



รายงานความก้าวหน้า 240-401 โครงการวิศวกรรมคอมพิวเตอร์ 1 ครั้งที่ 1/2563

ระบบตรวจจับแนวโน้มเกี่ยวกับโรคซึมเศร้าและการกลั่นแกล้งในสื่อสังคมออนไลน์

Depression and bullying detection system in social media

กัลญญ์ ตันบิน

6010110616

อาจารย์ที่ปรึกษาโครงการ

.....

(รศ.ดร.พิชญา ตัญทัย)

รายงานความก้าวหน้าโครงการวิศวกรรมคอมพิวเตอร์นี้เป็นส่วนหนึ่งของการศึกษา

ตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์

26 สิงหาคม 2563

สารบัญ

สารบัญ.....	ก
รายการรูปภาพ.....	ค
รายการตาราง.....	ง
รายการคำย่อ.....	จ
บทที่ 1 บทนำ.....	1
1.1 ที่มาและความสำคัญ.....	1
1.2 วัตถุประสงค์ของโครงการ.....	2
1.3 ขอบเขตของโครงการ.....	2
1.4 แผนการดำเนินงาน.....	3
บทที่ 2 ทฤษฎีและความรู้พื้นฐาน.....	5
2.1 โรคซึมเศร้า.....	5
2.2 การกลั่นแกล้งบนโลกออนไลน์.....	5
2.3 สถิติผู้ใช้ Facebook ในไทย.....	6
2.4 ซอฟต์แวร์ที่เกี่ยวข้องกับการพัฒนาระบบ.....	6
2.4.1 ระบบปฏิบัติการในมือถือ.....	6
2.4.2 ภาษาเขียนโปรแกรมและแพลตฟอร์มที่ใช้ในการพัฒนา.....	6
บทที่ 3 รายละเอียดการดำเนินงาน.....	10
3.1 ภาพรวมของระบบ.....	10
3.2 แบบจำลองโครงสร้างฐานข้อมูล (E-R Diagram).....	11
3.3 Use case diagram.....	12
3.4 NPL diagram.....	13
3.5 ส่วนประกอบของแอปพลิเคชัน (Mock up).....	14
บทที่ 4 ความก้าวหน้าการดำเนินงาน.....	21
4.1 คลังคำศัพท์.....	21
4.1.1 รายละเอียดการทดลอง.....	21
4.1.2 อุปสรรคในการพัฒนา.....	21
4.1.3 แนวทางแก้ไขปัญหา.....	21
4.2 พัฒนาส่วนการยืนยันตัวตนด้วยอีเมล.....	21
4.2.1 รายละเอียดการพัฒนา.....	21
4.2.2 อุปสรรคในการพัฒนา.....	23

4.2.3	แนวทางแก้ไขปัญหา.....	23
4.3	พัฒนาส่วนการยืนยันตัวตนด้วยเฟซบุ๊ก	23
4.3.1	รายละเอียดการพัฒนา	23
4.3.2	อุปสรรคในการพัฒนา	23
4.3.3	แนวทางแก้ไขปัญหา.....	23
4.4	พัฒนาส่วนการติดต่อกับผู้ใช้.....	23
4.4.1	รายละเอียดการพัฒนา	23
4.4.2	อุปสรรคในการพัฒนา	24
4.4.3	แนวทางแก้ไขปัญหา.....	24
4.5	การทดสอบใช้งาน PythaiNLP+LST20	25
4.5.1	รายละเอียด.....	25
4.5.2	ผลการทดลอง	25
4.5.3	วิเคราะห์ผล.....	26
4.6	สรุปผลงานที่มีความก้าวหน้า.....	26
บทที่ 5	สรุป.....	27
5.1	สรุปผลการดำเนินงาน	27
5.2	ปัญหาและอุปสรรค.....	27
5.3	งานที่จะดำเนินการต่อไป.....	27
บรรณานุกรม	28

รายการรูปภาพ

รูปที่ 1 จำนวนผู้ป่วยโรคซึมเศร้าตั้งแต่ 2555-2560	5
รูปที่ 2 แผนภาพแสดงความก้าวหน้า.....	9

รายการตาราง

ตารางที่ 1 แผนการดำเนินงาน.....	3
---------------------------------	---

รายการคำย่อ

API

Application Programming Interface

บทที่ 1

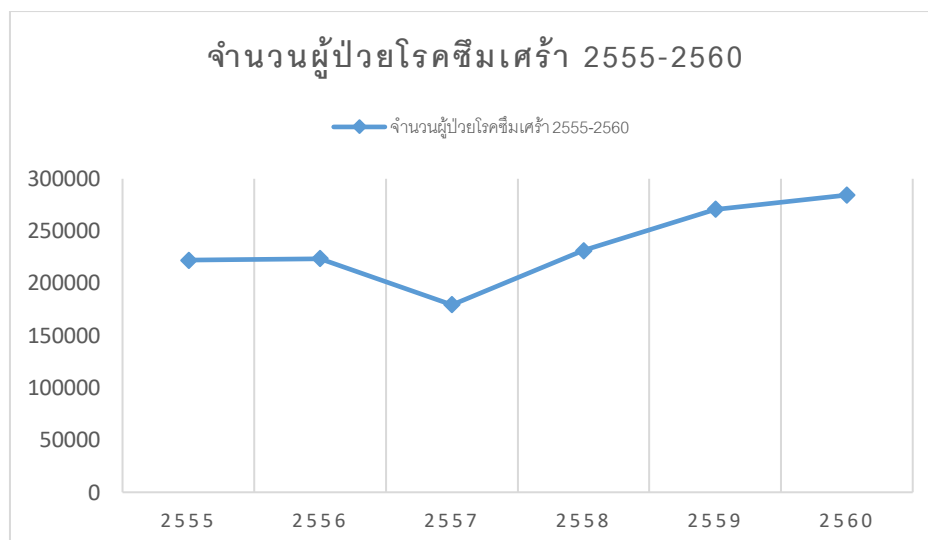
บทนำ

1.1 ที่มาและความสำคัญ

ในปัจจุบันมีการใช้งานสื่อสังคมออนไลน์เพิ่มมากขึ้น ทำให้ช่องทางการแสดงออกถึงอารมณ์และความรู้สึก มีแนวโน้มเพิ่มมากขึ้นตามการใช้งาน ทางผู้จัดทำโครงการจึงมีความสนใจที่จะใช้เทคโนโลยี เพื่อช่วยให้ผู้ใช้ตรวจจับแนวโน้มสถิติทางอารมณ์ด้วยคำพูดหรือข้อความที่เกี่ยวข้องกับโรคซึมเศร้าและการกลั่นแกล้งในสื่อสังคมออนไลน์

สำหรับโรคซึมเศร้านั้น นับเป็นปัญหาสำคัญทางสุขภาพ ถึงแม้จะสามารถรักษาให้หายขาดได้ แต่หากไม่ได้รับการรักษาอาจรุนแรงไปสู่การฆ่าตัวตายได้ โดยปัจจุบันผู้ป่วยโรคซึมเศร้าส่วนใหญ่ร้อยละ 62 อยู่ในวัยทำงานอายุ 25-59 ปี ร้อยละ 26.5 เป็นผู้ป่วยสูงอายุ 60 ปีขึ้นไป และร้อยละ 11.5 เป็นเยาวชนอายุ 15-24 โดยในแต่ละปีมีคนพยายามฆ่าตัวตายปีละ 53,000 คน เฉลี่ยชั่วโมงละ 6 คน และคนที่ฆ่าตัวตายสำเร็จมีถึง 4,000 คน [1] คิดเป็นร้อยละ 7.5 และปัญหาการกลั่นแกล้งของคนในสังคม ไม่ว่าจะเป็นเกิดขึ้นในกลุ่มเพื่อน สังคมที่ทำงาน หรือบนโลกออนไลน์เกิดขึ้นทุกวันและเป็นปัญหาที่กำลังได้รับความสนใจเป็นอย่างมาก หลายคนอาจมองว่าการกลั่นแกล้งเป็นเรื่องธรรมดา เป็นเรื่องสนุกสนานกันในกลุ่มเพื่อน แต่ปัจจุบันปัญหานี้ได้ส่งผลกระทบทางลบทั้งกับผู้ที่กระทำและผู้ที่ถูกกระทำ การกลั่นแกล้งกันไม่ใช่เรื่องล้อเล่นอีกต่อไป แต่เป็นภัยร้ายที่ใกล้ตัวและเป็นอันตรายมากกว่าที่คาดคิด [11]

โครงการนี้จึงมีวัตถุประสงค์เพื่อพัฒนาแอปพลิเคชันสำหรับเก็บข้อมูลจากผู้ใช้งาน เพื่อตรวจจับและนำมาวิเคราะห์ข้อความที่มีแนวโน้มเกี่ยวกับโรคซึมเศร้า และการกลั่นแกล้งในสื่อสังคมออนไลน์ ทำให้ตระหนักถึงแนวโน้มในการแสดงออกทางอารมณ์ของตนเอง เพื่อน และผู้ที่เกี่ยวข้องในสื่อสังคมออนไลน์ นอกจากนี้ยังช่วยแยกแยะและวิเคราะห์สถิติในภาพรวมด้วย เช่น ความถี่ในเชิงเวลาและปริมาณคำ



รูปที่ 1 จำนวนผู้ป่วยโรคซึมเศร้าตั้งแต่ 2555-2560 [1]

1.2 วัตถุประสงค์ของโครงการ

เพื่อพัฒนาแอปพลิเคชันที่สามารถตรวจจับข้อความที่เกี่ยวข้องกับโรคซึมเศร้าและการกลั่นแกล้งในสื่อสังคมออนไลน์และจัดกลุ่มที่มา ค่าทางสถิติ ความถี่ ความรุนแรงของข้อความ และแจ้งเตือน

1.3 ขอบเขตของโครงการ

1. อุปกรณ์รองรับระบบปฏิบัติการไอโอเอส เวอร์ชัน 9.0 ขึ้นไป และ ระบบปฏิบัติการแอนดรอยด์ เวอร์ชัน 4.4 ขึ้นไป
2. อุปกรณ์เคลื่อนที่มีฟังก์ชันการใช้งานพื้นฐาน โดยสามารถเชื่อมต่อสัญญาณอินเทอร์เน็ตไร้สาย
3. แอปพลิเคชันรองรับการตรวจจับข้อความเฉพาะภาษาไทยเท่านั้น
4. แอปพลิเคชันรองรับการตรวจจับข้อความเฉพาะจากสื่อออนไลน์เฟซบุ๊กเท่านั้น

1.4 แผนการดำเนินงาน

ตารางที่ 1 แผนการดำเนินงาน

ลำดับที่	รายละเอียด	ส.ค.				ก.ย.				ต.ค.				พ.ย.			
		W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4
1	ออกแบบฐานข้อมูล																
2	รวบรวมคำศัพท์																
3	พัฒนาส่วนยืนยันตัวตนและส่วนหน้า เมนูในแอปพลิเคชัน																
4	จัดหมวดหมู่คำศัพท์ที่เกี่ยวข้องโรค ซึมเศร้าและการกลั่นแกล้งในสื่อ																
5	พัฒนาฐานข้อมูล																
6	พัฒนาฟังก์ชันสำหรับประมวลผลข้อมูล																
7	พัฒนาแอปพลิเคชันส่วนแสดงผลและ ส่วนแจ้งเตือน																

ลำดับที่	รายละเอียด	ธ.ค.				ม.ค.				ก.พ.				มี.ค.			
		W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4
8	พัฒนาแอปพลิเคชันส่วนแสดงผลและส่วนแจ้งเตือน																
9	ทดสอบการทำงาน																
10	ปรับปรุงแก้ไข																
11	เตรียมนำเสนองาน																

บทที่ 2

ทฤษฎีและความรู้พื้นฐาน

2.1 โรคซึมเศร้า

โรคซึมเศร้าเป็นโรคทางอารมณ์ที่พบบ่อย โดยมีความชุกตลอดช่วงชีวิตถึง 12% พบในผู้หญิงมากกว่าผู้ชายและพบได้ในทุกช่วงอายุ โดยเฉพาะอย่างยิ่งเมื่อเกิดเหตุการณ์เลวร้ายที่ส่งผลกระทบต่อความรู้สึก เช่น การสูญเสีย ความผิดหวังหรือการหย่าร้าง การเป็นโรคนี้ไม่ได้หมายความว่าผู้ที่เป็คนนั้นจะเป็นคนอ่อนแอ ล้มเหลวหรือไม่มีความสามารถ เพราะมีหลักฐานทางการแพทย์ยืนยันว่าโรคซึมเศร้ามีสาเหตุส่วนหนึ่งจากการทำงานของระบบสมองที่ผิดปกติ ในปัจจุบันโรคซึมเศร้าสามารถรักษาได้ด้วยการใช้ยาและการรักษาทางจิตใจ หากไม่ได้รับการรักษาอาจเกิดผลกระทบต่อกรดำเนินชีวิตประจำวัน การทำงานและความสัมพันธ์กับคนรอบข้าง นำไปสู่ภาวะซึมเศร้าที่รุนแรงมากขึ้น เช่น มีอาการหลงผิด หูแว่ว มีความคิดทำร้ายตนเองหรือฆ่าตัวตาย [2]

2.2 การกลั่นแกล้งบนโลกออนไลน์

สาเหตุของการเกิด cyberbullying เกือบทั้งหมด มักเริ่มก่อตัวจากความขัดแย้ง ความเห็นต่าง หรือมีกรณีพิพาทระหว่าง 2 คน ลูกหลานจนเป็นชนวนของการกลั่นแกล้งกันในโลกออนไลน์ นอกจากนี้ผู้คนยังเข้าใจว่าโซเชียลมีเดียคือพื้นที่ส่วนตัว สามารถใช้ระบายความรู้สึกได้ ถ้อยคำที่ใช้โพสต์จึงออกแนวจนแรง โพสต์เสียดสี หรือสร้างความเสียหายให้อีกฝ่ายและแสดงอารมณ์แง่ลบออกมาได้อย่างเต็มที่โดยไม่ต้องเกรงใจ [3]

ประเภทของ Cyberbullying

1. การส่งข้อความนินทาผู้อื่นให้เสียหาย
2. การไล่บางคนออกจากกลุ่มออนไลน์ เช่น กลุ่มไลน์หรือเฟซบุ๊กกรุ๊ป
3. การแอบเข้าไปในใช้เฟซบุ๊กของคนอื่นและโพสต์ข้อความให้เจ้าของบัญชีเฟซบุ๊กเสียหาย และทำให้คนรอบตัวเข้าใจผิด
4. การว่ากล่าว ต่ำทอ ด้วยถ้อยคำหยาบคาย ตอกย้ำปมด้อยทำให้เสียความมั่นใจ
5. ส่งข้อความ รูป วิดีโอ หรืออะไรก็ตามที่ทำให้คนอื่นอับอายบนอินเทอร์เน็ต รวมถึงการข่มขู่
6. หยอกล้อ ยั่วโมโหจนอีกฝ่ายเผยความลับที่น่าอายของตัวเองบนโลกออนไลน์
7. เห็นการกลั่นแกล้งบนโลกออนไลน์แล้วเข้าไปร่วมด้วย

2.3 สถิติผู้ใช้ Facebook ในไทย

ข้อมูลจาก Facebook Audience Insights ระบุว่าผู้ใช้งานในช่วงอายุ 18-64 ปี ประเทศไทยมีผู้ใช้ Facebook ทั้งหมด 45 ล้านบัญชี โดยสามารถเป็นออกเป็นผู้ใช้งานเพศชายจำนวน 23 ล้านบัญชี และผู้ใช้งานเพศหญิงจำนวน 22 ล้านบัญชี ซึ่งถือว่ามียอดตราส่วนที่ใกล้เคียงกันมาก โดยปริมาณผู้ใช้ Facebook ในไทยแบ่งตามช่วงอายุได้เป็น [13]

- ผู้ใช้งานช่วงอายุ 18-24 ปี มีทั้งหมด 16.0 ล้านบัญชี แบ่งเป็นเพศชายมี 8.10 ล้านบัญชี และเพศหญิงมี 7.90 ล้านบัญชี และเพศหญิงมี 7.20 ล้านบัญชี
- ผู้ใช้งานช่วงอายุ 25-34 ปี มีทั้งหมด 15.0 ล้านบัญชี แบ่งเป็นเพศชายมี 7.80 ล้านบัญชี
- ผู้ใช้งานช่วงอายุ 35-44 ปี มีทั้งหมด 8.1 ล้านบัญชี แบ่งเป็นเพศชายมี 4.10 ล้านบัญชี และเพศหญิงมี 4.00 ล้านบัญชี
- ผู้ใช้งานช่วงอายุ 45-54 ปี มีทั้งหมด 4.3 ล้านบัญชี แบ่งเป็นเพศชายมี 2.10 ล้านบัญชี และเพศหญิงมี 2.20 ล้านบัญชี
- ผู้ใช้งานช่วงอายุ 55-64 ปี มีทั้งหมด 1.8 ล้านบัญชี แบ่งเป็นเพศชายมี 0.89 ล้านบัญชี และเพศหญิงมี 0.91 ล้านบัญชี

2.4 ซอฟต์แวร์ที่เกี่ยวข้องกับการพัฒนาระบบ

2.4.1 ระบบปฏิบัติการในมือถือ

1. IOS คือ ระบบปฏิบัติการของอุปกรณ์พกพาของ Apple เดิมมีชื่อว่า iPhone OS ทำขึ้นเพื่อสมาร์ทโฟน แท็บเล็ต และอุปกรณ์อื่นๆ โดย iOS เป็นระบบปฏิบัติการที่ควบคุมแอปพลิเคชัน การทำงานด้านต่างๆ สามารถใช้งานได้เฉพาะอุปกรณ์ของ apple เท่านั้น และแอปพลิเคชันที่พัฒนาขึ้นรองรับการใช้งานบนระบบปฏิบัติการ iOS [4]
2. ANDROID คือ ระบบปฏิบัติการสำหรับอุปกรณ์พกพา เช่น สมาร์ทโฟน แท็บเล็ต คอมพิวเตอร์เน็ตบุ๊ก ทำงานบนลินุกซ์เคอร์เนล และแอปพลิเคชันที่พัฒนาขึ้นรองรับการทำงานบนระบบปฏิบัติการ android [5]

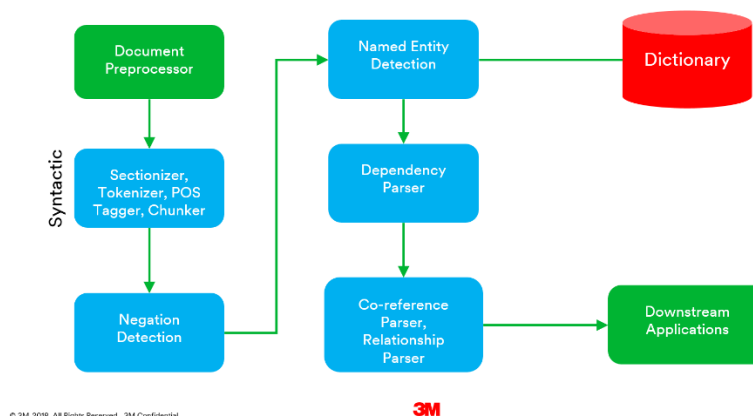
2.4.2 ภาษาเขียนโปรแกรมและแพลตฟอร์มที่ใช้ในการพัฒนา

1. Jupyter คือ เครื่องมือหนึ่งที่นิยมมากในงานด้าน Data Science ซึ่งต้องทำงานที่เกี่ยวข้องกับการจัดการข้อมูลเป็นจำนวนมาก ซึ่ง Jupyter Notebook ได้ออกแบบมาตรงตามจุดประสงค์การใช้งานไม่ว่าจะเป็น การเรียกใช้งาน library พร้อมทั้งเขียน code และดูผลได้ Jupyter Notebook นั้นถูกออกแบบมาให้ทำงานและอ่านได้ง่าย รวมไปถึงยังสามารถที่จะพิมพ์ตัวหนังสือภาษาไทยลงไปได้ [6]

2. Python คือ ภาษาหนึ่งที่ใช้ในการเขียนโปรแกรมคอมพิวเตอร์ คล้ายกับภาษา c, c++, Java หรือ Javascript ซึ่งข้อดีที่ทำให้เหมาะกับการเขียนคำสั่งให้คอมพิวเตอร์มีการเรียนรู้ตามที่เรต้องการ Python จึงเป็นภาษาที่เหมาะสมในการนำมาใช้เนื่องจาก [6]
 - Python เป็นภาษาที่ถูกพัฒนาขึ้นมา เพื่อต้องการให้อ่านง่าย คำสั่งที่เขียนขึ้นมาก็ไม่ซับซ้อนเท่ากับภาษาอื่นๆ
 - Python ถูกออกแบบให้สามารถทำงานได้ในทุกแพลตฟอร์ม ไม่ว่าจะเป็นใน Windows, Linux, macOS เป็นต้น
 - Python เป็น open source ที่เราสามารถนำมาพัฒนาเป็นโปรแกรม หรือคำสั่งของเราเองได้ ดังนั้นจึงมีคนเข้ามาพัฒนาโปรแกรมต่างๆ โดยใช้ภาษา python เป็นพื้นฐาน ทำให้มี Library ที่ถูกสร้างและพัฒนาขึ้นสำหรับการทำ Machine Learning และการทำ Data Visualization มากมาย โดยเราสามารถที่จะเรียกใช้ Library นั้นๆได้ฟรี ไม่ว่าจะเป็น numpy, pandas, sklearn, matplotlib และ pythainlp
3. Library: PythaiNLP เป็นไลบรารีภาษาไพทอนเพื่อการประมวลผลภาษาธรรมชาติ โดยเน้นการสนับสนุนภาษาไทย [7] โดยมีความสามารถดังนี้
 - ตัดคำภาษาไทย
 - ถอดเสียงภาษาไทยเป็นอักษรละตินและสัทอักษร
 - ระบุชนิดคำภาษาไทย
 - อ่านตัวเลขเป็นข้อความภาษาไทย
 - เรียงลำดับคำตามพจนานุกรมไทย
 - แก้ไขปัญหาการพิมพ์สลับเปลี่ยนภาษา
 - ตรวจสอบคำสะกดผิดในภาษาไทย
 - soundex ภาษาไทย
 - Thai WordNet wrapper
4. API: Graph api เป็นวิธีขั้นต้นในการรับและส่งออกข้อมูลไปยังและออกจากแพลตฟอร์ม Facebook API นี้เป็น API ที่อยู่บน HTTP ซึ่งเราสามารถสืบค้นข้อมูล โพสต์เรื่องราวใหม่ๆ จัดการโฆษณา อัปโหลดรูปภาพ และทำงานอื่นๆ API กราฟตั้งชื่อตามแนวคิดของ “กราฟสังคม” ซึ่งเป็นสิ่งที่ใช้แทนข้อมูลใน Facebook โดย API นี้ประกอบด้วย:
 - โหนด โดยพื้นฐานแล้วก็คืออ็อบเจกต์แต่ละตัว เช่น ผู้ใช้ รูปภาพ เพจ หรือความคิดเห็น
 - จุดเชื่อมโยง จุดเชื่อมโยงระหว่างคอลเลกชันของอ็อบเจกต์และอ็อบเจกต์เดี่ยว เช่น รูปภาพในเพจหรือความคิดเห็นในรูปภาพ
 - ฟิลด์ - ข้อมูลเกี่ยวกับอ็อบเจกต์ เช่น วันเกิดของผู้ใช้ หรือชื่อของเพจ

5. Natural Language Processing ทำหน้าที่เป็นอีกหนึ่งกลไกสำคัญ ที่ช่วยให้ AI สามารถประมวลผลจากคำสั่งต่างๆ ได้อย่างมีประสิทธิภาพ แม่นยำ รวมถึงแสดงผลการวิเคราะห์ต่างๆ ออกมาเป็นภาษามนุษย์ได้อย่างเป็นธรรมชาติ [8] โดยส่วนประกอบหลักในการทำ NLP แสดงดังรูปที่ 3

Typical Components of an NLP Application



รูปที่ 2 ส่วนประกอบหลักๆของการทำ NLP [12]

6. React Native คือ เครื่องมือที่สามารถพัฒนา mobile application ทั้ง iOS และ Android หรือก็คือ Cross Platform Technology โดยใช้ JavaScript เป็นหลักในการพัฒนา [9]
7. Firebase คือ Platform ที่รวบรวมเครื่องมือต่างๆ สำหรับจัดการในส่วนของ Backend หรือฝั่ง Server ซึ่งทำให้สามารถพัฒนา Mobile Application ได้อย่างมีประสิทธิภาพ และช่วยเวลาและค่าใช้จ่ายของการทำฝั่ง Server โดยมีทั้งเครื่องมือที่ฟรี และเครื่องมือที่มีค่าใช้จ่าย โดยเครื่องมือที่นำมาใช้ได้แก่
- Cloud Firestore ซึ่งนำมาจัดการในส่วนของ database โดย cloud firestore เป็น realtime database รุ่นใหม่ที่มาพร้อมการค้นหาและการปรับขนาดอัตโนมัติ ที่มีประสิทธิภาพมากขึ้น ปรับปรุงวิธีการเก็บข้อมูลใหม่เป็น collection
 - Authentication ซึ่งนำมาใช้จัดการการเข้าสู่ระบบของแอปพลิเคชัน โดยสามารถตรวจสอบได้หลายวิธี เช่น Email/Password, เบอร์โทรศัพท์, บัญชี Google, Facebook, Twitter, Github เป็นต้น และมีฐานข้อมูลเป็นของตัวเองไม่ต้องสร้างใหม่หรือออกแบบวิธีการเก็บ ซึ่งสามารถดูได้ว่าสมัครด้วยวิธีไหน สมัครเมื่อไหร่ และเข้าใช้ระบบครั้งสุดท้ายเมื่อไหร่

8. คลังข้อมูลคำศัพท์ LST20 Corpus เป็นคลังข้อมูลภาษาสำหรับการประมวลผลภาษาไทย ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ คลังข้อมูลนี้กำกับด้วยข้อมูลทางภาษา 5 ระดับ ได้แก่ ขอบเขตของคำ (word boundaries), ชนิดของคำ (part of speech), ขอบเขตและชนิดของชื่อเฉพาะ (named entities), ขอบเขตของประโยคย่อย (clause boundaries), และขอบเขตประโยคใหญ่ (sentence boundaries) คลังข้อมูลนี้ประกอบด้วย: 3,164,002 คำ, 288,020 ชื่อเฉพาะ, 248,181 ประโยคย่อย, และ 74,180 ประโยคใหญ่ โดยมีชนิดของคำทั้งสิ้นเพียง 16 แท็ก นอกจากนี้คลังข้อมูลนี้สร้างขึ้นจากการรวบรวมข่าว 3,745 ชิ้น ซึ่งกำกับชนิดของข่าวเอาไว้ 15 ชนิด ทำให้คลังข้อมูลนี้สามารถนำไปใช้ประโยชน์ในการพัฒนาระบบการประมวลผลภาษาไทยที่มีโครงสร้างซับซ้อนได้ [14]

11 August 2020

A Brief Specification of LST20 Corpus (NECTEC)

4

Executive Summary

NECTEC
a member of NSTDA

- LST20 Corpus
 - Dataset for training fundamental Thai language processing tasks
 - Featured linguistic information
 - **Word boundaries** for word segmentation
 - **Named entities** for named entity recognition
 - **Clause boundaries** for clause segmentation
 - **Sentence boundaries** for sentence segmentation
 - **News genres** for document classification
 - CoNLL-2003 format: tab-separated columns

Words	3,162,864
Named entities	288,020
Clauses	248,962
Sentences	74,180
Distinct words	46,692
Genres	15
News articles	3,745

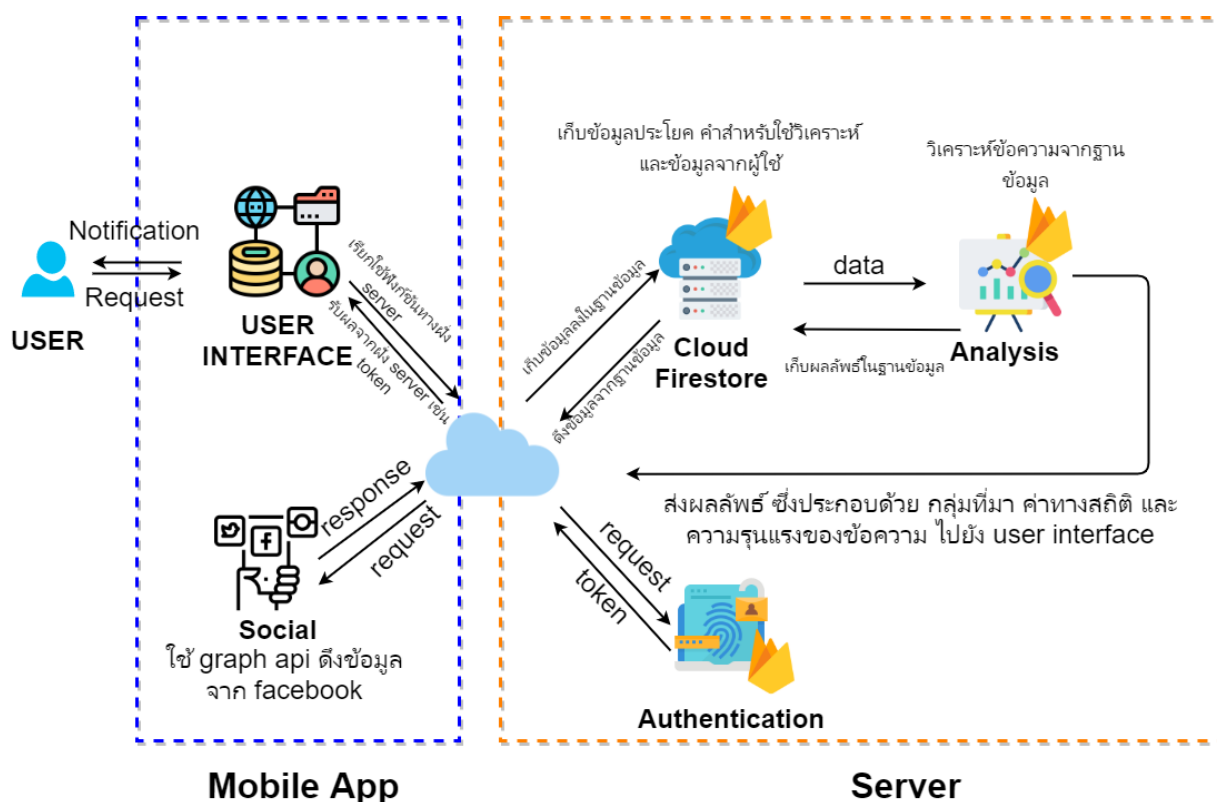
Available at
<https://aiforthai.in.th>

รูปที่ 3 สรุปสิ่งที่มีในคลังข้อมูล LST20 Corpus

บทที่ 3

รายละเอียดการดำเนินงาน

3.1 ภาพรวมของระบบ



รูปที่ 4 ภาพรวมการทำงานของระบบ

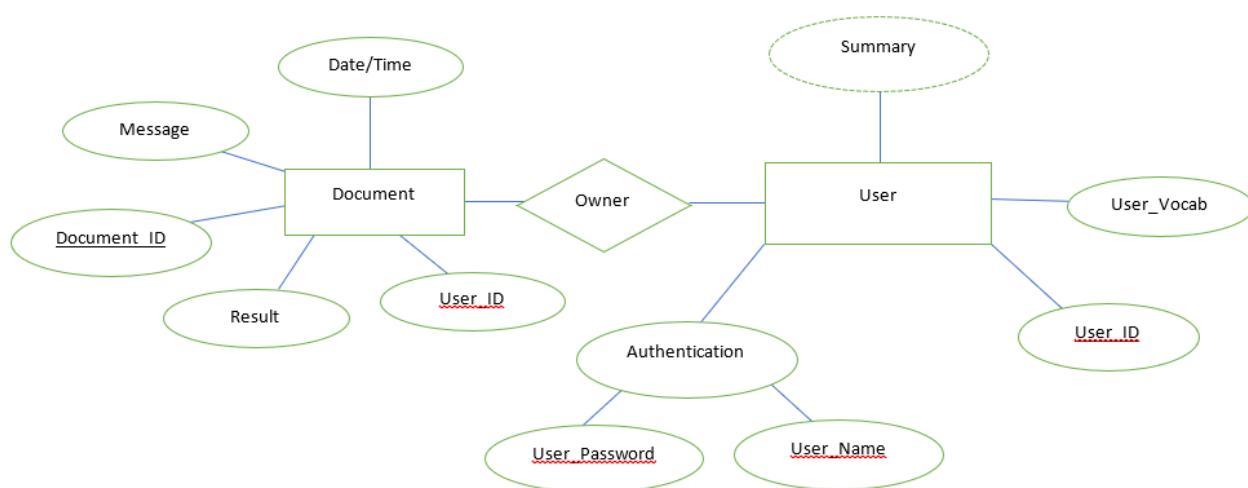
จากรูปที่ 4 ภาพรวมของระบบตรวจจับแนวโน้มเกี่ยวกับโรคซึมเศร้าและการกลั่นแกล้งในสื่อสังคมออนไลน์ประกอบด้วย

- Mobile application โดย React-native
- การยืนยันตัวตนผ่าน Firebase Authentication
- ส่วนประมวลผล
- จัดเก็บข้อมูลโดย Cloud Firestore

สำหรับระบบตรวจจับแนวโน้มเกี่ยวกับโรคซึมเศร้าและการการกลั่นแกล้งในสื่อสังคมออนไลน์นั้นมีจุดประสงค์เพื่อที่ผู้ใช้สามารถตรวจสอบข้อความบนสื่อออนไลน์(Facebook) เพื่อระบุถึงข้อความที่มีถ้อยคำรุนแรง หรือข้อความที่มีแนวโน้มเกี่ยวกับโรคซึมเศร้า และสามารถแจ้งเตือนเมื่อผู้ใช้งานใช้ข้อความที่มีความรุนแรงหรือมีแนวโน้มเกี่ยวกับโรคซึมเศร้า โดยแบ่งการแจ้งเตือนเป็น 3 ระดับ

- ระดับที่ 1 เมื่อข้อความมีระดับความรุนแรงเกินกว่า 50% ของข้อความนั้นๆ
- ระดับที่ 2 เมื่อในหนึ่งสัปดาห์มีระดับความรุนแรงรวมของทุกข้อความมากกว่า 50% ในสัปดาห์นั้นๆ
- ระดับที่ 3 เมื่อในหนึ่งเดือนมีระดับความรุนแรงรวมของทุกข้อความมากกว่า 50% ในเดือนนั้นๆ

3.2 แบบจำลองโครงสร้างฐานข้อมูล (E-R Diagram)



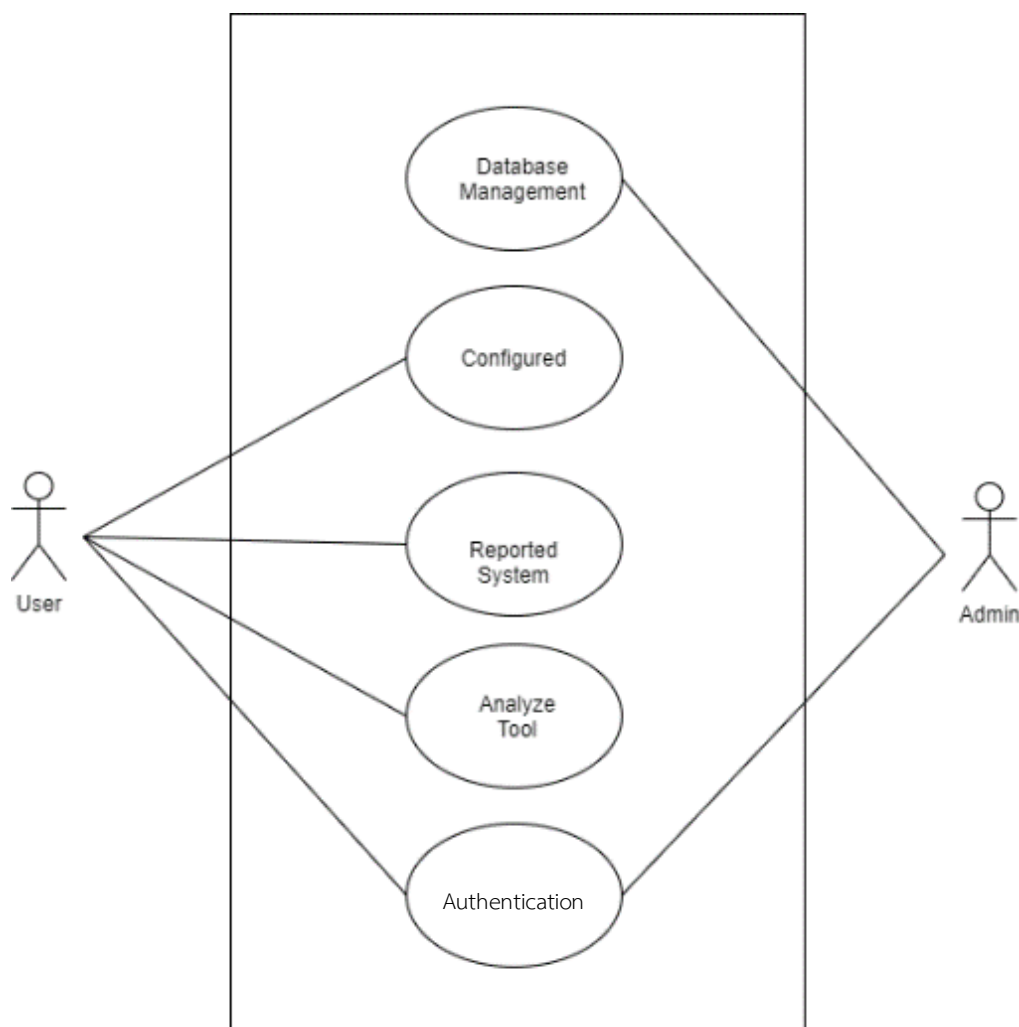
รูปที่ 5 แบบจำลองโครงสร้างฐานข้อมูล

จากรูปที่ 5 ภาพรวมโครงสร้างฐานข้อมูล ซึ่งประกอบด้วย 2 ส่วนหลัก ได้แก่

- ส่วนแรก User ที่จัดเก็บ User_ID สำหรับแยกแยะแต่ละผู้ใช้, จัดเก็บข้อมูลสำหรับยืนยันตัวตน ประกอบด้วย User_Email และ User_Password, จัดเก็บ User_Vocab ให้ผู้ใช้สามารถแก้ไขคำ เพิ่มคำ หรือให้คะแนนคำด้วยตนเอง, Summary ใช้แสดงภาพรวมของการวิเคราะห์ข้อมูลข้อความของผู้ใช้
- ส่วนที่สอง Document ซึ่งจะจัดเก็บ User_ID สำหรับระบุเจ้าของ Document นั้นๆ, Document_ID สำหรับระบุเอกสารนั้นๆ, Message ซึ่งเป็นข้อความที่ดึงมาจากผู้ใช้, Date/Time เพื่อระบุวันที่ของข้อความที่ผู้ใช้โพสต์, เก็บ Result ของการประมวลผลข้อความที่ได้รับมา

โดยความสัมพันธ์ระหว่าง User และ Document คือ User จะสามารถเป็นเจ้าของ Document ได้มากกว่า 1 Document

3.3 Use case diagram

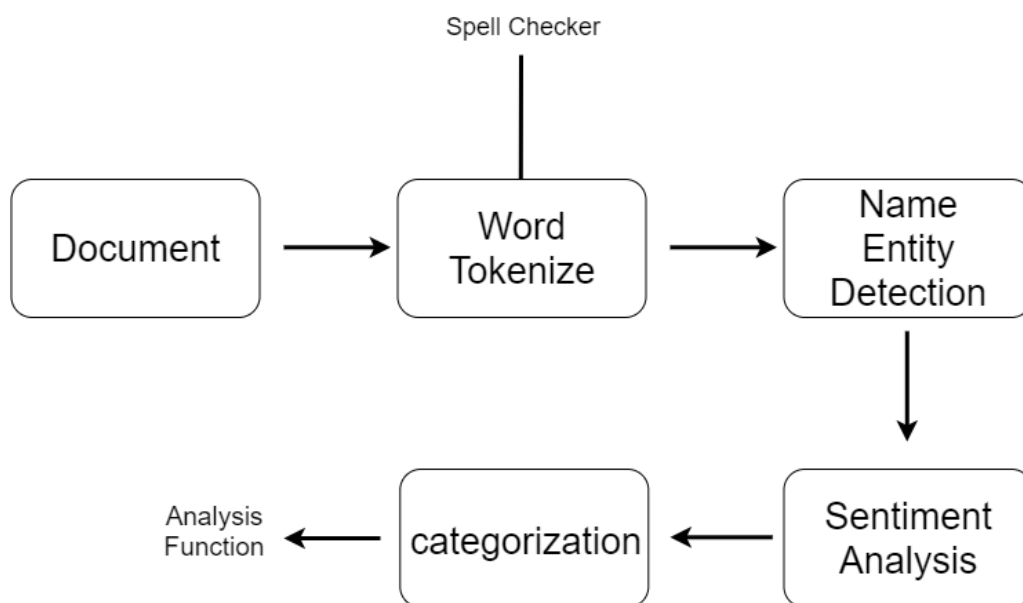


รูปที่ 6 การทำงานผู้ใช้งานระบบ

จากรูปที่ 6 ภาพรวมที่ผู้ใช้สามารถกระทำต่อระบบได้

- ส่วนแรก User สามารถใช้งานในส่วนของ Configured เพื่อตั้งค่าการใช้งานและจัดการคำศัพท์ได้
- ส่วนที่สอง User สามารถใช้งานในส่วนของ Reported System เพื่อรับการแจ้งเตือนได้
- ส่วนที่สาม User สามารถใช้งานในส่วนของ Analyze tool เพื่อวิเคราะห์ข้อความได้
- ส่วนที่สี่ User สามารถใช้งานในส่วนของ Authentication เพื่อยืนยันตัวตน
- ส่วนที่ห้า Admin สามารถใช้งานส่วน Authentication เพื่อจัดการข้อมูลได้
- ส่วนที่หก Admin สามารถใช้งานส่วน Database Management เพื่อจัดการข้อมูลได้

3.4 NPL diagram

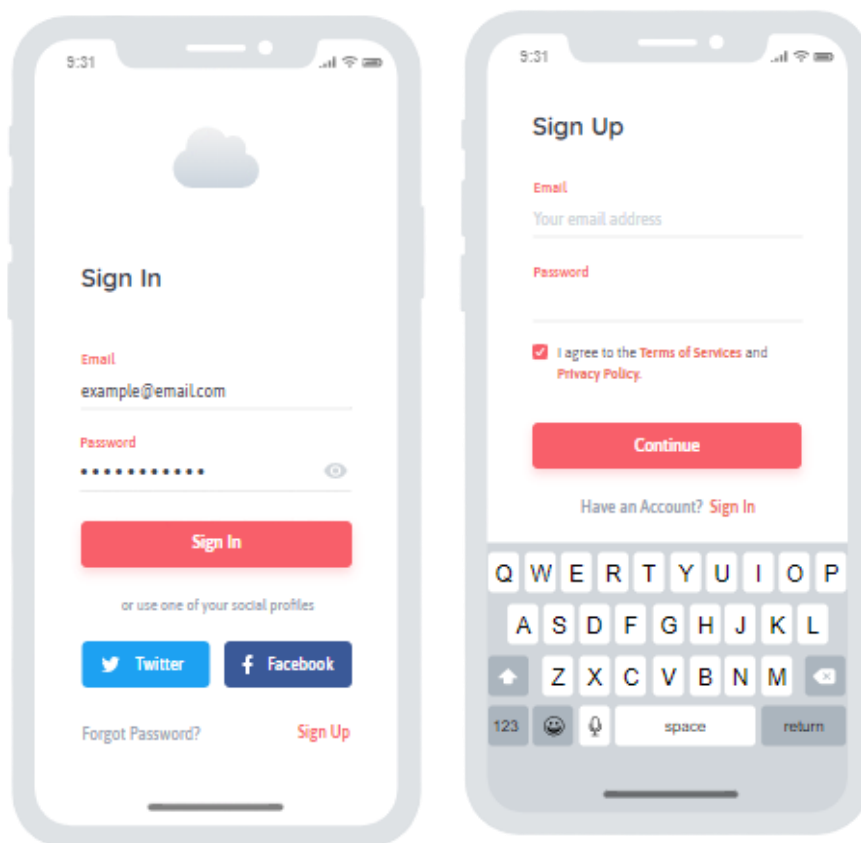


รูปที่ 7 กระบวนการ NLP

จากรูปที่ 7 ภาพรวมกระบวนการทำงานของ NLP

- ส่วนแรก เริ่มจากการนำข้อความจากใน Document เข้ากระบวนการ Word Tokenize(การตัดคำ) โดยมี spell checker ช่วยในกรณีที่มีการสะกดคำผิด
- เข้าสู่กระบวนการ Name Entity Detection สำหรับแยกแยะชนิดของคำ
- นำไปวิเคราะห์ความรู้สึกของคำในขั้นตอน Sentiment analysis
- เข้าสู่กระบวนการ Categorization(จำแนกประเภท) และส่งต่อไปยังส่วนวิเคราะห์ต่อไป

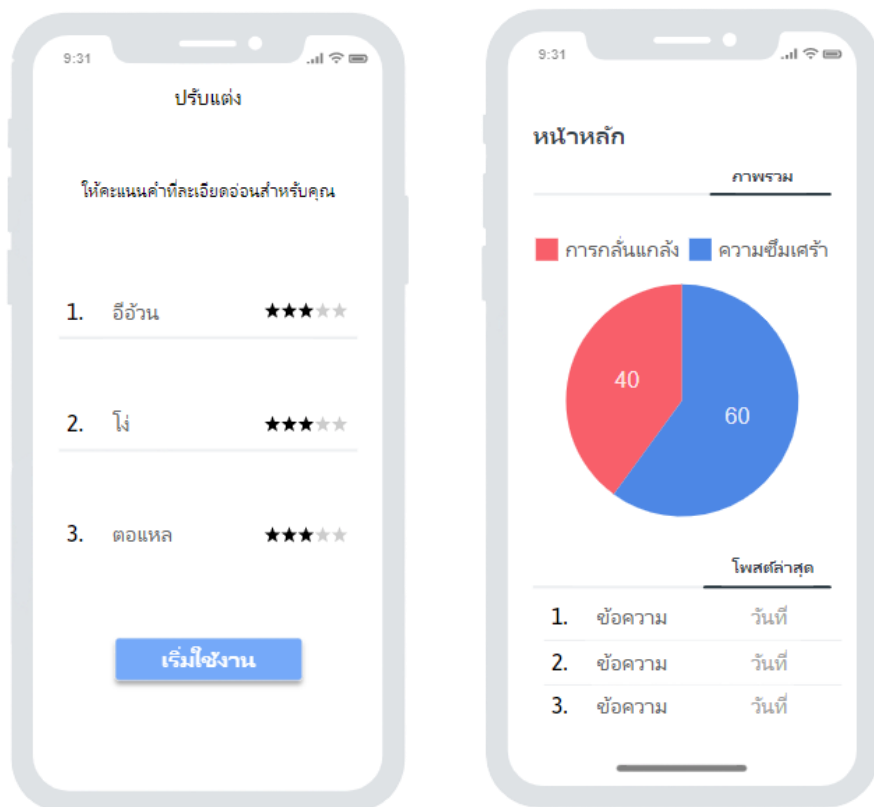
3.5 ส่วนประกอบของแอปพลิเคชัน (Mock up)



รูปที่ 8 หน้าเข้าสู่ระบบของแอปพลิเคชัน รูปที่ 9 หน้าสมัครเข้าแอปพลิเคชัน

จากรูปที่ 8 แสดงหน้าเข้าสู่ระบบของแอปพลิเคชันซึ่งผู้ใช้ที่มีบัญชีอยู่แล้วจะสามารถเข้าสู่ระบบผ่านหน้านี้ได้

จากรูปที่ 9 แสดงหน้าสมัครเข้าแอปพลิเคชันซึ่งจะให้ผู้ใช้ที่ยังไม่มีบัญชีสามารถสร้างบัญชีเพื่อเข้าสู่ระบบจากหน้านี้ได้

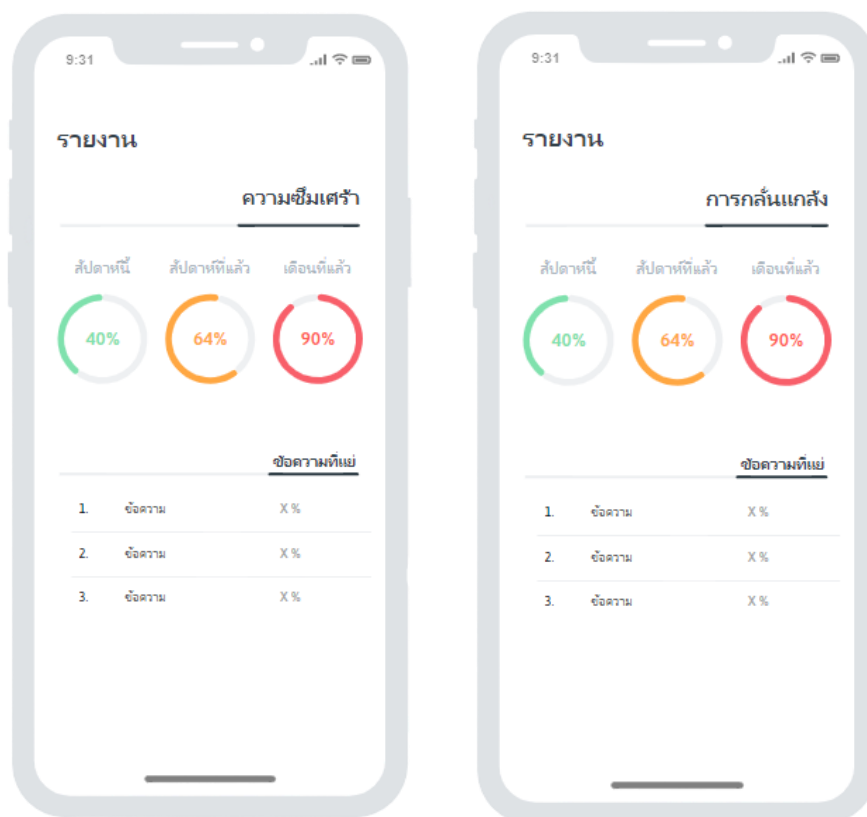


รูปที่ 10 หน้าปรับแต่งก่อนเข้าใช้งาน

รูปที่ 11 ภาพรวมการกลั่นแกล้งและความซึมเศร้า

จากรูปที่ 10 แสดงการปรับแต่งก่อนเข้าใช้งานซึ่งจะให้ผู้ใช้งานให้คะแนนคำที่มีความละเอียดอ่อนสำหรับตนเอง เพื่อนำไปวิเคราะห์ระดับการให้คะแนนคำ

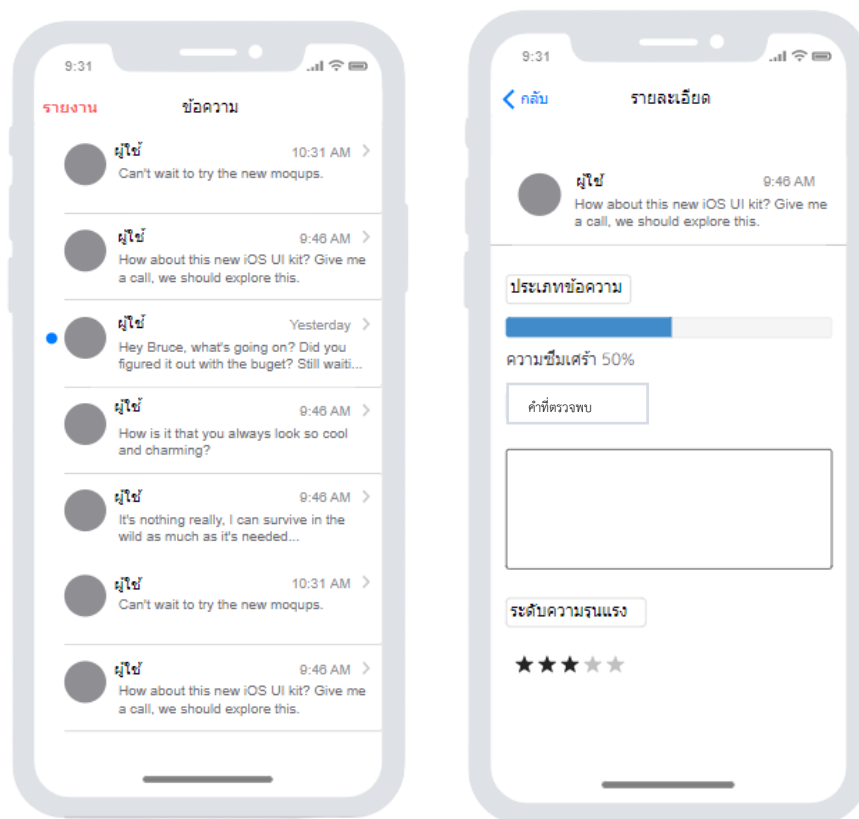
จากรูปที่ 11 แสดงภาพรวมอัตราส่วนระหว่างการกลั่นแกล้งและความซึมเศร้า รวมถึงโพสต์ล่าสุดของผู้ใช้ 3 อันดับ



รูปที่ 12 รายละเอียดรายงานของความซึมเศร้า รูปที่ 13 รายละเอียดรายงานของการกลั่นแกล้ง

จากรูปที่ 12 แสดงรายละเอียดรายงานของความซึมเศร้าซึ่งจะแบ่งเป็น 3 ส่วน ประกอบด้วยรายงานของสัปดาห์นี้ สัปดาห์ที่แล้ว และเดือนที่แล้ว รวมถึงแสดงอันดับข้อความ 3 อันดับ

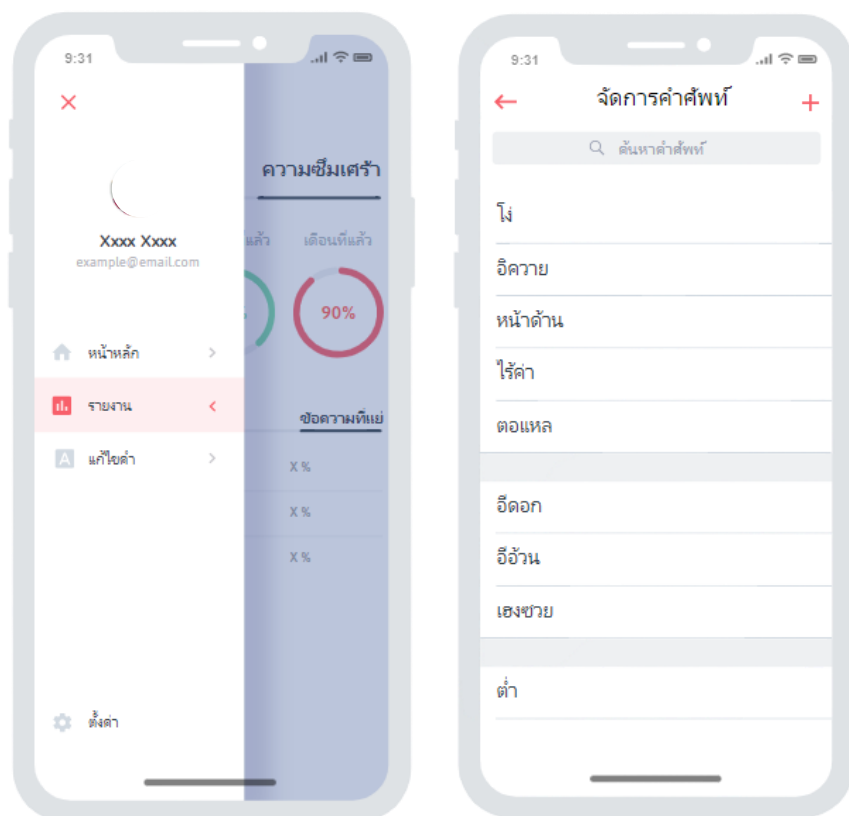
จากรูปที่ 13 แสดงรายละเอียดรายงานของการกลั่นแกล้งซึ่งจะแบ่งเป็น 3 ส่วน ประกอบด้วยรายงานของสัปดาห์นี้ สัปดาห์ที่แล้ว และเดือนที่แล้ว รวมถึงแสดงอันดับข้อความ 3 อันดับ



รูปที่ 14 ข้อความที่ผ่านการประมวลผล รูปที่ 15 รายละเอียดของการประมวลผลข้อความ

จากรูปที่ 14 แสดงข้อความที่ผ่านการประมวลผลมาทั้งหมดของผู้ใช้ โดยแต่ละข้อความสามารถเข้าป
 ติการประมวลผลข้อความอย่างรายละเอียดได้

จากรูปที่ 15 แสดงรายละเอียดของการประมวลผลข้อความแต่ละครั้งซึ่งจะบอกถึงประเภทของ
 ข้อความ ว่าเป็นข้อความประเภทความขี้เมรั้า หรือประเภทการกลั่นแกล้ง พร้อมทั้งแสดงคำที่ตรวจ
 พบในข้อความ และบอกถึงระดับความรุนแรงรวมของข้อความนั้น

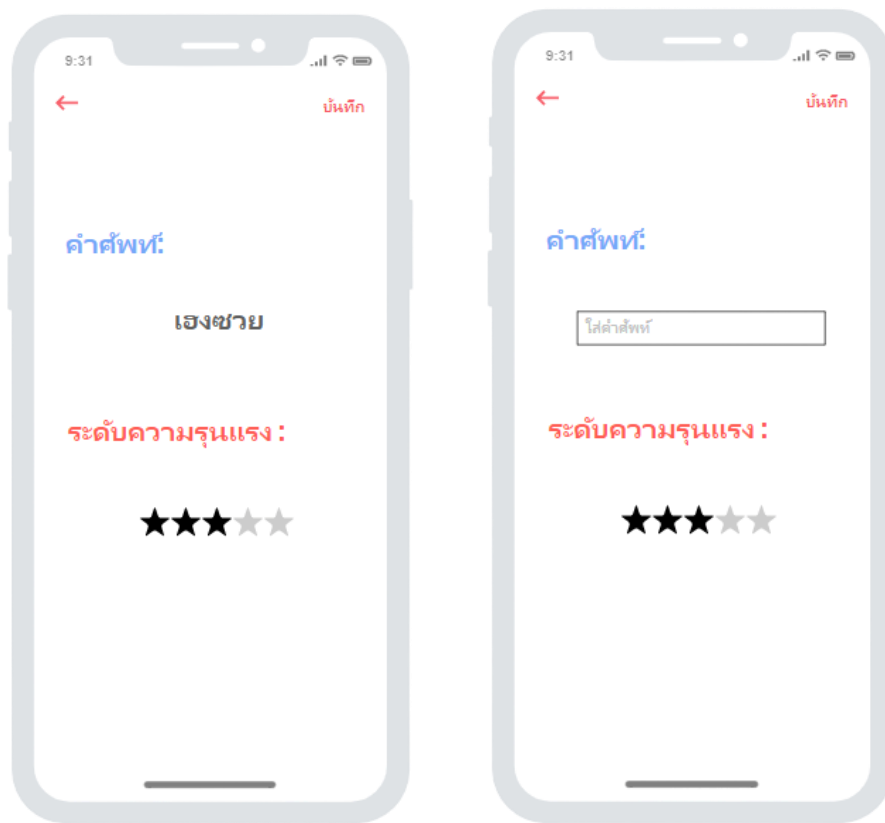


รูปที่ 16 หน้าเมนู

รูปที่ 17 หน้าจัดการคำสั่งพท์

จากรูปที่ 16 แสดงหน้าต่างเมนูซึ่งประกอบด้วย หน้าหลัก หน้ารายงาน หน้าแก้ไขค่า และหน้าตั้งค่า

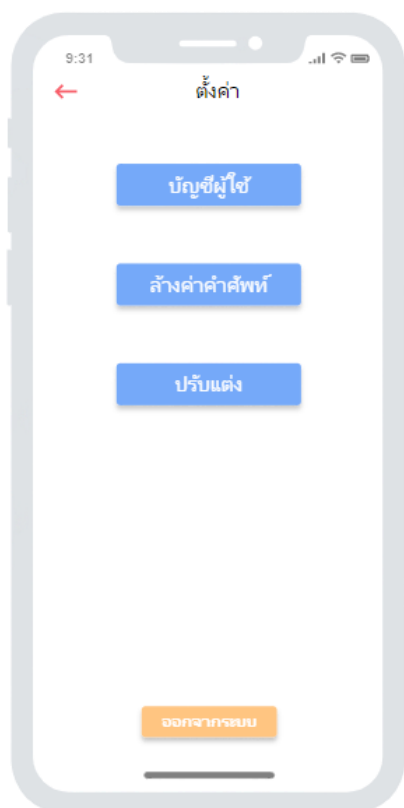
จากรูปที่ 17 แสดงหน้าการจัดการค่าจะให้ผู้ใช้งานสามารถดูคำสั่งพท์และเข้าไปแก้ไขระดับความรุนแรงได้



รูปที่ 18 หน้าแก้ไขความรุนแรงคำศัพท์ รูปที่ 19 หน้าเพิ่มคำศัพท์และให้ระดับความรุนแรง

จากรูปที่ 18 แสดงหน้าการให้คะแนนคำซึ่งผู้ใช้จะสามารถเปลี่ยนแปลงระดับความรุนแรงของคำได้

จากรูปที่ 19 แสดงหน้าการเพิ่มและให้คะแนนคำซึ่งผู้ใช้จะสามารถเพิ่มคำและกำหนดระดับความรุนแรงของคำนั้นๆได้ด้วยตนเอง



รูปที่ 20 หน้าตั้งค่า

จากรูปที่ 20 แสดงหน้าการตั้งค่าแอปพลิเคชันซึ่งประกอบด้วย หน้าจัดการบัญชีผู้ใช้ การล้างค่าคำศัพท์ให้กลับไปค่าตั้งต้น หน้าปรับแต่ง และปุ่มออกจากระบบ

บทที่ 4

ความก้าวหน้าการดำเนินงาน

4.1 คลังคำศัพท์

4.1.1 รายละเอียดการทดลอง

คลังข้อมูลคำศัพท์ภาษาไทยที่เกี่ยวข้องกับแนวโน้มเกี่ยวกับโรคซึมเศร้าและการกลั่นแกล้งในสื่อสังคมออนไลน์

4.1.2 อุปสรรคในการพัฒนา

คลังข้อมูลคำศัพท์ภาษาไทยที่เกี่ยวข้องกับแนวโน้มเกี่ยวกับโรคซึมเศร้าและการกลั่นแกล้งในสื่อสังคมออนไลน์

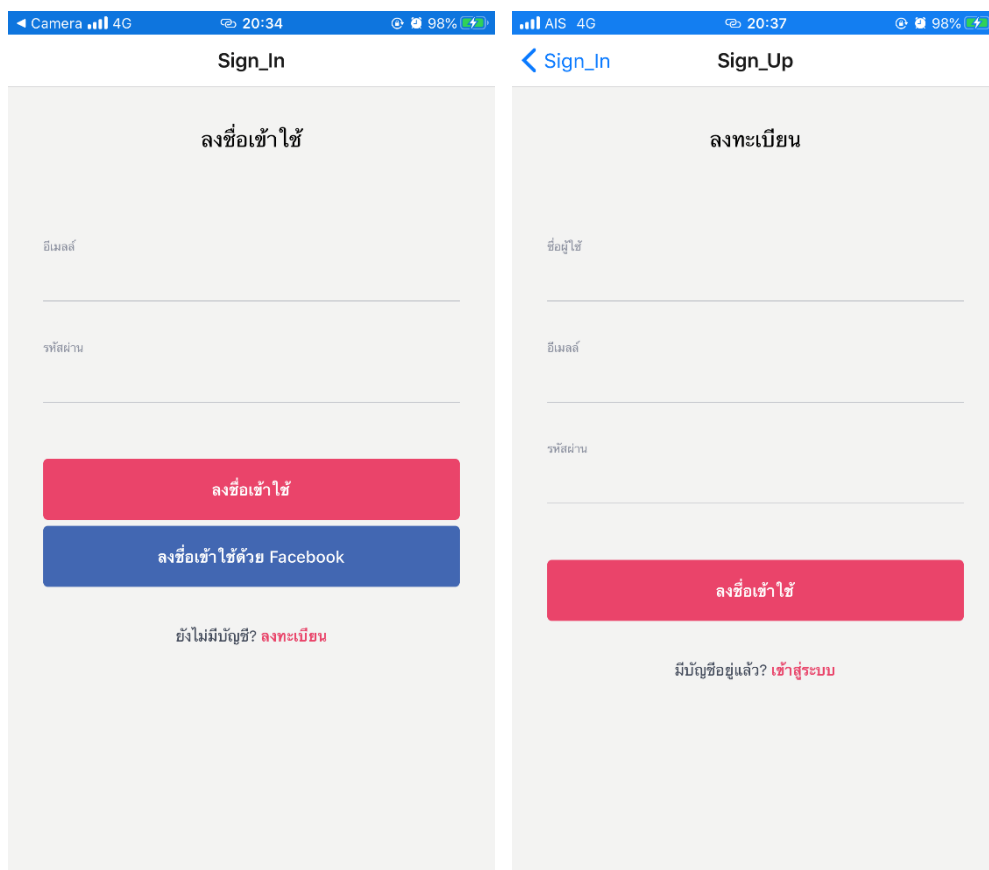
4.1.3 แนวทางแก้ไขปัญหา

ใช้คลังข้อมูลคำศัพท์ LST20 ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติซึ่งคลังข้อมูลนี้กำกับด้วยข้อมูลทางภาษา 5 ระดับ ได้แก่ ขอบเขตของคำ (word boundaries), ชนิดของคำ (part of speech), ขอบเขตและชนิดของชื่อเฉพาะ (named entities), ขอบเขตของประโยคย่อย (clause boundaries), และขอบเขตประโยคใหญ่ (sentence boundaries)

4.2 พัฒนาล่วงการยืนยันตัวตนด้วยอีเมลล์

4.2.1 รายละเอียดการพัฒนา

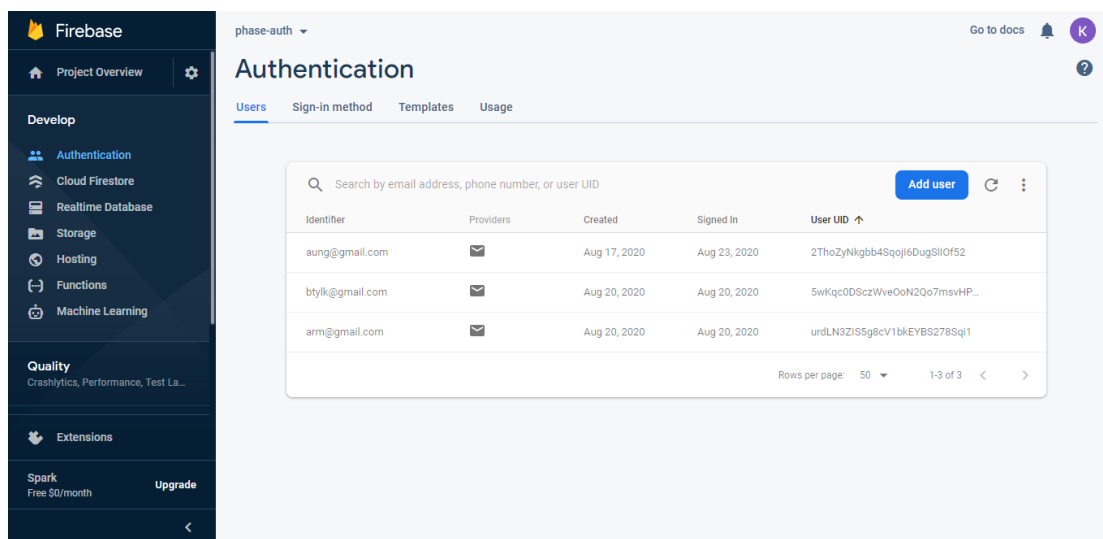
ส่วนยืนยันตัวตนก่อนใช้งานแอปพลิเคชัน โดย Firebase Authentication ประกอบด้วยหน้า Sign Up สำหรับบัญชีใช้งานแอปพลิเคชัน และหน้า Sign in สำหรับผู้ที่มีบัญชีเข้าใช้แอปพลิเคชันอยู่แล้ว ดังรูปที่ 21 และ รูปที่ 22



รูปที่ 21 แสดงหน้าลงชื่อเข้าใช้

รูปที่ 22 แสดงหน้าลงทะเบียน

เมื่อผู้ใช้งานยืนยันตัวตนสำเร็จ ข้อมูลจะถูกบันทึกเข้าไปใน database ของทาง Firebase ดังรูปที่ 23



รูปที่ 23 Firebase console แสดงผลการลงทะเบียนของผู้ใช้

4.2.2 อุปสรรคในการพัฒนา

บาง Module ที่จำเป็นต้องติดตั้งเพิ่มเติม ทำให้ module และ dependencies ที่มีอยู่เดิมหายไป และเกิด error ระหว่าง compile

4.2.3 แนวทางแก้ไขปัญหา

จำเป็นต้องลง module ที่หายไปซ้ำอีกครั้ง เพื่อให้ module และ ที่หายไป dependencies กลับมา เพื่อให้ทำงานได้ตามปกติ

4.3 พัฒนส่วนการยืนยันตัวตนด้วยเฟซบุ๊ก

4.3.1 รายละเอียดการพัฒนา

ส่วนยืนยันตัวตนก่อนใช้งานแอปพลิเคชันด้วยเฟซบุ๊ก ซึ่งจัดการโดย Firebase Authentication

4.3.2 อุปสรรคในการพัฒนา

Module ที่ใช้ในการพัฒนามีปัญหา เกิดการ error ในขั้นตอน compile

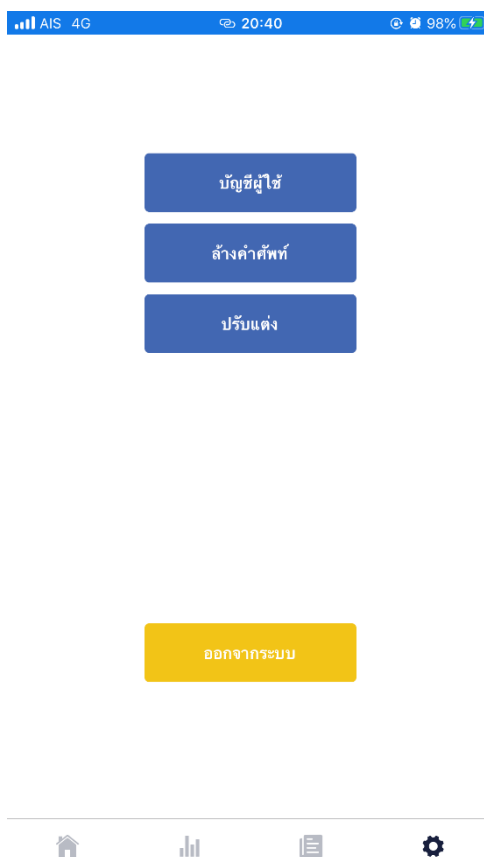
4.3.3 แนวทางแก้ไขปัญหา

อาจจำเป็นต้องหา module อื่นมาใช้แทน

4.4 พัฒนส่วนการติดต่อกับผู้ใช้

4.4.1 รายละเอียดการพัฒนา

ส่วนการติดต่อกับผู้ใช้ซึ่งประกอบหน้า หน้าหลัก หน้ารายงานผล หน้าจัดการคำศัพท์ หน้าตั้งค่า



รูปที่ 24 ส่วนการติดต่อกับผู้ใช้น้ำการตั้งค่า

4.4.2 อุปสรรคในการพัฒนา

- Module ที่ใช้ในการพัฒนามีปัญหา เกิดการ error ในขั้นตอน compile ทำให้หน้าหลักยังไม่สามารถพัฒนาต่อได้
- ส่วนการรายงานผลยังไม่สามารถพัฒนาได้เนื่องจากยังไม่มีส่วนประมวลผล
- หน้าจัดการคำศัพท์ยังไม่สามารถพัฒนาได้เนื่องจากข้อมูลคำศัพท์ยังแยกหมวดหมู่ไม่เสร็จและยังไม่ได้ใส่ข้อมูลลงใน database

4.4.3 แนวทางแก้ไขปัญหา

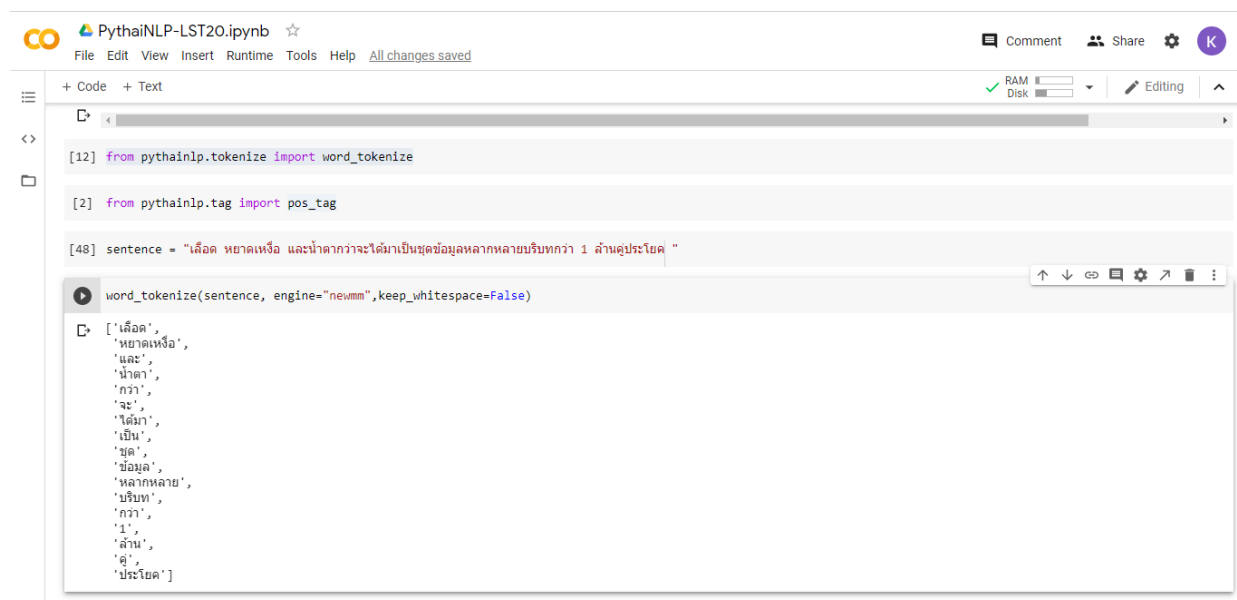
- หา module ที่ใช้ในการสร้าง pie chart ใหม่
- พัฒนาส่วนประมวลผล
- พัฒนาคำศัพท์ที่แยกประเภท ให้คะแนน ลงฐานข้อมูล

4.5 การทดสอบใช้งาน PythaiNLP+LST20

4.5.1 รายละเอียด

ทดสอบการใช้งาน library PythaiNLP คู่กับคลังคำศัพท์ LST20 ในการตัดคำ(Word Tokenize) และการแยกแยะชนิดของคำ

4.5.2 ผลการทดลอง



The screenshot shows a Jupyter Notebook interface with the following code and output:

```
[12] from pythainlp.tokenize import word_tokenize

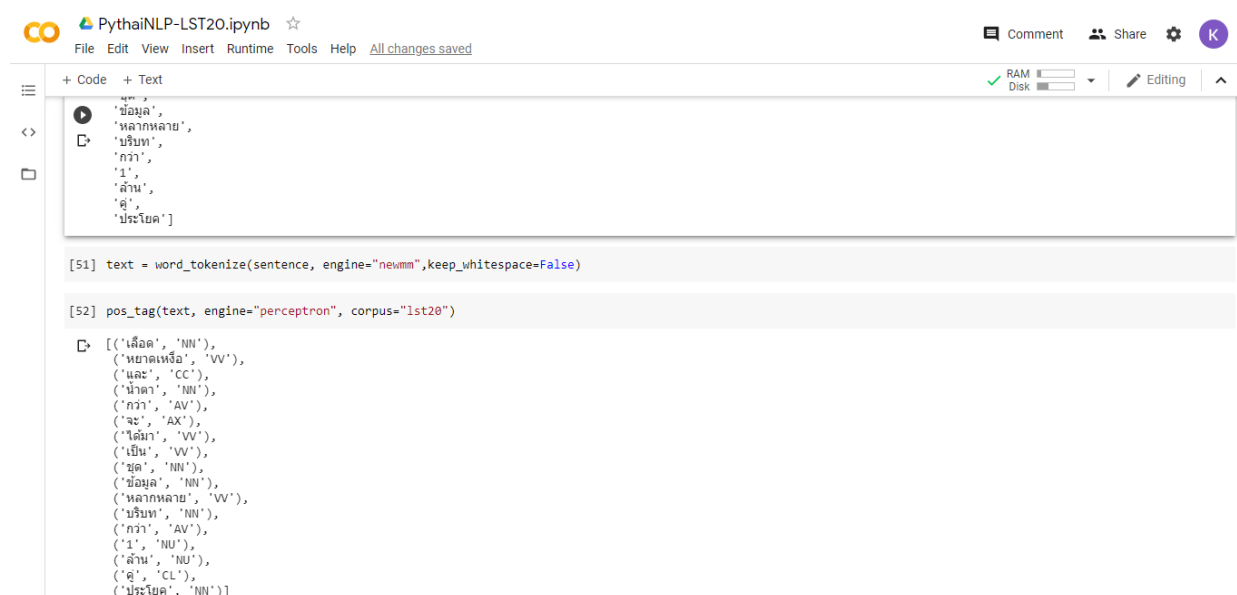
[2] from pythainlp.tag import pos_tag

[48] sentence = "เลือก หยาดเหงื่อ และน้ำตากว่าจะได้มาเป็นชุดข้อมูลหลากหลายบริบทกว่า 1 ล้านคู่ประโยค"
```

The output of the `word_tokenize` function is displayed as a list of tokens:

```
[ 'เลือก',
  'หยาดเหงื่อ',
  'และ',
  'น้ำตา',
  'กว่า',
  'จะ',
  'ได้มา',
  'เป็น',
  'ชุด',
  'ข้อมูล',
  'หลากหลาย',
  'บริบท',
  'กว่า',
  '1',
  'ล้าน',
  'คู่',
  'ประโยค']
```

รูปที่ 25 การตัดคำโดยใช้ PythaiNLP ร่วมกับ LST20



The screenshot shows a Jupyter Notebook interface with the following code and output:

```
[51] text = word_tokenize(sentence, engine="newmm", keep_whitespace=False)

[52] pos_tag(text, engine="perceptron", corpus="lst20")
```

The output of the `pos_tag` function is displayed as a list of tuples, where each tuple contains a token and its corresponding part-of-speech tag:

```
[ ('เลือก', 'NN'),
  ('หยาดเหงื่อ', 'VV'),
  ('และ', 'CC'),
  ('น้ำตา', 'NN'),
  ('กว่า', 'AV'),
  ('จะ', 'AX'),
  ('ได้มา', 'VV'),
  ('เป็น', 'VV'),
  ('ชุด', 'NN'),
  ('ข้อมูล', 'NN'),
  ('หลากหลาย', 'VV'),
  ('บริบท', 'NN'),
  ('กว่า', 'AV'),
  ('1', 'NU'),
  ('ล้าน', 'NU'),
  ('คู่', 'CL'),
  ('ประโยค', 'NN')]
```

รูปที่ 26 การแยกแยะชนิดคำโดยใช้ PythaiNLP ร่วมกับ LST20

4.5.3 วิเคราะห์ผล

จากการทดลองตัดคำและแยกแยะชนิดคำ พบว่ามีความแม่นยำค่อนข้างสูง อาจมีข้อผิดพลาดเล็กน้อยหากประโยคหรือคำที่ตัดมีความหมายกำกวม

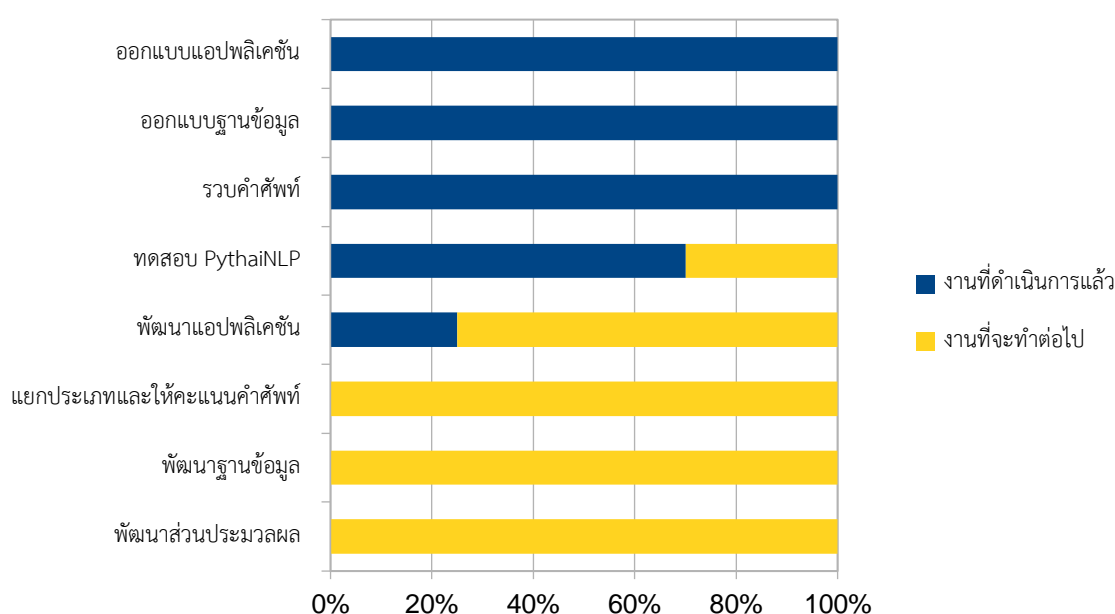
4.6 สรุปผลงานที่มีความก้าวหน้า

ในการรวบรวมคลังคำศัพท์ มีข้อมูลคำศัพท์มากเพียงพอต่อการนำไปใช้งาน แต่ยังคงต้องการนำไปแยกประเภทและให้คะแนนคำต่อไป

ส่วนการยืนยันตัวตนสามารถทำได้โดยการใช้อีเมลล์ และจะทำการเพิ่มการยืนยันตัวตนผ่านเฟซบุ๊กต่อไป

ส่วนการติดต่อกับผู้ใช้ได้พัฒนาในส่วนของเมนูเพื่อเข้าใช้งานในหมวดต่างๆ แต่ยังคงต้องรอการพัฒนาส่วนประมวลผลและฐานข้อมูลต่อไป เพื่อให้เมนูส่วนต่างๆสามารถใช้งานได้

ส่วนการประมวลผลข้อความ ได้มีการทดลองการการตัดคำและแยกแยะชนิดคำโดยใช้ PythaiNLP ร่วมกับคลังคำศัพท์ LST20



รูปที่ 27 แผนภาพแสดงความก้าวหน้า

บทที่ 5

สรุป

5.1 สรุปผลการดำเนินงาน

1. พัฒนา Mobile application ด้วย React native
2. พัฒนาระบบการยืนยันตัวตนด้วย Firebase Authentication ผ่านอีเมลล์
3. พัฒนาระบบการติดต่อกับผู้ใช้ ซึ่งต้องรอการพัฒนาส่วนประมวลผลและฐานข้อมูล เพื่อให้สามารถใช้งานได้
4. รวบรวมข้อมูลคำศัพท์
5. ทดลองการใช้ library PythaiNLP ร่วมกับ คลังคำศัพท์ LST20

5.2 ปัญหาและอุปสรรค

1. ในการพัฒนา Mobile application ด้วย React native มีความจำเป็นต้องลง Module เพิ่มสำหรับการพัฒนาส่วนต่างๆ ซึ่งบาง module ที่ใช้มีปัญหาระหว่างการ compile ทำให้ไม่สามารถพัฒนาในบางส่วนได้
2. ในการรวบรวมคลังคำศัพท์ ข้อมูลที่ได้มีจำนวนมากเกินไป จำเป็นต้องจำแนกให้เข้ากับจุดประสงค์ของโครงการจึงจะสามารถนำมาใช้ได้อย่างมีประสิทธิภาพ
3. ในการทดลองใช้ PythaiNLP ร่วมกับคลังข้อมูล LST20 พบว่ายังมีบางบริบทของประโยคที่นำไปทดลอง

5.3 งานที่จะดำเนินการต่อไป

1. คัดแยกและจำแนกประเภทของคำศัพท์จาก LST20 ให้สอดคล้องกับแนวโน้มเกี่ยวกับโรคซึมเศร้าและการกลั่นแกล้งในสังคมออนไลน์
2. จากข้อที่1 นำคำศัพท์ที่ผ่านการคัดแยกมาทำการให้คะแนนความรุนแรง
3. พัฒนาส่วนประมวลผล
4. พัฒนาส่วน database
5. พัฒนาส่วนติดต่อกับผู้ใช้ให้สมบูรณ์

บรรณานุกรม

[1] posttoday. (2563, มกราคม 26) “อึ้ง! สภาพัฒน์เผยคนไทยฆ่าตัวตายปีละ 4,000 คนส่วนใหญ่ ซึมเศร้า” [ออนไลน์]. แหล่งที่มา:

<https://www.posttoday.com/economy/news/572358>

[2] bangkokhospital. (2563, มกราคม 26) “โรคซึมเศร้า” [ออนไลน์]. แหล่งที่มา:

<https://www.bangkokhospital.com/th/disease-treatment/depression>

[3] มูลนิธิยุวพัฒน์. (2563, มกราคม 26) “การกลั่นแกล้ง (Bullying) ความรุนแรงในสังคม” [ออนไลน์]. แหล่งที่มา: <https://www.yuvabadhanafoundation.org/th/ข่าวสาร/บทความทั่วไป/การกลั่นแกล้ง-bullying-วัยรุ่น/>

[4] Modify. (2563, มกราคม 26) “iOS คืออะไร ประวัติความเป็นมา ใช้กับอุปกรณ์อะไร (ระบบปฏิบัติการ)” [ออนไลน์]. แหล่งที่มา: <https://www.modify.in.th/7945>

[5] similantechnology. (2563, มกราคม 26) “Android คืออะไร” [ออนไลน์]. แหล่งที่มา: <http://www.similantechnology.com/news&article/android.html>

[6] ปลั๊กไฟ. (2563, มกราคม 26) “JUPYTER NOTEBOOK” [ออนไลน์]. แหล่งที่มา: <https://ilog.ai/jupyter-notebook-for-data-science/>

[7] PyThaiNLP. (2563, มกราคม 26) “Thai Natural Language Processing in Python” [ออนไลน์]. แหล่งที่มา: <https://github.com/PyThaiNLP/pythainlp>

[8] DIGITAL VENTURES. (2563, มกราคม 26) “Natural Language Processing เทคโนโลยีเชื่อมโยงปัญญาประดิษฐ์กับมนุษย์ด้วยภาษา” [ออนไลน์]. แหล่งที่มา: <http://www.dv.co.th/blog-th/get-to-know-natural-language-processing-nlp/>

[9] Jedsada Saengow. (2563, มกราคม 26) “[React Native] คืออะไร ทำความรู้จักและเริ่มต้นสร้าง Project” [ออนไลน์]. แหล่งที่มา: <https://medium.com/jed-ng/react-native-ทำความรู้จัก-และเริ่มต้นสร้าง-project-91788ef6cac3>

[10] Sirawit. (2563, มกราคม 26) “ทำความรู้จัก Firebase และผลิตภัณฑ์ต่าง ๆ ในช่วงต้นปี 2019 กัน” [ออนไลน์]. แหล่งที่มา: <https://medium.com/@sirawit/firebase-คืออะไร-ทำความรู้จัก-firebase-ในช่วงต้นปี-2019-กัน-473a8e8699fb>

[11] Kapook. (2563, มกราคม 26) “Cyberbullying การกลั่นแกล้งบนโลกออนไลน์ ทำร้ายชาวเน็ตได้ง่ายกว่าที่คิด!” [ออนไลน์]. แหล่งที่มา: <https://health.kapook.com/view150050.html>

[12] Mr.P L. (2563, มกราคม 26) “NLP(Natural Language Processing) ศาสตร์(ไม่)ใหม่ ศาสตร์แห่งเจได: แยกประเภทอีเมลล์ด้วยพลังฟอร์ซ” [ออนไลน์]. แหล่งที่มา: <https://medium.com/mmp-li/nlp-natural-language-processing-ศาสตร์-ไม่-ใหม่-ศาสตร์แห่งเจได-แยกประเภทอีเมลล์ด้วยพลังฟอร์ซ-66b8bdff2e42>

[13] Wittawin.A. (2563, เมษายน 20) “สรุปสถิติผู้ใช้ Facebook ในไทย โดย Customer Insight” [ออนไลน์]. แหล่งที่มา: <https://www.thumbsup.in.th/2019/03/facebook-thailand-insight/>

[14] Prachya Boonkwan. (2563, สิงหาคม 10) “คลังข้อมูล LST20 Corpus” [ออนไลน์]. แหล่งที่มา: <https://www.facebook.com/dancearmy/posts/10157641945708284>