

ระบบจัดการและวิเคราะห์ข้อมูลจากสังคมออนไลน์
เพื่อจำแนกประเภทปัญหาเมือง

SOCIAL MEDIA DATA MANAGEMENT AND ANALYSIS
FOR URBAN PROBLEM CLASSIFICATION



สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2559

ระบบจัดการและวิเคราะห์ข้อมูลจากสังคมออนไลน์
เพื่อจำแนกประเภทปัญหาเมือง

SOCIAL MEDIA DATA MANAGEMENT AND ANALYSIS
FOR URBAN PROBLEM CLASSIFICATION



สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2559

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SOCIAL MEDIA DATA MANAGEMENT AND ANALYSIS
FOR URBAN PROBLEM CLASSIFICATION



COOPERATIVE EDUCATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (COMPUTER SCIENCE)
DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2016

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา

ระบบจัดการและวิเคราะห์ข้อมูลจากสังคมออนไลน์เพื่อจำแนกประเภทปัญหาเมือง

Social Media Data Management And Analysis For Urban Problem Classification

ชื่อนักศึกษา

นายกิตติพัฒน์ กิมขุนทด รหัสนักศึกษา 56050198

ปริญญา

วิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)

ภาควิชา

วิทยาการคอมพิวเตอร์


ปีการศึกษา

2559

อาจารย์ที่ปรึกษา

ผศ.กฤษฎา บุศรา

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์) ประจำปีการศึกษา 2559

คณะกรรมการสอบ	ลายมือชื่อ
ผศ.กฤษฎา บุศรา กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา	ระบบจัดการและวิเคราะห์ข้อมูลจากสังคมออนไลน์เพื่อจำแนกประเภทปัญหาเมือง
ชื่อนักศึกษา	นายกิตติพัฒน์ กิมขุนทด รหัสนักศึกษา 56050198
ปริญญา	วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา	วิทยาการคอมพิวเตอร์
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา	2559
อาจารย์ที่ปรึกษา	ผศ.กฤษฎา บุศรา

บทคัดย่อ

สื่อสังคมสังคมออนไลน์ เช่น Facebook และ Twitter ได้รับความนิยมอย่างมากและมีการใช้งานแบบเรียลไทม์ พบว่ามีข้อมูลในเครือข่ายสังคมจำนวนมากที่เกี่ยวกับปัญหาของเมือง เช่นปัญหาการระบายน้ำ อากาศเสีย รวมไปถึงการจราจรและสภาพผิวจราจร อย่างไรก็ตาม การนำข้อมูลจากสื่อสังคมออนไลน์มาใช้เพื่อให้เป็นประโยชน์สำหรับผู้อื่นต้องทำการคัดกรองและวิเคราะห์ข้อมูลเหล่านั้นก่อน เนื่องจากข้อมูลจากสื่อสังคมออนไลน์นี้มีรูปแบบที่หลากหลายและอาจมีค่าที่ไม่สุภาพไม่เหมาะสมแก่การนำข้อมูลมาเผยแพร่ ดังนั้นจึงต้องการอัลกอริทึมที่ใช้ในการสกัดและจำแนกประเภทข้อมูลปัญหาเมืองก่อน ข้อมูลปัญหาเมืองจากสื่อสังคมออนไลน์ ถูกสกัดโดยใช้การวิเคราะห์ประโยคซึ่งอาศัยการแบ่งประโยคออกเป็นคำที่มีความหมายหลังจากนั้นจึงทำการจำแนกประเภทของข้อมูลวิธีการนี้สามารถจำแนกประเภทข้อมูลปัญหาเมืองจำนวน 3,000 ทวิต โดยมีความแม่นยำเฉลี่ยร้อยละ 94.08 และระบบสามารถวิเคราะห์และจำแนกประเภทของข้อมูลอื่นๆได้อีกด้วย

คำสำคัญ : สังคมออนไลน์ ข้อมูลปัญหาเมือง การจัดหมวดหมู่ การวิเคราะห์ประโยค

Title	SOCIAL MEDIA DATA MANAGEMENT AND ANALYSIS FOR URBAN PROBLEM CLASSIFICATION
Students	Kittiphat Kuemkhunthod Student ID 56050198
Degree	Bachelor of Science (Computer Science)
Department	Computer Science
Faculty	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2016
Advisor	Asst.Prof.Kridsada Budsara

Abstract

Social media such as Facebook and Twitter social networking is popular and being used in real-time. Information found on social networks a lot about the problems of the city. Such as drainage problems, fire, including traffic jam and road surface. However, information from social media must perform the screening and analysis of that information before publishing and becoming useful for others. Because of social media, there are various forms and may contain words that are inappropriate, disrespectful to the data to publish. Therefore, we need the algorithm used in the extraction and classification for city problem Information. City problem information was extracted from Twitter using syntactic analysis and tokenization Then, the classification of the information. This method can classify 3,000 tweets into the classification with 94.08% accuracy and the system can analyze and classify of other information as well.

Keywords : Social Media, City Problem Information, Classification, Syntactic Analysis

กิตติกรรมประกาศ

การปฏิบัติงานโครงการสหกิจศึกษาที่ห้องปฏิบัติการวิจัยระบบขนส่งและจราจรอัจฉริยะ (ITS) ซึ่งอยู่ในเครือของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) อุทยานวิทยาศาสตร์ประเทศไทย (NSTDA) สำเร็จลงได้ด้วยดีเนื่องด้วยความอนุเคราะห์ช่วยเหลือจากอาจารย์ ผศ.กฤษฎา บุศรา ที่ให้คำปรึกษาและติดต่อประสานงานต่างๆ เพื่อให้การทำโครงการสหกิจศึกษา ลุล่วงไปด้วยดี ข้าพเจ้าจึงขอขอบคุณอาจารย์ไว้ ณ ที่นี้

ขอขอบพระคุณอาจารย์สาขาวิชาวิทยาการคอมพิวเตอร์ทุกท่านที่ได้ให้วิชาความรู้และให้คำปรึกษาทั้งในภาคทฤษฎีและภาคปฏิบัติมาตลอดระยะเวลา 4 ปีจนกระทั่งสัมฤทธิ์ ผลได้ด้วยดีทุกประการ

นอกจากนี้ข้าพเจ้ากราบขอบคุณ ดร. วสันต์ ภัทรอธิคม ดร. มนต์ศักดิ์ โช้เจริญธรรม ผู้ดูแลห้องปฏิบัติการวิจัยระบบขนส่งและจราจรอัจฉริยะ นางสุภรณ์ เหมือนหนู นายอรุณพล ก้อมมังกร นายนพพล กิรติวรนนท์ นายสิปปกรณ์ ไช้มุกต์ นายวัชร พุ่งวัชรากร และ นายณพงศ์ วาณิชยพงศ์ ที่คอยดูแล ให้คำแนะนำ และช่วยฝึกสอนข้าพเจ้า ซึ่งทำให้ข้าพเจ้ามีความรู้และประสบการณ์ในการทำงาน

ท้ายที่สุดนี้ข้าพเจ้าขอกราบขอบพระคุณบิดา มารดา ที่ได้ให้การสนับสนุนด้านทุนการศึกษา และคอยเป็นกำลังใจตลอดมา ข้าพเจ้าจึงใคร่ขอขอบพระคุณทุกท่านเป็นอย่างสูงไว้ ณ ที่นี้

กิตติพัฒน์ กิมขุนทด

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ช
สารบัญรูป	ซ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	2
1.5 เครื่องมือและภาษาที่ใช้ในงานวิจัย	3
1.6.1 ฮาร์ดแวร์ที่ใช้ในสหกิจศึกษา	3
1.6.2 ซอฟต์แวร์ที่ใช้ในสหกิจศึกษา	3
1.6.3 ภาษาที่ใช้ในสหกิจศึกษา	3
1.6 ระยะเวลาและแผนการดำเนินงาน	4
1.6.1 ระยะเวลาในการดำเนินงาน	4
1.6.2 แผนการดำเนินงาน	4
บทที่ 2 ทฤษฎีและเทคโนโลยีที่เกี่ยวข้อง	5
2.1 การเชื่อมต่อ API	5
2.1.1 รูปแบบ JSON	7
2.1.2 รูปแบบ XML	8
2.2 สื่อสังคมออนไลน์	9
2.2.1 ทวิตเตอร์ (Twitter)	9
2.2.2 เฟสบุ๊ก (Facebook)	10
2.3 การวัดประสิทธิภาพของโมเดล	11
2.3.1 การทดสอบประสิทธิภาพของโมเดล	11
2.3.2 การประเมินผลลัพธ์ของโมเดล	15

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

	หน้า
2.4 การตัดคำ	16
2.5 ภาษาที่ใช้งาน	18
2.5.1 ภาษา Python	18
2.5.2 ภาษา PHP	18
2.6 ซอร์ฟแวร์ที่ใช้งาน	19
2.5.1 MongoDB.....	19
2.6.2 ทาโบลัว (Tableau)	21
บทที่ 3 วิธีการดำเนินงานวิจัย	22
3.1 การวางแผนและการเตรียมการ	22
3.1.1 ศึกษาความเป็นไปได้และเก็บรวบรวมข้อมูลที่เกี่ยวข้องกับการดำเนินงาน	22
3.1.2 วิเคราะห์ความต้องการของระบบ	22
3.1.3 ออกแบบระบบ	23
3.2 ขั้นตอนการวิเคราะห์ระบบ	23
3.2.1 วิเคราะห์ระบบงานเดิม	23
3.2.2 วิเคราะห์ระบบงานใหม่	23
3.3 ขั้นตอนการออกแบบระบบ	24
3.3.1 วงจรการพัฒนาาระบบ	24
3.3.2 เครื่องมือที่ใช้ในการออกแบบ	25
3.4 ขั้นตอนการพัฒนาและทดสอบระบบ	38
บทที่ 4 ผลการดำเนินงานและการอภิปรายผล	39
4.1 การทดสอบการจำแนกประเภทของปัญหาเมือง	39
4.1.1 ข้อมูลที่นำมาทำการทดสอบการจำแนกประเภทของปัญหาเมือง	39
4.1.2 ผลลัพธ์จากการทดสอบการจำแนกประเภทของปัญหาเมือง	40
4.2 การทดสอบการจำแนกประเภทข้อมูลของเว็บเซอร์วิส	40
4.2.1 การทดสอบหน้าหลัก	41
4.2.2 การทดสอบหน้าใช้งานโมเดล	42
4.2.3 การทดสอบหน้าสตอปโมเดล	44
4.3 ผลการดำเนินงาน	46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

	หน้า
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	49
5.1 สรุปผลการดำเนินงานและการทดสอบ	49
5.1.1 สรุปผลการดำเนินงาน	49
5.1.2 สรุปผลการทดสอบเว็บเซอร์วิส	49
5.2 ข้อจำกัดในการพัฒนาระบบ	50
5.3 ข้อเสนอแนะในการพัฒนาระบบ	50
เอกสารอ้างอิง	51
ภาคผนวก	53
ภาคผนวก ก การติดตั้งโปรแกรม JetBrains PhpStorm 8.0.4	54
ภาคผนวก ข การติดตั้งโปรแกรม MongoDB	62
ภาคผนวก ค คู่มือการใช้งานเว็บเซอร์วิส (FASTTEXT WEB SERVICE)	67

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
3.1 อธิบายขั้นตอนการทำงานของกรรवरรวมและจำแนกข้อมูล	26
4.1 การทดสอบความถูกต้องในการจำแนกประเภทข้อมูล	46
4.2 การทดสอบการทำงานของหน้าหลัก	46
4.3 การทดสอบการทำงานของหน้าหลัก (ต่อ)	47
4.4 การทดสอบการทำงานของหน้าใช้งานโมเดล	47
4.5 การทดสอบการทำงานของหน้าทดสอบโมเดล	48
5.1 ผลการทดสอบเว็บเซอร์วิส	50



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 ตัวอย่างของการใช้งาน API	5
2.2 ตัวอย่างของภาษา JSON	7
2.3 ตัวอย่างของภาษา XML	8
2.4 กระบวนการของ K- fold Cross-validation Test	11
2.5 กระบวนการของ Self-Consistency Test	12
2.6 กระบวนการของ Split Test	13
2.7 กระบวนการรอบที่ 1 ของ Leave one out Test	14
2.8 กระบวนการรอบที่ 2 ของ Leave one out Test	14
2.9 Confusion Matrix	15
2.10 ตัวอย่างค่าของ MongoDB	19
2.11 ตัวอย่างการเพิ่มค่าของ MongoDB	20
2.12 ตารางเปรียบเทียบ MySQL และ MongoDB	20
3.1 Kimball 2008 Lifecycle	24
3.2 Work Flow Diagram การวิเคราะห์และจำแนกประเภทของข้อมูล	25
3.3 Use Case Diagram	27
3.4 Activity Diagram สำหรับการลงทะเบียนเพื่อใช้งาน	28
3.5 Activity Diagram สำหรับการใช้งาน หน้าหลัก	29
3.6 Activity Diagram สำหรับการใช้งาน หน้าใช้งานโมเดล	30
3.7 Activity Diagram สำหรับการใช้งาน หน้าทดสอบโมเดล	31
3.8 Activity Diagram สำหรับการใช้งาน หน้าวิธีการใช้งาน	32
3.9 Sequence Diagram สำหรับการลงทะเบียนเพื่อใช้งาน	33
3.10 Sequence Diagram สำหรับการใช้งาน หน้าหลัก	34
3.11 Sequence Diagram สำหรับการใช้งาน หน้าใช้งานโมเดล	35
3.12 Sequence Diagram สำหรับการใช้งาน หน้าทดสอบโมเดล	36
3.13 Sequence Diagram สำหรับการใช้งาน หน้าวิธีการใช้งาน	37
4.1 ตัวอย่างของข้อมูลปัญหาเมื่อนำมาจำแนกประเภท	39
4.2 Confusion matrix แสดงผลลัพธ์จากการทำนาย	40
4.3 ค่าความถูกต้องของการทำนาย	40
4.4 เลือกไฟล์ที่ต้องจำแนกประเภทของข้อมูล	41
4.5 เลือกตัวเลือกในการวิเคราะห์ผล	41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่	หน้า
4.6 วิเคราะห์และแสดงผลพีธีในการทำนาย	42
4.7 เลือกโมเดลที่ต้องการใช้เพื่อวิเคราะห์ผลลัพธ์	42
4.8 นำข้อความมาใส่ในช่องข้อความเพื่อวิเคราะห์ผลลัพธ์	43
4.9 วิเคราะห์ผลลัพธ์ในการทำนาย	43
4.10 เลือกโมเดลที่ต้องการใช้เพื่อวิเคราะห์ผลลัพธ์	44
4.11 นำข้อความมาใส่ในช่องข้อความเพื่อวิเคราะห์ผลลัพธ์	44
4.12 วิเคราะห์ผลลัพธ์ในการทำนาย	45
4.13 แก้ไขโมเดลและทดสอบใหม่	45
ก.1 Execute File Setup เพื่อเริ่มการติดตั้ง	54
ก.2 แสดงหน้าต่างเริ่มการติดตั้ง	54
ก.3 คลิก I Agree เพื่อทำการเริ่มการติดตั้ง	55
ก.4 เลือกตำแหน่งหรือพื้นที่ที่ต้องการติดตั้ง	55
ก.5 เลือกตัวเลือกในการติดตั้ง	56
ก.6 สร้างโฟลเดอร์ในการติดตั้ง	56
ก.7 แสดงรายละเอียดการติดตั้ง	57
ก.8 คลิก Finish เพื่อเสร็จสิ้นการติดตั้ง	57
ก.9 นำเข้าหรือตั้งค่าเวอร์ชัน	58
ก.10 ทำการใส่ชื่อและคีย์ของโปรแกรม	58
ก.11 แสดงรายละเอียดข้อตกลงในการใช้งาน	59
ก.12 ตั้งค่าเบื้องต้นในการเริ่มต้นการใช้งานโปรแกรม	59
ก.13 ยืนยันการตั้งค่าเบื้องต้นพร้อมกับทำการเริ่มต้นโปรแกรมใหม่	60
ก.14 แสดงหน้าจอตัวเลือกในการเริ่มต้นการใช้งาน	60
ก.15 แสดงหน้าจอเริ่มต้นใช้งาน	61
ข.1 ดาวน์โหลด MongoDB	62
ข.2 Execute File Setup.exe	62
ข.3 คลิก Next เพื่อทำการเริ่มการติดตั้ง	63
ข.4 แสดงรายละเอียดในการ Setup MongoDB	63
ข.5 เลือกตัวเลือกในการติดตั้ง	64
ข.6 สร้างโฟลเดอร์ในการติดตั้ง	64
ข.7 คลิก Install เพื่อเริ่มการติดตั้ง	65

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่	หน้า
ข.8 แสดงรายละเอียดการติดตั้ง	65
ข.9 คลิก Finish เพื่อเสร็จสิ้นการติดตั้ง	66
ค.1 หน้าแรกของเว็บเซอร์วิส	67
ค.2 หน้าแรกของเว็บเซอร์วิส เมนู เกี่ยวกับเรา	67
ค.3 หน้าแรกของเว็บเซอร์วิส เมนู เซอร์วิส	68
ค.4 หน้าแรกของเว็บเซอร์วิส เมนู ตัวอย่าง	68
ค.5 หน้าจอหลักขั้นตอนที่ 1	69
ค.6 หน้าจอหลักขั้นตอนที่ 2	69
ค.7 หน้าจอหลักขั้นตอนที่ 3	70
ค.8 หน้าใช้งานโมเดลขั้นตอนที่ 1	70
ค.9 หน้าใช้งานโมเดลขั้นตอนที่ 2	71
ค.10 หน้าทดสอบโมเดลขั้นตอนที่ 1	71
ค.11 หน้าทดสอบโมเดลขั้นตอนที่ 2	72
ค.12 หน้าทดสอบโมเดลขั้นตอนที่ 3	72
ค.13 หน้าวิธีการใช้งาน	73
ค.14 หน้าวิธีการใช้งาน	73

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติในเครือของเนคเทค (NECTEC) ภายใต้การทำงานของห้องปฏิบัติการวิจัยระบบขนส่งและจราจรอัจฉริยะ (ITS) ได้มีศูนย์ข้อมูลทางจราจรจากสื่อสังคมออนไลน์ (Social Media) ซึ่งมีจุดมุ่งหมาย เพื่อให้สามารถนำข้อมูลจากสื่อสังคมออนไลน์มาเป็นฐานข้อมูลในการพัฒนาต่อระบบอื่นๆและให้บริการ API สำหรับนักวิจัยหรือนักพัฒนาที่สนใจใช้ข้อมูลจากสื่อสังคมออนไลน์ (Social Media) เนื่องจากในปัจจุบันเทคโนโลยีสื่อสังคมออนไลน์ (Social Media) เช่น Facebook , Twitter เข้ามามีบทบาทในชีวิตประจำวันอย่างมาก ใช้กันอย่างแพร่หลาย และมีการใช้งานหลากหลายรูปแบบขึ้นอยู่กับความสนใจของแต่ละบุคคลว่าชอบหรือรักที่จะทำอะไรกับสังคมนั้นๆ ไม่ว่าจะเป็นการ โพสต์แสดงความคิดเห็น รูปภาพ แชร์บทความ แชร์สถานที่และอีกมากมาย เพื่ออธิบายสิ่งที่กำลังกระทำอยู่ในขณะนั้น รวมถึงเผยแพร่เว็บไซต์ต่างๆ การเติบโตของสื่อสังคมออนไลน์เช่น Facebook , Twitter เริ่มจากการใช้งานเป็นช่องทางการสื่อสารส่งข้อความสั้นๆ โพสต์หรือแชทสื่อต่างๆระหว่างกัน โดยจะมีทั้งการโพสต์เหตุการณ์ที่เกิดขึ้นในขณะนั้นหรือเกี่ยวกับเรื่องราวที่สนใจและจะมีการแชร์เหตุการณ์หรือเรื่องราวที่สนใจร่วมกัน หลังจากนั้นเริ่มมีการใช้เป็นช่องทางในการดำเนินกิจกรรมในด้านต่างๆ ไม่ว่าจะเป็นด้านธุรกิจ รวมไปถึงด้านวงการข่าว เช่นสำนักข่าวที่มีชื่อเสียงระดับสากลหลายแห่งต้องเปิดบล็อก (Blog) หรือ Twitter ควบคู่ไปกับการส่งข่าวตามวิธีมาตรฐาน ตัวอย่างเช่น CNN , New York Time เป็นต้น ซึ่งช่องทางใหม่ของการนำเสนอข่าวดังกล่าวมีชื่อเรียกละเอียดว่า @CNN , @NYTimes ตามลำดับ ซึ่งในประเทศไทยได้มีองค์กรหรือหน่วยงานที่นำเสนอข่าวสารในด้านการบันเทิง ด้านกีฬา ด้านเศรษฐกิจ และด้านปัญหาเมือง โดยปัญหาเมืองนี้จะเป็นปัญหาที่เกี่ยวข้อง ด้านการจราจร ด้านการระบายน้ำ ด้านถนนชำรุดต่างๆ รวมไปถึงปัญหาด้านอัคคีภัย ปัญหาเมืองส่วนใหญ่จะมีองค์กรหรือหน่วยงานที่เกี่ยวข้องเป็นศูนย์กลางในการรายงานข่าวสารจากสื่อสังคมออนไลน์ เช่น JS100RADIO FM91TRAFFICPRO เป็นต้น นอกจากนี้ผู้ใช้ทั่วไปยังสามารถเปลี่ยนแปลงบทบาทจากการเป็นพลเมืองผู้บริโภคข่าวกลายเป็นสื่อมวลชนผู้รายงานข่าวได้ด้วย โดยทางห้องปฏิบัติการวิจัยระบบขนส่งและจราจรอัจฉริยะ (ITS) มีข้อมูลด้านจราจร (Traffic) แล้ว

ปัญหาพิเศษนี้จะรวบรวมปัญหาเมืองในด้านอื่นๆที่นอกเหนือจากปัญหาด้านจราจรที่ทางห้องปฏิบัติการวิจัยระบบขนส่งและจราจรอัจฉริยะ (ITS) ได้มีการรวบรวมแล้ว โดยจะรวบรวมปัญหาเมืองจากแหล่งข้อมูลสื่อสังคมออนไลน์ (Social Media) มาจัดการวิเคราะห์ประโยค คัดกรอง และจัดหมวดหมู่ของข้อมูล พร้อมทั้งพัฒนาระบบการจัดการและวิเคราะห์ข้อมูลในการจำแนกประเภทของข้อมูล (Text Classifications)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 วัตถุประสงค์ของงานวิจัย

- 1) เพื่อพัฒนาระบบการจัดการและวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์ในการจำแนกประเภทปัญหาของเมือง
- 2) เพื่อพัฒนาเว็บเซอร์วิสที่ให้บริการจัดการและวิเคราะห์ข้อมูลในการจำแนกประเภทของข้อมูล (Text Classifications)

1.3 ขอบเขตของงานวิจัย

- 1) ศึกษาเครื่องมือและเทคโนโลยีในการพัฒนาระบบการจัดการและวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์ในการจำแนกประเภทของข้อมูล
- 2) พัฒนาระบบงานดังต่อไปนี้
 - 2.1) พัฒนาระบบการจัดการและวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์ในการจำแนกประเภทของปัญหาเมือง
 - 2.2) พัฒนาเว็บเซอร์วิสที่ให้บริการจัดการและวิเคราะห์ข้อมูลในการจำแนกประเภทของข้อมูล (Text Classifications)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

ในส่วน of ประโยชน์ที่คาดว่าจะได้รับ สามารถจำแนกออกเป็น 2 ส่วน คือ

1.4.1 ประโยชน์ต่อผู้ใช้งาน

- 1) ผู้ใช้งานระบบสามารถนำเข้าข้อมูลเพื่อจำแนกประเภทของข้อมูลได้ โดยจะช่วยอำนวยความสะดวกต่อผู้ใช้ ลดระยะเวลา ลดค่าใช้จ่ายในการพัฒนาระบบ
- 2) ผู้ใช้งานระบบสามารถแก้ไขต้นแบบ (Classification Model) ที่ใช้ในการจำแนกประเภทของข้อมูลเพื่อเพิ่มประสิทธิภาพในการใช้งานต้นแบบ (Classification Model)
- 3) ผู้ใช้งานระบบสามารถตรวจสอบความถูกต้องของการจำแนกประเภทของข้อมูล รวมไปถึงประสิทธิภาพของต้นแบบ (Classification Model)

1.4.2 ประโยชน์ต่อองค์กร

- 1) องค์กรมีระบบรวบรวมข้อมูลการรายงานปัญหาของเมืองจากแหล่งข้อมูลสื่อสังคมออนไลน์ (Social Media)
- 2) องค์กรมีระบบให้บริการจัดการและวิเคราะห์ข้อมูลในการจำแนกประเภทของข้อมูล
- 3) องค์กรมีระบบที่สามารถแก้ไขต้นแบบ (Classification Model) ที่ใช้ในการจำแนกประเภทของข้อมูลเพื่อเพิ่มประสิทธิภาพในการใช้งานต้นแบบ (Classification Model)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5 เครื่องมือและภาษาที่ใช้ในงานวิจัย

1.5.1 ฮาร์ดแวร์ที่ใช้ในสหกิจศึกษา

- 1) เครื่องคอมพิวเตอร์ใช้เป็นเครื่องควบคุม
 - 1.1) หน่วยประมวลผลกลาง (CPU) Intel(R) Core (TM) i5-3317U @1.70GHz
 - 1.2) หน่วยความจำหลัก (RAM) 4GB
 - 1.3) หน่วยความจำสำรอง 1 TB
 - 1.4) จอภาพ (Display)
 - 1.5) การ์ดจอ (Graphics Card)
 - 1.6) เมาส์ (Mouse)
 - 1.7) แป้นพิมพ์ (Keyboard)

1.5.2 ซอฟต์แวร์ที่ใช้ในสหกิจศึกษา

- 1) MongoDB ใช้ในการจัดการข้อมูล ค้นหาข้อมูล แก้ไข ข้อมูล เพิ่มข้อมูล ลบข้อมูล ในฐานข้อมูลที่ใช้งาน
- 2) PhpStorm 10 IDE ใช้ในการสร้างส่วนติดต่อกับผู้ใช้
- 3) โปรแกรม Adobe Illustrator ใช้ในการช่วยออกแบบตกแต่งรูปภาพ
- 4) โปรแกรม Adobe Photoshop CS6 ใช้ในการช่วยออกแบบตกแต่งรูปภาพ
- 5) โปรแกรม Tableau Desktop เป็น Business Intelligence Tools ใช้ในการช่วยสร้างรายงาน Dashboard ในการวิเคราะห์ข้อมูล
- 6) โปรแกรม Google Chrome เพื่อใช้เป็น Web Browser

1.5.3 ภาษาที่ใช้ในสหกิจศึกษา

- 1) ภาษา PHP (PHP Hypertext Preprocessor) เป็นภาษาจำพวก scripting language ใช้ในการประมวลผลและติดต่อกับเซิร์ฟเวอร์ (Server)
- 2) ภาษา HTML (Hyper Text Markup Language) เป็นภาษาที่ใช้ในการเขียนเว็บไซต์เพื่อแสดงข้อความ รูปภาพหรือวัตถุอื่น ๆ ผ่านโปรแกรมเว็บเบราว์เซอร์ (Browser)
- 3) ภาษา Java Script เพื่อช่วยในการเขียนเว็บแอปพลิเคชันให้สามารถแสดงเนื้อหาที่มีการเปลี่ยนแปลงได้ตามเงื่อนไขที่แตกต่างกันเพื่อตอบโต้กับผู้ใช้ได้มากยิ่งขึ้น
- 4) ภาษา Python เป็นภาษาที่ใช้ในการประมวลผล ทดสอบวิเคราะห์และจำแนกประเภท สามารถประมวลผลทางด้านวิทยาศาสตร์ และวิศวกรรมศาสตร์ได้อย่างมีประสิทธิภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.6 ระยะเวลาและแผนการดำเนินงาน

1.6.1 ระยะเวลาในการดำเนินงาน

- 1) ระยะเวลาตั้งแต่วันที่ 1 มิถุนายน 2559 – 30 ธันวาคม 2559

1.6.2 แผนการดำเนินงาน

- 1) ศึกษาขอบเขตปัญหาพิเศษและข้อมูลต่าง ๆ ที่เกี่ยวข้อง รวมไปถึงโปรแกรมและอุปกรณ์ที่นำมาใช้
- 2) ศึกษาข้อมูลทางด้านภาษาที่จะนำมาใช้ในการพัฒนาเว็บเซอร์วิส
- 3) วิเคราะห์ปัญหาและออกแบบส่วนติดต่อผู้ใช้ของเว็บเซอร์วิสสำหรับให้บริการ
- 4) พัฒนาและทดสอบเว็บเซอร์วิสกับอุปกรณ์ที่นำมาใช้
- 5) ประเมินความถูกต้องเพื่อนำมาปรับปรุงและพัฒนาเว็บเซอร์วิส
- 6) จัดทำคู่มือการใช้งานนำเสนอปัญหาพิเศษ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

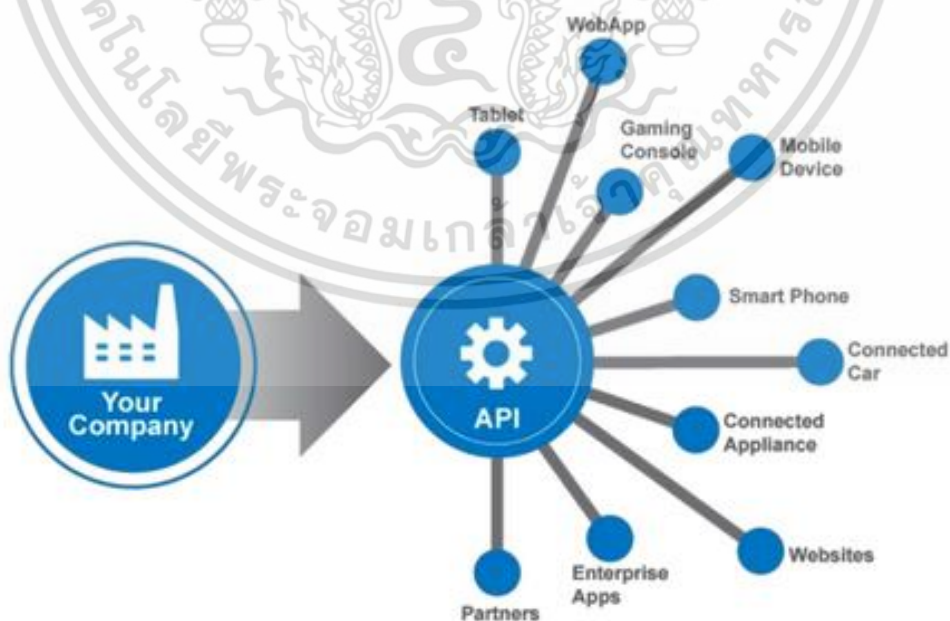
บทที่ 2

ทฤษฎีและเทคโนโลยีที่เกี่ยวข้อง

การวิจัยระบบจัดการและวิเคราะห์ข้อมูลจากสังคมออนไลน์เพื่อจำแนกประเภทปัญหาเมือง เพื่อนำข้อมูลที่ผ่านการวิเคราะห์และจำแนกประเภทของข้อมูลแล้วมาเป็นฐานข้อมูลในการพัฒนาระบบที่เกี่ยวข้องกับปัญหาของเมืองต่อไป พร้อมทั้งพัฒนาเว็บเซอร์วิสให้บริการการจัดการและวิเคราะห์ข้อมูลในการจำแนกประเภทของข้อมูล ผู้พัฒนาได้ศึกษาหลักการทฤษฎีต่างๆ และเทคโนโลยีที่เกี่ยวข้องที่สามารถประยุกต์ใช้งานได้ โดยแบ่งเป็นหัวข้อต่างๆ ดังต่อไปนี้

- 2.1 การเชื่อมต่อ API (Application Programming Interface)
- 2.2 สื่อสังคมออนไลน์ (Social Media)
- 2.3 การวัดประสิทธิภาพของโมเดล
- 2.4 การตัดคำ (Tokenization)
- 2.5 ภาษาที่ใช้งาน (Programmer Language)
- 2.6 ซอร์ฟแวร์ที่ใช้งาน (Software)

2.1 การเชื่อมต่อ API (Application Programming Interface)



รูปที่ 2.1 ตัวอย่างของการใช้งาน API

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Application Programming Interface เป็นช่องทางในการเชื่อมต่อช่องทางหนึ่งที่จะเชื่อมต่อกับเว็บไซต์ผู้ให้บริการ API เป็นตัวกลางที่ทำให้โปรแกรมประยุกต์เชื่อมต่อกับโปรแกรมประยุกต์อื่นหรือเป็นตัวกลางที่เชื่อมการทำงานเข้ากับ ระบบปฏิบัติการ ซึ่งผู้ให้บริการ API ทำหน้าที่ช่วยในการเข้าถึงข้อมูลต่างๆ หรือการนำข้อมูลต่างๆ ภายในเว็บไซต์ผู้ให้บริการ API ตามที่เว็บไซต์ผู้ให้บริการ กำหนดขอบเขตของข้อมูลไว้ ตัวอย่างเช่น Twitter มีหลายเว็บไซต์ที่มีการเชื่อมโยงข้อมูลกับ twitter ทั้งเป็นการอ่านข้อมูลจาก twitter หรือ ส่งข้อมูลเข้า twitter ซึ่งล้วนอาศัยการเชื่อมต่อแลกเปลี่ยนข้อมูลกันด้วย API และอีกหนึ่งตัวอย่าง เช่น Google Maps API คือบริการของ Google อีกรูปแบบหนึ่งที่เราสามารถนำข้อมูลของ Google Maps ที่ทาง Google ให้บริการโดยส่วนมากจะนำมาใช้กับเว็บไซต์ ของบริษัทฯหรือเว็บไซต์ห้างร้านต่างๆ เพื่อเป็นอีกช่องทางที่ให้ลูกค้ารู้ว่าบริษัทฯ หรือห้างร้านนั้นอยู่ที่ใด แต่อย่างไรก็ดี API นั้นไม่ได้มีมาตรฐานที่ตายตัวว่าจะต้องใช้งานหรือต้องเขียนคำสั่งที่เหมือนกันทั้งหมด นั่นก็คือถ้ามี API ของ 10 ระบบ ก็ไม่จำเป็นต้องมีวิธีเชื่อมต่อเหมือนกันทั้งหมด ในบางที่อาจจะมีการเชื่อมต่อ API ทั้ง 10 แบบ เพราะว่าแต่ละบริษัท ต่างมีการพัฒนา API ในแนวทางที่ต่างกัน ดังนั้นจึงเป็นหน้าที่ของผู้ที่ต้องการเรียกใช้ API ที่จะต้องศึกษาว่าจะเชื่อมต่อข้อมูลจากผู้ให้บริการ API ได้อย่างไรและจะส่งข้อมูลไปได้อย่างไร

API (Application Programming Interface) ทำหน้าที่ช่วยในการเข้าถึงข้อมูลต่าง ๆหรือจะเป็นการนำข้อมูลต่างๆออกจากเว็บไซต์ หรือจะเป็นการส่งข้อมูลเข้าไปยังเว็บไซต์ โดยเจ้าของเว็บไซต์ที่มี API จะกำหนดขอบเขตในการเข้าถึงบริการต่าง ๆของทางเว็บไซต์ โดยประโยชน์ของ API สามารถสรุปได้ดังนี้

1) ช่วยในการพัฒนาเว็บไซต์หรือแอปพลิเคชัน (Application) ได้ง่ายและรวดเร็ว ซึ่ง API จะเป็นตัวช่วยที่นักพัฒนาไม่ต้องเข้าไปแก้ไขคำสั่ง (Code) ทำให้สะดวกสบายในการใช้งาน

2) ไม่จำเป็นต้องเข้าหน้าเว็บหลัก ก็มีข้อมูลของเว็บหลัก จากเว็บที่ดึง API โดยแบ่งเป็น API ที่ขึ้นกับภาษา (language-dependent API) คือ API ที่สามารถการเรียกใช้จากโปรแกรมที่เขียนขึ้นด้วยภาษาเพียงภาษาใดภาษาหนึ่ง และ API ไม่ขึ้นกับภาษา (language-independent API) คือ API ที่สามารถเรียกได้จากโปรแกรมหลายๆภาษา

3) ช่วยให้นักพัฒนาเว็บไซต์หรือเจ้าของเว็บไซต์สามารถฐานผู้ชมเว็บไซต์ให้มากขึ้น

4) API สามารถรับส่งข้อมูลข้าม Server ได้

API ถือเป็นกลุ่มของฟังก์ชันหรือคลาส (Class) ที่ระบบปฏิบัติการ (OS) หรือผู้ให้บริการ สร้างขึ้นมา เพื่อรองรับการเรียกขอข้อมูล จากโปรแกรมอื่น ๆ ทั้งนี้ API สามารถใช้งานได้กับภาษาในการเขียนโปรแกรมที่รองรับเท่านั้น ซึ่งมันจะถูกจัดทำให้อยู่ในรูปแบบ (Syntax) หรือ (Element) ที่สามารถนำไปใช้ได้อย่างสะดวก โดย API มี 2 รูปแบบที่นิยมใช้กันคือ JSON และ XML

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.1 รูปแบบ JSON (JavaScript Object Notation)

JSON หรือ Java Script Object Notation เป็นรูปแบบหรือฟอร์แมต (Format) ของข้อมูลที่ใช้สำหรับแลกเปลี่ยนข้อมูลที่มีขนาดเล็ก ซึ่งสามารถทำความเข้าใจได้ง่าย และสามารถถูกสร้างและอ่านโดยเครื่องได้ง่าย โดย JSON ถูกกำหนดภายใต้ภาษา JavaScript (JavaScript Programming Language) JSON เป็นรูปแบบข้อมูลตัวอักษรที่มีความเป็นอิสระอย่างสมบูรณ์ แต่จะมีหลักการการเขียนที่คุ้นเคยกับนักเขียนโปรแกรมภาษาต่างๆ ได้ ไม่ว่าจะเป็น ภาษา C, C++, C#, Java, JavaScript, Perl, Python และอื่นๆ คุณสมบัติเหล่านี้ทำให้ JSON เป็นภาษาที่ใช้สำหรับแลกเปลี่ยนข้อมูลที่มีความสมบูรณ์แบบ คือจะมีความกระชับ สั้นและเข้าใจง่ายกว่า XML อยู่มาก ฟอร์แมต JSON นั้นอยู่ในรูปข้อความธรรมดา (Plain text) ส่วนใหญ่จะใช้ในการส่งข้อมูลเพื่อจัดเก็บเข้าเซิร์ฟเวอร์ (Server)

ในการทำงานหลายอย่างที่ต้องใช้ภาษา JavaScript มักจะพบ JSON เข้ามาเกี่ยวข้องด้วย ยกตัวอย่างการทำงานกับ script หลายๆตัวที่มีการเรียกข้อมูลแบบ AJAX มักจะส่งข้อความตอบกลับมาในรูปแบบ JSON ส่วนมาตรฐานของฟอร์แมต JSON คือ RFC 4627 มี Internet media type เป็น application/json และมีนามสกุลของไฟล์เป็น .json ปัจจุบัน JSON นิยมใช้ในเว็บแอปพลิเคชัน โดยเฉพาะ AJAX โดย JSON เป็นฟอร์แมตทางเลือกในการส่งข้อมูล นอกเหนือไปจาก XML ซึ่งนิยมใช้กันอยู่แต่เดิม

JSON จะใช้สัญลักษณ์ [] แทนอาร์เรย์ (Array) และใช้สัญลักษณ์ { } แทนแฮช (hash) หรือ (Associate Array) โดยแต่ละสมาชิกคั่นด้วยสัญลักษณ์ comma (,) และแต่ละชื่อสมาชิกคั่นด้วยสัญลักษณ์ colon (:). ลักษณะของข้อมูลจะเป็นดังรูป 2.2

```
{
  "primary_source" : "Twitter",
  "sub_type" : "ฝนตก",
  "main_type" : "สภาพอากาศ",
  "event_begin" : "2014-02-15 06:45:48",
  "geomtype" : "point",
  "news_id" : "504529",
  "description" : "ซักจะมาเป็นเม็ด #ฝนตก #สวณรถไฟ 6.45น."
}
```

รูปที่ 2.2 ตัวอย่างของภาษา JSON

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.1 รูปแบบ XML (Extensible Markup Language)

XML หรือ Extensible Markup Language คือภาษาหนึ่งที่ใช้ในการแสดงผลข้อมูล ซึ่งภาษาที่ใช้กำหนดรูปแบบของคำสั่งภาษา HTML หรือที่เรียกว่า Meta Data ซึ่งจะใช้สำหรับกำหนดรูปแบบของคำสั่ง Markup ต่าง ๆ ถ้าเปรียบเทียบกับภาษา HTML จะแตกต่างกันตรงที่ภาษา HTML ถูกออกแบบมาเพื่อการแสดงผลอย่างเดียวนั้น เช่น ให้แสดงผลตัวอักษรตัวเล็ก ตัวหนา ตัวเอียง เหมือนที่แสดงผลในเว็บเพจทั่วไป แต่ภาษา XML นั้นถูกออกแบบมาเพื่อเก็บข้อมูล โดยจะจัดเก็บทั้งข้อมูลและโครงสร้างของข้อมูลนั้นๆไว้ด้วยกัน สำหรับการแสดงผลก็จะใช้ภาษาเฉพาะซึ่งก็คือ XSL หรือ (Extensible Stylesheet Language)

ภาษา XML มีโครงสร้างที่ประกอบด้วยแท็กเปิด <tag> และแท็กปิด </tag> เช่นเดียวกับภาษา HTML แต่ภาษา XML สามารถสร้างแท็กรวมทั้งกำหนดโครงสร้างของข้อมูลได้เอง ซึ่งความสามารถนี้ภาษา HTML ไม่สามารถทำได้เพราะภาษา HTML ถูกกำหนดแท็กตายตัวโดย W3C หรือ World Wide Web Consortium อาจกล่าวได้ว่า XML เป็นส่วนเสริมของ HTML เพราะ XML ไม่สามารถแสดงผลได้ในตัวของมันเอง หากต้องการแสดงผลที่ถูกต้อง จะต้องมีการใช้ร่วมกับภาษาอื่น เช่น HTML, JSP, PHP, ASP หรือภาษาอื่น ๆ ที่สนับสนุน XML จะมีนามสกุลเป็น .XML สามารถสร้างขึ้นโดยใช้โปรแกรมประเภท Text Editor ใดๆก็ได้ เช่น Notepad, Editplus, Dreamweaver, MS Word เป็นต้น

สิ่งที่ถือได้ว่าเป็นจุดเด่นของ XML นั้นจะเป็นความสะดวกในการจัดการด้านระบบการติดต่อกับผู้ใช้จากโครงสร้างของข้อมูล สามารถนำข้อมูลจากหลายแหล่งมาแสดงผลและประมวลผลร่วมกันได้ ไม่ว่าจะเป็นข้อมูลลูกค้า รายการสั่งซื้อ ผลการวิจัย รายการรับชำระเงิน ข้อมูลเวชระเบียน รายการสินค้าหรือข้อมูลสารสนเทศอื่นๆ ก็สามารถแปลงให้เป็น XML ได้ และในส่วนของข้อมูลสามารถปรับให้เป็น HTML ได้ สำหรับประโยชน์ในการใช้งานนั้น สามารถนำมาใช้สำหรับการเข้าถึงระบบข้อมูลขนาดใหญ่ใช้กับระบบเครือข่ายในองค์กร หรืออินเทอร์เน็ตเพื่อดูข้อมูลหรือเรียกใช้ข้อมูลที่ทำให้การแสดงผลทางหน้าจอที่รวดเร็ว

```

</item>
▼ <item>
  <title>อุณหภูมิต่ำเดือนเมษายน 2559 ทำลายสถิติของประเทศไทย</title>
  <pubDate>Wed, 11 May 2016 12:36:54 +0700</pubDate>
  <guid>5105</guid>
  <link>\uploads\weatherclimate\aprilbreakrecord.pdf</link>
  <author>กรมอุตุนิยมวิทยา</author>
  ▼ <description>
    <![CDATA[ ลมวันที่: 11 พฤษภาคม 2559<br /> ]]>
  </description>
</item>

```

รูปที่ 2.3 ตัวอย่างของภาษา XML

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 สื่อสังคมออนไลน์ (Social Media)

จะกล่าวถึงสื่อสังคมออนไลน์ต้องมีคำสองคำนี้เข้ามาเกี่ยวข้องคือ คำว่าโซเชียลมีเดีย (Social Media) และโซเชียลเน็ตเวิร์ค (Social Network) เพราะทั้งสองคำนี้มาพร้อมกับความก้าวหน้าของเทคโนโลยีคอมพิวเตอร์และเครือข่าย ซึ่งปัจจุบันเราสามารถอ่านหนังสือพิมพ์ผ่านทางอินเทอร์เน็ต ที่เราเรียกว่าหนังสือพิมพ์ออนไลน์ ดูหนังผ่านทางอินเทอร์เน็ต ขายสินค้าผ่านทางอินเทอร์เน็ตเรียนหนังสือผ่านทางอินเทอร์เน็ต นอกจากนั้น ยังสามารถนำเอารูปภาพ วิดีโอ ของตนเองเผยแพร่ทางอินเทอร์เน็ต (Internet) ให้ผู้คนในสังคมออนไลน์ (Online Community) ได้ชม ดังนั้นการเผยแพร่สื่อในรูปแบบต่างๆดังกล่าวนี้เข้าไปในเครือข่ายอินเทอร์เน็ต (Internet) จึงเป็นที่มาของคำว่าโซเชียลมีเดีย (Social Media) และโซเชียลเน็ตเวิร์ค (Social Network)

ความหมายของโซเชียลมีเดีย (Social Media) แบ่งเป็นสองส่วน คือ คำว่า “Social” หมายถึง สังคม ในที่นี้จะหมายถึงสังคมออนไลน์ ซึ่งมีขนาดใหญ่มากในปัจจุบัน และคำว่า “Media” หมายถึง สื่อ ซึ่งก็คือเนื้อหา เรื่องราว บทความ วิดีโอ เพลง รูปภาพ เป็นต้น ดังนั้นคำว่าโซเชียลมีเดีย (Social Media) จึงหมายถึงสื่อสังคมออนไลน์ที่มีการตอบสนองทางสังคมได้หลายทิศทาง โดยผ่านเครือข่ายอินเทอร์เน็ต (Internet)

พื้นฐานการเกิดโซเชียลมีเดีย (Social Media) มาจากความต้องการของมนุษย์หรือบุคคลที่ต้องการติดต่อสื่อสารหรือมีปฏิสัมพันธ์กัน จากเดิมมีเว็บในยุค 1.0 ซึ่งก็คือเว็บที่แสดงเนื้อหาอย่างเดียว บุคคลแต่ละคนจะไม่สามารถติดต่อหรือโต้ตอบกันได้ แต่เมื่อเทคโนโลยีเว็บพัฒนาเข้าสู่ยุค 2.0 ก็มีการพัฒนาเว็บไซต์ที่เรียกว่าเว็บแอปพลิเคชัน (Web Application) คือเว็บไซต์มีแอปพลิเคชันหรือโปรแกรมต่างๆ ที่มีการโต้ตอบกับผู้ใช้งานมากขึ้น ผู้ใช้งานสามารถโต้ตอบกันได้ผ่านหน้าเว็บ

2.2.1 ทวิตเตอร์ (Twitter)

เป็นบริการเครือข่ายสังคมออนไลน์จำพวกไมโครบล็อก (Micro Blogging) โดยผู้ใช้งานสามารถส่งข้อความได้ความยาวไม่เกิน 140 ตัวอักษร ว่ากำลังทำอะไรอยู่ในช่วงเวลานั้นหรือทวิต (Re-tweet) ข่าวสารที่น่าสนใจของบุคคลอื่น และข้อความที่ส่งถึงกันนั้นมีศัพท์เรียกว่าทวิต (Tweets) ซึ่งเปรียบเหมือนเสียงกริ่งอยู่ตลอดเวลา ข้อความที่จะส่งนั้นต้องเป็นข้อความธรรมดา (Plain text) เท่านั้นจะแทรกคำสั่งโปรแกรมอะไรไม่ได้ ทวิตเตอร์ (Twitter) ก่อตั้งขึ้นเมื่อเดือนมีนาคม พ.ศ. 2549 โดย แจ็ก ดอร์ซี, บิช สโตน และอีวาน วิลเลียมส์ เจ้าของบริษัทออบวิโออูส และต่อมาในเดือนกรกฎาคม ก็ได้เปิดตัวเว็บไซต์สังคมออนไลน์ ซึ่งต่อมาได้รับความนิยมจากทั่วโลก โดยมีผู้ใช้ที่ลงทะเบียนเข้าใช้มากกว่า 500 ล้านคนในปี พ.ศ. 2555 รวมไปถึงมีทวิต (Tweets) มากกว่า 340 ล้านทวิตต่อวัน และมีการค้นหาข้อมูลผ่านเว็บไซต์มากกว่า 1,600 ล้านครั้งต่อวัน นับตั้งแต่วันเปิดตัวทวิตเตอร์เป็นเว็บไซต์ 1 ใน 10 อันดับที่มีผู้เข้าใช้งานมากที่สุด โดยได้รับการขนานนามว่าเป็นเว็บไซต์สำหรับการส่งบริการข้อความสั้นบนอินเทอร์เน็ต ซึ่งในภายหลัง นอกจากที่จะสามารถทวิต (Tweets) บนเว็บไซต์แล้ว ได้มีการเปิดให้ใช้งานการส่งทวิต ด้วยการส่งบริการข้อความสั้น (SMS) และบนโปรแกรมประยุกต์ ในโทรศัพท์มือถือ และสมาร์ตโฟน ตัวระบบซอฟต์แวร์ของทวิตเตอร์ เดิมพัฒนาเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ด้วย รูบี้ออนเรลส์ (ruby on rails) จนเมื่อราวสิ้นปี ค.ศ. 2008 หรือ พุทธศักราช 2551 จึงได้เปลี่ยนมาใช้ภาษาสกาลา (Scala) บนแพลตฟอร์มจาวา (Java) ปัจจุบันมีการประยุกต์ใช้ทวิตเตอร์ (Twitter) สำหรับดึงข้อมูลเพื่อตรวจสอบเหตุการณ์หรือกิจกรรมต่างๆ โดยมีความแตกต่างระหว่างขนาดเล็กและขนาดใหญ่ เหตุการณ์หรือกิจกรรมขนาดเล็กเช่น การจราจร รถชน ไฟไหม้ เหตุการณ์ท้องถิ่น โดยปกติขนาดเล็กจะต้องมีความแม่นยำทางภูมิศาสตร์และมีความชัดเจนของช่วงเวลา ในทางกลับกัน เหตุการณ์หรือกิจกรรมขนาดใหญ่เช่น แผ่นดินไหว พายุ การเลือกตั้ง โดยขนาดใหญ่จะต้องมีจำนวนที่มากและมีข้อมูลทางภูมิศาสตร์เป็นวงกว้าง เป็นต้น

ทวิตเตอร์ (Twitter) มีการสร้าง API ขึ้นมาเพื่อให้ผู้ใช้งานสามารถนำข้อมูลจากทวิตเตอร์ (Twitter) ไปใช้งานได้ แต่ทางทวิตเตอร์ (Twitter) ก็ได้มีข้อจำกัดในการใช้งาน API คือ อัตราของ API จะถือเป็นหลักพื้นฐานต่อผู้ใช้ซึ่งขึ้นต่อโทเค็น (Token) การเข้าถึงในการควบคุมของผู้ใช้ โดยอัตราเชื่อมต่อ Twitter API จำกัดที่ 180 คำสั่งต่อการค้นหาและต้องรออีก 15 นาทีจึงจะสามารถเรียกคำสั่งใหม่ได้ เมื่อโปรแกรมค้นหาเกินขีดจำกัดของอัตราที่กำหนด Twitter API จะส่งค่า HTTP 429 หรือหมายความว่าเมื่ออัตราการเชื่อมต่อมากเกินไปกลับมาในรหัสการตอบสนอง

2.2.2 เฟสบุ๊ก (Facebook)

เป็นบริการบนอินเทอร์เน็ตบริการหนึ่ง ที่จะทำให้ผู้ใช้สามารถติดต่อสื่อสารและร่วมทำกิจกรรมใดกิจกรรมหนึ่งหรือหลายๆกิจกรรมกับผู้ใช้เฟสบุ๊ก (Facebook) คนอื่นๆได้ ไม่ว่าจะเป็นการตั้งประเด็นถามตอบในเรื่องที่สนใจ โปสต์รูปภาพ โปสต์คลิปวิดีโอ เขียนบทความหรือบล็อก (Blog) แชนทคุยกัน เล่นเกมส์แบบเป็นกลุ่ม และยังสามารถทำกิจกรรมอื่นๆ ผ่านแอปพลิเคชันเสริม (Applications) ที่มีอยู่อย่างมากมาย ซึ่งแอปพลิเคชัน (Applications) ดังกล่าวได้ถูกพัฒนาอยู่เรื่อยๆ โดยสำนักงานใหญ่อยู่ที่ เมนโลพาร์ก รัฐแคลิฟอร์เนีย ซึ่งเฟสบุ๊ก (Facebook) เป็นชื่อที่ใช้ในการสื่อสารหรือป้ายบอกทางแก่เหล่านักศึกษาในบางมหาวิทยาลัยอเมริกัน เฟสบุ๊ก (Facebook) ก่อตั้งเมื่อวันอังคารที่ 4 กุมภาพันธ์ ค.ศ. 2004 โดยมาร์ก ซักเคอร์เบิร์ก และเหล่าเพื่อนในมหาวิทยาลัยฮาร์วาร์ด (Harvard University) พร้อมด้วยสมาชิกเพื่อนผู้ก่อตั้งคือ Eduardo Saverin, Andrew McCollum, Dustin Moskovitz และ Chris Hughes เริ่มแรกเว็บไซต์นี้ทำเพื่อให้นักศึกษาภายในมหาวิทยาลัยฮาร์วาร์ด (Harvard University) ใช้งานจึงมีการเข้าขมอย่างจำกัด แต่ภายหลังได้ทำการขยายเพิ่มจำนวนในมหาวิทยาลัย ในพื้นที่บอสตัน (Boston) ไอวิลีก (Ivy League) และมหาวิทยาลัยสแตนฟอร์ด (Stanford University) และค่อยๆรับรองมหาวิทยาลัยอื่นต่างๆ และต่อมาก็รับรองโรงเรียนมัธยมศึกษา โดยเฟสบุ๊ก (Facebook) ให้การอนุญาตให้เยาวชนอายุต่ำกว่า 13 ปีทั่วโลกสามารถสมัครสมาชิกได้ภายในเว็บไซต์ โดยไม่ต้องอ้างอิงหลักฐานใด ๆ

เฟสบุ๊ก (Facebook) มีบริการ API เพื่อให้สามารถนำข้อมูลต่างๆจากเฟสบุ๊ก (Facebook) มาใช้ โดยผู้ใช้จะต้องมีเฟสบุ๊กแอฟ (Facebook App) ทำหน้าที่เป็นตัวกลางในการเชื่อมต่อแอปพลิเคชัน (Application) และเว็บไซต์กับเฟสบุ๊ก (Facebook) ซึ่งจะทำให้นักพัฒนาสามารถส่งข้อมูลหรือดึงข้อมูลจากเฟสบุ๊ก (Facebook) มาใช้งานที่แอปพลิเคชัน (Application) หรือ

เว็บไซต์ได้ ผ่านทาง Facebook SDK แบบต่างๆ ประกอบด้วย IOS, Android, PHP, Javascript และอื่นๆ เป็นต้น การใช้งาน SDK เหล่านี้ นักพัฒนาจำเป็นต้องมีแอปไอดี (App ID) ของตนเองก่อน โดยแอปไอดี (App ID) จะใช้อ้างอิงกับเฟสบุ๊คแอป (Facebook App) ที่สร้างขึ้น และการเข้าใช้งาน เฟสบุ๊คแอป (Facebook App) ผ่านทาง SDK จะถูกเก็บเป็นสถิติให้นักพัฒนาสามารถตรวจสอบ วิเคราะห์ข้อมูลได้

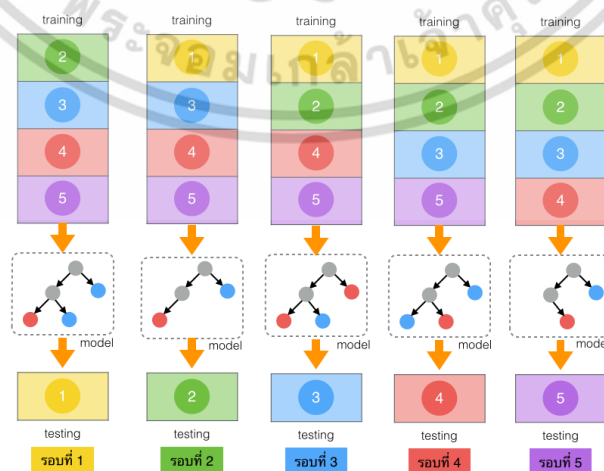
2.3 การวัดประสิทธิภาพของโมเดล

การวัดประสิทธิภาพของโครงข่ายประสาทเทียม เป็นการวัดประสิทธิภาพของโมเดล (Model) ที่ได้จากขั้นตอนการสร้างโครงข่าย ซึ่งสามารถแบ่งได้ ดังนี้

2.3.1 การทดสอบประสิทธิภาพของโมเดล

การทดสอบประสิทธิภาพของโมเดล จะใช้หลักการในการแบ่งข้อมูลเพื่อทำการทดสอบ ซึ่งสามารถแบ่งได้ดังนี้

1) K- fold Cross-validation Test วิธีนี้เป็นวิธีที่นิยมในการทำงานวิจัยเพื่อใช้ในการทดสอบประสิทธิภาพของโมเดล (Model) เนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัดประสิทธิภาพด้วยวิธี Cross-validation นี้จะทำการแบ่งข้อมูลออกเป็นหลายส่วน โดยจะแสดงด้วยค่า k ตัวอย่างเช่น 5-fold cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หรือ 10-fold cross-validation คือ การแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล (Model) การทำงานจะทำงานไปเช่นนี้จนครบจำนวนที่แบ่งไว้ เช่น การทดสอบด้วยวิธี 5-fold cross-validation แบ่งข้อมูลเทรนนิ่งดาต้า (Training data) ออกเป็น 5 ส่วนที่มีจำนวนเท่ากัน หลังจากนั้นทำการทดสอบประสิทธิภาพของโมเดล 5 ครั้ง ดังนี้



รูปที่ 2.4 ตัวอย่างกระบวนการของ K- fold Cross-validation Test

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกิจกรรมเพื่อการศึกษาเท่านั้น เมื่อผู้ดูแลเห็นเบาะแสหรือข้อผิดพลาดในการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รอบที่ 1 ใช้ข้อมูลส่วนที่ 2, 3, 4 และ 5 สร้างโมเดลและใช้โมเดลทำนายข้อมูลส่วนที่ 1 เพื่อทำการทดสอบ

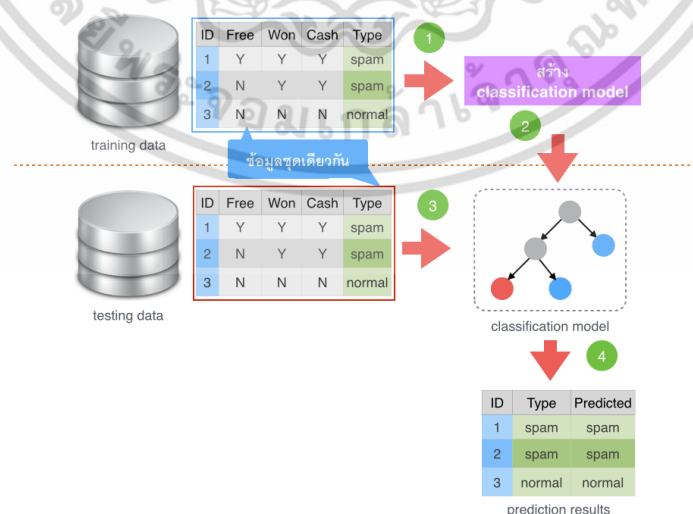
รอบที่ 2 ใช้ข้อมูลส่วนที่ 1, 3, 4 และ 5 สร้างโมเดลและใช้โมเดลทำนายข้อมูลส่วนที่ 2 เพื่อทำการทดสอบ

รอบที่ 3 ใช้ข้อมูลส่วนที่ 1, 2, 4 และ 5 สร้างโมเดลและใช้โมเดลทำนายข้อมูลส่วนที่ 3 เพื่อทำการทดสอบ

รอบที่ 4 ใช้ข้อมูลส่วนที่ 1, 2, 3 และ 5 สร้างโมเดลและใช้โมเดลทำนายข้อมูลส่วนที่ 4 เพื่อทำการทดสอบ

รอบที่ 5 ใช้ข้อมูลส่วนที่ 1, 2, 3 และ 4 สร้างโมเดลและใช้โมเดลทำนายข้อมูลส่วนที่ 5 เพื่อทำการทดสอบ

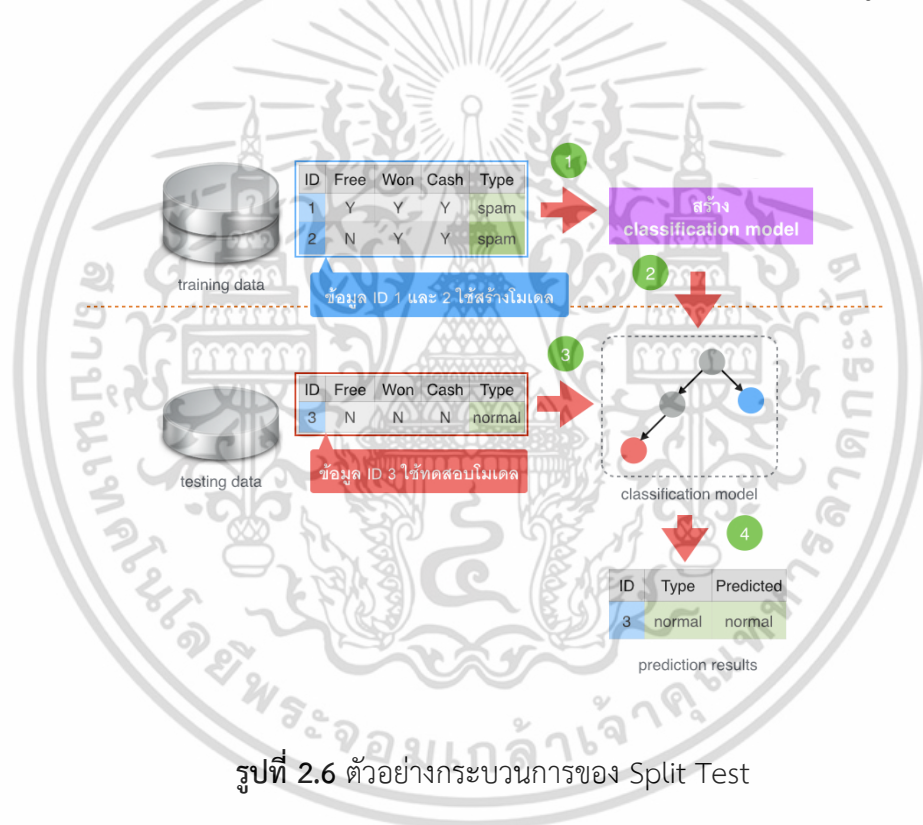
2) Self-Consistency Test หรือเรียกว่า Use Training Set นี้เป็นวิธีการที่ง่ายที่สุด นั่นคือ ข้อมูลที่ใช้ในการสร้างโมเดล (Model) และข้อมูลที่ใช้ในการทดสอบโมเดล (Model) เป็นข้อมูลชุดเดียวกัน กระบวนการนี้เริ่มจากสร้างโมเดล (Model) ด้วยข้อมูลเทรนนิ่งดาต้า (Training data) หลังจากนั้นนำโมเดล (Model) ที่สร้างได้มาทำนายข้อมูลเทรนนิ่งดาต้าชุดเดิม (Training data) ตัวอย่างเช่น นำข้อมูลเทรนนิ่งดาต้า (Training data) ในตาราง มาสร้างโมเดล (Model) และทดสอบโมเดล (Model) เป็นต้น การวัดประสิทธิภาพด้วยวิธีนี้จะให้ผลการวัดประสิทธิภาพที่มีค่าสูงมาก โดยอาจจะเข้าใกล้ 100% อันเนื่องมาจากเป็นข้อมูลชุดเดิมที่ระบบได้ทำการเรียนรู้มาแล้ว แต่ผลการวัดที่ได้ไม่เหมาะที่จะนำไปรายงานในงานวิจัยต่างๆ ซึ่งวิธีการนี้เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพเพื่อดูแนวโน้มของโมเดล (Model) ที่สร้างขึ้น ถ้าได้ผลการวัดที่น้อยแสดงว่าโมเดล (Model) ไม่เหมาะสมกับข้อมูล จึงไม่ควรจะนำไปทดสอบด้วยวิธีการแบ่งข้อมูลแบบต่างๆ



รูปที่ 2.5 ตัวอย่างกระบวนการของ Self-Consistency Test

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

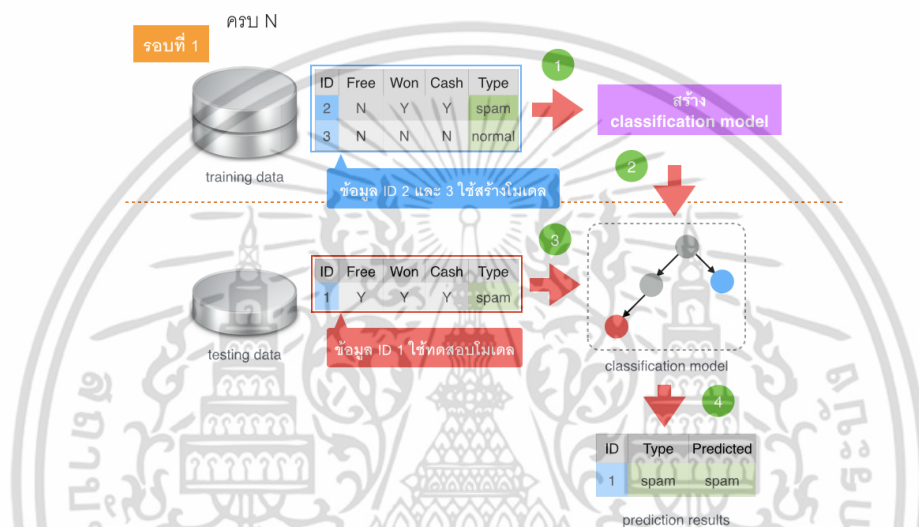
3) Split Test เป็นการแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วน เช่น 70% ต่อ 30% หรือ 80% ต่อ 20% โดยข้อมูลส่วนที่หนึ่ง 70% หรือ 80% จะใช้ในการสร้างโมเดล (Model) และข้อมูลส่วนที่สอง 30% หรือ 20% จะใช้ในการทดสอบประสิทธิภาพของโมเดล (Model) แบ่งตัวอย่างเช่น ข้อมูลเทรนนิ่งดาต้า (Training data) ในตารางออกเป็น 2 ตัวอย่างสำหรับการสร้างโมเดล (Model) และข้อมูล 1 ตัวอย่างใช้ในการทดสอบประสิทธิภาพของโมเดล (Model) เป็นต้น แต่การทดสอบแบบ Split Test นี้ทำการสุ่มข้อมูลเพียงครั้งเดียวซึ่งในบาง ครั้งถ้าการสุ่มข้อมูลที่ใช้ในการทดสอบที่มีลักษณะคล้ายกับข้อมูลที่ใช้สร้างโมเดล (Model) ทำให้ผลการวัดประสิทธิภาพได้ออกมาดี ในทางตรงข้ามถ้าการสุ่มข้อมูลที่ใช้ในการทดสอบที่มีลักษณะแตกต่างกับข้อมูลที่ใช้สร้างโมเดล (Model) มากทำให้ผลการวัดประสิทธิภาพได้ออกมาไม่ดี ดังนั้นการเลือกใช้วิธี Split Test นี้ควรทำการสุ่มหลายๆ ครั้ง แต่ข้อดีของวิธีการนี้คือใช้เวลาในการสร้างโมเดล (Model) น้อยซึ่งเหมาะกับชุดข้อมูลที่มีขนาดใหญ่



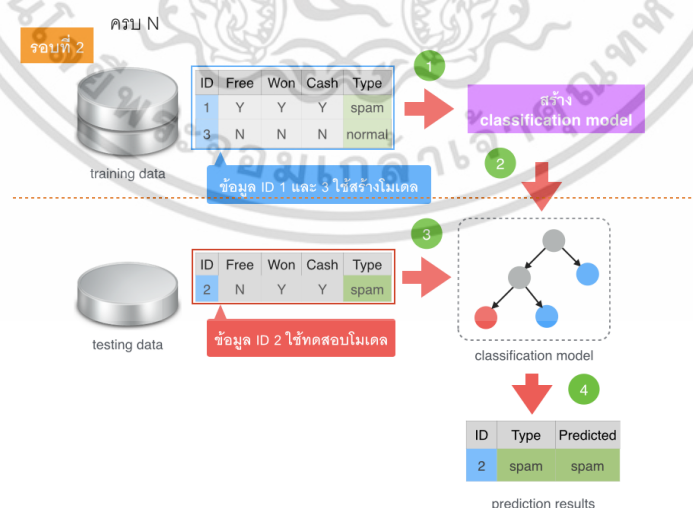
4) Leave one out Test เป็นการเลือกข้อมูลครั้งละหนึ่งตัวจากข้อมูลทั้งหมด โดยจะเลือกหนึ่งตัวต่อหนึ่งรอบ การทำงานก็จะทำงานตามจำนวนข้อมูลทั้งหมด N รอบ หากทำงาน N รอบ นั้นหมายความว่าโมเดล (Model) จะต้องได้รับการเรียนรู้ N ครั้ง โดยจะทำการดึงข้อมูลออกมาทั้งหมด N ตัว โดยที่ไม่ซ้ำกันเลย แสดงว่าข้อมูลแต่ละตัวมีโอกาสได้ถูกทำนาย ซึ่งจะส่งผลที่ดีเพราะว่าต้องการแสดงให้เห็นว่า การทดสอบแบบนี้ครอบคลุมจริง ๆ จะเขียนเป็นขั้นตอนได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 4.1) จากข้อมูลเริ่มต้น คือข้อมูล N ตัว
- 4.2) ทำการเลือกข้อมูลออกมา 1 ตัวจากข้อมูลทั้งหมด N ตัว
- 4.3) จำนวนข้อมูลทั้งหมดที่เหลืออยู่จากขั้นตอนที่ 2 คือ N-1 ตัว หรือที่เรียกว่า เทรนนิ่งดาต้า (Training Data)
- 4.4) สร้างโมเดล (Model) โดยใช้ เทรนนิ่งดาต้า (Training Data) แล้วทำการทดสอบด้วยข้อมูล 1 ตัว ที่ตัดออกมา
- 4.5) เริ่มขั้นตอนที่ 1 ใหม่ จนกว่าการทำงานจะครบ N รอบ



รูปที่ 2.7 ตัวอย่างกระบวนการรอบที่ 1 ของ Leave one out Test



รูปที่ 2.8 ตัวอย่างกระบวนการรอบที่ 2 ของ Leave one out Test

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2 การประเมินผลลัพธ์ของโมเดล (Confusion Matrix)

โมเดล (Model) ควรได้รับการประเมินเพื่อให้มั่นใจว่าโมเดล (Model) ที่สร้างนั้นมีประสิทธิภาพในการทำงานและผ่านเกณฑ์การทดสอบ เทคนิคการประเมินทั้งหมดนี้มาจากผลลัพธ์ของขั้นตอนการสร้างโมเดล (Model) โดยพิจารณาจากค่าความถูกต้องต่างๆจาก Confusion Matrix ดังนี้

		PREDICTED	
		(YES)	(NO)
ACTUAL	(YES)	True Positive (TP)	False Negative (FN)
	(NO)	False Positive (FP)	True Negative (TN)

รูปที่ 2.9 Confusion Matrix

อธิบายค่าต่างๆในตาราง Confusion Matrix ได้ดังนี้

True Positive (TP) หมายถึง ค่าของคลาสเป้าหมายคือ YES และตัวแบบทำนายว่า YES
 False Negative (FN) หมายถึง ค่าของคลาสเป้าหมายคือ YES และตัวแบบทำนายว่า NO
 True Negative (TN) หมายถึง ค่าของคลาสเป้าหมายคือ NO และตัวแบบทำนายว่า NO
 False Positive (FP) หมายถึง ค่าของคลาสเป้าหมายคือ NO และตัวแบบทำนายว่า YES

1) Correctly and Incorrectly Classified Instance คือ ค่าที่ทำนายถูกต้องและค่าที่ทำนายผิดพลาดในการจำแนกกลุ่ม

$$\text{ค่าที่ทำนายถูกต้อง (Correctly Classified Instance)} = TP + TN$$

$$\text{ค่าที่ทำนายผิดพลาด (Incorrectly Classified Instance)} = FP + FN$$

2) True Positive Rate (TPR) คือ ค่าจากการทำนายด้วยตัวแบบที่ตรงกับความเป็นจริงคำนวณจากจำนวนข้อมูลที่ทำนายถูกในคลาสนั้น/จำนวนข้อมูลทั้งหมดในคลาสนั้น

$$\text{TPR ของคลาสเป้าหมาย YES} = TP/(TP + FN)$$

$$\text{TPR ของคลาสเป้าหมาย NO} = TN/(FP + TN)$$

3) False Positive Rate (FPR) คือ ค่าของข้อมูลที่ไม่ได้อยู่ในคลาสนั้นแต่ตัวแบบทำนายว่าอยู่ในคลาสนั้น คำนวณจากจำนวนข้อมูลที่ทำนายผิดในคลาสนั้นที่สนใจ/จำนวนข้อมูลทั้งหมดในคลาสนั้น

$$\text{FPR ของคลาสเป้าหมาย YES} = FP/(FP + TN)$$

$$\text{FPR ของคลาสเป้าหมาย NO} = FN/(TP + FN)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) Precision คือค่าของตัวแบบที่ทำนายได้ถูกต้อง คำนวณจากจำนวนข้อมูลที่ทำนายถูกในคลาสนั้น/จำนวนข้อมูลทั้งหมดที่ทำนายให้ผลลัพธ์เดียวกันในคลาสนั้น

$$\text{Precision ของคลาสเป้าหมาย YES} = \text{TP}/(\text{TP} + \text{FP})$$

$$\text{Precision ของคลาสเป้าหมาย NO} = \text{TN}/(\text{TN} + \text{FN})$$

5) Recall คือค่าจากการทำนายด้วยตัวแบบที่ตรงกับความเป็นจริง คำนวณจากจำนวนข้อมูลที่ทำนายถูกในคลาสนั้น/จำนวนข้อมูลทั้งหมดในคลาสนั้น

$$\text{Recall ของคลาสเป้าหมาย YES} = \text{TP}/(\text{TP} + \text{FN})$$

$$\text{Recall ของคลาสเป้าหมาย NO} = \text{TN}/(\text{FP} + \text{TN})$$

2.4 การตัดคำ (Tokenization)

เนื่องจากความมุ่งหวังที่จะให้คอมพิวเตอร์สามารถประมวลผลภาษาไทยได้อย่างมีประสิทธิภาพ นั้น มีปัญหาเบื้องต้นคือ ลักษณะการเขียนภาษาไทย จะเขียนติดต่อกันเป็นสายอักขระ โดยไม่มีเครื่องหมายวรรคตอนแสดงการแบ่งคำดังเช่น ภาษาอังกฤษ ซึ่งเป็นอุปสรรคอย่างหนึ่งที่ต้องการศึกษาวิจัย และพัฒนา เพื่อให้คอมพิวเตอร์สามารถคำนวณเพื่อแบ่งสายอักขระไทยออกเป็นคำๆ ซึ่งจะส่งผลให้การทำงานของคอมพิวเตอร์ ในการค้นหาคำใดๆทำได้ถูกต้อง

ปัจจุบันนี้ มีการนำคอมพิวเตอร์เข้าไปใช้งานในด้านต่างๆ เป็นจำนวนมาก งานทางด้านจัดการกับภาษาไทย ได้แก่ การแปลงข้อความเป็นเสียง (Text - To - Speech) การตรวจตัวสะกด (Spell checking) การแปลภาษาด้วยเครื่องคอมพิวเตอร์ (Machine Translation) ฯลฯ ก็เป็นงานอีกด้านหนึ่งที่ได้นำคอมพิวเตอร์เข้าไปช่วย ในการทำงาน สิ่งแรกที่จะต้องจัดการคือ ต้องทำให้คอมพิวเตอร์รู้จักคำในภาษาไทยให้ได้ เช่น ถ้ามีข้อความว่า "ฉันนั่งตากลมที่หน้าบ้าน" จะต้องทำให้คอมพิวเตอร์รู้ว่า เป็น "ฉัน นั่ง ตาก ลม ที่ หน้า บ้าน" ไม่ใช่ "ฉัน นั่ง ตา ลม ที่ หน้า บ้าน" ลักษณะการเขียนภาษาไทยนั้น จะเขียนติดต่อกัน โดยไม่มีเครื่องหมายวรรคตอนคั่นระหว่างคำ ดังเช่นในภาษาอังกฤษ ซึ่งใช้ช่องว่าง (Space) คั่นระหว่างคำ จึงทำให้การโปรแกรมให้คอมพิวเตอร์ รู้จักคำในภาษาไทยนั้น มีความยุ่งยากกว่าในภาษาอังกฤษ

ดังนั้น จึงได้พัฒนาและค้นหาวิธีการต่างๆ ที่จะช่วยให้คอมพิวเตอร์รู้จักคำในภาษาไทย ซึ่งในปัจจุบัน มีอยู่หลายแนวคิดดังนี้

1) วิธีการตัดคำแบบยาวที่สุด (Longest Matching)

เมื่อต้องการโปรแกรมให้คอมพิวเตอร์รู้จักคำในภาษาไทย ก็ต้องทำการค้นหาคำโดยเริ่มจากตัวอักษรซ้ายสุดของข้อความนั้นไปยังตัวอักษรถัดไป จนกว่าจะพบว่าเป็นคำที่มีอยู่ในพจนานุกรม หลังจากนั้นก็ค้นหาคำถัดไป จนกว่าจะจบข้อความ ในกรณีที่พบว่าเป็นคำในพจนานุกรมจากจุดเริ่มต้นเดียวกัน จะเลือกคำที่ยาวที่สุด ตัวอย่างเช่น การแบ่งคำในประโยค "ฉันนั่งตากลมที่หน้าบ้าน" จะเริ่มจากตัวอักษร ฉ และคำแรกที่แบ่งได้คือ "ฉัน" หลังจากนั้น ก็ค้นหาตัวอักษรถัดไปและ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นำมาเปรียบเทียบคำในพจนานุกรม ก็จะแบ่งคำว่า "นั่ง" เป็นคำต่อไป ตัวอักษรถัดไป คือ ต จากตัวอักษรนี้ จะได้คำว่า "ตา" กับคำว่า "ตาก" แนวคิดนี้ให้เลือกคำที่ยาวที่สุด ที่ค้นพบ จึงเลือกคำว่า "ตาก" หลังจากนั้น ก็จะค้นหา และเปรียบเทียบต่อไป ซึ่งจะได้ผลลัพธ์ออกมาคือ "ฉันทั่ง ตาก ลม ที่ หน้า บ้าน"

2) วิธีการตัดคำแบบสอดคล้องมากที่สุด (Maximal Matching)

วิธีการตัดคำแบบนี้ เป็นการหาวิธีในการตัดคำที่สามารถจะเป็นไปได้ทั้งหมด เช่น เมื่อมีข้อความว่า "ไปหามเหสี" ก็จะตัดคำได้ 2 แบบ คือ 1. ไป หาม เห สี และ 2. ไป หาม เห สี

วิธีการนี้จะให้เลือกข้อความที่แบ่งแล้ว มีจำนวนค่าน้อยที่สุดคือ แบบที่ 2 ซึ่งมีจำนวนคำที่ตัดได้ 3 คำ ในขณะที่แบบที่ 1 มี 4 คำ ส่วนในกรณีที่มีจำนวนคำที่เท่ากัน ก็จะใช้วิธีการตัดคำแบบยาวที่สุด (Longest Matching) เข้ามาช่วย เช่น ข้อความว่า "ฉันทั่งตากลมที่หน้าบ้าน" ซึ่งจะสามารถแบ่งคำได้ 2 แบบ คือ

1. ฉันทั่ง ตาก ลม ที่ หน้า บ้าน
2. ฉันทั่ง ตา ลม ที่ หน้า บ้าน

ทั้ง 2 แบบมีจำนวนคำที่เท่ากัน จึงเลือกแบบที่ 1 โดยเปรียบเทียบจากคำที่ต่างกันที่ตัดได้นั้น (ตา / ตาก) จะเห็นได้ว่า ตากมีตัวอักษรมากกว่า

3) วิธีการตัดคำแบบคำนวณเชิงสถิติเพื่อหาความเป็นไปได้ (probabilistic Model)

วิธีการนี้นำเอาค่าสถิติการเกิดของคำและลำดับของหน้าที่ของคำ (part of speech) เข้ามาช่วยในการคำนวณหาความน่าจะเป็น เพื่อที่จะใช้เลือกแบบที่มีโอกาสเกิดมากที่สุด วิธีการนี้สามารถจะตัดคำได้ดีกว่า 2 แบบแรก แต่ข้อจำกัดของวิธีการนี้คือ จะต้องมีความรู้ข้อมูลที่มีการตัดคำที่ถูกต้อง และกำหนดหน้าที่ของคำให้ เพื่อที่จะได้นำไปใช้ในการสร้างสถิติ

4) วิธีการตัดคำแบบคุณลักษณะ (Feature - Based Approach)

วิธีการนี้จะพิจารณาจากบริบท (context words) และการเกิดร่วมกันของคำ หรือหน้าที่ของคำ (collocation) เข้ามาช่วยในการตัดคำตัวอย่างเช่น

"ตากลม" ถ้าพบคำว่า "แป้ว" ในบริบทก็จะสามารถตัดคำได้ว่า "ตา" "ลม"

"มากกว่า" ถ้าในบริบทที่ตามมาเป็นตัวเลขก็สามารถตัดคำได้ว่า "มา" "กว่า"

วิธีการนี้จำเป็นที่จะต้องมีความรู้ข้อมูลเป็นจำนวนมาก และจะต้องมีการเรียนรู้การสร้างคำในบริบท หรือการเกิดร่วมกันของคำแต่ละคำ เพื่อให้มีข้อมูลที่จะนำมาใช้ในการตัดคำ

2.5 ภาษาที่ใช้งาน (Programmer Language)

2.5.1 ภาษา Python

Python เป็นภาษาระดับสูงภาษาหนึ่ง ที่มีความสามารถสูงถูกสร้างขึ้นในปี 1989 โดย Guido van Rossum ซึ่งถูกพัฒนาขึ้นมาโดยไม่ยึดติดกับแพลตฟอร์ม กล่าวคือสามารถรันภาษา Python ได้ทั้งบนระบบ Unix, Linux, Windows NT, Windows 2000, Windows XP หรือแม้แต่ระบบ FreeBSD อีกอย่างหนึ่งภาษาตัว นี้เป็นภาษาลักษณะ Open Source เหมือนอย่าง PHP แต่ด้วยข้อดีหลายประการของภาษา Python ทำให้มีผู้นิยมใช้มากขึ้นเรื่อยๆ ซึ่งพอสรุปข้อดีของภาษา Python ได้ดังนี้

- 1) ง่ายต่อการเรียนรู้ โดยภาษา Python มีโครงสร้างของภาษาไม่ซับซ้อนเข้าใจง่าย ซึ่งโครงสร้างภาษา Python จะคล้ายกับภาษา C มาก เพราะภาษา Python สร้างขึ้นมาโดยใช้ภาษา C ทำให้ผู้ที่คุ้นเคยภาษา C อยู่แล้วใช้งานภาษา Python ได้ไม่ยาก นอกจากนี้โดยตัวภาษาเองมีความยืดหยุ่นสูงทำให้การจัดการกับงานด้านข้อความ และ Text File ได้เป็นอย่างดี
- 2) ไม่ต้องเสียค่าใช้จ่ายใดๆ ทั้งสิ้น เพราะตัวแปลภาษา Python อยู่ภายใต้ลิขสิทธิ์ GNU
- 3) ใช้ได้หลายแพลตฟอร์ม ในช่วงแรกภาษา Python ถูกออกแบบใช้งานกับระบบ Unix อยู่ก็จริง แต่ในปัจจุบันได้มีการพัฒนาตัวแปลภาษา Python ให้สามารถใช้กับระบบปฏิบัติการอื่นๆ อาทิ เช่น Linux, Windows 95/98/ME, Windows NT, Windows 2000, OS/2
- 4) ภาษา Python ถูกสร้างขึ้นโดยได้รวบรวมเอาส่วนดีของภาษาต่างๆ เข้ามาไว้ด้วยกัน อาทิ เช่น ภาษา C, C++, Java, Perl
- 5) ภาษา Python เป็นภาษาประเภท Server side Script คือการทำงานของภาษา Python จะทำงานด้านฝั่ง Server แล้วส่งผลลัพธ์กลับมายัง Client ทำให้มีความปลอดภัยสูง
- 6) ใช้พัฒนา Web Service โดยที่ภาษา Python สามารถนำมาพัฒนาเว็บเซอร์วิส รวมทั้งใช้บริหารการสร้างเว็บไซต์สำเร็จรูปที่เรียกว่า Content Management Framework (CMF) ตัวอย่าง CMF ที่มีชื่อเสียงมากและเบื้องหลังทำงานด้วย python คือ Plone

2.5.2 ภาษา PHP

PHP เป็นภาษาคอมพิวเตอร์ประเภทโอเพนซอร์สสำหรับพัฒนาเว็บเพจเมื่อเครื่องบริการได้รับคำร้องจากผู้ใช้ก็จะส่งให้กับตัวแปล ภาษาทำหน้าที่ประมวลผลและส่งข้อมูลกลับไปยังเครื่องของผู้ใช้ที่ร้องขอในรูปแบบ HTML ภาพ หรือแฟ้ม digital อื่นๆ ลักษณะของภาษามีคำสั่งมาจาก ภาษาซี ภาษาจาวาและ ภาษาเพิร์ล ซึ่ง ภาษา PHP นั้นง่ายต่อการเรียนรู้ ซึ่งเป้าหมายหลักของภาษา นี้ คือให้นักพัฒนาเว็บไซต์สามารถเขียนเว็บเพจ ที่มีความตอบโต้ได้อย่างรวดเร็ว PHP เป็นภาษา จำพวกภาษาสคริปต์คำสั่งต่างๆจะเก็บอยู่ในไฟล์ที่เรียกว่า สคริปต์ และเวลาใช้งานต้องอาศัยตัวแปล ชุดคำสั่ง ตัวอย่างของภาษาสคริปต์เช่น JavaScript, Perl เป็นต้น ลักษณะของ PHP ที่แตกต่างจาก

ภาษาสคริปต์แบบอื่น ๆ คือ PHP ได้รับการพัฒนาและออกแบบมาเพื่อใช้งานในการสร้างเอกสารแบบ HTML โดย สามารถ สอดแทรกหรือแก้ไขเนื้อหาได้โดยอัตโนมัติ) Visual Studio 2008

ลักษณะเด่นของ PHP สามารถสรุปได้ดังนี้

- 1) ใช้งานได้ฟรี
- 2) PHP เป็นโปรแกรมวิ่งข้าง Sever ดังนั้นขีดความสามารถไม่จำกัด
- 3) Conlatfun นั่นคือPHP วิ่งบนเครื่อง UNIX,Linux,Windows ได้หมด
- 4) เรียนรู้ง่าย เนื่องจาก PHP ผังเข้าไปใน HTML และใช้โครงสร้างและไวยากรณ์

ภาษาต่างๆ

5) เร็วและมีประสิทธิภาพ โดยเฉพาะเมื่อใช้กับ Apach Xerve เพราะไม่ต้องใช้โปรแกรมจากภายนอก

- 6) ใช้ร่วมกับ XML ได้ทันที
- 7) ใช้กับระบบแฟ้มข้อมูลได้
- 8) ใช้กับข้อมูลตัวอักษรได้อย่างมีประสิทธิภาพ
- 9) ใช้กับโครงสร้างข้อมูล แบบ Scalar,Array,Associative array
- 10) ใช้กับการประมวลผลภาพได้

2.6 ซอร์ฟแวร์ที่ใช้งาน (Software)

2.6.1 MongoDB

MongoDB เป็น open-source document database โดยเป็นฐานข้อมูลแบบ NoSQL คือไม่มีความสัมพันธ์ (Relation) ของตารางแบบ SQL ทั่วไป แต่จะเก็บข้อมูลเป็นแบบ JSON (JavaScript Object Notation) แทน การบันทึกข้อมูลทุกๆ record ใน MongoDB เราจะเรียกว่า Document ซึ่งจะเก็บค่าเป็น key และ value จะเหมือนกับ JSON

```
{
  "_id" : ObjectId("57d244a022af641548e6cc88"),
  "primary_source" : "Twitter",
  "main_type" : "สภาพอากาศ",
  "media_path" : null,
  "event_begin" : "2014-02-15 06:33:43",
  "description" : "RT @onjndii รถชนแกวพญาไท ผ่นตกนานมาก"
}
```

รูปที่ 2.10 ตัวอย่างค่าของ MongoDB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การเก็บข้อมูล Document ใน MongoDB จะถูกเก็บไว้ใน Collections (เปรียบเทียบกับ Table ใน Relational Database ทั่วไป) แต่แตกต่างกันที่ collection ไม่จำเป็นที่จะต้องมีการมี schema เหมือนกันก็สามารถบันทึกข้อมูลได้ ใน MongoDB ข้อมูล document ที่เก็บใน collection จะมีคีย์ `_id` ทำหน้าที่เปรียบเสมือน primary key อยู่ด้วย และ MongoDB ไม่รองรับการ join หรือ SQL

Schemaless คือการไม่ต้องกำหนดโครงสร้างใดๆให้มันเหมือน SQL ปกติทั่วไป เช่น collection Tweet มีเก็บข้อมูลดังรูป 2.10 ต่อมาเราสามารถเพิ่มการเก็บ location เข้ามาได้เลย ดังรูป

```
{
  "_id" : ObjectId("57d244a022af641548e6cc88"),
  "primary_source" : "Twitter",
  "main_type" : "สภาพอากาศ",
  "media_path" : null,
  "event_begin" : "2014-02-15 06:33:43",
  "description" : "RT @onjndii รกชนแถวพญาไท ฝนตกนานมาก",
  "location" : [100.546488, 13.7854165]
}
```

รูปที่ 2.11 ตัวอย่างการเพิ่มค่าของ MongoDB

ภาพเปรียบเทียบ MySQL และ MongoDB

MySQL	MongoDB
Table	Collection
Row	Document
Column	Field
Joins	Embedded documents, linking

รูปที่ 2.12 ตารางเปรียบเทียบ MySQL และ MongoDB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.2 Tableau Software

Tableau เป็นซอฟต์แวร์ (Software) ที่สามารถนำข้อมูลจำนวนมากที่มีหลากหลายในองค์กรมาทำการวิเคราะห์เพื่อเป็นข้อมูลเชิงธุรกิจช่วยให้มีข้อมูลเชิงลึกเพิ่มความสามารถในการตัดสินใจอย่างรวดเร็วและชาญฉลาด โดยรองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และสามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน ไม่ว่าจะเป็งานในส่วนรับผิดชอบของผู้บริหารระดับสูง (Top Management) นักวางแผนกลยุทธ์ (Strategic Planning) นักขายมีอาชีพ (Sales Manager) และผู้ที่อยู่ในกลุ่มธุรกิจอุตสาหกรรมกลุ่มธุรกิจนำเข้าและส่งออก หน่วยงานรัฐวิสาหกิจ หรือองค์ภาครัฐ รวมถึง สถาบันการศึกษาที่สามารถใช้ซอฟต์แวร์ทาโบลิว (Tableau Software) ได้

Tableau มีทั้งแบบเดสทอป (Tableau Desktop) และเซิร์ฟเวอร์ (Tableau Server) ใช้งานได้ง่าย มีระบบรักษาความปลอดภัยของข้อมูลสูง รองรับเหมืองข้อมูลไม่จำกัด (Data scalability) และประเมินผลได้รวดเร็ว (Downloads in minutes)

คุณสมบัติและจุดเด่นของซอฟต์แวร์ สามารถสรุปได้ดังนี้

- 1) รูปแบบการใช้งานที่ง่ายเพียงคลิกเมาส์ก็สามารถที่จะเปลี่ยนแปลงรายงานได้ตามความต้องการ
- 2) รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และสามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL Server, Microsoft Power pivot เป็นต้น
- 3) รูปแบบการนำเสนอรายงานที่สวยงาม เข้าใจง่าย และง่ายในการนำเสนอต่อผู้บริหาร
- 4) การเข้าถึงหรือการใช้งาน สามารถทำได้โดยง่าย ตั้งแต่ผู้บริหารระดับสูง ถึง พนักงานระดับปฏิบัติการ ทั้งนี้ทั้งนั้น สามารถเพิ่มทักษะและแนวคิดให้กับพนักงานระดับปฏิบัติการ ให้มีแนวคิดเชิงสถิติและการประยุกต์ใช้งานมากขึ้น
- 5) สิ่งที่สำคัญ คือ Engine ของ Tableau Software นั้น พัฒนาจาก VizQL Technology เป็นเทคโนโลยีที่ช่วยให้เห็นภาพของข้อมูลขนาดโตก็ได้เพียงการลากและวาง (Drag and Drop) โดยใช้ข้อมูลจากฐานข้อมูลตอบสนองในรูปแบบของกราฟิก

บทที่ 3

วิธีการดำเนินงานวิจัย

ในการจัดทำระบบจัดการและวิเคราะห์ข้อมูลจากสังคมออนไลน์เพื่อจำแนกประเภทปัญหาเมือง เป็นการศึกษาพัฒนาระบบสารสนเทศเพื่อจัดการและวิเคราะห์ข้อมูลจากสังคมออนไลน์ โดยที่ผู้จัดทำต้องการมุ่งเน้นในเรื่องประสิทธิภาพในการทำงาน และความสะดวกในการใช้งาน ซึ่งผู้จัดทำมีวิธีการดำเนินงานดังนี้

3.1 การวางแผนและการเตรียมการ

การดำเนินการพัฒนาระบบจัดการและวิเคราะห์ข้อมูลจากสังคมออนไลน์เพื่อจำแนกประเภทปัญหาเมืองนี้ ได้มีการวางแผนและการเตรียมการเพื่อจัดการพัฒนาโดยแบ่งเป็นขั้นตอนดังนี้

3.1.1 ศึกษาความเป็นไปได้และเก็บรวบรวมข้อมูล

ศึกษาความเป็นไปได้และเก็บรวบรวมข้อมูลจากสังคมออนไลน์ที่เกี่ยวกับปัญหาเมืองและปัญหาอื่น ๆ ที่นำมาทดสอบ รวมไปถึงศึกษาเครื่องมือที่ใช้ในการพัฒนาระบบ

1) ศึกษาการเก็บข้อมูลจากสังคมออนไลน์ และศึกษาการวิเคราะห์รูปแบบของประโยคเพื่อให้เข้าใจรูปแบบข้อความว่าควรมีลักษณะอย่างไร ในข้อความสามารถระบุสิ่งใดได้บ้าง

2) ศึกษาข้อมูลเกี่ยวกับภาษา PHP ซึ่งเป็นภาษาคอมพิวเตอร์ประเภทโอเพนทอร์ซอร์สำหรับพัฒนาเว็บเพจเมื่อเครื่องบริการได้รับคำสั่งจากผู้ใช้ก็จะส่งให้กับตัวแปล ภาษาทำหน้าที่ประมวลผลและส่งข้อมูลกลับไปยังเครื่องของผู้ใช้ที่ร้องขอในรูปแบบ HTML ภาพ หรือแฟ้ม digital อื่นๆ ที่นำภาษา Php มาใช้นั้นเพราะเล็งเห็นว่า สามารถใช้ร่วมกับภาษาอื่น ๆ ได้หลายภาษา เช่น JavaScript Bootstrap CSS HTML MVC เป็นต้น นอกจากนี้มี รูปแบบและการใช้งานออบเจ็คที่ง่ายขึ้น

3) ศึกษาโปรแกรม JetBrains PhpStorm 8.0.4 เป็นชุดโปรแกรมที่นำไปใช้เขียนโปรแกรมคอมพิวเตอร์ด้วยภาษาต่าง ๆ เช่น ภาษา Php, Drupal เป็นต้น ซึ่งในที่นี้ผู้จัดทำต้องการใช้ภาษา Php ในการพัฒนาระบบ

3.1.2 วิเคราะห์ความต้องการของระบบ

วิเคราะห์ความต้องการของระบบ (Requirement Collection and Analysis) เพื่อจำแนกถึงปัญหาและความต้องการออกเป็นกลุ่มๆ โดยกำหนดขอบเขตของระบบที่จะพัฒนาขึ้น

1) ศึกษาขั้นตอนการจำแนกประเภทของข้อมูลจากสังคมออนไลน์ โดยศึกษาเครื่องมือที่ใช้มีหลักการทำงานอย่างไร มีประสิทธิภาพในการนำประเภทข้อมูลที่มีโครงสร้างของข้อมูลซับซ้อนในระดับใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) ศึกษาขั้นตอนการวัดประสิทธิภาพในการจำแนกประเภทของข้อมูลจากสังคมออนไลน์ เพื่อประเมินความถูกต้องของขั้นตอนการทำงานและเครื่องมือที่ใช้ โดยศึกษาหลักในการวัดประสิทธิภาพของความถูกต้องในการนำประเภทข้อมูล ข้อผิดพลาดรวมไปถึงปัญหาที่ส่งผลให้ประสิทธิภาพลดลง

3.1.3 ออกแบบระบบ

ออกแบบระบบ (Design) เป็นขั้นตอนที่นำเอาปัญหาและความต้องการต่าง ๆ ที่จำแนกไว้ในขั้นต้น เพื่อใช้ในการออกแบบระบบงาน แบ่งออกเป็นขั้นตอนย่อย ๆ ไว้ ดังนี้

- 1) ทำการออกแบบโครงสร้างในการเก็บข้อมูลจากสื่อสังคมออนไลน์
- 2) ออกแบบแผนภาพกระแสนงานแสดงขั้นตอนลำดับการทำงานของกรรวบรวมและจำแนกประเภทของข้อมูลจากสังคมออนไลน์ (Work Flow Diagram)
- 3) ออกแบบแผนภาพ Use Case Diagram
- 4) ออกแบบแผนภาพแสดงกิจกรรมการทำงานสำหรับผู้ใช้งานระบบ (Activity Diagram)
- 5) ออกแบบแผนภาพแสดงกิจกรรมการทำงานสำหรับผู้ใช้งานระบบ (Sequence Diagram)

3.2 ขั้นตอนการวิเคราะห์ระบบ

3.2.1 วิเคราะห์ระบบงานเดิม

เป็นระบบการตรวจจับข้อมูลจากสื่อสังคมออนไลน์ (Social Media) ซึ่งสื่อสังคมออนไลน์ที่ระบบเดิมใช้ คือ ทวิตเตอร์ (Twitter) โดยจะเน้นการสกัดข้อมูลปัญหาเมืองในด้านจราจรเท่านั้น เพื่อรวบรวมและจำแนกข้อมูลปัญหาเมืองมาใช้เป็นฐานข้อมูลในการพัฒนาต่อยอดสร้างระบบอื่น ๆ รวมถึงเป็นฐานข้อมูลจากสื่อสังคมออนไลน์เพื่อเผยแพร่ให้กับนักวิจัยหรือนักพัฒนาที่สนใจผ่านทางบริการ API โดยระบบเดิมนี้มีข้อมูลไม่ครอบคลุมปัญหาเมืองด้านอื่นๆ เช่น ปัญหาด้านอัคคีภัย ปัญหาด้านการระบายน้ำ ปัญหาด้านความปลอดภัย ปัญหาด้านทางเท้า เป็นต้น จึงทำให้มีฐานข้อมูลปัญหาเมืองจากสื่อสังคมออนไลน์เพียงด้านเดียวเท่านั้น

3.2.2 วิเคราะห์ระบบงานใหม่

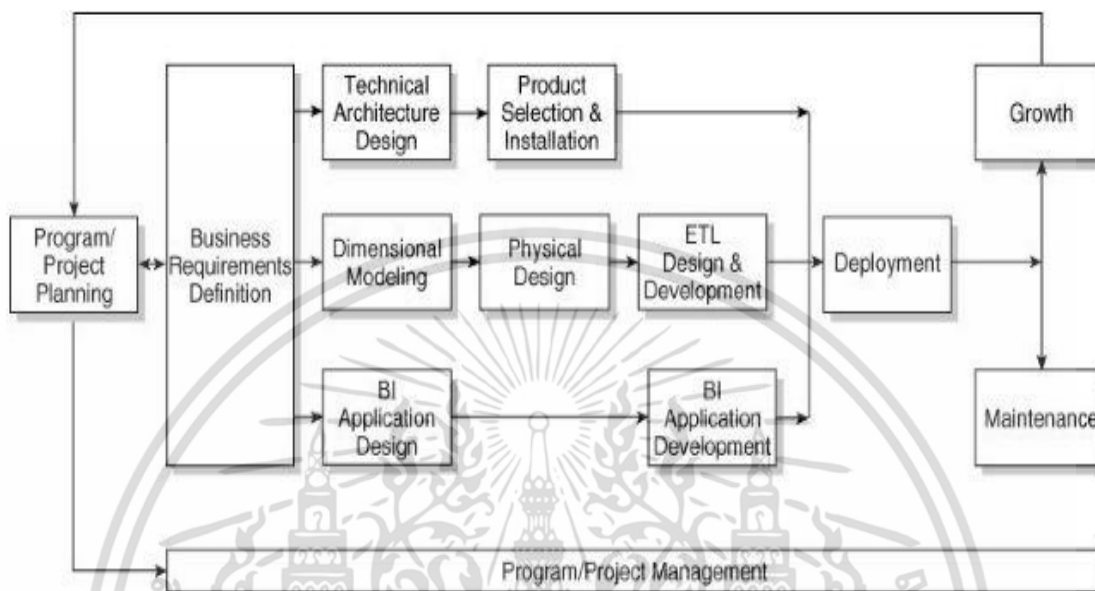
เป็นระบบการรวบรวมและจำแนกข้อมูลปัญหาเมืองจากสื่อสังคมออนไลน์ (Social Media) โดยจะขยายขอบเขตการสกัดข้อมูลปัญหาเมืองเพิ่มเติมจากระบบเดิมที่มีเพียงปัญหาด้านจราจรเท่านั้น เพื่อให้ข้อมูลปัญหาเมืองที่ได้จากสื่อสังคมออนไลน์ครอบคลุมปัญหาเมืองในปัจจุบันมากขึ้น ได้แก่ ปัญหาด้านอัคคีภัย ปัญหาด้านการระบายน้ำ ปัญหาด้านความปลอดภัย ปัญหาด้านทางเท้า ปัญหาด้านอุบัติเหตุ และปัญหาด้านมลภาวะ ข้อมูลเหล่านี้จะเป็นฐานข้อมูลจากสื่อสังคมออนไลน์ เพื่อเผยแพร่ให้กับนักวิจัยหรือนักพัฒนาที่สนใจผ่านทางบริการ API รวมถึงเป็นฐานข้อมูลในการพัฒนาระบบที่เกี่ยวข้องกับปัญหาของเมือง พร้อมทั้งพัฒนาเว็บไซต์ให้บริการจัดการและวิเคราะห์ข้อมูลในการจำแนกประเภทของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 ขั้นตอนการออกแบบระบบ

3.3.1 วงจรการพัฒนาาระบบ

ขั้นตอนในการพัฒนาระบบ ผู้พัฒนาใช้ Kimball 2008 ในการพัฒนา ดังรูปที่ 3.1



รูปที่ 3.1 Kimball 2008 Lifecycle

1) Program/Project Planning and Management

เป็นขั้นตอนการกำหนดขอบเขตงาน รวมถึงระยะเวลาที่ใช้ในการดำเนินงาน

2) Business Requirements

เป็นขั้นตอนการกำหนดความต้องการของระบบและวิเคราะห์ปัญหาที่อาจเกิดขึ้น กำหนดขอบเขตของแหล่งที่มาข้อมูลและขอบเขตข้อมูลปัญหาเบื้องต้นในด้านต่างๆ รวมถึงการจัดลำดับความสำคัญและพิจารณาปัจจัยที่ส่งผลกระทบต่อการทำงาน

3) Technology Track

เป็นขั้นตอนการเลือกเครื่องมือที่ใช้ในการดำเนินงาน ติดตั้งใช้งานเพื่อตอบสนองความต้องการตามระบบที่ได้กำหนดไว้ ซึ่งเครื่องมือที่เลือกใช้ในการจำแนกประเภทของข้อมูลคือ ฟาสต์เท็กซ์ (Fasttext) และเครื่องมืออื่นๆที่ระบุไว้ในบทที่ 2

4) Data Track

เป็นขั้นตอนการติดตามข้อมูล การออกแบบรูปแบบมิติข้อมูลเป้าหมาย รวมถึงการนำข้อมูลที่ได้นั้นมาทำ ETL โดยมีกระบวนการต่างๆ เช่น การกรองข้อมูล (Data Cleansing) การแยกคำ (Tokenization) การจัดรูปแบบข้อมูล (Data Transformation) เพื่อใช้ในการวิเคราะห์และจำแนกประเภทข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) Business Intelligence Track

เป็นขั้นตอนการติดตามผลลัพธ์ในการทดลอง โดยสามารถจัดทำในรูปแบบการแสดงผล (Dashboard) เพื่อวิเคราะห์ผลลัพธ์และวัดประสิทธิภาพในการทดลอง

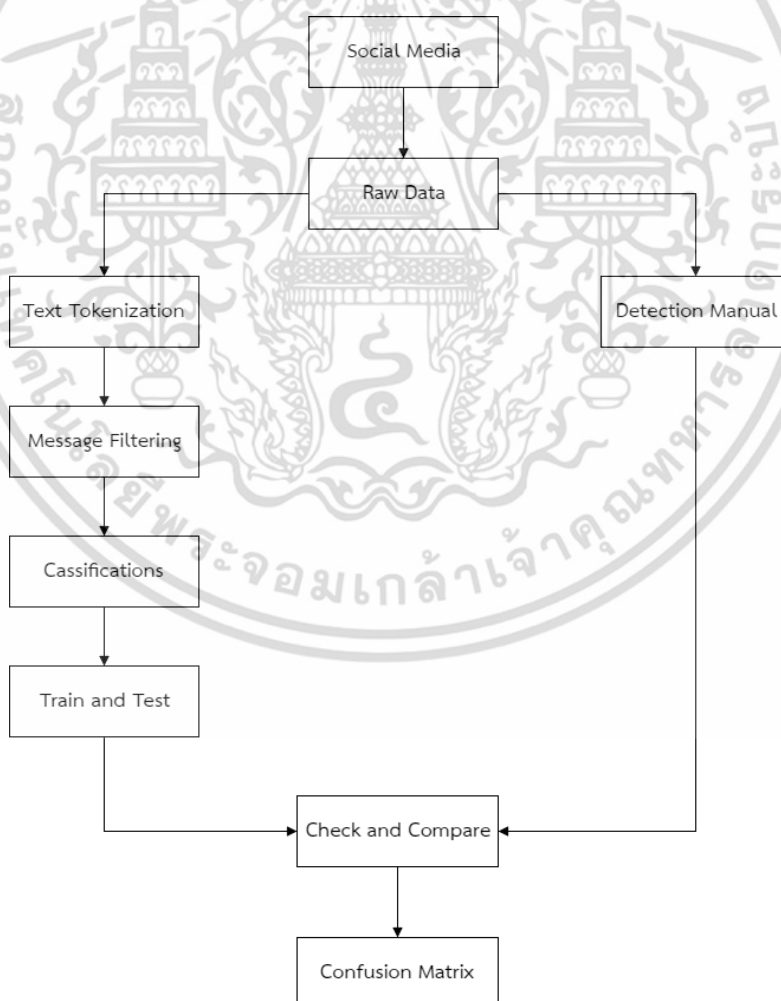
6) Deployment, Maintenance, and Growth

เป็นขั้นตอนการบำรุงรักษาและพัฒนาระบบให้มีประสิทธิภาพ สามารถใช้งานได้ตลอดเวลา

3.3.2 เครื่องมือที่ใช้ในการออกแบบ

1) Work Flow Diagram

แผนภาพกระแสนงาน เป็นการออกแบบขั้นตอนลำดับการทำงานของกรรวบรวมและจำแนกประเภทของข้อมูลจากสังคมออนไลน์และการทำงานของเว็บเซอร์วิสที่ให้บริการจำแนกประเภทของข้อมูลในส่วนของหน้าหลัก หน้าใช้งานโมเดล หน้าทดสอบโมเดล แสดงดัง รูปที่ 3.2 และ 3.3 ตามลำดับ



รูปที่ 3.2 Work Flow Diagram การวิเคราะห์และจำแนกประเภทของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

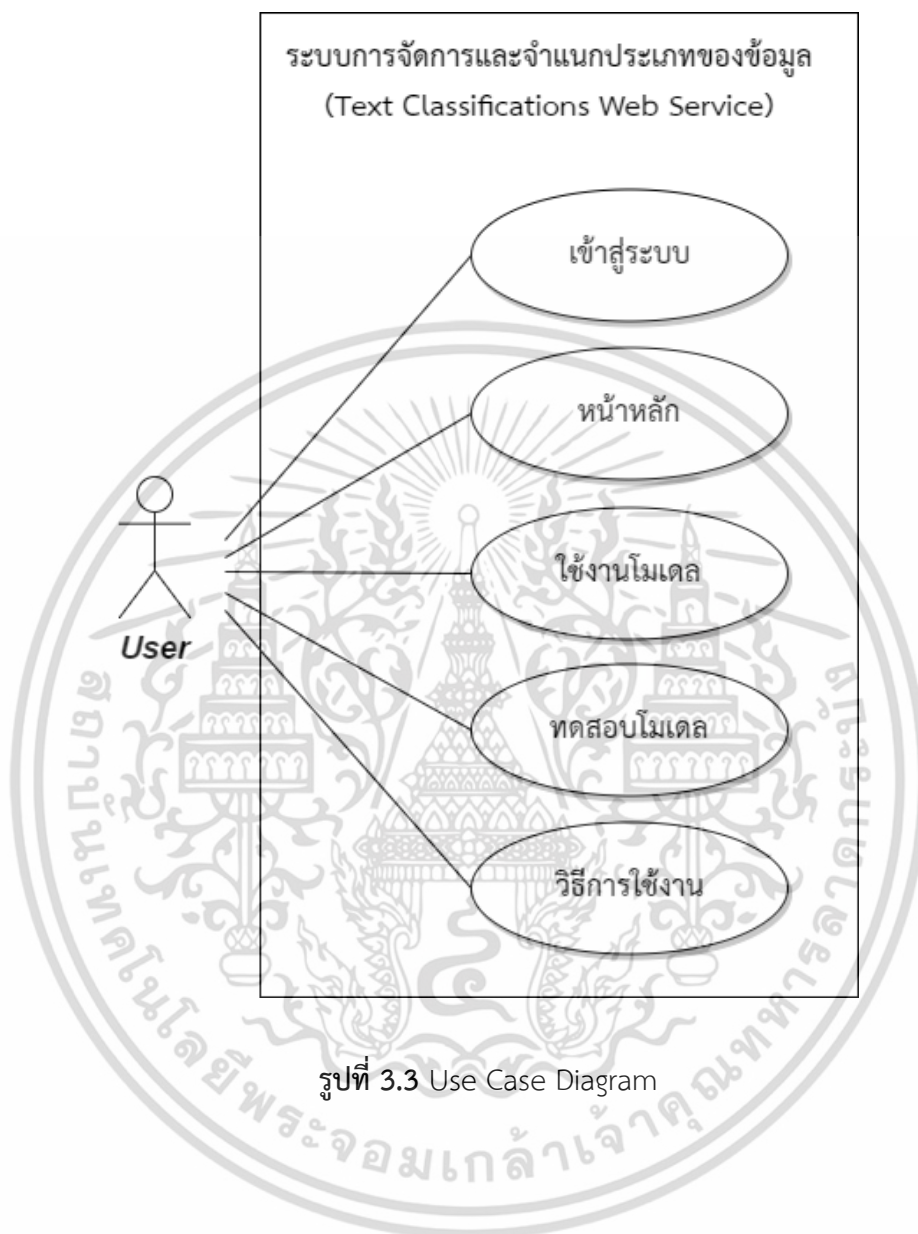
ตารางที่ 3.1 อธิบายขั้นตอนการทำงานของกระบวนการรวบรวมและจำแนกข้อมูล

Social Media :	สื่อสังคมออนไลน์ในที่นี้คือ Twitter และ Facebook
Raw Data :	ข้อมูลดิบที่ได้มาจากข้อความใน Twitter และ Facebook โดยข้อมูลนี้จะเป็นข้อมูลดิบที่เกี่ยวข้องกับปัญหาเมือง
Detection Manual :	การทำผลเฉลย จัดทำจากข้อมูลดิบที่ได้มาแล้วจำแนกข้อความนั้น ออกเป็นปัญหาเมืองประเภทต่างๆ โดยเฉลยทั้งหมดจะทำโดยมนุษย์เป็นพื้นฐาน
Text Tokenization :	ระบบตัดคำ ข้อมูลดิบที่ได้มาจะต้องผ่านระบบตัดคำโดยแยกคำที่มีความหมายออกจากประโยคเพื่อระบุประเภทของคำต่างๆ
Message Filtering :	เป็นการกรอกข้อมูลที่ไม่เกี่ยวข้องหรือในประโยคมีคำหยาบคายประกอบอยู่ออกไป
Classifications :	สร้างโมเดลที่ใช้เป็นต้นแบบในการทดสอบการจำแนกประเภทของปัญหาเมือง
Train and Test :	เป็นการแบ่งข้อมูลออกเป็นส่วนๆ เพื่อในข้อมูลแต่ละส่วนมาทำการเรียนรู้และทดสอบการจำแนกประเภทของข้อมูล
Check and Compare :	การตรวจสอบผลลัพธ์ โดยตรวจสอบจากผลเฉลยว่าตรงกับผลลัพธ์ที่ได้จากระบบจำแนกประเภทของข้อมูลหรือไม่
Confusion Matrix :	การทำตารางประเมินคุณภาพ เพื่อวัดประสิทธิภาพในการแนกประเภทของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) Use Case Diagram

แผนภาพแสดงขั้นตอนการระหว่างผู้ใช้งาน แสดงดังรูปที่ 3.3



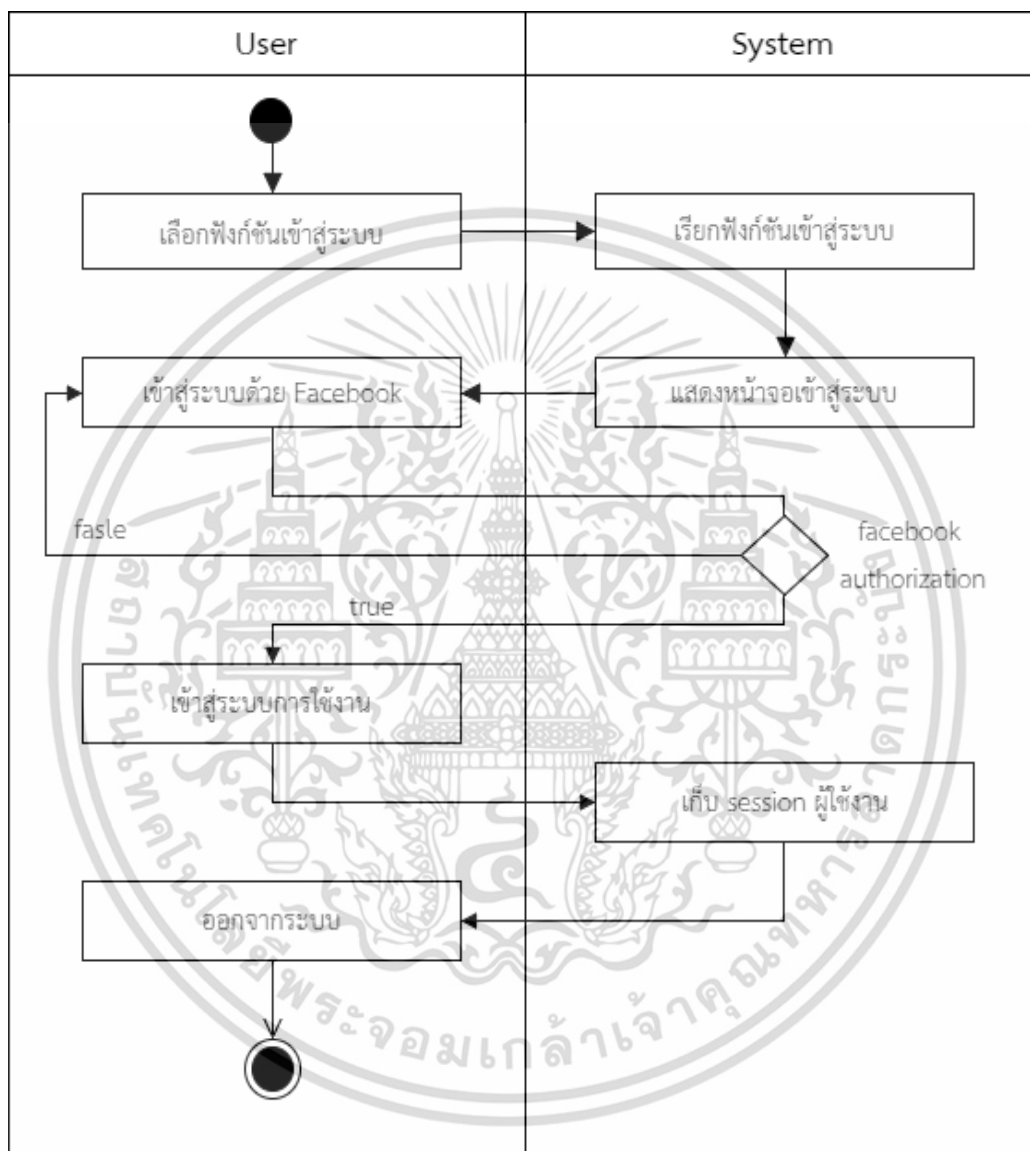
รูปที่ 3.3 Use Case Diagram

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) Activity Diagram

แผนภาพกิจกรรมการทำงานสำหรับผู้ใช้งานระบบจัดการและวิเคราะห์ข้อมูลจากสังคมออนไลน์เพื่อจำแนกประเภทข้อมูล มีรายละเอียดดังนี้

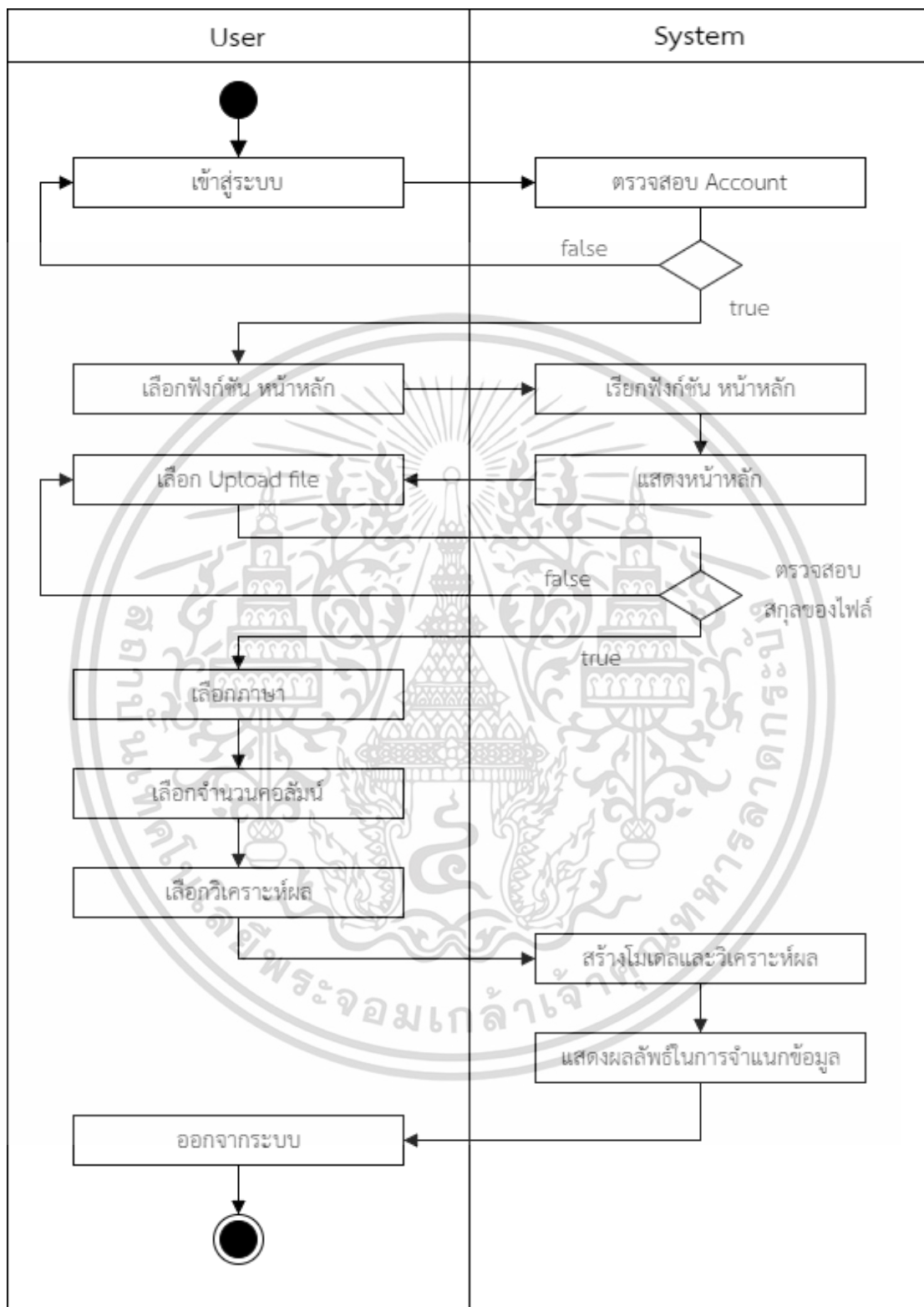
3.1) การลงทะเบียนเพื่อใช้งาน แสดงดังรูปที่ 3.4



รูปที่ 3.4 Activity Diagram สำหรับการลงทะเบียนเพื่อใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

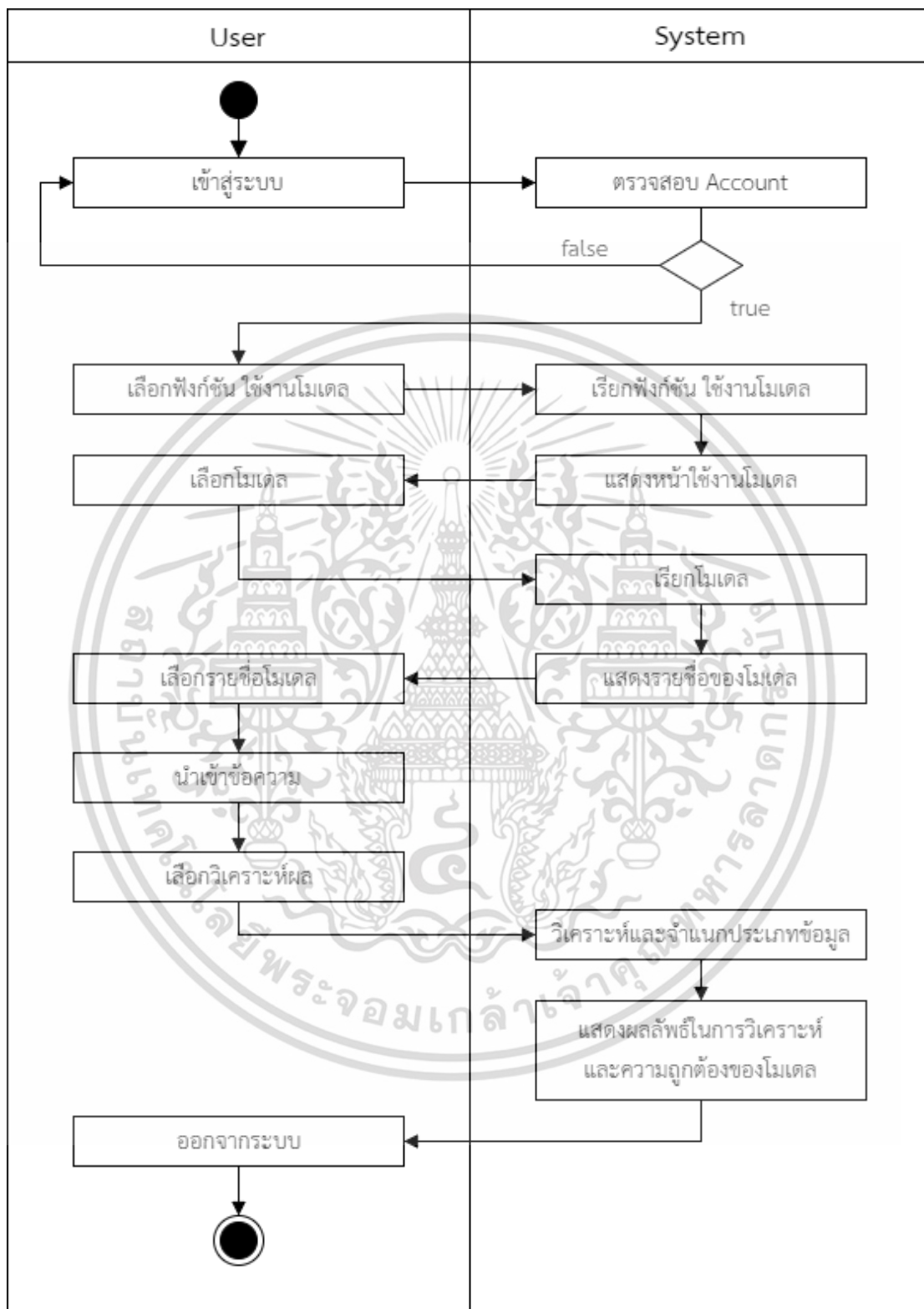
3.2) การใช้งาน หน้าหลัก แสดงดังรูปที่ 3.5



รูปที่ 3.5 Activity Diagram สำหรับการใช้งาน หน้าหลัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

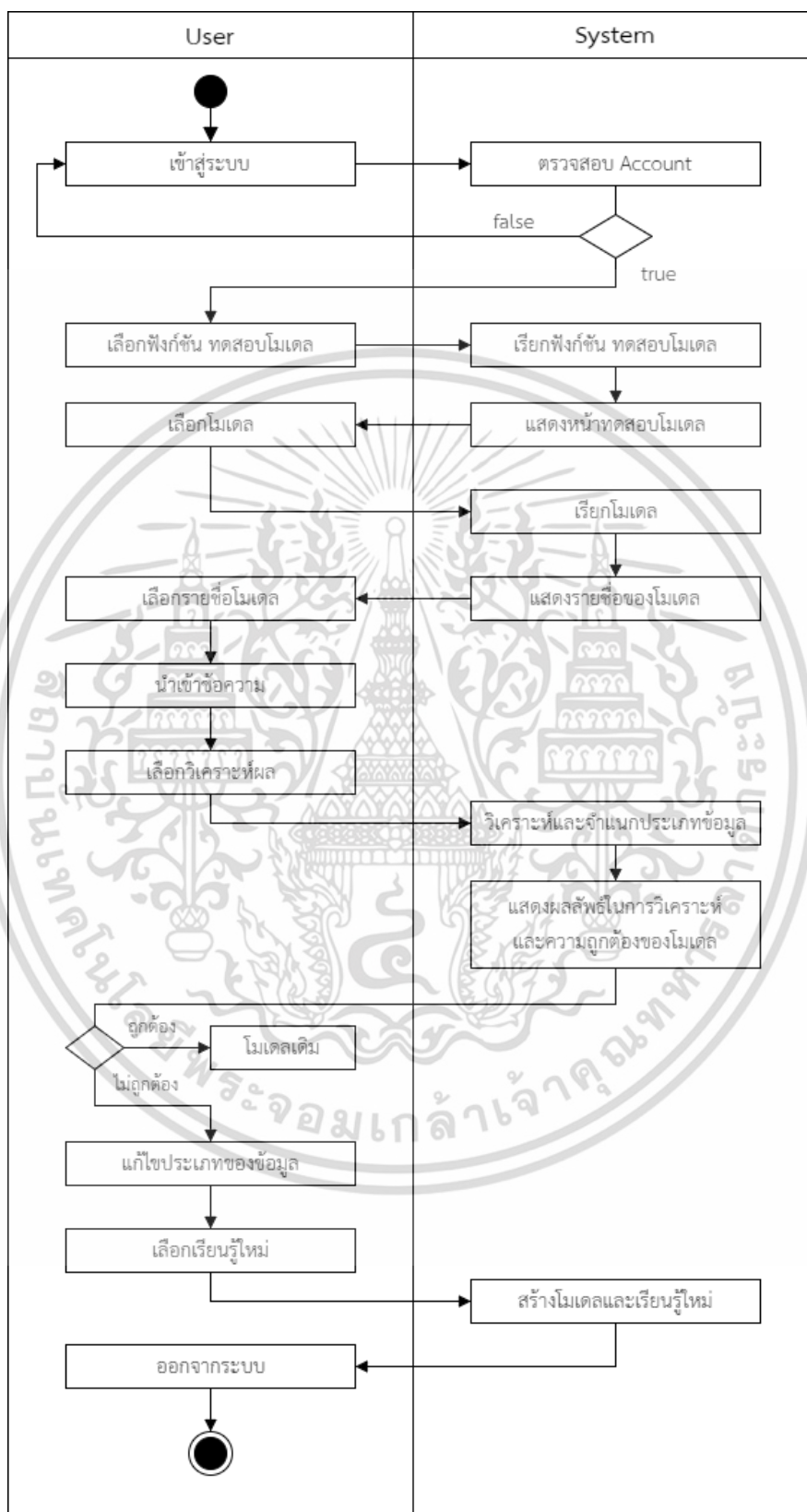
3.3) การใช้งาน หน้าใช้งานโมเดล แสดงดังรูปที่ 3.6



รูปที่ 3.6 Activity Diagram สำหรับการใช้งาน หน้าใช้งานโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

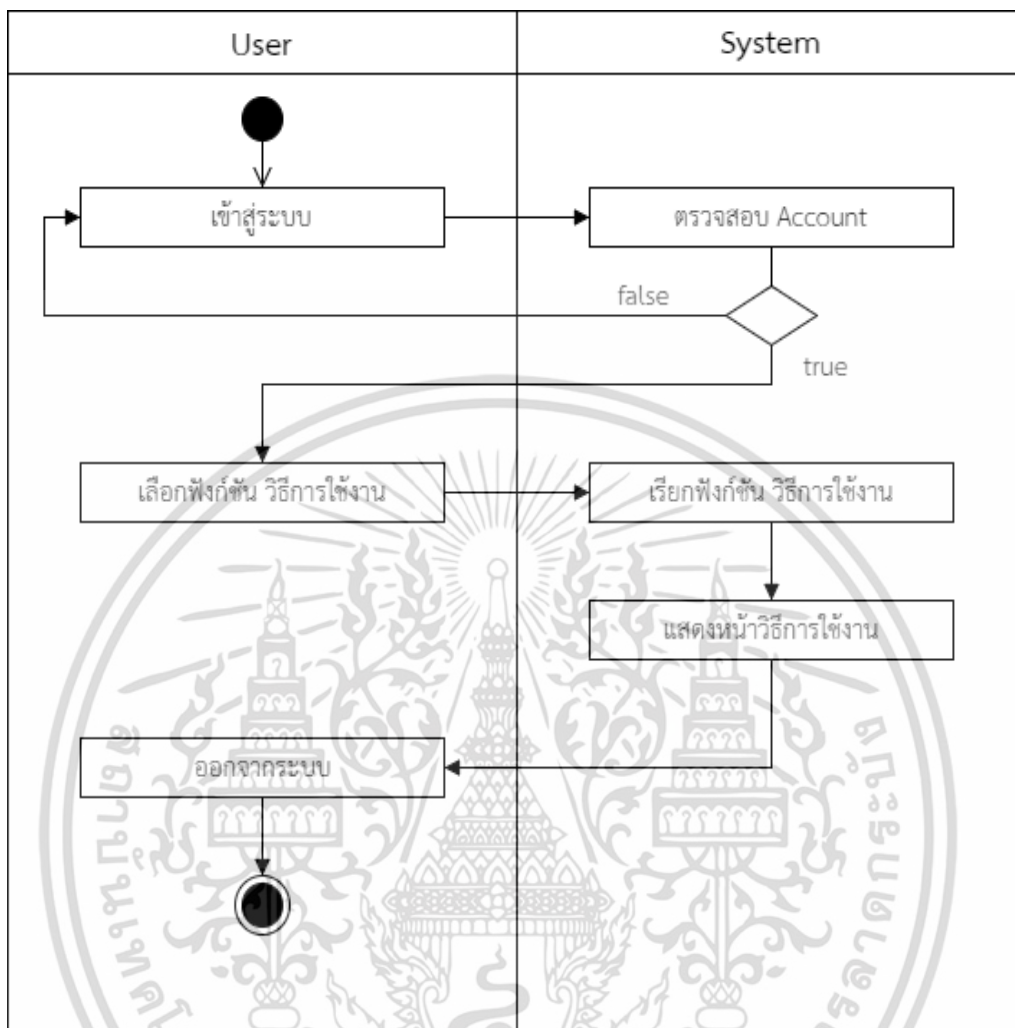
3.4) การใช้งาน หน้าทดสอบโมเดล แสดงดังรูปที่ 3.7



รูปที่ 3.7 Activity Diagram สำหรับการใช้งาน หน้าทดสอบโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อผู้ดูแลระบบเห็นประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5) การใช้งาน หน้าวิธีการใช้งาน แสดงดังรูปที่ 3.8



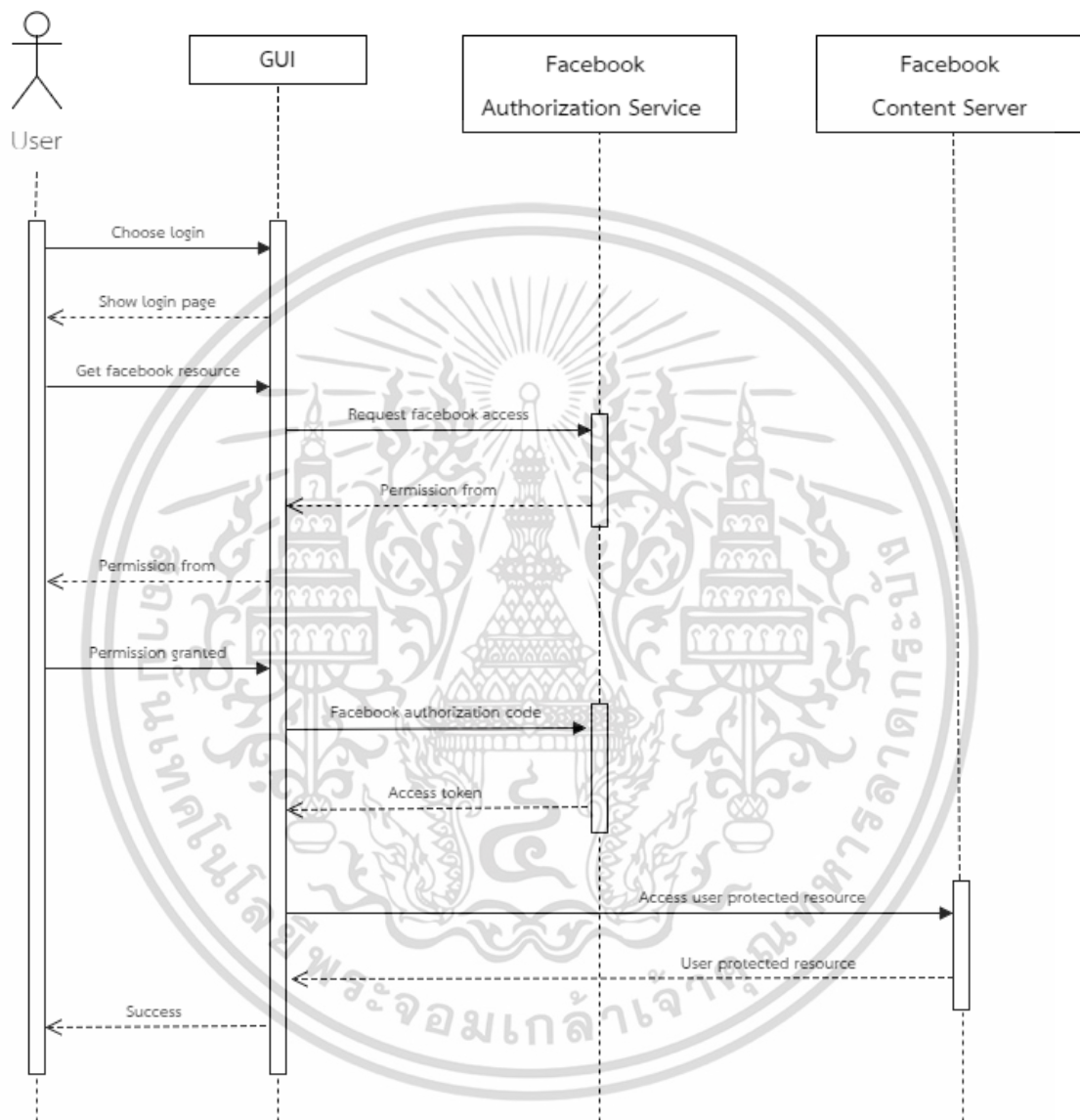
รูปที่ 3.8 Activity Diagram สำหรับการใช้งาน หน้าวิธีการใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) Sequence Diagram

แผนภาพกิจกรรมการทำงานสำหรับผู้ใช้งานระบบจัดการและวิเคราะห์ข้อมูลจากสังคมออนไลน์เพื่อจำแนกประเภทข้อมูล มีรายละเอียดดังนี้

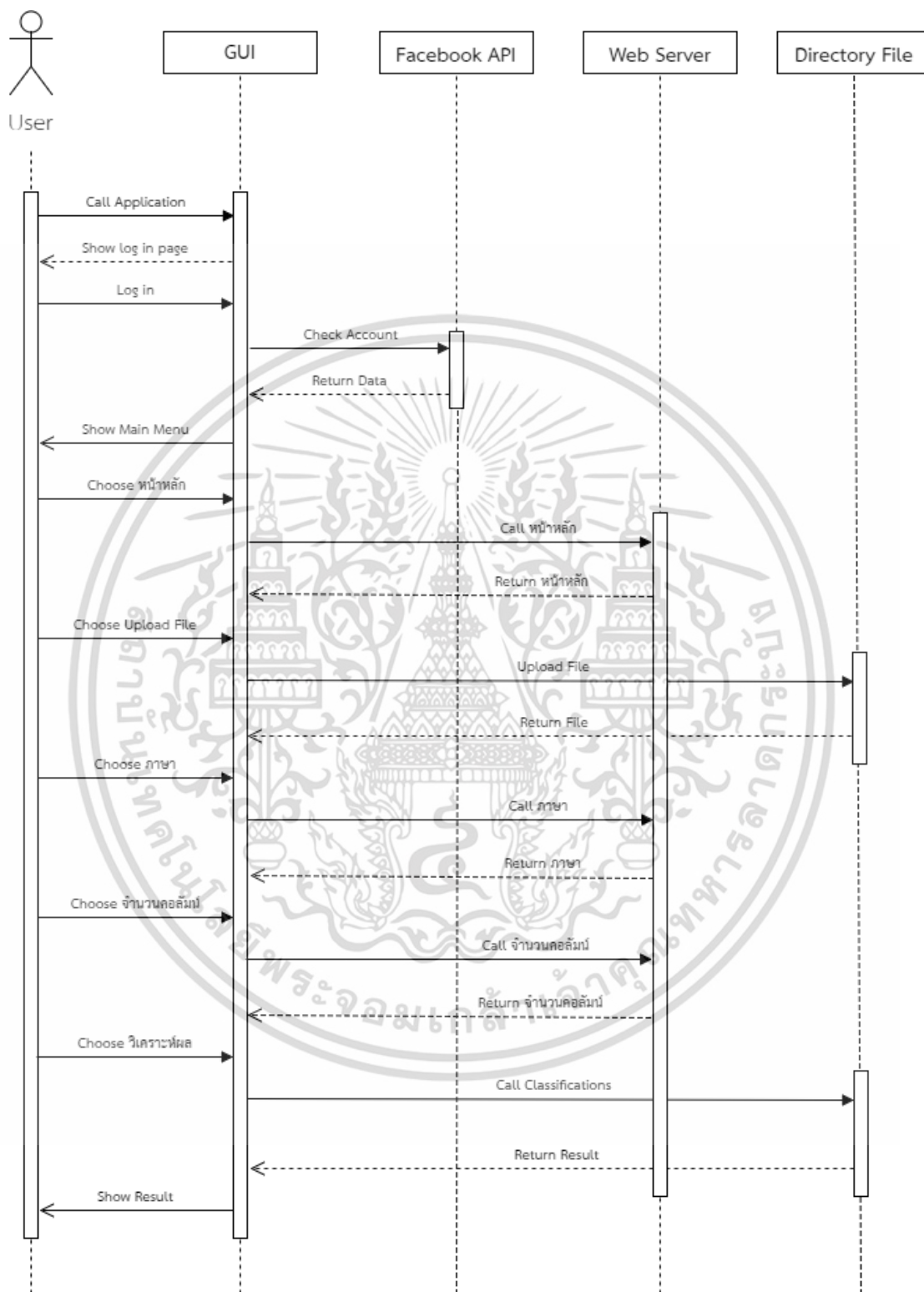
4.1) การลงทะเบียนเพื่อใช้งาน แสดงดังรูปที่ 3.9



รูปที่ 3.9 Sequence Diagram สำหรับการลงทะเบียนเพื่อใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

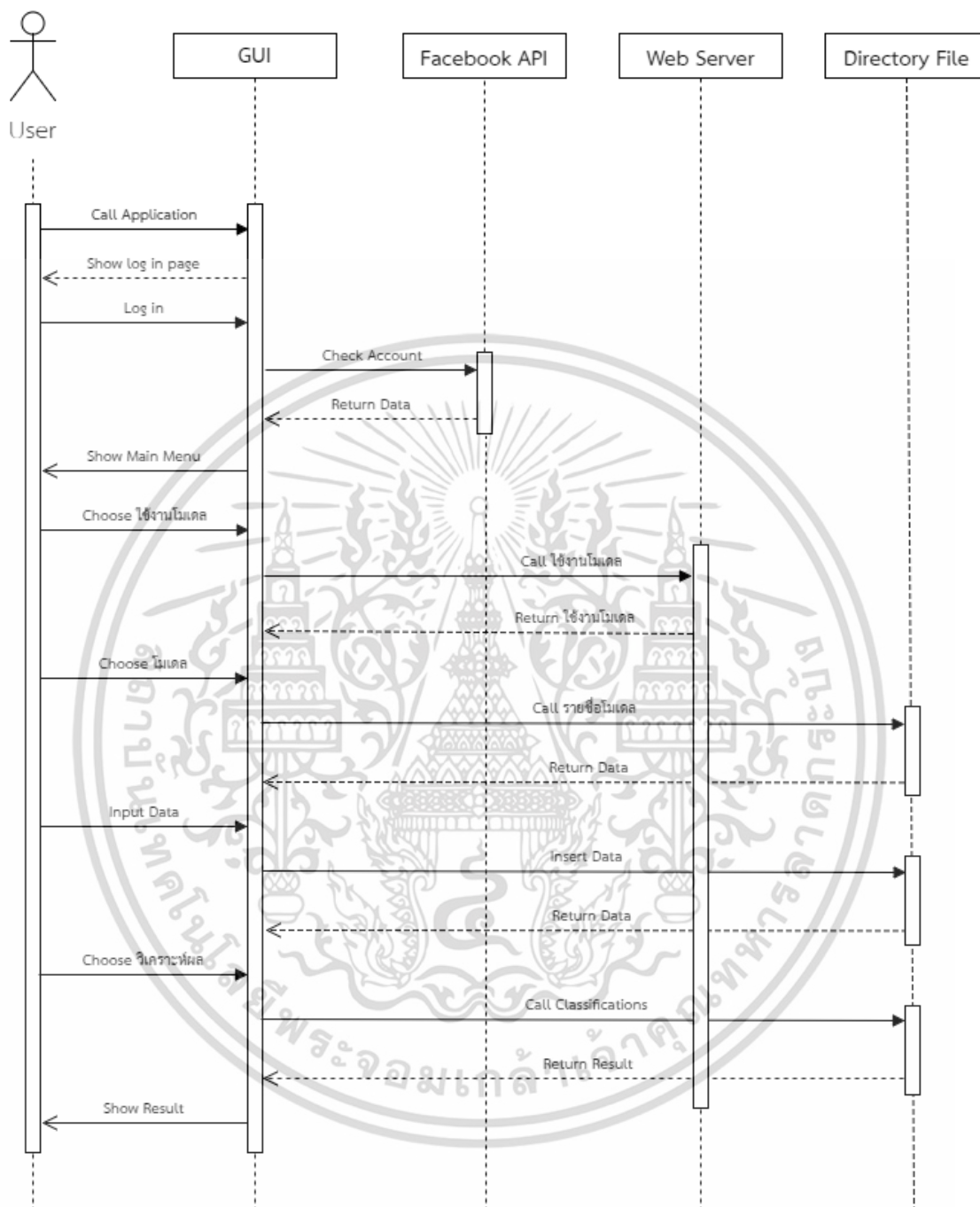
4.2) การใช้งาน หน้าหลัก แสดงดังรูปที่ 3.10



รูปที่ 3.10 Sequence Diagram สำหรับการใช้งาน หน้าหลัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

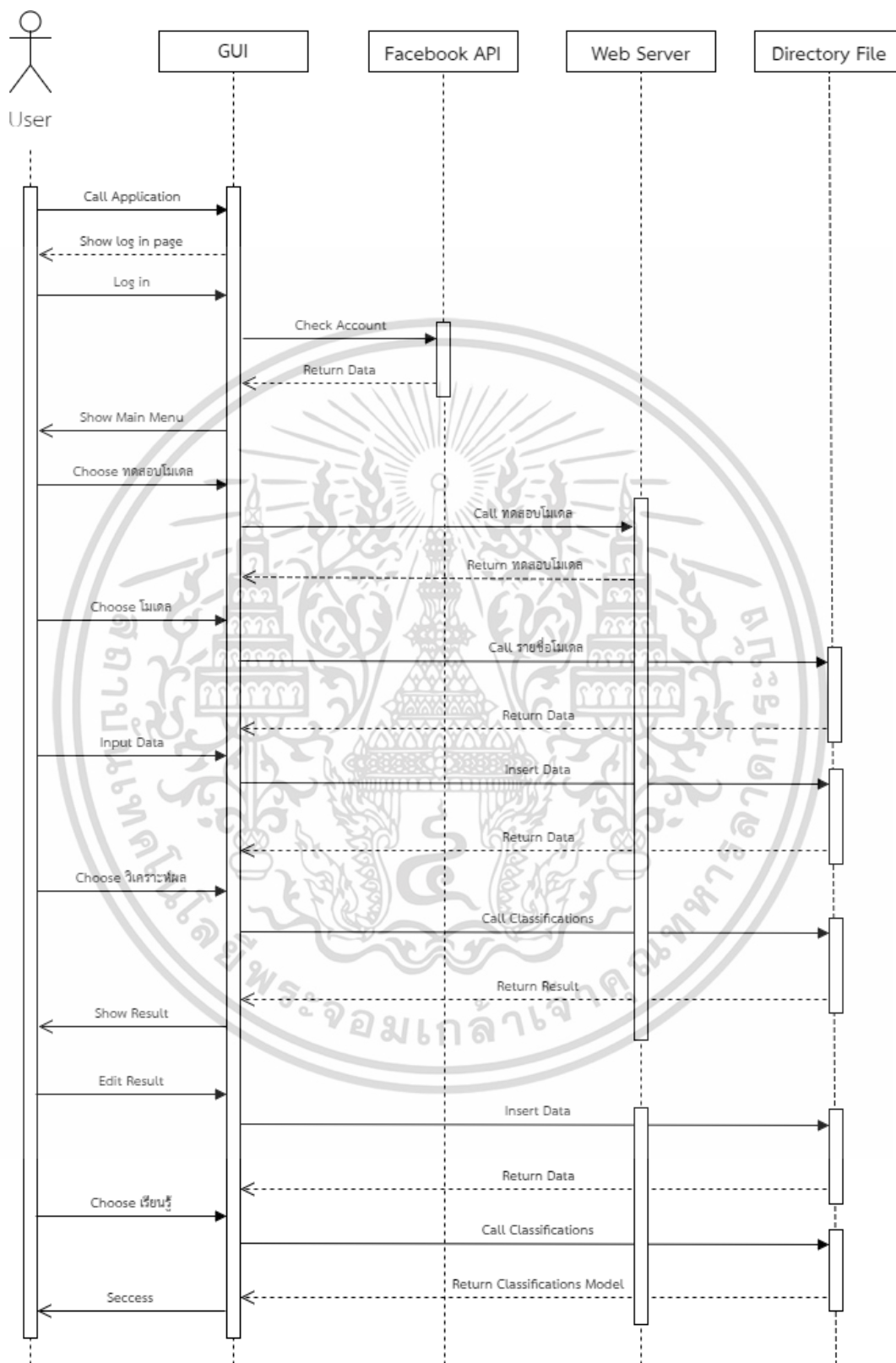
4.3) การใช้งาน หน้าใช้งานโมเดล แสดงดังรูปที่ 3.11



รูปที่ 3.11 Sequence Diagram สำหรับการใช้งาน หน้าใช้งานโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

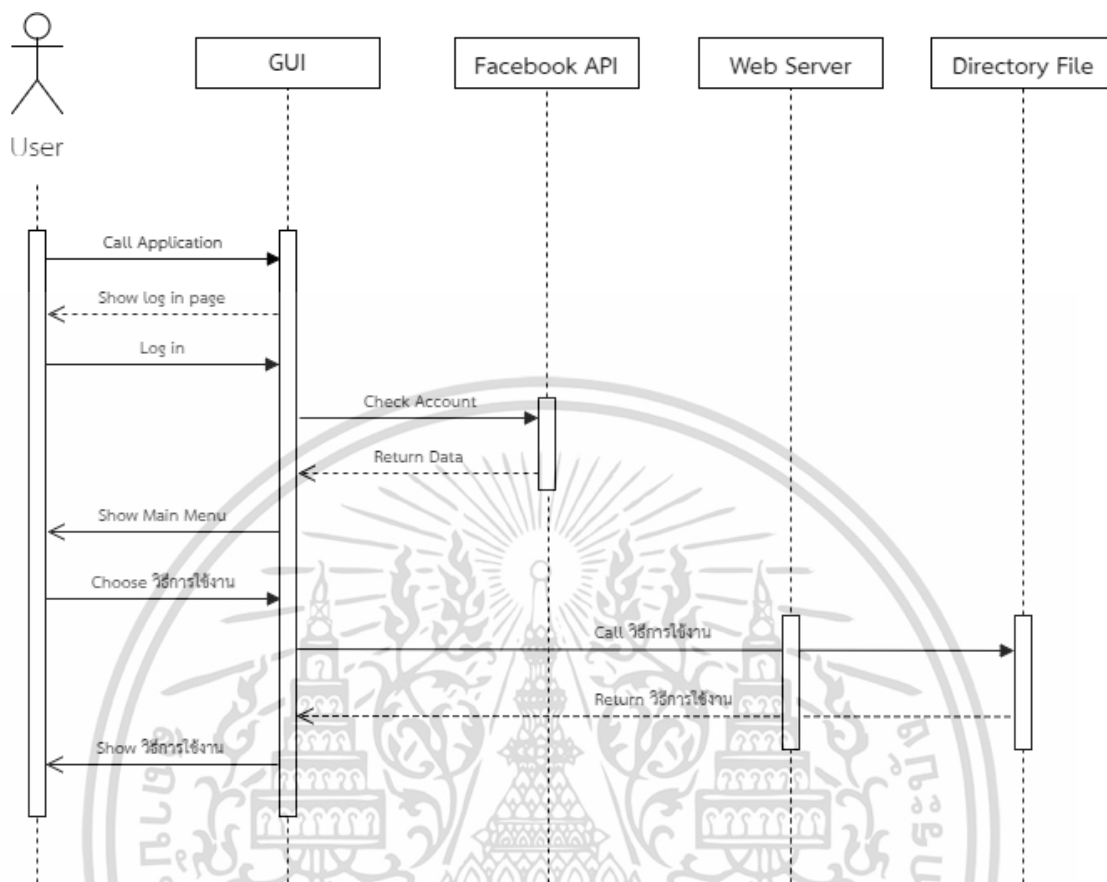
4.4) การใช้งาน หน้าทดสอบโมเดล แสดงดังรูปที่ 3.12



รูปที่ 3.12 Sequence Diagram สำหรับการใช้งาน หน้าทดสอบโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5) การใช้งาน หน้าวิธีการใช้งาน แสดงดังรูปที่ 3.13



รูปที่ 3.13 Sequence Diagram สำหรับการใช้งาน หน้าวิธีการใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 ขั้นตอนการพัฒนาและทดสอบระบบ

จากขั้นตอนศึกษาและรวบรวมข้อมูลระบบ วิเคราะห์ระบบ ออกแบบระบบ จนถึงขั้นตอนการออกแบบหน้าจอสำหรับผู้ใช้งาน ทำให้ทราบถึงวิธีการและขั้นตอนต่างๆในการพัฒนาเว็บเซอวิส สำหรับให้บริการการจัดการและวิเคราะห์ข้อมูลในการจำแนกประเภทของข้อมูล ซึ่งในการพัฒนาเว็บเซอวิสสำหรับให้บริการการจัดการและวิเคราะห์ข้อมูลในการจำแนกประเภทของข้อมูล ใช้โปรแกรมในการพัฒนาคือโปรแกรม JetBrains PhpStorm 8.0.4 และภาษาที่ใช้ในการพัฒนาระบบการรวบรวมและจำแนกข้อมูลคือภาษา Python ส่วนภาษาที่ใช้ในการพัฒนาเว็บเซอวิสสำหรับให้บริการการจัดการและวิเคราะห์ข้อมูลในการจำแนกประเภทของข้อมูล คือ ภาษา Php และภาษา JavaScript เป็นต้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการดำเนินงานและการอภิปรายผล

บทนี้จะเป็นการกล่าวถึงการทดสอบความถูกต้องของการจำแนกประเภทของข้อมูลปัญหาเมืองที่ได้จากสื่อสังคมออนไลน์ และทดสอบเว็บเซอร์วิสที่ให้บริการจำแนกประเภทของข้อมูล ซึ่งเว็บเซอร์วิสนี้ได้ถูกพัฒนาตามทีออกแบบไว้ในบทที่ 3

4.1 การทดสอบการจำแนกประเภทของปัญหาเมือง

ข้อมูลที่นำมาทดสอบได้มาจากสื่อสังคมออนไลน์ที่เกี่ยวข้องกับปัญหาของเมือง ประกอบด้วย ปัญหาด้านการระบายน้ำ จำนวน 400 record ปัญหาด้านอัคคีภัย 500 record ปัญหาด้านอัคคีภัย 500 record ปัญหาด้านทางเท้า 500 record ปัญหาด้านความปลอดภัย 500 record และปัญหาด้านผิวจราจร 500 record รวมทั้งสิ้น 2400 record ซึ่งผลการทดสอบนี้สามารถเปรียบเทียบกับผลเฉลี่ยที่จัดทำขึ้นได้ดังต่อไปนี้

4.1.1 ข้อมูลที่นำมาทำการทดสอบการจำแนกประเภทของปัญหาเมือง

ข้อมูลที่นำมาทดสอบได้มาจากสื่อสังคมออนไลน์ โดยรูปแบบของข้อมูลต้องประกอบด้วย คอลัมน์ข้อความและคอลัมน์ผลเฉลี่ยของประเภทข้อความ ดังรูปที่ 4.1

Description	Actual
00.10 น.สภ.เมืองนครปฐม แจ้งปิด.เพชรเกษม ขาเข้า ช่องทางด่วนหน้าบึงชินนครปฐม ซ่อมถนนทรุดตัว ให้ใช้ช่องคู่ช. ผิวจราจร	
00.12 น.ฝนตกหนักเขตหลักสี่ ถ.วิภาวดีฯ หน้า #ThaiPBS มีน้ำท่วมข้างสูง /กทม.รายงานปริมาณฝนสูงสุดเขตหลักสี่ 90 การระบายน้ำ	
00.22 น.ถนนประชาร่วมใจ(มีนบุรี) ตลอดสาย มีต้นไม้ล้มหลายจุด ไฟส่องทางดับมืดตลอดเส้นทาง เจ้าหน้าที่อยู่ระหว่า. ความปลอดภัย	
(00.25)#แจ้งเตือน จนท.เปลี่ยนหลอดไฟส่องทาง บนยกระดับดอนเมือง-โทลเวย์ ขาออก ตั้งแต่แยกหลักสี่ จนถึงสนาม. ความปลอดภัย	
00:29 เพลงใหม่อาคาร บริษัทวิริยะประกันภัย จำกัด โกลด์แยกมีสทิน ถ.รามคำแหง ล่าสุดเพลงสงบแล้ว	อัคคีภัย
00.30 น.ถ.แจ้งวัฒนะ มawangเวียนบางเขน น้ำท่วมข้างสูง #อาสาสมัครจราจร #สายตรวจจราจรไทยพีบีเอส #ThaiPBS การระบายน้ำ	
00.30น. ถนนสุขุมวิท ตั้งแต่อนุสาวรีย์ข้างสามเศียร - หน้า สภ.สาโรจน์เหนือ น้ำท่วมข้าง สูง 10-20 ซม. รถเล็กควรถกหัก. การระบายน้ำ	
(00.32)#เพลงใหม่ รถเบนซ์ สีดำ ที่ลานจอดรถฟุตแลนดท์หัวหมาก ถ.รามคำแหง จนท.ใช้ถังดับเพลิงดับไฟเรียบร้อยแล้ว	อัคคีภัย
(00.33)#เตือนระวัง ไฟส่องทางดับ ทั้ง 2 ฝั่ง บนสะพานข้ามแยกศรีอยุธยา มุ่งหน้าถนนจตุรทิศ เตือนผู้ใช้รถใช้ถนนโปรด. ความปลอดภัย	

รูปที่ 4.1 ตัวอย่างของข้อมูลปัญหาเมืองที่นำมาจำแนกประเภท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 ผลลัพธ์จากการทดสอบการจำแนกประเภทของปัญหาเมือง

จากผลลัพธ์ของการทดสอบสามารถนำมาเปรียบเทียบกับผลเฉลยเพื่อหาความถูกต้องของการจำแนกประเภทได้ดังนี้

- 1) การระบายน้ำ ถูกต้อง 456 จาก 500 Record
- 2) อัคคีภัย ถูกต้อง 489 จาก 500 Record
- 3) ทางเท้า ถูกต้อง 474 จาก 500 Record
- 4) ความปลอดภัย ถูกต้อง 392 จาก 400 Record
- 5) ผิวจราจร ถูกต้อง 447 จาก 500 Record

		Predicted				
		การระบายน้ำ	อัคคีภัย	ทางเท้า	ความปลอดภัย	ผิวจราจร
Actual	การระบายน้ำ	456	7	13	3	21
	อัคคีภัย	2	489	1	2	6
	ทางเท้า	1	1	474	2	22
	ความปลอดภัย	0	0	3	392	5
	ผิวจราจร	9	7	20	17	447

รูปที่ 4.2 Confusion matrix แสดงผลลัพธ์จากการทำนาย

	การระบายน้ำ	อัคคีภัย	ทางเท้า	ความปลอดภัย	ผิวจราจร
Precision	97.44%	97.02%	92.76%	94.23%	89.22%
Recall	91.20%	97.80%	94.80%	98.00%	89.40%
accuracy					94.08%

รูปที่ 4.3 ค่าความถูกต้องของการทำนาย

4.2 การทดสอบการจำแนกประเภทข้อมูลของเว็บไซต์

เว็บไซต์ที่ให้บริการจำแนกประเภทของข้อมูล สามารถอธิบายหลักการทำงานของเว็บไซต์นี้มีเมนูทั้งหมด 3 ส่วน ดังนี้

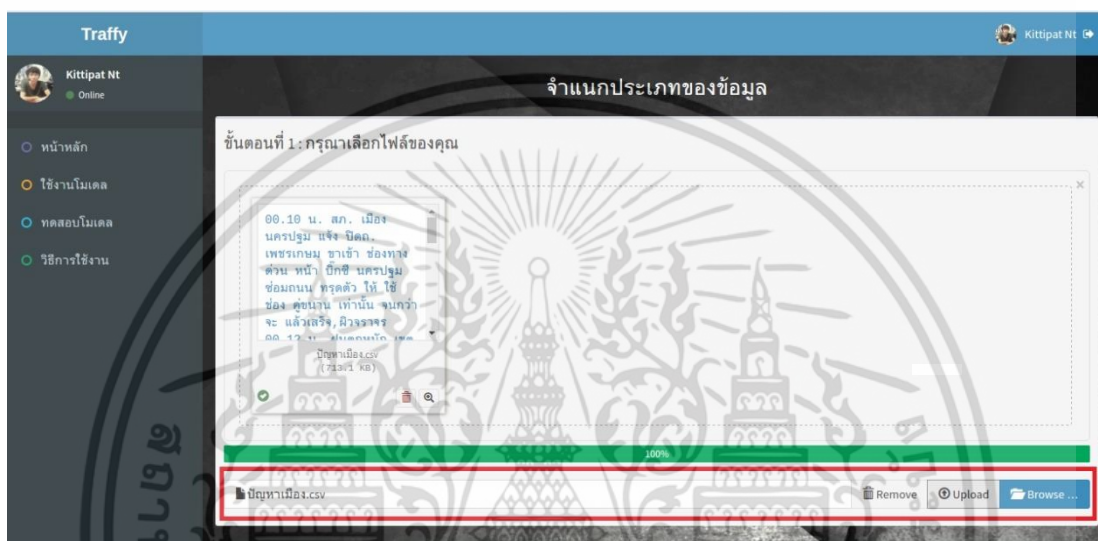
- 1) หน้าหลัก
- 2) หน้าใช้งานโมเดล
- 3) หน้าทดสอบโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.1 หน้าหลัก

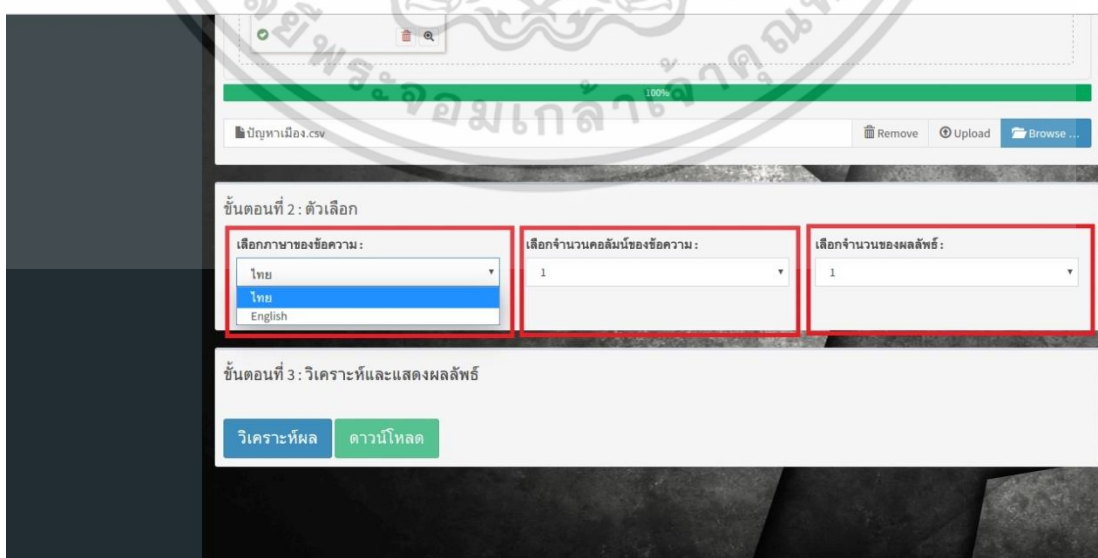
ในหน้านี้เป็นหน้าหลักในการจำแนกประเภทของข้อมูล โดยจะมีขั้นตอนในการจำแนกประเภทของข้อมูล 3 ขั้นตอน คือ ส่วนที่ 1 นำเข้าข้อมูลโดยรูปแบบของข้อมูลต้องประกอบด้วย คอลัมน์ข้อความและคอลัมน์ผลเฉลยของประเภทข้อความ ส่วนที่ 2 เลือกตัวเลือกในการวิเคราะห์ผลการทำนาย และส่วนสุดท้าย ทำการวิเคราะห์และแสดงผลการทำนายผลการทำนาย

1) เลือกไฟล์ที่ต้องจำแนกประเภทของข้อมูล



รูปที่ 4.4 เลือกไฟล์ที่ต้องจำแนกประเภทของข้อมูล

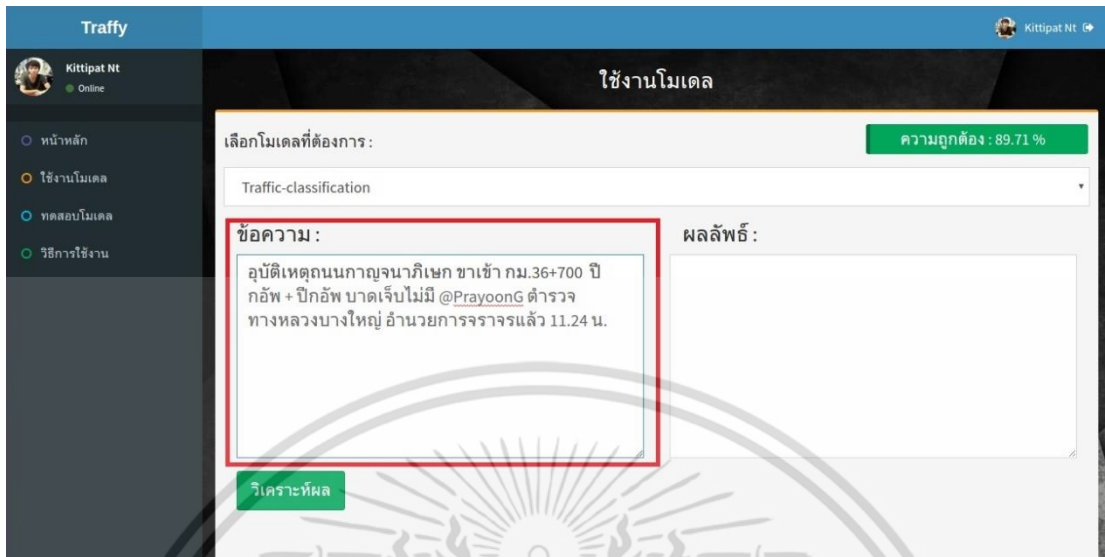
2) เลือกตัวเลือกในการวิเคราะห์ผล



รูปที่ 4.5 เลือกตัวเลือกในการวิเคราะห์ผล

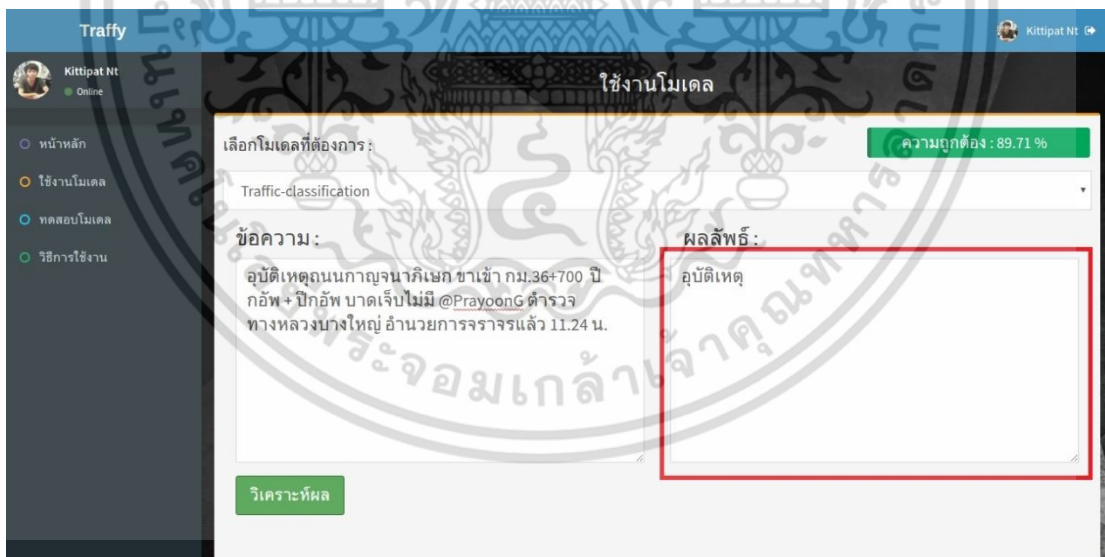
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) นำข้อความมาใส่ในช่องข้อความเพื่อวิเคราะห์ผลลัพธ์



รูปที่ 4.8 นำข้อความมาใส่ในช่องข้อความเพื่อวิเคราะห์ผลลัพธ์

3) วิเคราะห์ผลลัพธ์ในการทำนาย



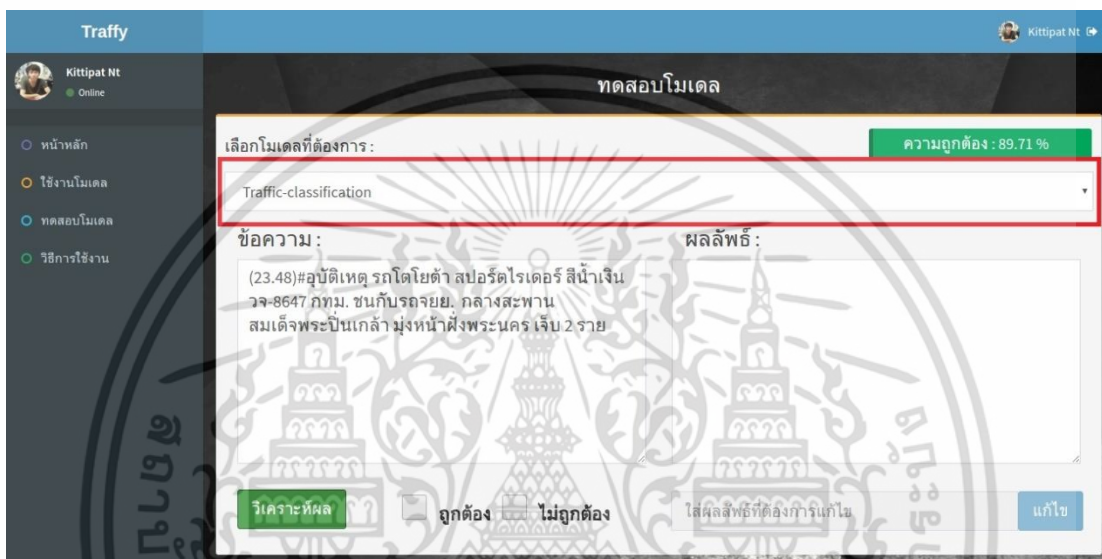
รูปที่ 4.9 วิเคราะห์ผลลัพธ์ในการทำนาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3 หน้าการใช้งานโมเดล

ในหน้านี้สามารถเลือกโมเดลในการทดสอบได้ว่าต้องการใช้งานโมเดลใด แล้วนำข้อความมาใส่ในช่องข้อความจากนั้นทำการวิเคราะห์ผลเพื่อจำแนกประเภทของข้อมูลพร้อมทั้งแสดงความถูกต้องของโมเดลที่ใช้ พร้อมทั้งสามารถแก้ไขโมเดลได้ถ้าหากผลลัพธ์จากการทำนายไม่ถูกต้องและทำการสร้างโมเดลใหม่

1) เลือกโมเดลที่ต้องการใช้เพื่อวิเคราะห์ผลลัพธ์



รูปที่ 4.10 เลือกโมเดลที่ต้องการใช้เพื่อวิเคราะห์ผลลัพธ์

2) นำข้อความมาใส่ในช่องข้อความเพื่อวิเคราะห์ผลลัพธ์



รูปที่ 4.11 นำข้อความมาใส่ในช่องข้อความเพื่อวิเคราะห์ผลลัพธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) วิเคราะห์ผลลัพธ์ในการทำงาน

Traffy

Kittipat NT Online

หน้าหลัก
ใช้งานโมเดล
ทดสอบโมเดล
วิธีการใช้งาน

ทดสอบโมเดล

เลือกโมเดลที่ต้องการ: ความถูกต้อง : 89.71 %

Traffic-classification

ข้อความ:
(23.48)#อุบัติเหตุ รถโตโยต้า สปอร์ตโรดอร์ สีน้ำเงิน วจ-8647 กทม. ชนกับรถจยย. กลางสะพาน สมเด็จพระปิ่นเกล้า มุ่งหน้าฝั่งพระนคร เจ็บ 2 ราย

ผลลัพธ์:
อุบัติเหตุ

วิเคราะห์ผล ถูกต้อง ไม่ถูกต้อง แก้ไข

รูปที่ 4.12 วิเคราะห์ผลลัพธ์ในการทำงาน

4) แก้ไขโมเดลและทดสอบใหม่

Traffy

Kittipat NT Online

หน้าหลัก
ใช้งานโมเดล
ทดสอบโมเดล
วิธีการใช้งาน

ทดสอบโมเดล

เลือกโมเดลที่ต้องการ: ความถูกต้อง : 89.71 %

Traffic-classification

ข้อความ:
(23.48)#อุบัติเหตุ รถโตโยต้า สปอร์ตโรดอร์ สีน้ำเงิน วจ-8647 กทม. ชนกับรถจยย. กลางสะพาน สมเด็จพระปิ่นเกล้า มุ่งหน้าฝั่งพระนคร เจ็บ 2 ราย

ผลลัพธ์:
อุบัติเหตุ

วิเคราะห์ผล ถูกต้อง ไม่ถูกต้อง แก้ไข

รูปที่ 4.13 แก้ไขโมเดลและทดสอบใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 ผลการดำเนินงาน

ตารางที่ 4.1 การทดสอบความถูกต้องในการจำแนกประเภทข้อมูล

อธิบายการทดสอบ	จำนวนประเภท ของข้อมูล	จำนวนของ ข้อมูล	ความถูกต้อง
จำแนกประเภทของข้อมูลจรรยาจร	6	3,000	ร้อยละ 89.73
จำแนกประเภทของข้อมูลยานพาหนะที่ เกิดอุบัติเหตุ	22	46,928	ร้อยละ 97.54
จำแนกประเภทของข้อมูลผู้ที่ได้บาดเจ็บ หรือเสียชีวิตจากอุบัติเหตุ	4	46,928	ร้อยละ 99.42
จำแนกประเภทของข้อมูล keywords ที่ ได้จาก Abstract ของวารสาร	10	3,000	ร้อยละ 84.53

ตารางที่ 4.2 การทดสอบการทำงานของหน้าหลัก

ขั้นตอน	อธิบายการทดสอบ	ผลทดสอบ	ผู้ทดสอบ	ผ่าน / ไม่ผ่าน
1.	นำเข้าไฟล์ข้อมูล	เลือกไฟล์ข้อมูลที่ต้องการ นำมาจำแนกประเภท ข้อมูลได้	ณพงศ์	ผ่าน
2.	ตรวจสอบ ไฟล์ข้อมูลที่นำเข้า	สามารถตรวจสอบได้หาก ไฟล์ข้อมูลไม่ตรงกับ รูปแบบที่ต้องการ	ณพงศ์	ผ่าน
3.	การตัดคำ Tokenization	สามารถ Tokenization ได้	ณพงศ์	ผ่าน
4.	การจำแนกประเภท ของข้อมูล	จำแนกประเภทของข้อมูล ได้ตามประเภทข้อมูลที่ ผู้ใช้กำหนด	ณพงศ์	ผ่าน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 การทดสอบการทำงานของหน้าหลัก (ต่อ)

ขั้นตอน	อธิบายการทดสอบ	ผลทดสอบ	ผู้ทดสอบ	ผ่าน / ไม่ผ่าน
5.	ดาวน์โหลดไฟล์ข้อมูลพร้อมผลการจำแนกประเภทข้อมูล	สามารถดาวน์โหลดไฟล์ข้อมูลพร้อมผลการจำแนกประเภทข้อมูล	ณพงศ์	ผ่าน

ตารางที่ 4.4 การทดสอบการทำงานของหน้าใช้งานโมเดล

ขั้นตอน	อธิบายการทดสอบ	ผลทดสอบ	ผู้ทดสอบ	ผ่าน / ไม่ผ่าน
1.	การเลือกโมเดล	มีให้เลือกได้ และสามารถโชว์ทั้งหมดทุกแผนกได้	ณพงศ์	ผ่าน
2.	นำเข้าข้อมูล	นำเข้าข้อมูลเพื่อวิเคราะห์ผลการจำแนกประเภทข้อมูลได้	ณพงศ์	ผ่าน
3.	การจำแนกประเภทของข้อมูล	จำแนกประเภทของข้อมูลได้ตามประเภทข้อมูลที่กำหนดในแต่ละโมเดล	ณพงศ์	ผ่าน
4.	แสดงความถูกต้องของโมเดลที่ใช้	แสดงความถูกต้องของโมเดลที่ใช้ได้	ณพงศ์	ผ่าน
5.	แสดงผลลัพธ์ของการทำนาย	แสดงผลลัพธ์ของการทำนายจากการนำเข้าข้อมูลได้	ณพงศ์	ผ่าน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 การทดสอบการทำงานของหน้าทดสอบโมเดล

ขั้นตอน	อธิบายการทดสอบ	ผลทดสอบ	ผู้ทดสอบ	ผ่าน / ไม่ผ่าน
1.	การเลือกโมเดล	มีให้เลือกได้ และสามารถ โชว์ทั้งหมดทุกแผนกได้	ณพงศ์	ผ่าน
2.	นำเข้าข้อมูล	นำเข้าข้อมูลเพื่อวิเคราะห์ ผลการจำแนกประเภท ข้อมูลได้	ณพงศ์	ผ่าน
3.	การจำแนกประเภท ของข้อมูล	จำแนกประเภทของข้อมูล ได้ตามประเภทข้อมูลที่ กำหนดในแต่ละโมเดล	ณพงศ์	ผ่าน
4.	แสดงความถูกต้อง ของโมเดลที่ใช้	แสดงความถูกต้องของ โมเดลที่ใช้ได้	ณพงศ์	ผ่าน
5.	แสดงผลลัพธ์ของ การทำนาย	แสดงผลลัพธ์ของการ ทำนายจากการนำเข้า ข้อมูลได้	ณพงศ์	ผ่าน
6.	ตรวจสอบและแก้ไข ความถูกต้อง	ตรวจสอบและแก้ไขความ ถูกต้องของผลลัพธ์จาก การทำนายได้	ณพงศ์	ผ่าน
7.	สร้างโมเดลใหม่	สร้างโมเดลใหม่เมื่อมีการ แก้ไขผลลัพธ์จากการ ทำนายได้	ณพงศ์	ผ่าน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการดำเนินงานและข้อเสนอแนะ

5.1 สรุปผลการดำเนินงานและการทดสอบ

5.1.1 สรุปผลการดำเนินงาน

สหกิจนี้มีจุดมุ่งหมายหลัก คือการพัฒนากระบวนการจัดการและวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์ (Social Media) ในการจำแนกประเภทของข้อมูล เพื่อจำแนกประเภทของข้อมูลที่ได้จากสื่อสังคมออนไลน์ (Social Media) เนื่องจากข้อมูลจากสื่อสังคมออนไลน์ (Social Media) นั้นมีปริมาณมากและหลากหลายจะอยู่ในรูปแบบ Big data คือจะเป็นกลุ่มของข้อมูลที่มีรูปแบบหลากหลายเป็นข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) จึงจำเป็นที่จะต้องนำข้อมูลเหล่านี้มาผ่านกระบวนการจัดการและวิเคราะห์ข้อมูลเพื่อนำมาใช้ประโยชน์ต่อไป เช่น จำแนกประเภทของข้อมูลตามสิ่งที่สนใจ ในที่นี้ คือจำแนกประเภทของข้อมูลที่เกี่ยวข้องกับปัญหาของเมืองที่กำหนดไว้ โดยใช้หลักการวิเคราะห์ประโยค (Syntactic Analysis) เพื่อทำการหารูปแบบของข้อความประเภทปัญหาของเมืองแต่ละประเภทและทำการจัดหมวดหมู่หรือจำแนกประเภท (Classifications) ข้อความนั้นๆให้อยู่ในหมวดหมู่เดียวกัน จากผลการทดลอง จำแนกประเภทของข้อมูลปัญหาเมือง มีความแม่นยำในการทำนายโดยเฉลี่ยร้อยละ 94.08 ผู้ใช้สามารถนำต้นแบบ (Classifications Model) ไปใช้ทดสอบข้อมูลปัญหาเมืองได้ รวมไปถึงสามารถนำไปใช้ในการจำแนกประเภทของข้อมูลที่เกี่ยวข้องกับปัญหาเมืองอื่นๆเพิ่มเติมได้ โดยทำการเพิ่มจำนวนของประเภทปัญหาเมืองอื่นๆให้กับต้นแบบ (Classifications Model) เพื่อเพิ่มประสิทธิภาพในการจำแนกประเภทให้ครอบคลุมมากขึ้น

5.1.2 สรุปผลการทดสอบเว็บเซอร์วิส

ผลการประเมินจากการพัฒนาเว็บเซอร์วิส ในการดำเนินงานโดยนำข้อมูลจากสื่อสังคมออนไลน์ประเภทอื่นๆมาทดสอบในการจำแนกประเภทของข้อมูล ได้แก่ จำแนกประเภทของข้อมูลจราจร จำแนกประเภทของข้อมูลยานพาหนะที่เกิดอุบัติเหตุ จำแนกประเภทของข้อมูลผู้ที่ได้บาดเจ็บหรือเสียชีวิตจากอุบัติเหตุ และจำแนกประเภทของข้อมูล keywords ที่ได้จาก Abstract ของวารสาร มีผลลัพธ์ดังตารางที่ 5.1

ตารางที่ 5.1 ผลการทดสอบเว็บเซอร์วิส

อธิบายการทดสอบ	จำนวนประเภทของข้อมูล	จำนวนของข้อมูล	ความถูกต้อง
จำแนกประเภทของข้อมูลจราจร	6	3,000	ร้อยละ 89.73
จำแนกประเภทของข้อมูลยานพาหนะที่เกิดอุบัติเหตุ	22	46,928	ร้อยละ 97.54
จำแนกประเภทของข้อมูลผู้ที่บาดเจ็บหรือเสียชีวิตจากอุบัติเหตุ	4	46,928	ร้อยละ 99.42
จำแนกประเภทของข้อมูล keywords ที่ได้จาก Abstract ของวารสาร	10	3,000	ร้อยละ 84.53

การทดสอบเว็บเซอร์วิสในการให้บริการจำแนกประเภทของข้อมูล ที่ต้องการข้อมูลที่มีผลเฉลยของประเภทข้อความและข้อความ ซึ่งเว็บเซอร์วิสที่พัฒนาขึ้นนั้น สามารถทำงานจริงได้ถูกต้อง แสดงข้อมูลได้รวดเร็ว สามารถอำนวยความสะดวกในการจำแนกประเภทของข้อมูลต่างๆได้ตามที่ผู้ใช้งานต้องการ

5.2 ข้อจำกัดในการพัฒนาระบบ

จากความสามารถของเว็บเซอร์วิสที่ได้กล่าวมานั้น ยังมีความสามารถที่ยังมีข้อจำกัด ดังนี้

- 1) ในการใช้งานเว็บเซอร์วิสนี้ ไม่สามารถนำเข้าไฟล์ที่มีขนาดใหญ่หลายๆได้
- 2) การทำงานในการทำนายผลลัพธ์จะทำงาน n รอบตามจำนวนของข้อมูล ถ้าข้อมูลมีขนาดใหญ่จะส่งผลให้การทำงานช้า
- 3) ไม่สามารถทำนายผลลัพธ์จากข้อมูลที่มีรูปแบบหลากหลายไม่ซ้ำกันได้

5.3 ข้อเสนอแนะในการพัฒนาระบบ

จากที่ได้กล่าวถึงความสามารถและข้อจำกัดของเว็บเซอร์วิส ทางผู้พัฒนาเห็นว่าบางส่วน ควรมีความสามารถที่จะมีการพัฒนาต่อไป เพื่อความสะดวกในการใช้งานมากยิ่งขึ้น ดังนี้

- 1) ควรศึกษาการใช้งาน Classifications ด้วยวิธีการแบบอื่นเพื่อใช้ในการเปรียบเทียบประสิทธิภาพจากการทำนาย เช่น การจำแนกประเภทของข้อมูลด้วย Deep learning โดยใช้ TensorFlow หรือ H2O เป็นต้น
- 2) เพิ่มจำนวนประเภทข้อมูลที่ใช้ในการทดสอบเพิ่มมากขึ้น
- 3) พัฒนาเว็บเซอร์วิสให้สามารถจำแนกประเภทของข้อมูลที่มีรูปแบบซับซ้อน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

- โชติพันธุ์ หล่อเลิศสุนทร. 2554. **คู่มือเรียนเขียนโปรแกรม Python (ภาคปฏิบัติ)**. กรุงเทพมหานคร. เอ-บุ๊ก ดิสทริบิวชัน.
- ธัญพัฒน์ วงศ์รัตน์. 2013. **คู่มือพัฒนาเว็บแอปพลิเคชันด้วย PHP & AJAX+jQuery ฉบับ Workshop +CD**. กรุงเทพมหานคร. ซีเอ็ดยูเคชั่น.
- เอกสิทธิ์ พัทธวงศ์ศักดิ์. 2558. การแบ่งข้อมูลเพื่อนำมาทดสอบประสิทธิภาพของโมเดล. [ออนไลน์] สืบค้นจาก : <http://dataminingtrend.com/2014/data-mining-techniques/cross-validation/> 12 สิงหาคม 2559.
- เอกสิทธิ์ พัทธวงศ์ศักดิ์. 2557. **An Introduction to Data Mining Techniques (ฉบับภาษาไทย)**. กรุงเทพมหานคร. หสม. ดาด้า คิวบ์.
- Armand Joulin. 2016. **Bag of Tricks for Efficient Text Classification**. [Online]. Available : <https://arxiv.org/abs/1607.01759> 10 August 2016.
- Chai Phonbopit. 2558. **MongoDB คืออะไร? + สอนวิธีใช้งานเบื้องต้น**. [ออนไลน์] สืบค้นจาก : <https://devahoy.com/posts/getting-started-with-mongodb/> 9 สิงหาคม 2559.
- EdouardGrave. 2016. **fastText : Library for fast text representation and classification**. [Online]. Available : <https://github.com/facebookresearch/fastText> 10 August 2016.
- Konpat Ta Preechakul. 2558. **ว่าด้วยเรื่องการวัดประสิทธิภาพ**. [ออนไลน์] สืบค้นจาก : <https://medium.com/o-v-e-r-f-i-t-t-e-d/ว่าด้วยเรื่องการวัดประสิทธิภาพ-49c81cd3316#acqmxzmf4> 12 สิงหาคม 2559.
- Lexto. 2537. **Thai Lexeme Tokenizer by Sansarn (NECTEC)**. [Online]. Available : <http://www.sansarn.com/lexto/> 15 สิงหาคม 2559.
- Napong Wanichayapong. Wasan Pattara-Atikom. Wasawat Pruthipunyaskul. 2011. **Social-based traffic information extraction and classification**. [Online]. Available : <http://ieeexplore.ieee.org/abstract/document/6060036/> 8 August 2016.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง(ต่อ)

Piotr Bojanowski. Edouard Grave. 2016. **Enriching Word Vectors with Subword Information**. [Online]. Available : <https://arxiv.org/abs/1607.04606>. 10 August 2016.

Takeshi Sakaki. Makoto Okazaki. Yutaka Matsuo. 2010. **Earthquake Shakes Twitter Users: Realtime Event Detection by Social Sensors**. [Online]. Available : <http://dl.acm.org/citation.cfm?id=1772777> 8 August 2016.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

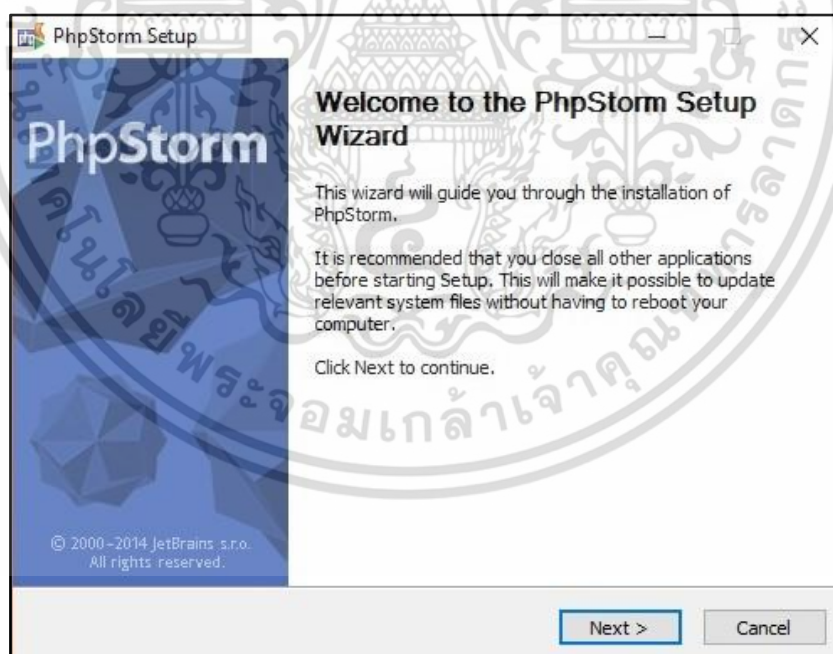
การติดตั้งโปรแกรม JetBrains PhpStorm 8.0.4

1) ทำการ Execute File Setup.exe ดังรูปที่ ก.1



รูปที่ ก.1 Execute File Setup เพื่อเริ่มการติดตั้ง

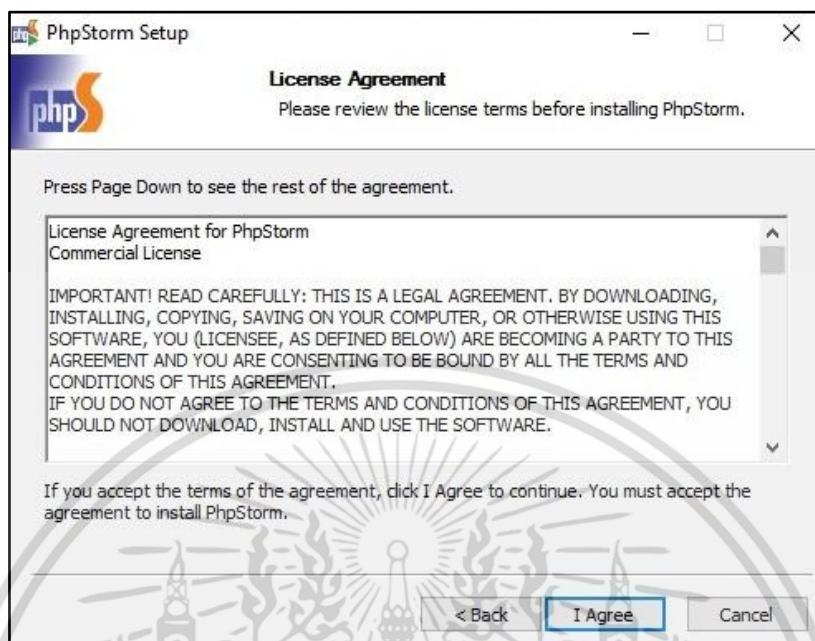
2) จะปรากฏหน้าจอสำหรับ Setup PhpStorm ให้ทำการคลิกที่ Next เพื่อเข้าสู่ขั้นตอนต่อไป ดังรูปที่ ก.2



รูปที่ ก.2 แสดงหน้าต่างเริ่มการติดตั้ง

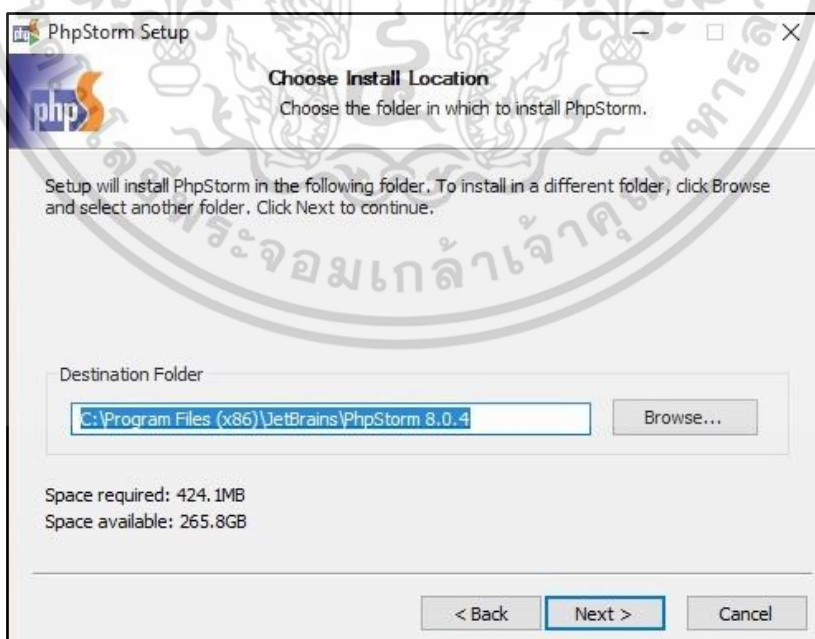
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) จะปรากฏหน้าจอสำหรับ Setup PhpStorm ให้ทำการคลิกที่ I Agree เพื่อทำการเริ่มการติดตั้งต่อไป ดังรูปที่ ก.3



รูปที่ ก.3 คลิก I Agree เพื่อทำการเริ่มการติดตั้ง

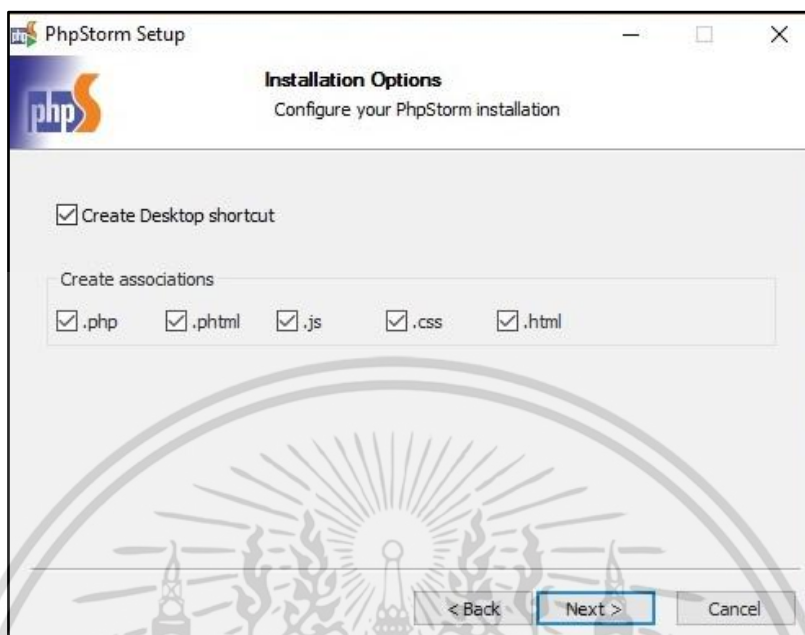
4) จะปรากฏหน้าจอสำหรับทำการเลือกตำแหน่งหรือพื้นที่ที่ต้องการติดตั้ง PhpStorm หลังจากนั้นทำการคลิกที่ Next เพื่อเข้าสู่ขั้นตอนต่อไป ดังรูปที่ ก.4



รูปที่ ก.4 เลือกตำแหน่งหรือพื้นที่ที่ต้องการติดตั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) จะปรากฏหน้าจอสำหรับตัวเลือกในการติดตั้ง หลังจากนั้นทำการคลิกที่ Next เพื่อเข้าสู่ขั้นตอนต่อไป ดังรูปที่ ก.5



รูปที่ ก.5 เลือกตัวเลือกในการติดตั้ง

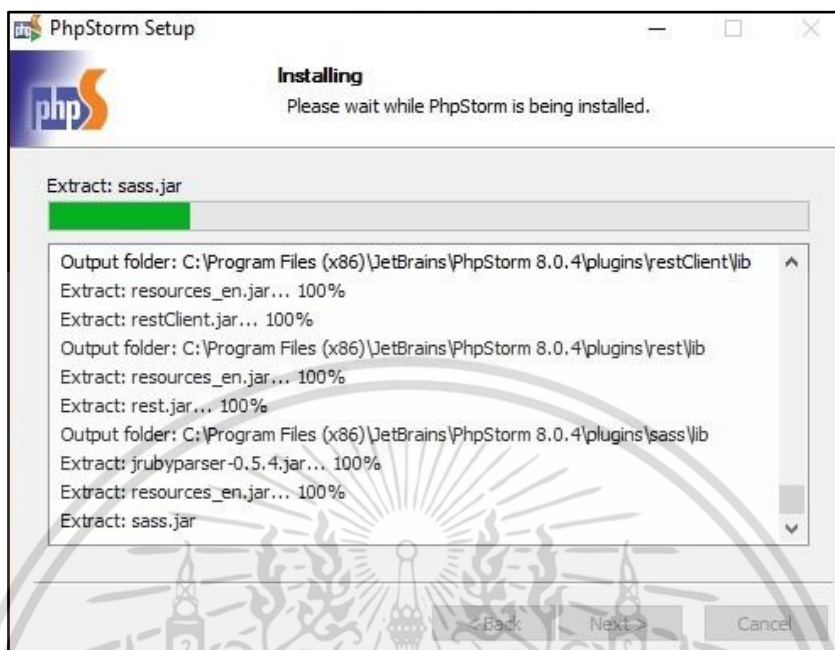
6) จะปรากฏหน้าจอสำหรับทำการสร้างโฟลเดอร์ในการติดตั้ง PhpStorm หลังจากนั้นทำการคลิกที่ Install เพื่อเริ่มการติดตั้ง ดังรูปที่ ก.6



รูปที่ ก.6 สร้างโฟลเดอร์ในการติดตั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7) จะปรากฏหน้าจอสำหรับแสดงรายละเอียดการติดตั้ง ให้รอจนกว่าโปรแกรมจะติดตั้งเสร็จสมบูรณ์ ดังรูปที่ ก.7



รูปที่ ก.7 แสดงรายละเอียดการติดตั้ง

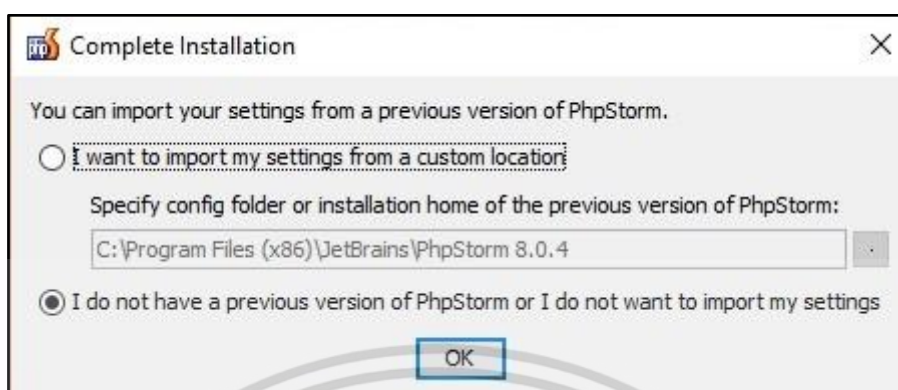
8) จะปรากฏหน้าจอสำหรับรายงานเมื่อทำการติดตั้ง PhpStorm เสร็จสมบูรณ์แล้ว หลังจากนั้นทำการคลิกที่ Finish เพื่อเสร็จสิ้นการติดตั้งและเริ่มต้นการใช้งาน ดังรูปที่ ก.8



รูปที่ ก.8 คลิก Finish เพื่อเสร็จสิ้นการติดตั้ง

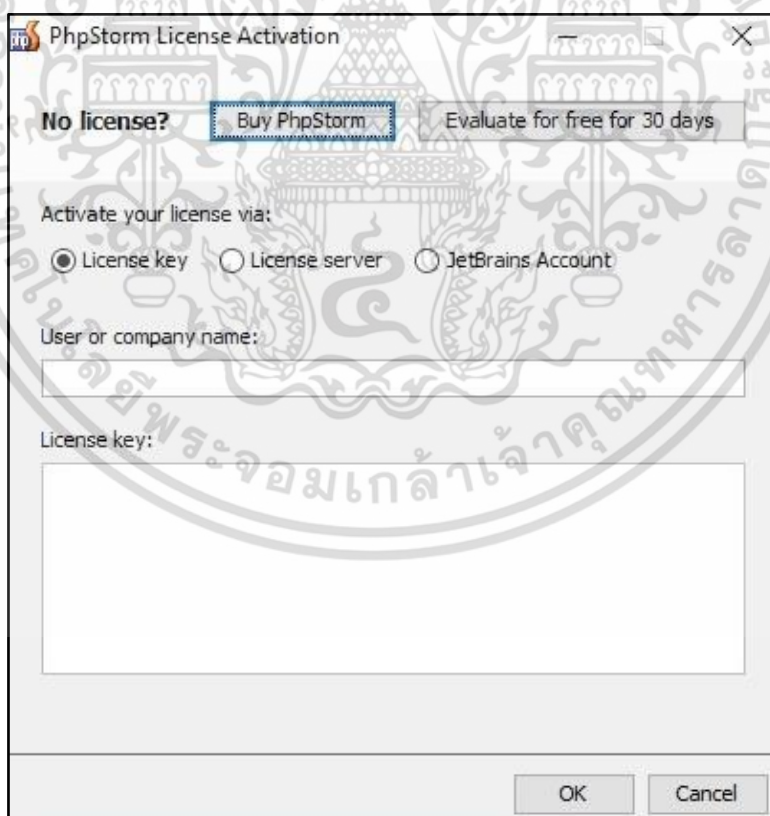
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9) จะปรากฏหน้าจอสำหรับนำเข้าหรือตั้งค่าเวอร์ชันของ PhpStorm เพื่อเข้าสู่ขั้นตอนต่อไป ดังรูปที่ ก.9



รูปที่ ก.9 นำเข้าหรือตั้งค่าเวอร์ชัน

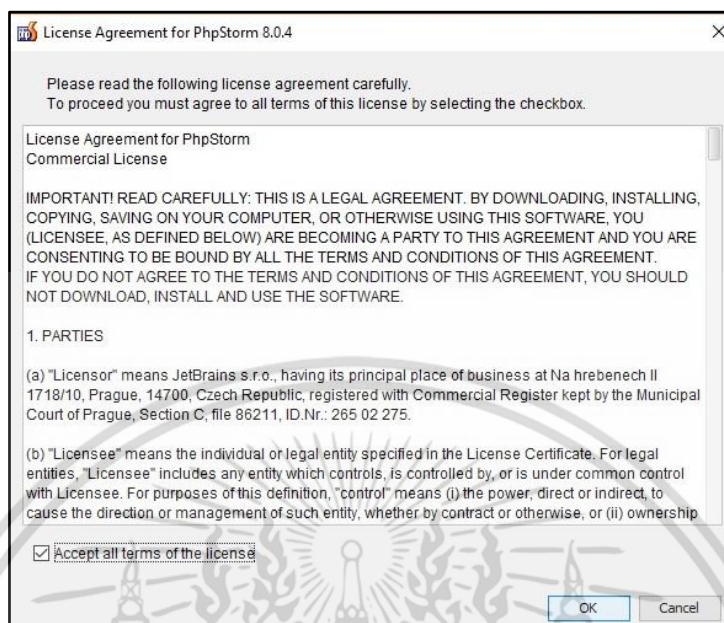
10) จะปรากฏหน้าจอสำหรับใส่คีย์ของโปรแกรม ให้ทำการใส่ชื่อและคีย์ของโปรแกรม หลังจากนั้นทำการคลิกที่ OK เพื่อเข้าสู่ขั้นตอนต่อไป ดังรูปที่ ก.10



รูปที่ ก.10 ทำการใส่ชื่อและคีย์ของโปรแกรม

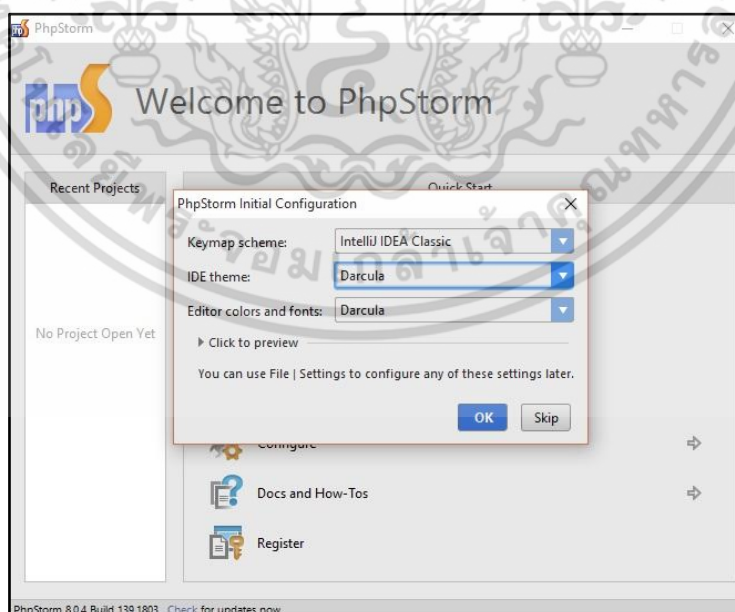
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

11) จะปรากฏหน้าต่างสำหรับแสดงรายละเอียดข้อตกลงในการใช้งานของ PhpStorm เพื่อเข้าสู่ขั้นตอนต่อไป ดังรูปที่ ก.11



รูปที่ ก.11 แสดงรายละเอียดข้อตกลงในการใช้งาน

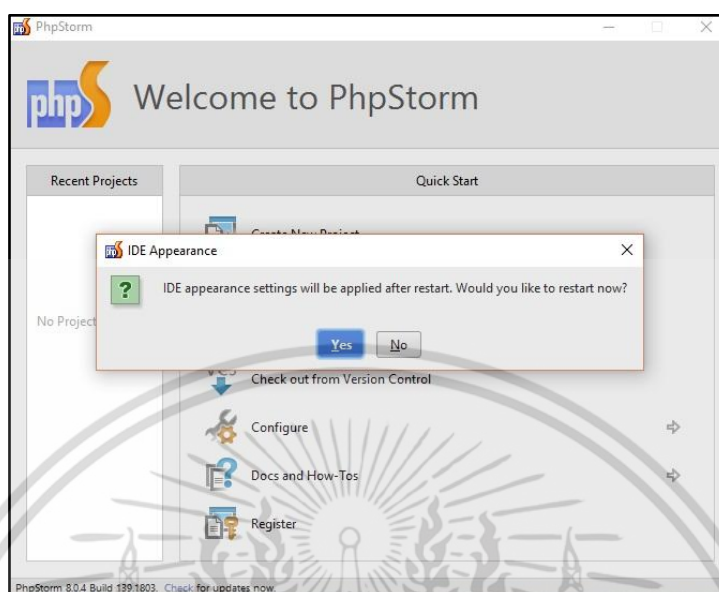
12) จะปรากฏหน้าต่างสำหรับตั้งค่าเบื้องต้นในการเริ่มต้นการใช้งานโปรแกรม หลังจากนั้นทำการคลิกที่ OK เพื่อเข้าสู่ขั้นตอนต่อไป ดังรูปที่ ก.12



รูปที่ ก.12 ตั้งค่าเบื้องต้นในการเริ่มต้นการใช้งานโปรแกรม

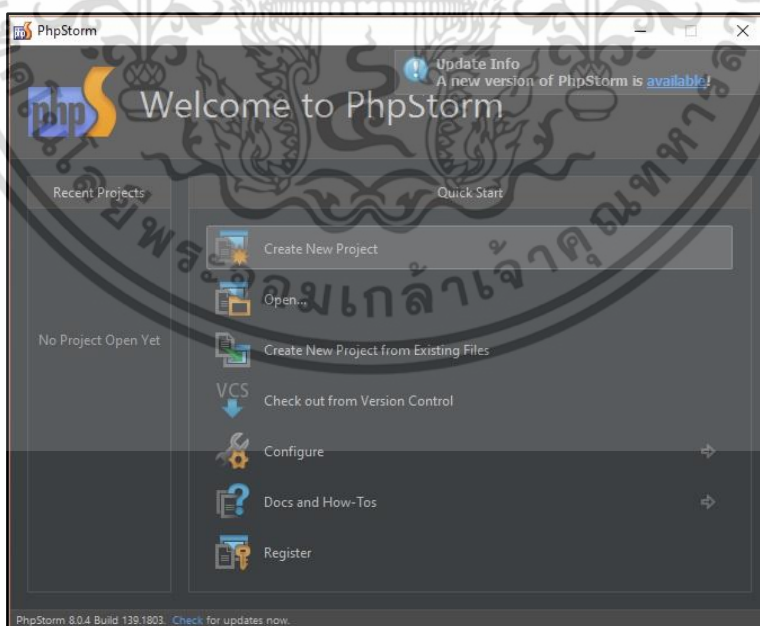
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

13) จะปรากฏหน้าจอสำหรับยืนยันการตั้งค่าเบื้องต้นพร้อมกับทำการเริ่มต้นโปรแกรมใหม่ หลังจากนั้นทำการคลิกที่ Yes เพื่อใช้งานโปรแกรม ดังรูปที่ ก.13



รูปที่ ก.13 ยืนยันการตั้งค่าเบื้องต้นพร้อมกับทำการเริ่มต้นโปรแกรมใหม่

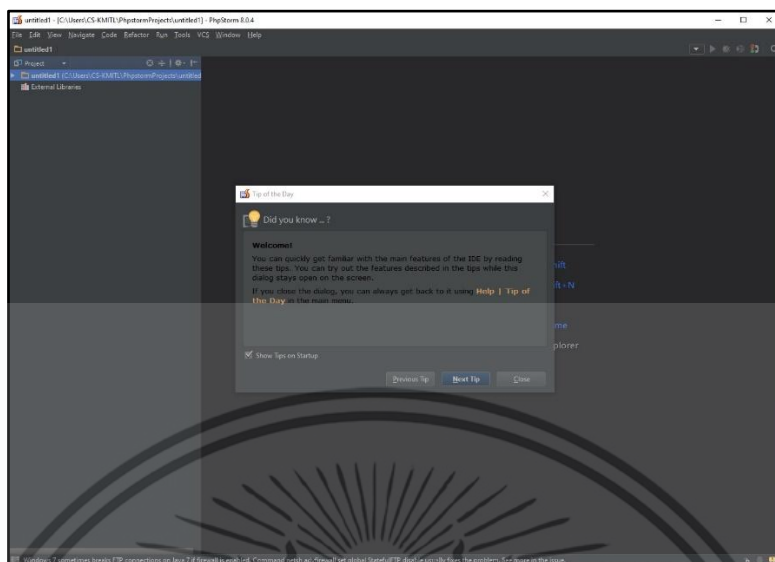
14) จะปรากฏหน้าจอสำหรับตัวเลือกในการเริ่มต้นการใช้งาน หลังจากนั้นทำการคลิกที่ Create New Project เพื่อสร้างโปรเจ็คใหม่สำหรับพัฒนาโปรแกรม ดังรูปที่ ก.14



รูปที่ ก.14 แสดงหน้าจอตัวเลือกในการเริ่มต้นการใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

15) จะปรากฏหน้าจอสำหรับพร้อมเริ่มต้นใช้งานในการพัฒนาโปรเจ็ค ดังรูปที่ ก.14



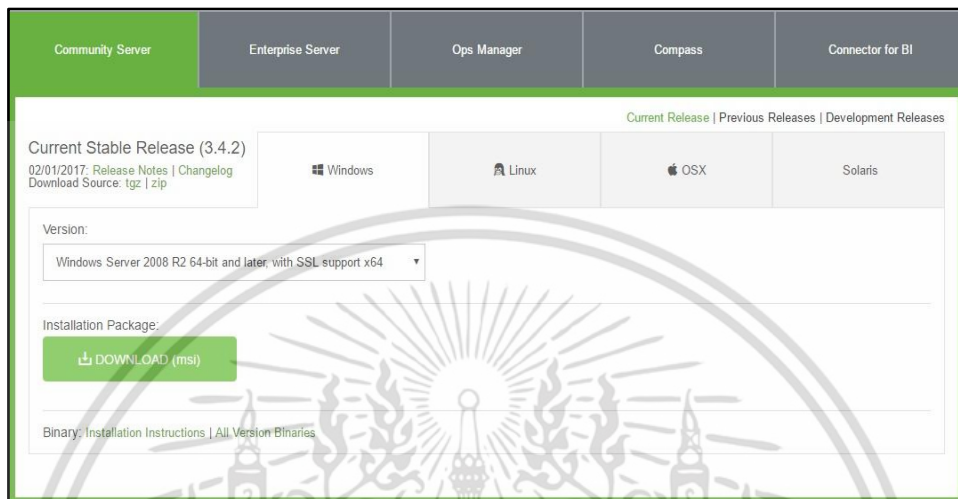
รูปที่ ก.15 แสดงหน้าจอเริ่มต้นใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

การติดตั้งโปรแกรม MongoDB

- 1) ทำการดาวน์โหลด MongoDB ตามเวอร์ชันที่ต้ง ดังรูปที่ ข.1



รูปที่ ข.1 ทำการดาวน์โหลด MongoDB

- 2) ทำการ Execute File Setup.exe ดังรูปที่ ข.2



รูปที่ ข.2 ทำการ Execute File Setup.exe

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) จะปรากฏหน้าจอสำหรับ Setup MongoDB ให้ทำการคลิกที่ Next เพื่อทำการเริ่มการติดตั้งต่อไป ดังรูปที่ ข.3



รูปที่ ข.3 คลิก Next เพื่อทำการเริ่มการติดตั้ง

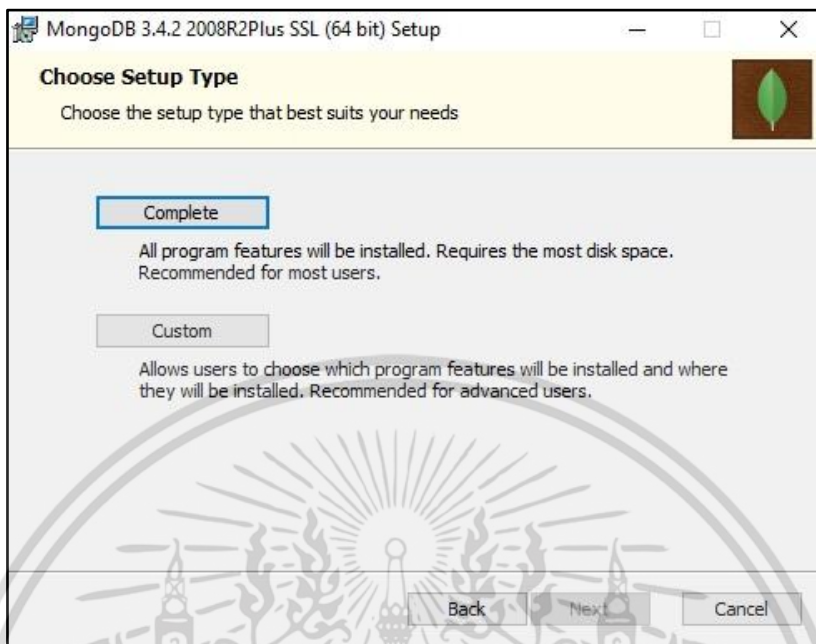
4) จะปรากฏหน้าจอสำหรับแสดงรายละเอียดในการ Setup MongoDB หลังจากนั้นทำการคลิกที่ Next เพื่อเข้าสู่ขั้นตอนต่อไป ดังรูปที่ ข.4



รูปที่ ข.4 แสดงรายละเอียดในการ Setup MongoDB

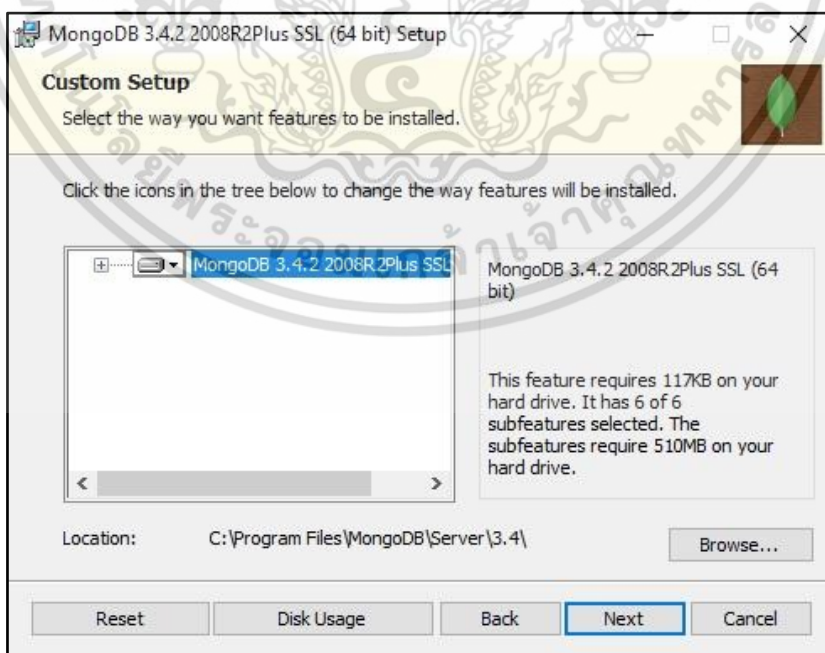
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) จะปรากฏหน้าจอสำหรับตัวเลือกในการติดตั้ง ในที่นี้เลือก Custom หลังจากนั้นทำการคลิกที่ Next เพื่อเข้าสู่ขั้นตอนต่อไป ดังรูปที่ ข.5



รูปที่ ข.5 เลือกตัวเลือกในการติดตั้ง

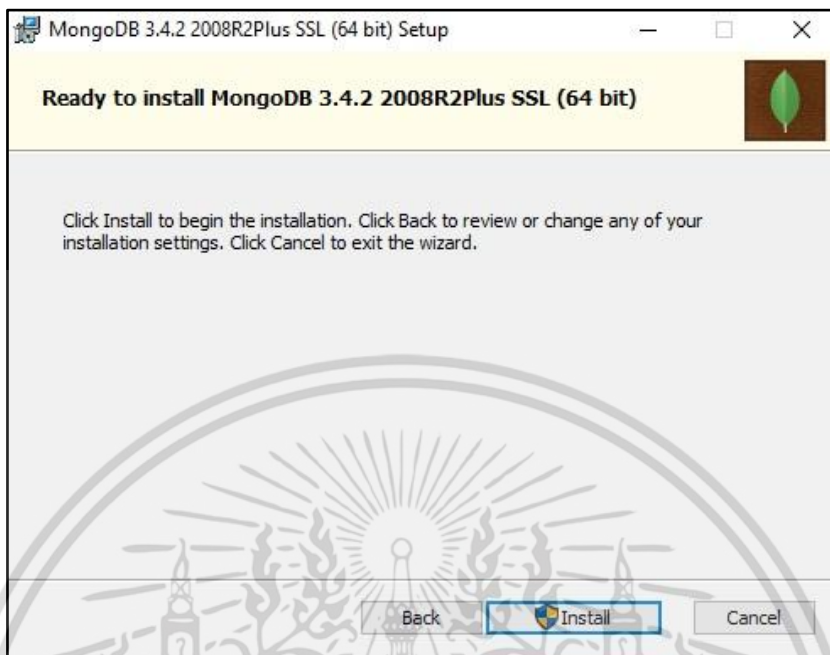
6) จะปรากฏหน้าจอสำหรับทำการสร้างโฟลเดอร์ในการติดตั้ง PhpStorm หลังจากนั้นทำการคลิกที่ Next เพื่อเริ่มการติดตั้ง ดังรูปที่ ข.6



รูปที่ ข.6 สร้างโฟลเดอร์ในการติดตั้ง

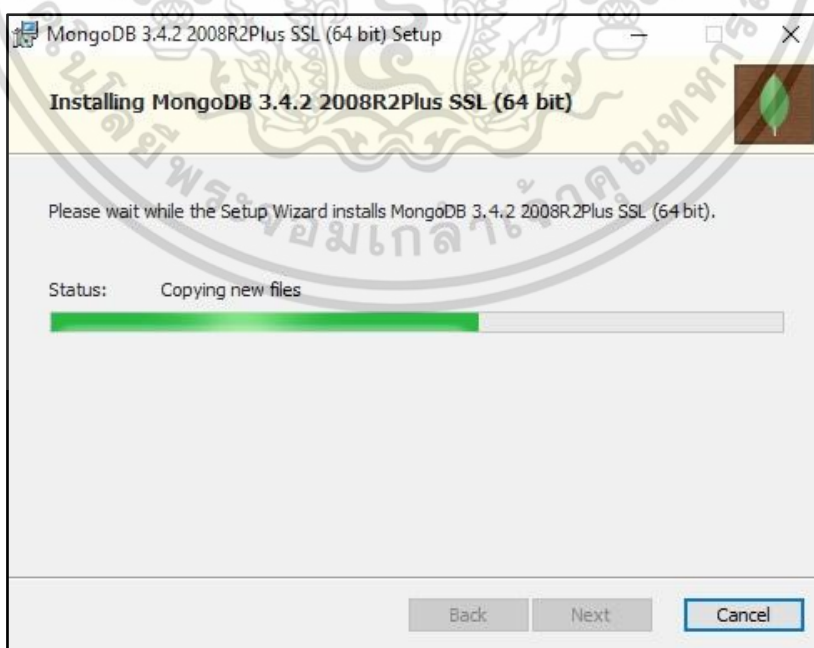
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7) จะปรากฏหน้าจอสำหรับติดตั้งโปรแกรม หลังจากนั้นทำการคลิกที่ Install เพื่อเริ่มการติดตั้ง ดังรูปที่ ข.7



รูปที่ ข.7 คลิก Install เพื่อเริ่มการติดตั้ง

8) จะปรากฏหน้าจอสำหรับแสดงรายละเอียดการติดตั้ง ให้รอจนกว่าโปรแกรมจะติดตั้งเสร็จสมบูรณ์ ดังรูปที่ ข.8



รูปที่ ข.8 แสดงรายละเอียดการติดตั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9) จะปรากฏหน้าจอสำหรับรายงานเมื่อทำการติดตั้งเสร็จสมบูรณ์แล้ว หลังจากนั้นทำการคลิกที่ Finish เพื่อเสร็จสิ้นการติดตั้งและเริ่มต้นการใช้งาน ดังรูปที่ ข.9



รูปที่ ข.9 คลิก Finish เพื่อเสร็จสิ้นการติดตั้ง

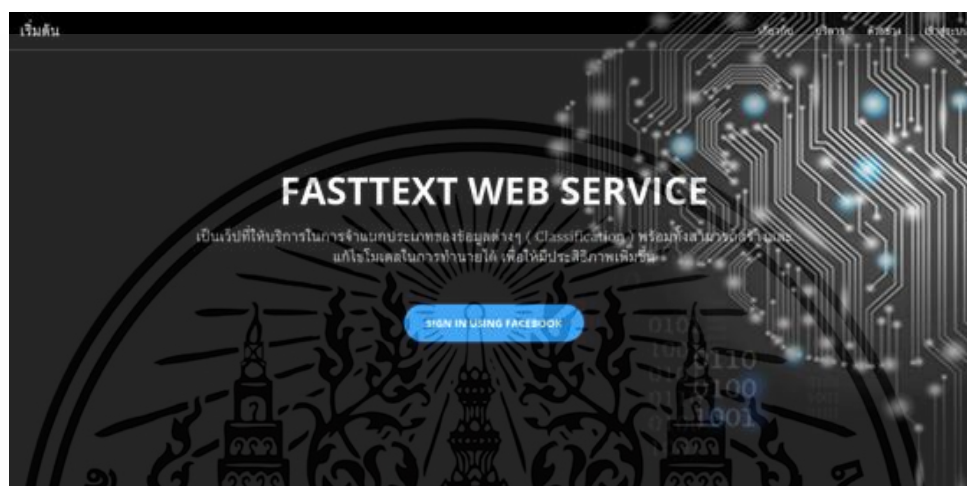
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค

คู่มือการใช้งานเว็บเซอร์วิส (FASTTEXT WEB SERVICE)

1) หน้าแรกของเว็บเซอร์วิส

หน้าแรกของเว็บเซอร์วิส เป็นหน้าจอที่อธิบายว่าเว็บเซอร์วิสนี้ให้บริการเกี่ยวกับอะไร มีเซอร์วิสอะไรบ้าง โดยจะมีเมนูตัวเลือกดังนี้ เกี่ยวกับเรา , เซอร์วิส , ตัวอย่าง และ เข้าสู่ระบบ



รูปที่ ค.1 หน้าแรกของเว็บเซอร์วิส

2) หน้าแรกของเว็บเซอร์วิส เมนู เกี่ยวกับเรา

ในหน้า เกี่ยวกับเรา เป็นหน้าจอที่อธิบายว่าเว็บเซอร์วิสนี้เป็นเว็บที่ใช้ทำอะไร ให้บริการเกี่ยวกับอะไรและทำอะไรได้บ้าง



รูปที่ ค.2 หน้าแรกของเว็บเซอร์วิส เมนู เกี่ยวกับเรา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) หน้าแรกของเว็บเซอร์วิส เมนู เซอร์วิส

ในหน้า เซอร์วิส เป็นหน้าจอที่อธิบายว่าเว็บเซอร์วิสนี้มีเซอร์วิสอะไรที่ให้บริการบ้าง คือจำแนกประเภทของข้อมูล ใช้งานโมเดล และทดสอบโมเดล



รูปที่ ค.3 หน้าแรกของเว็บเซอร์วิส เมนู เซอร์วิส

4) หน้าแรกของเว็บเซอร์วิส เมนู ตัวอย่าง

ในหน้า ตัวอย่าง เป็นหน้าจอที่แสดงภาพตัวอย่างของการใช้งานเว็บเซอร์วิส

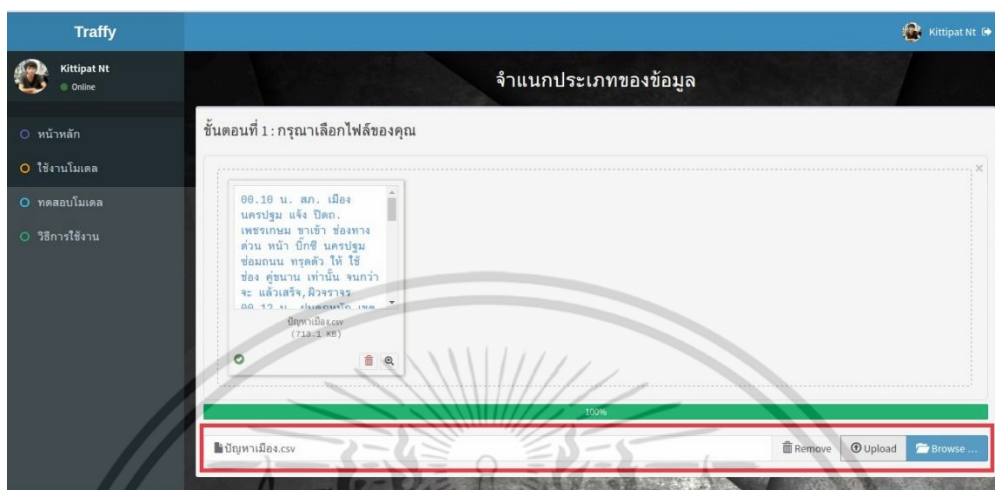


รูปที่ ค.4 หน้าแรกของเว็บเซอร์วิส เมนู ตัวอย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) หน้าจอหลัก

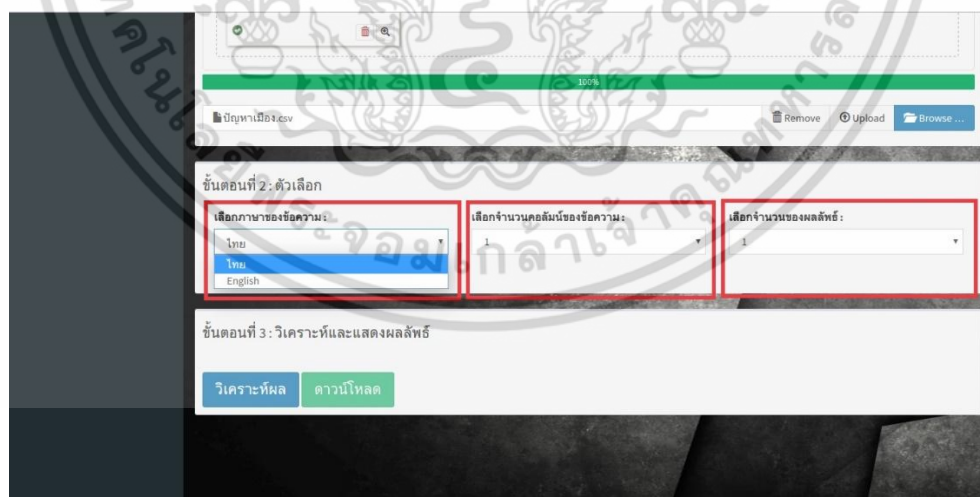
ในหน้านี้ เป็นขั้นตอนที่ 1 ผู้ใช้สามารถนำไฟล์ CSV ที่มีรูปแบบตามที่กำหนด คือจะต้องมีคอลัมน์ที่เป็นข้อความและคอลัมน์ที่เป็นผลเฉลย มาทำการจำแนกประเภทของข้อมูลได้



รูปที่ ค.5 หน้าจอหลักขั้นตอนที่ 1

6) หน้าจอหลัก

ในหน้านี้ เป็นขั้นตอนที่ 2 ผู้ใช้สามารถเลือกตัวเลือกในการวิเคราะห์ผลที่ต้องการ โดยจะมีตัวเลือกภาษาของข้อมูลและคอลัมน์ตามไฟล์ CSV ที่มีรูปแบบตามที่กำหนดที่ผู้ใช้โฟลด์

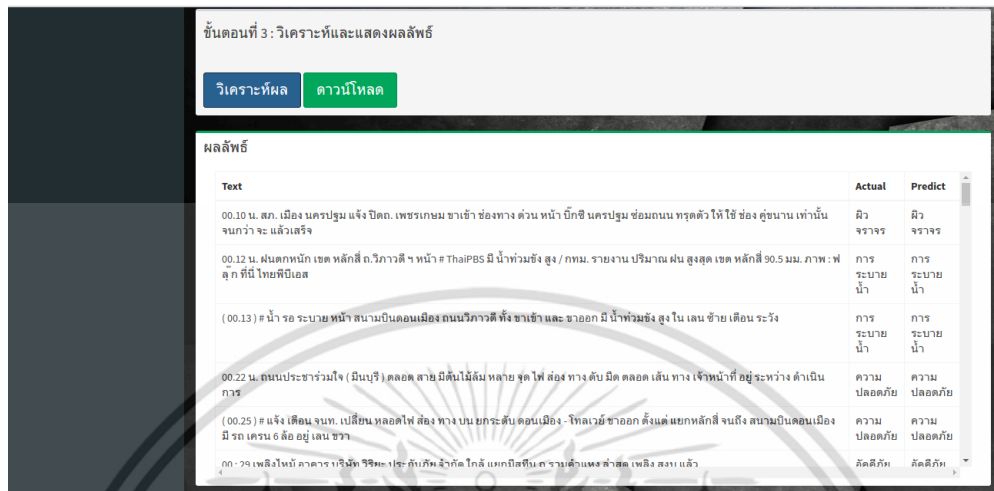


รูปที่ ค.6 หน้าจอหลักขั้นตอนที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7) หน้าจอหลัก

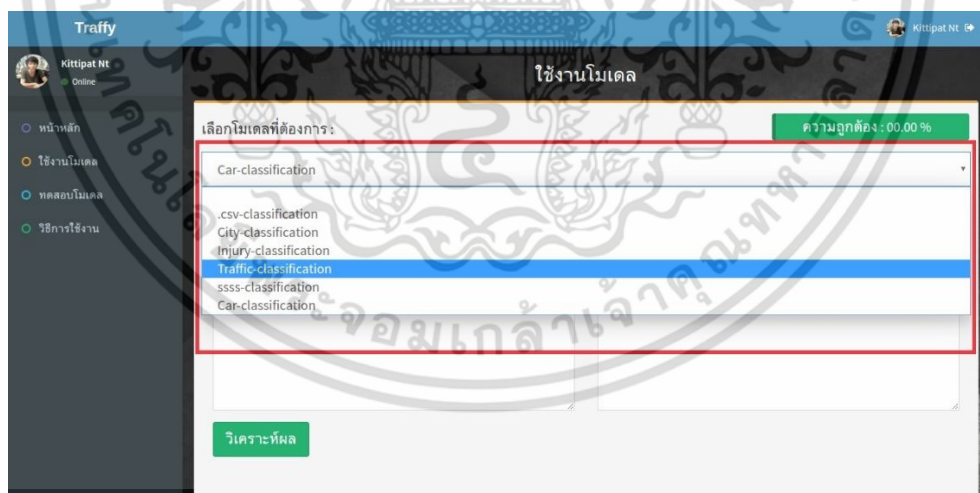
ในหน้านี้ เป็นขั้นตอนที่ 3 ผู้ใช้สามารถทำการวิเคราะห์และแสดงผลลัพธ์ หลังจากเลือกตัวเลือกในการวิเคราะห์ผลต้องที่ต้องการแล้ว



รูปที่ ค.7 หน้าจอหลักขั้นตอนที่ 3

8) หน้าใช้งานโมเดล

ในหน้านี้ เป็นขั้นตอนที่ 1 ผู้ใช้สามารถเลือกโมเดลในการทดสอบได้ว่าต้องการใช้งานโมเดลใดในการจำแนกประเภทของข้อมูล

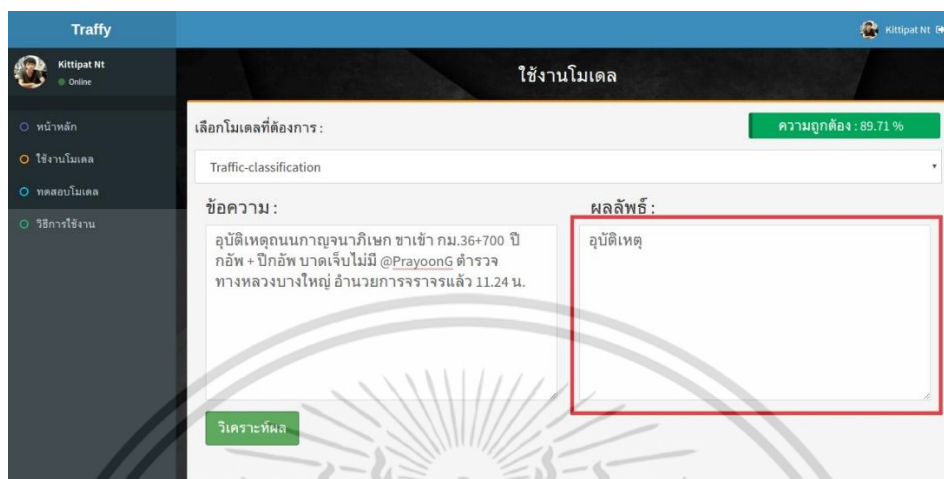


รูปที่ ค.8 หน้าใช้งานโมเดลขั้นตอนที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9) หน้าใช้งานโมเดล

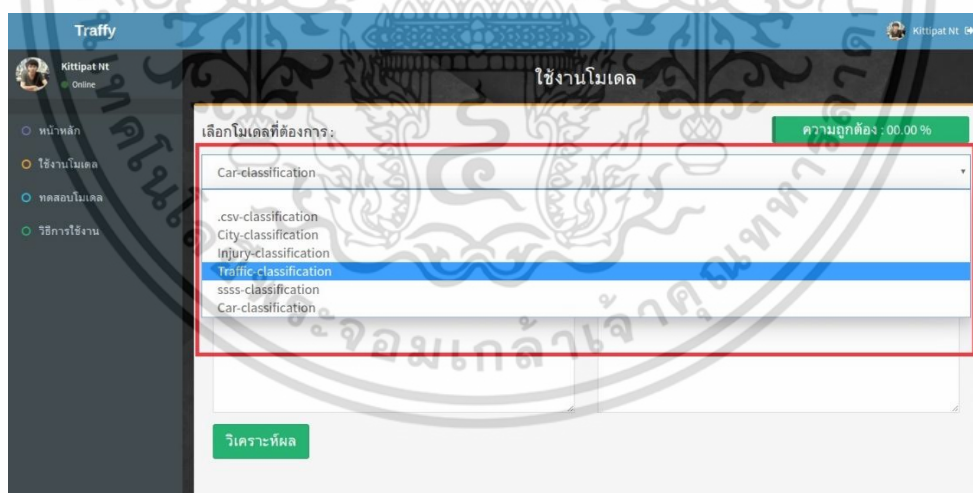
ในหน้านี้ เป็นขั้นตอนที่ 2 ผู้ใช้สามารถนำข้อความมาใส่ในช่องข้อความ จากนั้นทำการวิเคราะห์ผลเพื่อจำแนกประเภทของข้อมูลพร้อมทั้งแสดงความถูกต้องของโมเดลที่ใช้



รูปที่ ค.9 หน้าใช้งานโมเดลขั้นตอนที่ 2

10) หน้าทดสอบโมเดล

ในหน้านี้ เป็นขั้นตอนที่ 1 ผู้ใช้สามารถเลือกโมเดลในการทดสอบได้ว่าต้องการใช้งานโมเดลใดในการจำแนกประเภทของข้อมูล

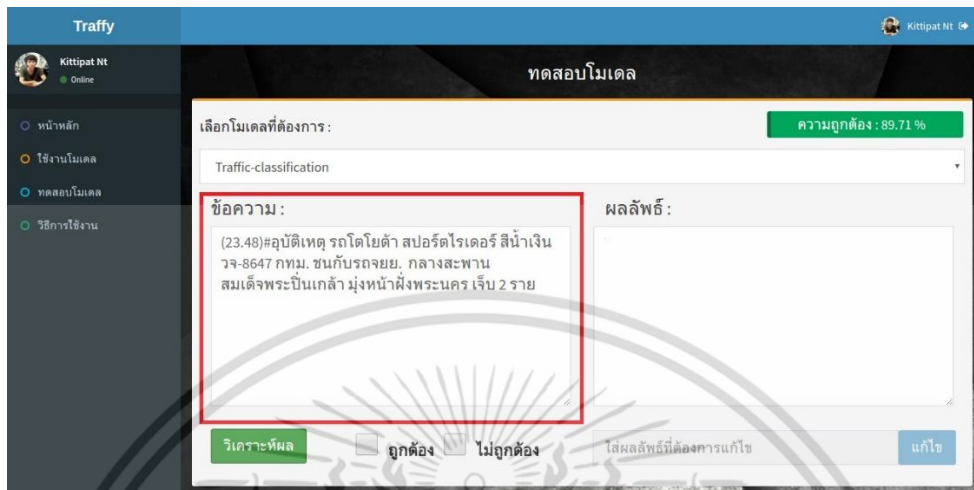


รูปที่ ค.10 หน้าทดสอบโมเดลขั้นตอนที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

11) หน้าทดสอบโมเดล

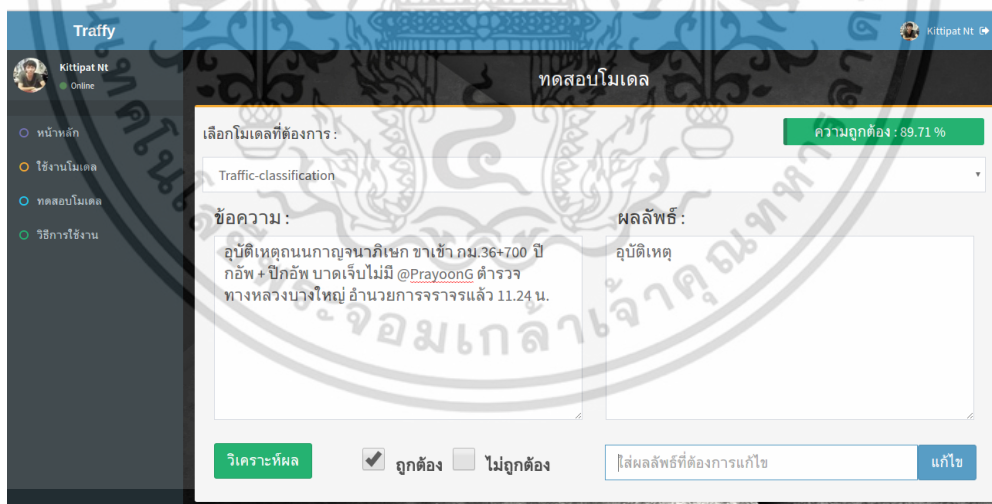
ในหน้านี้ เป็นขั้นตอนที่ 2 ผู้ใช้สามารถนำข้อความมาใส่ในช่องข้อความ จากนั้นทำการวิเคราะห์ผลเพื่อจำแนกประเภทของข้อมูลพร้อมทั้งแสดงความถูกต้องของโมเดลที่ใช้



รูปที่ ค.11 หน้าใช้งานโมเดลขั้นตอนที่ 2

12) หน้าทดสอบโมเดล

ในหน้านี้ เป็นขั้นตอนที่ 3 ผู้ใช้สามารถเลือกโมเดลในการทดสอบได้เมื่อผลของการทำนายไม่ถูกต้องพร้อมทั้งสามารถให้โมเดลเรียนรู้ใหม่ได้



รูปที่ ค.12 หน้าทดสอบโมเดลขั้นตอนที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

13) หน้าวิธีการใช้งาน

ในหน้า วิธีการใช้งานนี้จะอธิบายขั้นตอนในการใช้งานระบบโดยแบ่งออกเป็นสามขั้นตอน ดังนี้ขั้นตอนที่หนึ่งคือการเตรียมข้อมูลในไฟล์ตามรูปแบบที่กำหนด เลือกตัวเลือกในการวิเคราะห์ ผลลัพธ์ การวิเคราะห์และแสดงผลลัพธ์

วิธีการใช้งาน

ขั้นตอนที่ 1 : เลือกไฟล์สกุล Comma Separated Values (CSV) โดยต้องมีรูปแบบข้อมูลข้อมูล ดังนี้

การเตรียมข้อมูลจะต้องประกอบด้วยคอลัมน์ที่เป็นข้อความ (Text) ซึ่งมีได้มากกว่าหนึ่งคอลัมน์ และคอลัมน์ที่เป็นผลเฉลยของข้อความ (text) ซึ่งจะใช้ผลเฉลยนี้มาเป็นต้นแบบของการจำแนกประเภท (label) โดยไฟล์ข้อมูลนี้ต้องมี Header ไฟล์

ตัวอย่าง

ถ.วิภาวดีรังสิต ช้างแยกหลักสี่ แยกบางเขน มีทั้งหมดสูงประมาณ 10-20 ซม. ในช่องทางด่วนและคู่ขนาน	น้ำท่วมขัง
06:23 น. บนทางด่วนชั้น 2 แจ้งวัฒนะมุ่งหน้าลงประตูน้ำและลงพระราม 6 รถมากเคลื่อนตัวช้าๆตามเลนเลนบนแจ้งวัฒนะ	จราจรติดขัด
ถ.อโศกดินแดงช่วงแยกโบสถ์แม่พระ > ใต้ทางด่วนดินแดงแยกโบสถ์แม่พระเล็กน้อย รถเก๋งจอดเลี้ยวขวาช่องทางขวา	รถเสีย
06:28 น. ถนนมอเคอร์เวย์ (ท.ล. 7) ค่าระดับกับข้างรถติดที่ขาเข้าและขาออก...	จราจรติดขัด
แยกบางนาเลี้ยวซ้ายไปลำโพงรถแทรกเลนตรงเลนจอดคนออกป้าย วันว่างสวนเลนคนข้ามไม่ใช้สะพาน	ขนส่งสาธารณะ
07:55 น. ถนนวิภาวดีรังสิตฝั่งขาออกตรงข้าม นสพ. เติบูนวิดิ้งตรงมีน้ำท่วมขังช่องทางซ้าย	น้ำท่วมขัง
09 :06 ถนนพระราม 2 ขาออก ไปต่างระดับบริเวณนาฬิกา 77 รถพ่วง 18 ล้อเสียหลักพลิกคว่ำแดง	อุบัติเหตุ

รูปที่ ค.13 หน้าวิธีการใช้งาน

ขั้นตอนที่ 2 : เลือกตัวเลือกของการทำนาย

- ภาษาของข้อความ
- จำนวนคอลัมน์ของข้อความ
- จำนวนผลลัพธ์

ระบุคอลัมน์ที่ใช้เป็นข้อความ เป็นข้อความภาษาไทยหรือภาษาอังกฤษ เพื่อนำมาใช้ในการตัดคำ
ระบุจำนวนของคอลัมน์ที่เป็นข้อความ (Text) สามารถมีได้มากกว่าหนึ่งคอลัมน์
ระบุจำนวนของผลลัพธ์ที่ต้องการทำนาย

ขั้นตอนที่ 3 : ทำการวิเคราะห์และแสดงผลลัพธ์

ทำการวิเคราะห์ผลลัพธ์ จะแสดงตารางข้อมูลพร้อมทั้งผลลัพธ์จากการทำนาย และสามารถดาวน์โหลดไฟล์ผลลัพธ์เป็น ไฟล์สกุล Comma Separated Values (CSV)

รูปที่ ค.14 หน้าวิธีการใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้