319
macro avg	0.74	0.74	0.74	29319
weighted avg	0.74	0.74	0.74	29319



รูป 4.8 เมทริกซ์ความสับสน Decision Tree ชุดทดสอบ



รูป 4.9 Decision Tree AUC ชุดตรวจสอบ



รูป 4.10 Decision Tree AUC ชุดทดสอบ

จากผลการทดลองต้นไม้ตัดสินใจ พบว่าสำหรับข้อมูลชุดตรวจสอบโมเดลมีค่าความแม่นยำอยู่ที่ 100% และสำหรับข้อมูลชุดทดสอบโมเดลมีค่าความแม่นยำอยู่ที่ 74% ซึ่งมีความต่างกันค่อนข้างมาก หมายความว่าโมเดลต้นไม้ตัดสินใจสามารถทำนายสิ่งที่เคยเรียนรู้แล้วได้ดี แต่ทำนายสิ่งที่ไม่เคยเรียนยังไม่ดี

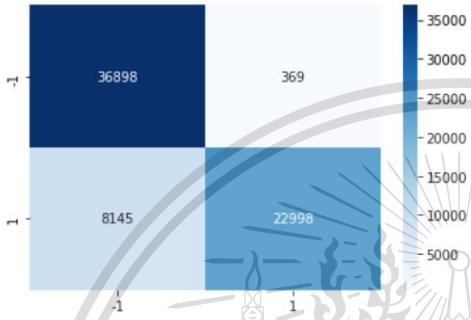
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.2.3 ผลการทดลอง Random Forest

- **ชุดตรวจสอบ**

--- Validate Set ---

	precision	recall	f1-score	support
-1	0.82	0.99	0.90	37267
1	0.98	0.74	0.84	31143
accuracy			0.88	68410
macro avg	0.90	0.86	0.87	68410
weighted avg	0.89	0.88	0.87	68410

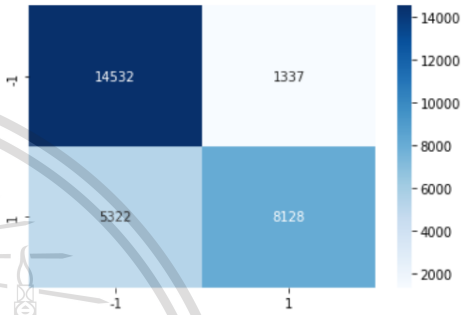


รูป 4.11 เมทริกซ์ความสับสน Random Forest ชุดตรวจสอบ

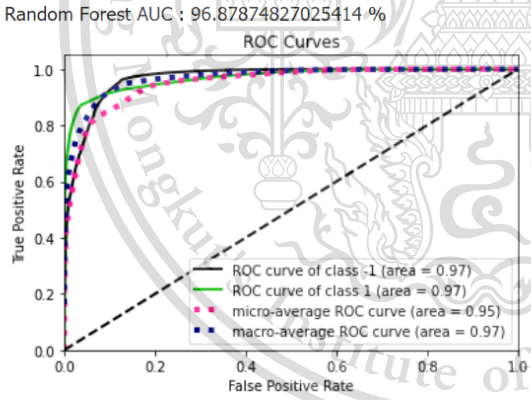
- **ชุดทดสอบ**

--- Test Set ---

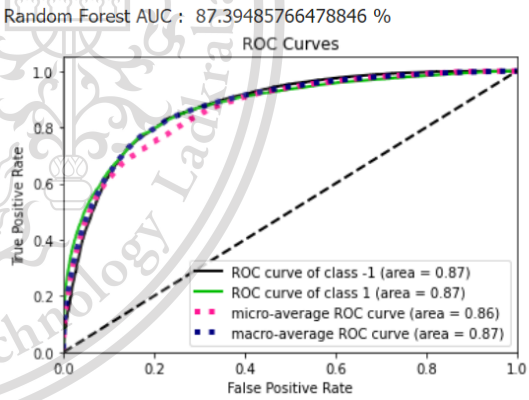
	precision	recall	f1-score	support
-1	0.73	0.92	0.81	15869
1	0.86	0.60	0.71	13450
accuracy			0.77	29319
macro avg	0.80	0.76	0.76	29319
weighted avg	0.79	0.77	0.77	29319



รูป 4.12 เมทริกซ์ความสับสน Random Forest ชุดทดสอบ



รูป 4.13 Random Forest AUC ชุดตรวจสอบ



รูป 4.14 Random Forest AUC ชุดทดสอบ

จากผลการทดลอง Random Forest พบว่าสำหรับข้อมูลชุดตรวจสอบโมเดลมีค่าความแม่นยำอยู่ที่ 88% และสำหรับข้อมูลชุดทดสอบโมเดลมีค่าความแม่นยำอยู่ที่ 77% และจากข้อมูลทั้ง 2 ชุด โมเดลมีความผิดพลาดในการทำนายคลาส 1 มากกว่าคลาส -1 อยู่ค่อนข้างมาก

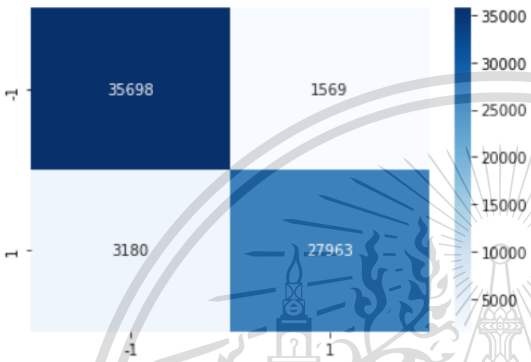
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.2.4 ผลการทดลองซัพพอร์ทเวกเตอร์แมชชีน

- **ชุดตรวจสอบ**

--- Validate Set ---

	precision	recall	f1-score	support
-1	0.92	0.96	0.94	37267
1	0.95	0.90	0.92	31143
accuracy			0.93	68410
macro avg	0.93	0.93	0.93	68410
weighted avg	0.93	0.93	0.93	68410



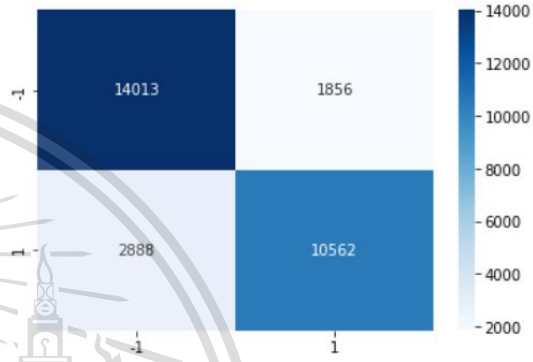
รูป 4.15 เมทริกซ์ความสับสน SVM ชุดตรวจสอบ

รูป 4.17 SVM AUC ชุดตรวจสอบ

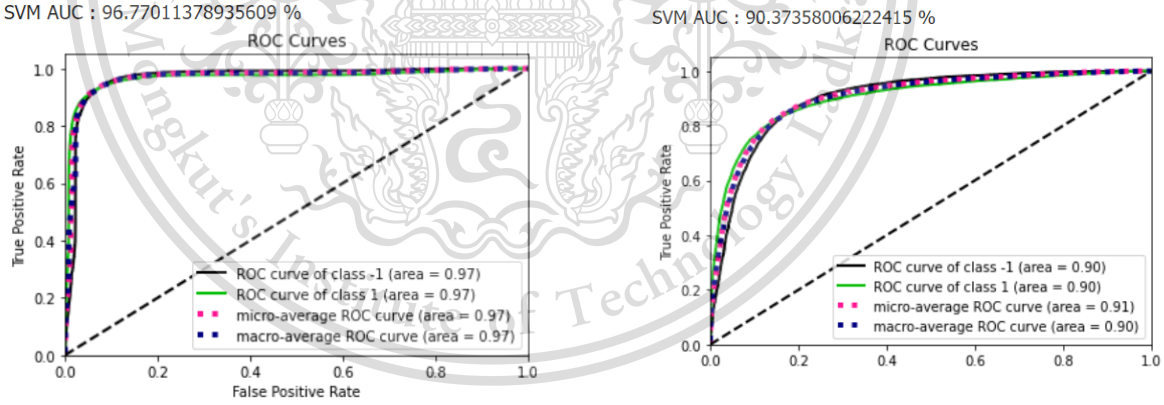
- **ชุดทดสอบ**

--- Test Set ---

	precision	recall	f1-score	support
-1	0.83	0.88	0.86	15869
1	0.85	0.79	0.82	13450
accuracy			0.84	29319
macro avg	0.84	0.83	0.84	29319
weighted avg	0.84	0.84	0.84	29319



รูป 4.16 เมทริกซ์ความสับสน SVM ชุดทดสอบ



รูป 4.18 SVM AUC ชุดทดสอบ

จากผลการทดลองซัพพอร์ทเวกเตอร์แมชชีน พบว่าสำหรับข้อมูลชุดตรวจสอบโมเดลมีค่าความแม่นยำอยู่ที่ 93% และสำหรับข้อมูลชุดทดสอบโมเดลมีค่าความแม่นยำอยู่ที่ 84% ซึ่งค่อนข้างสูง และจากข้อมูลทั้ง 2 ชุด โมเดลมีความผิดพลาดในการทำนายคลาส 1 มากกว่าคลาส -1 ไม่สูงมาก

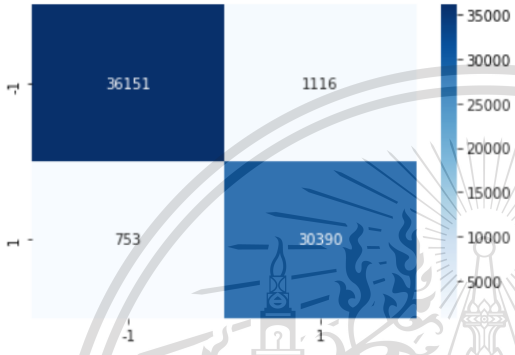
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.2.5 ผลการทดลองโครงข่ายประสาทแบบ Long short-term memory

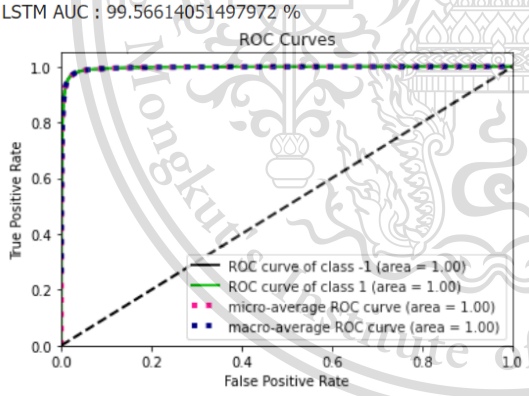
- **ชุดตรวจสอบ**

--- Validate Set ---

	precision	recall	f1-score	support
-1	0.98	0.97	0.97	37267
1	0.96	0.98	0.97	31143
accuracy			0.97	68410
macro avg	0.97	0.97	0.97	68410
weighted avg	0.97	0.97	0.97	68410



รูป 4.19 เมทริกซ์ความสับสน LSTM ชุดตรวจสอบ

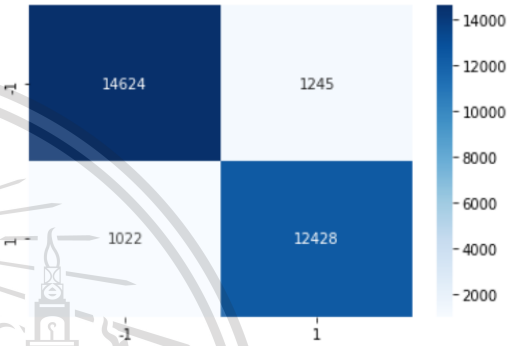


รูป 4.21 LSTM AUC ชุดตรวจสอบ

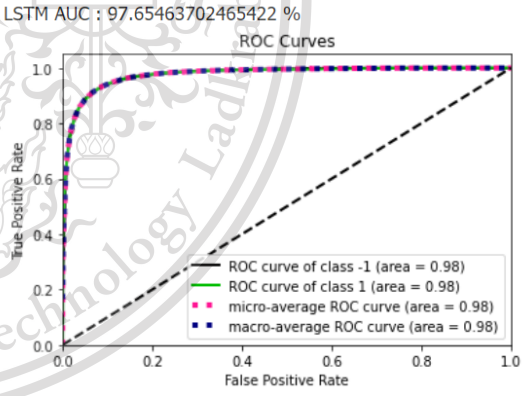
- **ชุดทดสอบ**

--- Test Set ---

	precision	recall	f1-score	support
-1	0.93	0.92	0.93	15869
1	0.91	0.92	0.92	13450
accuracy			0.92	29319
macro avg	0.92	0.92	0.92	29319
weighted avg	0.92	0.92	0.92	29319



รูป 4.20 เมทริกซ์ความสับสน LSTM ชุดทดสอบ



รูป 4.22 LSTM AUC ชุดทดสอบ

จากผลการทดลองโครงข่ายประสาทแบบ Long short-term memory พบว่าสำหรับข้อมูลชุดตรวจสอบโมเดลมีค่าความแม่นยำอยู่ที่ 97% และสำหรับข้อมูลชุดทดสอบโมเดลมีค่าความแม่นยำอยู่ที่ 92% ซึ่งค่อนข้างดีมาก และมีความผิดพลาดในการทำนายคลาส 1 และคลาส -1 อยู่ค่อนข้างต่ำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.3 สรุปผลการทดลองโมเดล

Model	Validate Set		Test Set	
	ACC	AUC	ACC	AUC
Multinomial Naïve Bayes	86	92.46	78	84.41
Decision Tree	**100**	**99.99**	74	73.27
Random Forest	88	96.87	77	87.39
Support Vector Machine	93	96.77	84	90.37
LSTM	97	99.56	**92**	**97.65**

ตาราง 4.1 สรุปค่าความแม่นยำของทุกโมเดล

Model	Validate Set			Test Set		
	P	R	F1	P	R	F1
Multinomial Naïve Bayes	0.87	0.85	0.86	0.80	0.78	0.78
Decision Tree	**1.00**	**1.00**	**1.00**	0.74	0.74	0.74
Random Forest	0.90	0.87	0.87	0.80	0.76	0.76
Support Vector Machine	0.94	0.93	0.93	0.84	0.84	0.84
LSTM	0.97	0.98	0.97	**0.92**	**0.92**	**0.93**

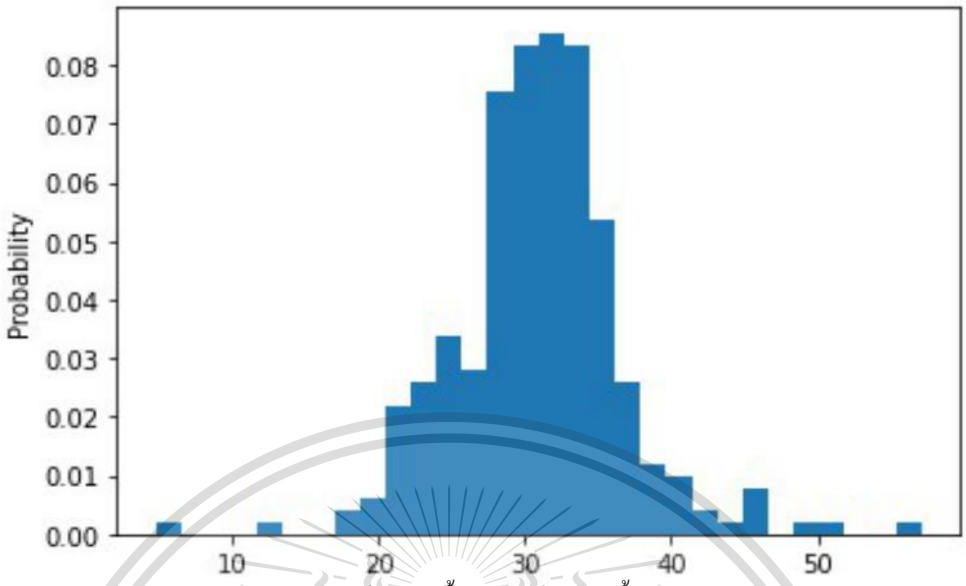
ตาราง 4.2 สรุปค่า P, R และ F1 ของทุกโมเดล

4.4 การพิจารณาโมเดลที่เหมาะสม

จากผลการทดลองโมเดลใน 4.2.1-4.2.5 จะเห็นว่า โมเดลโครงข่ายประสาทแบบ Long Short-Term Memory (LSTM) เป็นโมเดลที่มีประสิทธิภาพโดยรวมสูงที่สุด มีค่าความแม่นยำของชุดทดสอบที่สูงกว่าโมเดลประเภทอื่น มีความสามารถในการทำนายสิ่งที่ยังไม่เคยเรียนได้ค่อนข้างดี และมีค่าความผิดพลาดในการทำนายแต่ละคลาสค่อนข้างต่ำ ดีกว่าโมเดลประเภทอื่นที่ความผิดพลาดในการทำนายแต่ละคลาสต่างกันค่อนข้างมาก ดังนั้นโมเดล LSTM จึงมีความเหมาะสมที่จะนำไปใช้งานจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.5 ผลการทดลองทดสอบโมเดลกับชุดข้อมูล CLPsych 2015



รูป 4.23 การทดลองทดสอบโมเดลกับชุดข้อมูล CLPsych 2015

จากทดลองนำโมเดลไปทดสอบกับชุดข้อมูล CLPsych 2015 ไปเพื่อทำนายทวิต (ข้อความ) ของแต่ละบัญชีทวิตเตอร์เพื่อหาอัตราส่วนข้อความขั้วลบต่อข้อความทั้งหมดได้ผลดังแสดงในรูปที่ 4.23 จะเห็นได้ว่าผู้ใช้ที่มีอาการของโรคซึมเศร้าส่วนมากจะมีอัตราส่วนข้อความขั้วลบต่อข้อความทั้งหมดอยู่ในช่วง 30% ค่อนข้างมาก จากข้อมูลนี้ทำให้สามารถคำนวณหาค่า Q1,Q2 และ Q3 ได้อยู่ที่ Q1 = 28.14 / Q2 = 31 / Q3 = 33.75 ซึ่งค่า Quartile จะถูกนำไปเป็นเกณฑ์ในการประเมินโรคซึมเศร้าต่อไป

4.6 เปรียบเทียบโครงงานกับงานวิจัยที่เกี่ยวข้อง

4.6.1 Classification of Depression on Social Media Using Text Mining

จากโครงงานเรื่อง Classification of Depression on Social Media Using Text Mining ของ Nikie Jo Elauria Deocampo (2019) ได้มีผลการทดลองซึ่งเป็นค่า AUC ของข้อมูลชุด Validate ในระดับ Tweet-level ดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Model	AUC (Validate Set) (%)
Naive Bayes	93.79
ต้นไม้ตัดสินใจ	**98.56**
ซัพพอร์ทเวกเตอร์แมชชีน	50.00
วิธีการเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbour)	81.46
Random Forest	49.10

ตาราง 4.3 ผลการทดลอง Classification of Depression on Social Media Using Text Mining

จากตาราง 4.3 โมเดลที่ดีที่สุดคือโมเดลต้นไม้ตัดสินใจซึ่งมีค่า AUC ของข้อมูลชุด Validate อยู่ที่ 98.56% เมื่อทำการเปรียบเทียบงานของเราในตารางที่ 4.1 พบว่างานของเรามีโมเดลที่มีค่า AUC ของข้อมูลชุด Validate สูงกว่างานนี้อยู่ 2 โมเดล ได้แก่ โมเดลต้นไม้ตัดสินใจซึ่งมีค่า AUC ของข้อมูลชุด Validate อยู่ที่ 99.99% และโมเดลโครงข่ายประสาทเทียม LSTM ซึ่งมีค่า AUC ของข้อมูลชุด Validate อยู่ที่ 99.56%

4.6.2 Monitoring Tweets for Depression to Detect At-risk Users

จากโครงงานเรื่อง Monitoring Tweets for Depression to Detect At-risk Users ของ Zunaira Jamil และคณะ (2017) ได้มีผลการทดลองซึ่งเป็นค่าความแม่นยำ Precision Recall และ F1 ของข้อมูลชุด Test ในระดับ Tweet-level ดังต่อไปนี้

Model	Accuracy (%)	Precision	Recall	F1
Baseline	**94.69**	**1.0000**	0.0111	0.0219
Exp1-Original	93.37	NA	0.0000	NA
Exp1-SMOTE	78.16	0.1706	0.5939	**0.2650**
Exp1-Undersample	61.02	0.1237	**0.8020**	0.2144
Exp2-Original	93.03	0.2222	0.0203	0.0372
Exp2-SMOTE	77.11	0.1124	0.3553	0.1707
Exp2-Undersample	61.43	0.1219	0.7766	0.2107

ตาราง 4.4 ผลการทดลอง Monitoring Tweets for Depression to Detect At-risk Users

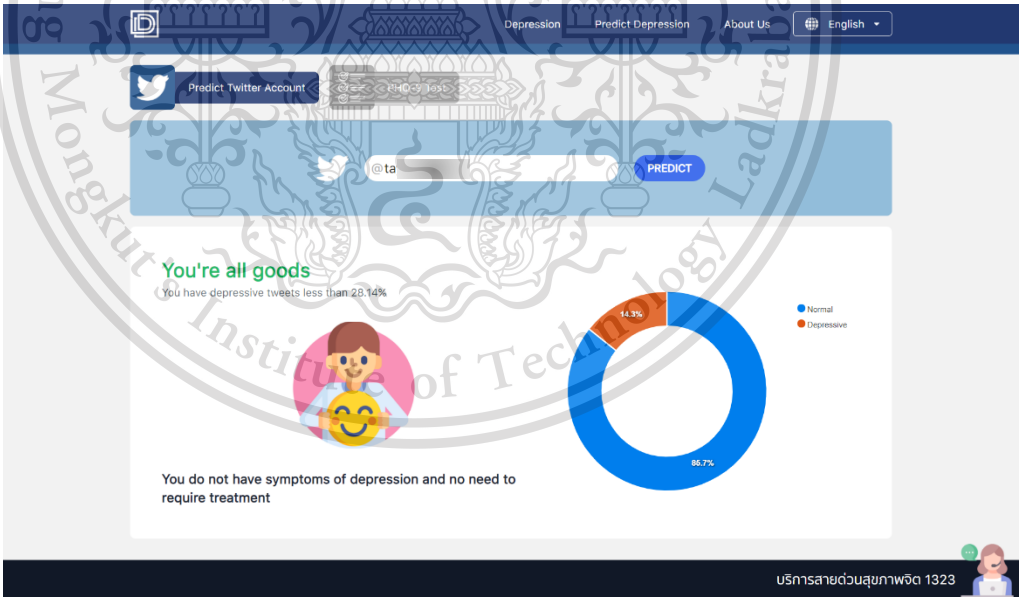
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จากตาราง 4.4 พบว่าโมเดลที่มีค่าความแม่นยำ และ Precision สูงที่สุดคือ Baseline ซึ่งมีค่าความแม่นยำอยู่ที่ 94.96% และค่า Precision อยู่ที่ 1.000 โมเดลที่มีค่า Recall สูงที่สุดคือ Exp1-Undersample ซึ่งมีค่า Recall อยู่ที่ 0.8020 และโมเดลที่มีค่า F1 สูงที่สุดคือ Exp1-SMOTE ซึ่งมีค่า F1 อยู่ที่ 0.2650 เมื่อทำการเปรียบเทียบงานของเราในตารางที่ 4.2 พบว่าโมเดลของเราที่มีประสิทธิภาพสูงที่สุด (โครงข่ายประสาทเทียม LSTM) ของเรานั้นมีประสิทธิภาพสูงกว่าในค่าของ Recall และ F1 โดยมีค่า Recall อยู่ที่ 0.92 และค่า F1 อยู่ที่ 0.93 แต่ในส่วนของค่าความแม่นยำ และ Precision นั้น โมเดลของเรามีประสิทธิภาพที่ด้อยกว่า โดยมีค่าความแม่นยำอยู่ที่ 92% และค่า Precision อยู่ที่ 0.92

4.7 ผลการทดสอบเว็บแอปพลิเคชัน

4.7.1 ผลการทดสอบคุณสมบัติของเว็บแอปพลิเคชัน

1.) หน้าทำนายโรคซึมเศร้าด้วยบัญชีทวิตเตอร์ ผู้ใช้ทำการกรอกชื่อบัญชีทวิตเตอร์ และกดปุ่มทำนายผล หากชื่อบัญชีทวิตเตอร์ถูกต้องระบบจะทำการทำนาย และแสดงผลลัพธ์การทำนาย (R01, R02) ดังแสดงในรูปที่ 4.24

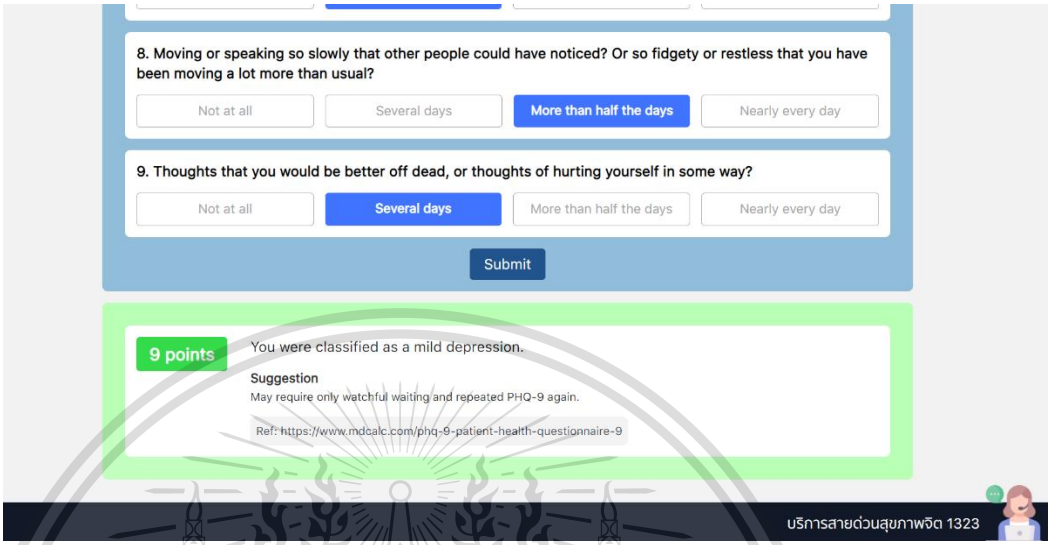


รูป 4.24 การทดสอบเว็บหน้าทำนายโรคซึมเศร้าด้วยบัญชีทวิตเตอร์

หากมีข้อผิดพลาดเกิดขึ้น ได้แก่ ชื่อบัญชีทวิตเตอร์ไม่มีอยู่จริง ปริมาณข้อความในไทม์ไลน์ผู้ใช้ไม่เพียงพอต่อการทำนาย และไม่สามารถเชื่อมต่อกับ API ได้ ระบบจะแสดงผลตามข้อผิดพลาดที่เกิดขึ้น

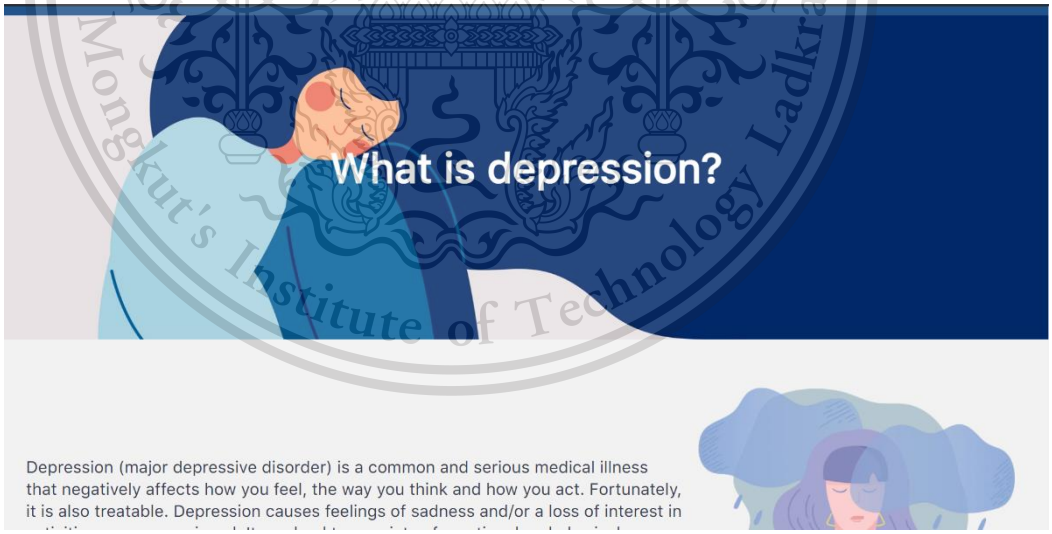
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.) หน้าทำแบบทดสอบโรคซึมเศร้า PHQ-9 ผู้ใช้ทำการตอบคำถามตามความจริงให้ครบทุกข้อ แล้วกดปุ่มส่งคำตอบ ระบบจะแสดงคะแนน และผลการทำแบบทดสอบ **(R03, R04)** ดังแสดงในรูปที่ 4.25



รูป 4.25 การทดสอบเว็บหน้าทำแบบทดสอบ PHQ-9

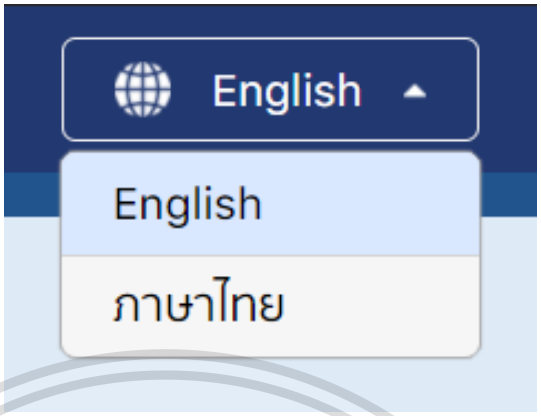
3.) หน้าทำข้อมูลโรคซึมเศร้า ผู้ใช้สามารถอ่านรายละเอียด และศึกษาข้อมูลเกี่ยวกับโรคซึมเศร้าได้ **(R05)** ดังแสดงในรูปที่ 4.26



รูป 4.26 การทดสอบเว็บหน้าข้อมูลโรคซึมเศร้า

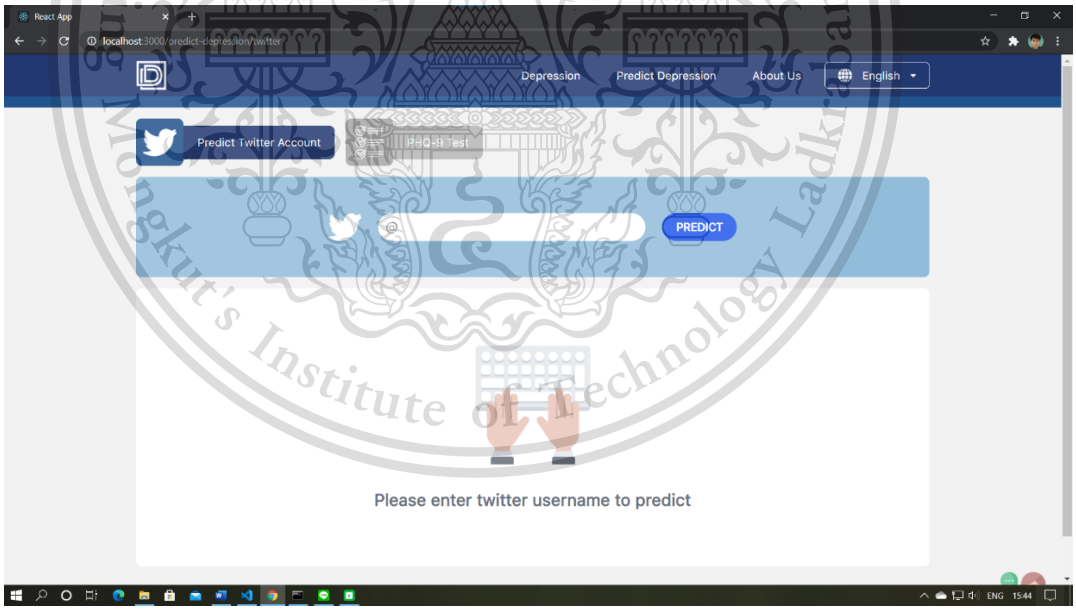
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.) ที่แถบนำทาง ผู้ใช้ทำการกดที่ Drop-down ภาษา และเลือกภาษาที่ต้องการ ระบบจะทำการแสดงหน้าเว็บตามภาษาที่ผู้ใช้เลือก **(R06)** ดังแสดงในรูปที่ 4.27



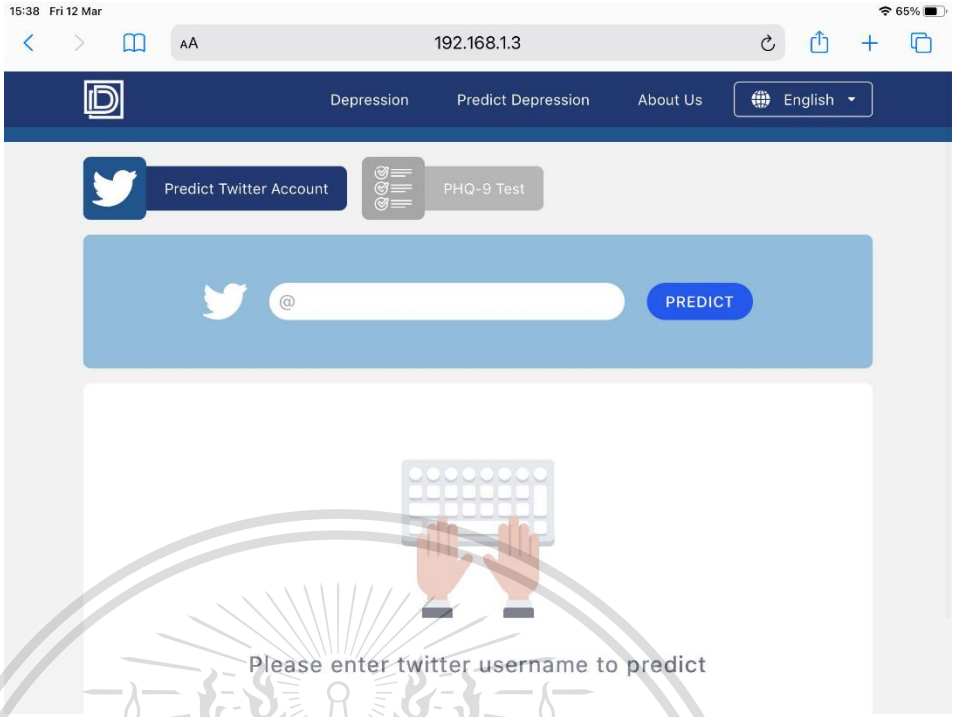
รูป 4.27 การทดสอบการเปลี่ยนภาษา

5.) หน้าเว็บจะทำการปรับรูปแบบส่วนติดต่อผู้ใช้ให้เหมาะสมกับอุปกรณ์ที่ผู้ใช้ใช้งาน **(R07, R08)** ดังแสดงในรูปที่ 4.28 และรูปที่ 4.29

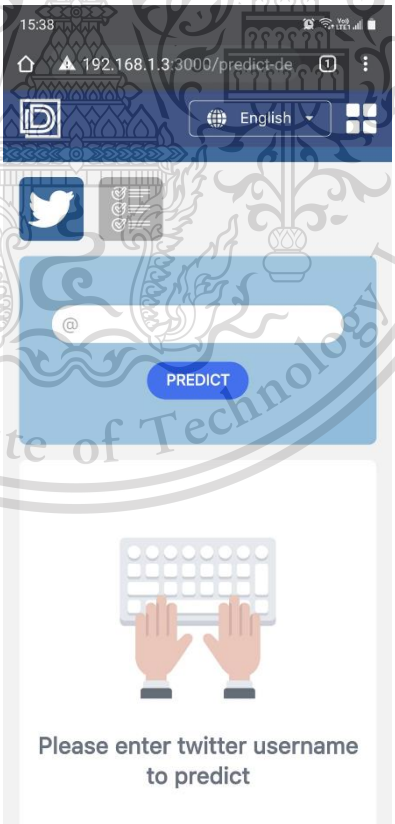


รูป 4.28 การทดสอบใช้งานบนเดสก์ท็อป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 4.29 การทดสอบใช้งานบนแท็บเล็ต



รูป 4.30 การทดสอบใช้งานบนสมาร์ทโฟน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

6.) ระบบจะไม่เปิดเผยข้อมูลข้อความทวีตบนไทม์ไลน์ของผู้ใช้ ตามข้อปฏิบัติของ Twitter Developer Policy ที่มีข้อบังคับข้อตกลงในการใช้บริการ Twitter API ที่ไม่ให้นำข้อมูลที่ดึงจากผู้ใช้ไปเผยแพร่ ทำให้ระบบออกแบบมาโดยไม่มีการเก็บข้อมูลทวิตที่นำมาวิเคราะห์ทั้งหมด และข้อมูลที่ใช้บริการ Twitter API ดึงข้อมูลมาจะเป็นความลับทั้งหมด

4.7.2 ผลการทดลองการทำนายผลจากไทม์ไลน์ทวิตเตอร์ของผู้ใช้

ทำการทดลองโดยทำการสุ่มผู้ใช้ทวิตเตอร์ 15 คน แล้วป้อนชื่อลงไปในเว็บแอพพลิเคชั่นเพื่อทำการทำนายผลจากไทม์ไลน์ทวิตเตอร์ของผู้ใช้ แล้วบันทึกผลแนวโน้มการป่วยเป็นโรคซึมเศร้าของผู้ใช้ (ทำการทดลอง ณ วันที่ 19 เมษายน 2564)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ผู้ใช้*	ผลการทดลองการทำนายผลจากไทม์ไลน์ทวิตเตอร์ของผู้ใช้				อัตราส่วนข้อความปกติ (%)	อัตราส่วนข้อความซึมเศร้า (%)
	ไม่มีอาการซึมเศร้า	ซึมเศร้าเล็กน้อย	ซึมเศร้าปานกลาง	ซึมเศร้ามาก		
ผู้ใช้ 1	✓				82.70	17.30
ผู้ใช้ 2	✓				76.30	23.70
ผู้ใช้ 3		✓			69.70	30.40
ผู้ใช้ 4	✓				78.40	21.60
ผู้ใช้ 5				✓	66.00	34.00
ผู้ใช้ 6				✓	54.60	45.40
ผู้ใช้ 7	✓				86.30	13.80
ผู้ใช้ 8	✓				85.30	14.70
ผู้ใช้ 9	✓				78.50	21.50
ผู้ใช้ 10	✓				80.60	19.40
ผู้ใช้ 11	✓				73.80	26.20
ผู้ใช้ 12	✓				90.30	9.70
ผู้ใช้ 13	✓				73.50	26.50
ผู้ใช้ 14		✓			69.30	30.70
ผู้ใช้ 15	✓				83.90	16.10

ตาราง 4.5 ผลผลการทดลองการทำนายผลจากไทม์ไลน์ทวิตเตอร์ของผู้ใช้

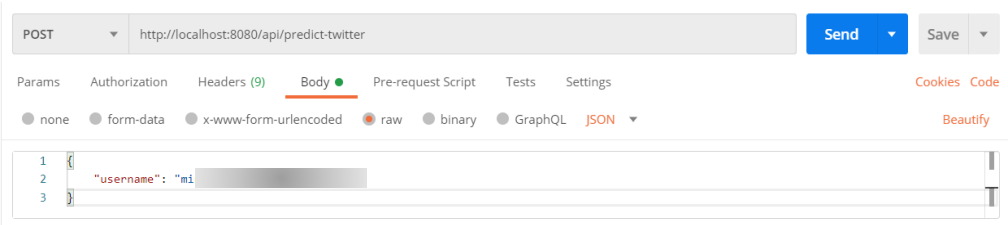
*ชื่อผู้ใช้ไม่สามารถเปิดเผยได้เนื่องจาก ต้องทำให้เป็นไปตามข้อบังคับของบัญชีผู้พัฒนาทวิตเตอร์ ซึ่งไม่ให้เปิดเผยข้อมูลของผู้ใช้

จากตาราง 4.8 พบว่าจากผู้ใช้ทั้งหมด 15 คน มีผู้ใช้ที่ไม่มีอาการซึมเศร้า 11 คน ผู้ใช้ที่มีแนวโน้มซึมเศร้าเล็กน้อย 2 คน ผู้ใช้ที่มีแนวโน้มซึมเศร้าปานกลาง 0 คน และมีผู้ใช้ที่มีแนวโน้มซึมเศร้ามาก 2 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.7.3 ผลการทดสอบส่วนต่อประสานโปรแกรม

ในส่วนของการทดสอบส่วนต่อประสานโปรแกรมทำได้โดยใช้โปรแกรม Postman ทดสอบการส่ง POST Request ไปที่ Back-end Server โดยกำหนด Body ในรูปแบบ JSON ดังแสดงในรูปที่ 4.31

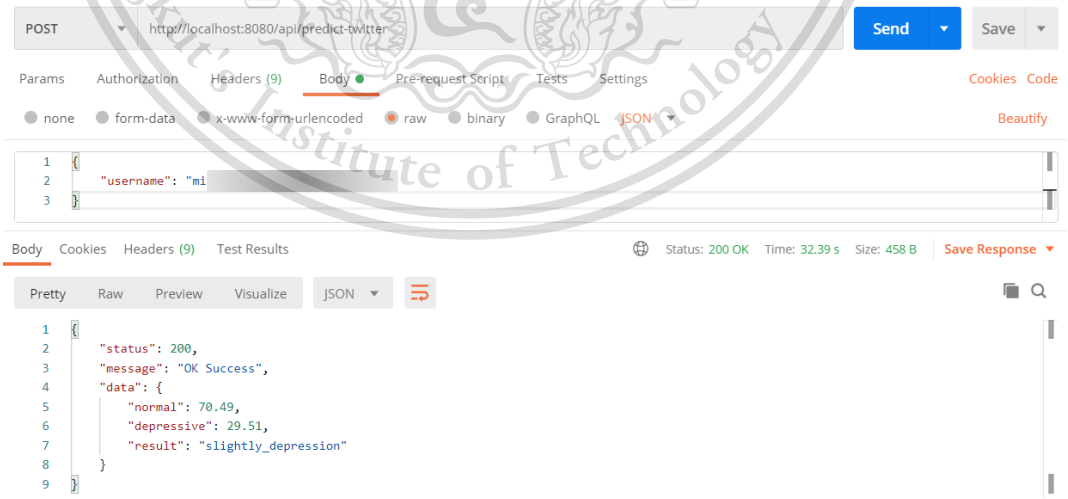


รูป 4.31 ตัวอย่าง POST Request ไปที่ส่วนต่อประสานโปรแกรม (ก)

เมื่อ Back-end Server ได้รับ Request เข้ามาแล้ว จะต้องทำการเรียก Process ขึ้นมาทำงาน โดยจะแสดงผลใน Terminal ของฝั่ง Back-end ตามรูปที่ 4.32

รูป 4.32 Back-end Terminal เมื่อมี Request เข้ามา

เมื่อ Process เสร็จสิ้น จะทำการ Response ผลลัพธ์ JSON กลับไป ดังแสดงในรูปที่ 4.33



รูป 4.33 ตัวอย่าง POST Request ไปที่ส่วนต่อประสานโปรแกรม (ข)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 5
สรุป

จากการทำปริญญานิพนธ์กลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์ จึงได้ทราบถึงปัญหาที่เกิดขึ้นระหว่างการจัดทำ และการทดสอบเว็บแอปพลิเคชัน ดังนั้นจึงมีผลสรุปและข้อเสนอแนะของการพัฒนากลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์ เพื่อที่นำมาเป็นแนวทางในการพัฒนาต่อไป

5.1 สรุปผลที่ได้จากการทดลอง

กลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์ได้ทำการสร้างโมเดลทำนายภาวะโรคซึมเศร้า โมเดลใช้ข้อมูลการทวีตของบุคคลทั่วไปที่ใช้สื่อสังคมออนไลน์ “ทวิตเตอร์” โดยดึงข้อมูลจากการสตรีมทวิตเตอร์ด้วยส่วนต่อประสานทวิตเตอร์ (Twitter API) ฝึกฝนโมเดลด้วยอัลกอริทึมทางการเรียนรู้ของเครื่อง (Machine Learning) และ การเรียนรู้เชิงลึก (Deep Learning) จากหลายๆ อัลกอริทึม ซึ่งผลลัพธ์ที่ดีที่สุดคืออัลกอริทึมทางการเรียนรู้เชิงลึก “LSTM” ที่มีความแม่นยำในการทำนายขั้วของข้อความอยู่ที่ 92% และใช้ชุดข้อมูล CLPsych 2015 Shared Task เป็นตัวกำหนดเกณฑ์ในการทำนายผลจากไทม์ไลน์ผู้ใช้

กลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์ได้พัฒนาเว็บแอปพลิเคชันโดยมีวัตถุประสงค์เพื่อความสะดวกในการใช้งานของผู้ใช้ เว็บแอปพลิเคชันมีฟีเจอร์หลักในงานทำนายโรคซึมเศร้าจากบัญชีผู้ใช้ทวิตเตอร์ และมีฟีเจอร์รองเป็นแบบทดสอบโรคซึมเศร้า PHQ-9 และเกร็ดความรู้โรคซึมเศร้า เว็บไซต์ได้ทำการออกแบบเป็นแบบ Responsive และรองกับการเปลี่ยนภาษาของเว็บไซต์เป็นภาษาอังกฤษ

ปัจจุบันสื่อสังคมออนไลน์ได้เข้ามาปฏิวัติวิถีชีวิตของคนในปัจจุบัน ทำให้ผู้คนเชื่อมถึงกันและแสดงออกกันมากขึ้น ซึ่งการแสดงออกบนสื่อสังคมออนไลน์ได้บ่งชี้ถึงอาการที่ไม่สามารถรู้ได้ในสภาวะปกติ จากการทดลองเห็นได้ชัดเจนว่าผู้ที่มีภาวะของโรคซึมเศร้าจะมีการแสดงออกซ่อนอยู่ในข้อความที่ทวิตลงในสื่อสังคมออนไลน์ “ทวิตเตอร์” ไม่ว่าความกังวล ความโดดเดี่ยว หรือความเศร้า ซึ่งอาการเหล่านี้ส่งต่อข้อความที่ทวิต โครงงานกลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์จึงหาอาการของโรคซึมเศร้าจะข้อความเหล่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

5.2 ปัญหาที่พบ

จากปัญหาที่พบในพัตนากลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์นี้ พบปัญหาที่เกิดขึ้นคือ

- ชุดข้อมูล CLP ที่ได้มามัน Imbalance และไม่ Generalized จากการที่มี Chunk ของไฟล์บางส่วนหายไป
- ทดลองแก้ปัญหาชุดข้อมูล CLP ด้วยวิธี Up-sample และ Down-sample พร้อมทำ Cross-validation แต่ตัวชุดข้อมูลยังคงไม่ Generalized อยู่
- การเตรียมและทำความสะอาดข้อมูลทวิตเตอร์เป็นอะไรที่ค่อนข้างท้าทาย
- ทวิตเตอร์จำกัดข้อความทวิตสูงสุด 200 ข้อความต่อรีเควสซึ่งทำให้ต้องใช้การวนลูปหรือ Recursive ตามรีเควสทำให้อาจใช้เวลาเพิ่มขึ้นในการดึงข้อมูลไทม์ไลน์ผู้ใช้
- Twitter Developer ที่นำมาใช้ไม่ใช่บัญชี Enterprise ทำให้มีการจำกัดรีเควสได้สูงสุด 100000 ครั้งต่อวัน
- การตรวจสอบไอดีทวีตนั้นมีตัวเลขที่เกินกว่า Int และ Float 64 บิตจึงต้องมีการใช้ String encoding มาช่วย

5.3 แนวทางในการพัฒนาต่อ

จากการพัฒนากลไกการวิเคราะห์พฤติกรรมทางอารมณ์จากข้อมูลสื่อสังคมออนไลน์ เพื่อการต่อยอดในอนาคตจึงเสนอแนวทางในการพัฒนาต่อดังนี้

- พัฒนาโมเดลให้รองรับภาษาไทย
- ฝึกฝนโมเดลด้วยข้อมูลจากสื่อสังคมออนไลน์อื่นๆ เช่น เฟซบุ๊ก อินสตาแกรม
- ใช้ข้อมูลรูปเป็นปัจจัยเสริมในการช่วยฝึกสอนโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บรรณานุกรม

[1] J. Clement. "**Worldwide digital population as of January 2020**". [Online]. statista. 2020. Available : https://www.statista.com/statistics/617136/digital-population-worldwide/

[2] "**Social Media: Impact on Human Behavior and Society**". [Online]. MBAVILLA.COM. 2019. Available : https://mbavilla.com/written-ability-test-wat/social-media-impact-on-human-behavior-and-society/

[3] Graciyal, Guna. (2018). "**Freedom of Expression in Social Media: A Political Perspective**": A Political Perspective.

[4] Orehek, Edward & Human, Lauren. (2017). **"Self-Expression on Social Media: Do Tweets Present Accurate and Positive Portraits of Impulsivity, Self-Esteem, and Attachment Style?".** Personality and Social Psychology Bulletin. 43. 60-70. 10.1177/0146167216675332.

[5] "**Sprout Social**". [Online]. Available : https://sproutsocial.com/

[6] "**Awario**". [Online]. Available : https://awario.com/?utm_expid=.yr42E1bWTVOgti6CpGnqSw.0&utm_referrer=https%3A%2F%2Fawario.com%2F

[7] "**Snaplytic**". [Online]. Available : https://www.snaplytics.io/

[8] "**Google Analytics**". [Online]. Available : https://analytics.google.com/analytics/web/provision/#/provision

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use. Forbidden to modify the content, and cite the document when use.

[9] “6 สถิติเกี่ยวกับ **“โรคซึมเศร้า” ที่คุณอาจไม่เคยรู้”**. [Online]. Available : https://www.sanook.com/health/7137/

[10] สถิตย์ วงศ์สุรประกิต และ สมพร สันติประสิทธิ์กุล. **“สถานการณ์โรคซึมเศร้าในกลุ่มเสี่ยงโดยการตรวจคัดกรองในเขตอำเภอเมือง จังหวัดเชียงราย”**. Thai Journal of Nursing Council Vol. 27 No.3 July-September (2012).

[11] Scherer K. R. **“On the nature and function of emotion: a component process approach**”. Approaches to emotion, pp. 293–317, 1984

[12] **“Sentiment Analysis Explained”.** [Online]. Available : https://www.lexalytics.com/technology/sentiment-analysis

[13] **“Depression-Sentiment-Analysis-with-Twitter-Data”.** https://github.com/AshwanthRamji/Depression-Sentiment-Analysis-with-Twitter-Data

[14] คำนาย อภิปรัชญาสกุล. (2557). “**คู่มือซอฟต์แวร์การวางแผนทรัพยากรองค์กร (Enterprise Resource Planning Software Handbook)”** (พิมพ์ครั้งที่ 2). กรุงเทพฯ: โฟกัสมีเดีย แอนด์ พับลิชชิ่ง.

[15] Saed Sayad. **“Naive Bayesian”.** [Online]. Available : http://www.saedsayad.com/naive_ bayesian.htm

[16] ชินพัฒน์ แก้วชินพร. การจำแนกประเภทข้อมูลด้วยเทคนิคต้นไม้ ตัดสินใจและการจัดกลุ่ม [ปริญญาวิทยาศาสตร์บัณฑิต]. กรุงเทพฯ :สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง; 2553.

[17] เอกสิทธิ์ พัชรวงศ์ศักดา. (2557). “**การสร้างโมเดล Ensemble แบบต่าง ๆ”.** [Online]. Available : http://dataminingtrend.com/2014/data-mining-techniques/ensemble-model

[18] Jaruwit Pratancheewin. **“เรียนรู้และทำความเข้าใจเรื่อง Support Vector Machine (SVM) คืออะไร”** [Online]. Available : https://www.glurgeek.com/education/support-vector-machine/

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

[19] โอม ศรนิล. (2556). **การออกแบบและพัฒนาคลังข้อมูล.** กรุงเทพฯ: บางกอกบล๊อก.

[20] **การทำนายที่ผิดพลาดและถูกต้องของAi ด้วย Confusion Matrix, Accuracy, Precision, Recall, F1-score.** [Online]. Available : https://medium.com/@oirofate/การทำนายที่ผิดพลาดและถูกต้องของai-ด้วย-confusion-matrix-accuracy-precision-recall-f1-score-a941d4130273

[21] **"Python คืออะไร - ภาษา python ใช้ทำอะไร"**. [Online]. Available : https://www.aosoft.co.th/article/322/Pythonคืออะไรภาษา-pythonใช้ทำอะไร.html

[22] **"Node.js คืออะไร ?"**. [Online]. Available : https://www.softmelt.com/article.php?id=582

[23] **"Express.js เอ็กเพรส ดอทเจเอส คืออะไร"**. [Online]. Available : https://www.mindphp.com/คู่มือ/73คืออะไร/3874-what-is-express-js.html

[24] Taha Sufiyan. **"What is React?".** [Online]. Available : https://www.simplilearn.com/what-is-react-article

[25] Colby Fayock. **"What is Tailwind CSS and How Can I Add it to my Website or React App?".** [Online]. Available : https://www.freecodecamp.org/news/what-is-tailwind-css-and-how-can-i-add-it-to-my-website-or-react-app/

[26] **"Getting Started with Redux"**. [Online]. Available : https://redux.js.org/introduction/getting-started#getting-started-with-redux

[27] Anawat Wannapila. **"ทำความรู้จักกับ Axios — Promise based HTTP client แบบหมดเปลือก"**. [Online]. Available : https://blog.nextzy.me/axios-based-http-promise-de13980aec55

[28] Mishra V & Garg T. "**A systematic study on predicting depression using text analytics**". J. Fundam. Appl. Sci., 2018, 10(2), 293-307

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

[29] Munmun De, Choudhury, Michael Gamon, Scott Counts,and Eric Horvitz. "**Predicting Depression via Social Media"**. Microsoft Research, Redmond WA 98052, 2013

[30] Lin, L. yi .. "**Association Between Social Media Use And Depression Among U.S. Young Adults"**. Depress Anxiety 33, 323–331 (2016).

[31] 1.Choi, S. W., Schalet, B., Cook, K. F. & Cella, D.. "**Establishing a common metric for depressive symptoms: Linking the BDI-II, CES-D, and PHQ-9 to PROMIS Depression**". Psychological Assessment 26, 513–527 (2014).

[32] M. Ghiassi, J. Skinner & D. Zimbra. "**Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network".** Expert Systems with Applications Volume 40, Issue 16, 15 November 2013, Pages 6266-6282

[33] ALRashdi, R. & O'Keefe, S.. "**Deep Learning and Word Embeddings for Tweet Classification for Crisis Response"**. arXiv (2019).

[34] Husseini Orabi, A., Buddhitha, P., Husseini Orabi, M. & Inkpen, D.. "**Deep Learning for Depression Detection of Twitter Users"**. in Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic (Association for Computational Linguistics, 2018). doi:10.18653/v1/w18-0609

[35] Akshi Kumar et al., (2019). **"Anxious Depression Prediction in Real-time Social Data"**. International Conference on Advanced Engineering, Science, Management and Technology – 2019 (ICAESMT19).

[36] Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). "**A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage"**. Computers in Human Behavior, 28(2), 561–569. doi:10.1016/j.chb.2011.11.001

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

[37] Leis, A., Ronzano, F., Mayer, M., Furlong, L., & Sanz, F. (2019). "**Detecting Signs of Depression in Tweets in Spanish: Behavioral and Linguistic Analysis"**. Journal Of Medical Internet Research, 21(6), e14199. doi:10.2196/14199

[38] Nikie Jo Elauria Deocampo (2019). **"Classification of Depression on Social Media Using Text Mining".** https://github.com/niquejoe/Classification-of-Depression-on-Social-Media-Using-Text-Mining

[39] Zunaira Jamil *et al*., (2017). **"Monitoring Tweets for Depression to Detect At-risk Users"**. Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality. doi:10.18653/v1/W17-3104

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.