



การพัฒนารูปแบบการรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่
กรุงเทพมหานครจากข้อมูลทวีตเตอร์ด้วยวิธีการเรียนรู้เชิงลึก
The Collection of Road Traffic Incidents in Bangkok
from Twitter Data based on Deep Learning Algorithm

กร พวงนาค
KORN PUANGNAK

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร วิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน คณะวิศวกรรมศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

2565

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

ชื่อวิทยานิพนธ์ การพัฒนารูปแบบการรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่
กรุงเทพมหานครจากข้อมูลทวีเตอร์ด้วยวิธีการเรียนรู้เชิงลึก
ชื่อ นามสกุล กร พวงนาค
ชื่อปริญญา วิศวกรรมศาสตรดุษฎีบัณฑิต (วิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน)
สาขาวิชา วิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน
คณะ วิศวกรรมศาสตร์
อาจารย์ที่ปรึกษา ดร.ณัฐวรพล รัชสิริวัชรบุล

คณะกรรมการสอบวิทยานิพนธ์ได้ให้ความเห็นชอบวิทยานิพนธ์ฉบับนี้แล้ว

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร.ประมุข อุณหเลขกะ)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ณัฐพงศ์ พันธุ์นะ)

..... กรรมการ

(ดร.ปริญญ์ บุญกนิษฐ)

..... กรรมการ

(ดร.เทอดพงษ์ แดงสี)

..... กรรมการ

(ดร.ณัฐวรพล รัชสิริวัชรบุล)

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็น
ส่วนหนึ่งของการศึกษาตามหลักสูตร วิศวกรรมศาสตรดุษฎีบัณฑิต สาขาวิชาวิศวกรรมการจัดการ
อุตสาหกรรมเพื่อความยั่งยืน มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

..... คณบดีคณะวิศวกรรมศาสตร์

(ผู้ช่วยศาสตราจารย์ ดร.ณัฐพงศ์ พันธุ์นะ)

วันที่ 27 เดือน กันยายน พ.ศ. 2565

ชื่อวิทยานิพนธ์	การพัฒนารูปแบบการรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่กรุงเทพมหานครจากข้อมูลทวีตเตอร์ด้วยวิธีการเรียนรู้เชิงลึก
ชื่อ นามสกุล	กร พวงนาค
ชื่อปริญญา	วิศวกรรมศาสตรดุษฎีบัณฑิต (วิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน)
สาขาวิชา และคณะ	วิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน คณะวิศวกรรมศาสตร์
ปีการศึกษา	2565

บทคัดย่อ

การเฝ้าระวังและรายงานอุบัติเหตุเพื่อการบริหารจัดการจราจรในปัจจุบัน ดำเนินการโดยการเฝ้าระวังและตรวจสอบโดยการดูภาพเหตุการณ์จากกล้องวงจรปิด ซึ่งเป็นวิธีการที่ตรวจหาเหตุการณ์อย่างล่าช้า และต้องใช้จำนวนคนปฏิบัติงานจำนวนมาก ด้วยปัจจุบันเทคโนโลยีการสื่อสารและสังคมออนไลน์เข้ามา มีบทบาทต่อการรายงานเหตุดังกล่าว จึงเกิดงานวิจัยในการตรวจสอบเหตุการณ์อุบัติเหตุที่สนใจ จากข้อความที่รายงานผ่านสื่อสังคมออนไลน์สำหรับภาษาไทย ที่ในปัจจุบันยังไม่มีการวิจัยหรือแพลตฟอร์มในการทำงานสนับสนุนการทำงานรูปแบบนี้อย่างชัดเจน

งานวิจัยนี้นำเสนอการพัฒนาการเรียนรู้อัจฉริยะแบบโครงข่ายประสาทเทียม เพื่อการแก้ปัญห การจำแนกรูปแบบอุบัติการณ์และการระบุความรุนแรงของอุบัติการณ์จากข้อความบนสื่อสังคมออนไลน์ ภาษาไทย โดยใช้รูปแบบการเรียนรู้เชิงลึกแบบ MLP, CNN, Bi-LSTM และ LSTM+CNN ซึ่งออกแบบ โดยแบ่งการศึกษาออกเป็น 3 รูปแบบ ประกอบด้วย การตรวจสอบ รูปแบบการระบุการเกิดอุบัติการณ์ จราจร ที่สามารถระบุการรายงานว่าเป็นข่าวสารทั่วไปหรือการรายงานข้อมูลทางด้านจราจร รูปแบบ การระบุรูปแบบของอุบัติการณ์ ประกอบไปด้วย สภาพจราจร อุบัติเหตุ การเกิดภัยพิบัติ ทางชำรุด หรือ อื่น ๆ นอกเหนือจากรูปแบบที่กล่าวถึง และรูปแบบการระบุระดับความรุนแรงของอุบัติการณ์ ประกอบไปด้วย ระดับปกติ ระดับกลาง ระดับปิดกั้นช่องจราจรหรือหยุดนิ่ง โดยผลการทดลองแสดงให้เห็นความสามารถ ของการเรียนรู้แบบ CNN+LSTM ที่มีผลลัพธ์ในการตรวจจับการเกิดอุบัติการณ์ได้ดีที่สุดเท่ากับร้อยละ 93.44 และ CNN มีผลลัพธ์ในการระบุรูปแบบของอุบัติการณ์ได้ดีที่สุดเท่ากับร้อยละ 85.29 และวิธีการ LSTM สามารถระบุระดับความรุนแรงของอุบัติการณ์ได้ดีที่สุดเท่ากับร้อยละ 88.53

คำสำคัญ : โครงข่ายประสาทเทียม, การเรียนรู้เชิงลึก, ทวีตเตอร์, อุบัติการณ์, สื่อสังคมออนไลน์

Thesis Title	The Collection of Road Traffic Incidents in Bangkok from Twitter Data based on Deep Learning Algorithm
Author	Korn Puangnak
Degree	Doctor of Engineering (Sustainable Industrial Management Engineering)
Major Program	Sustainable Industrial Management Engineering Faculty of Engineering
Academic Year	2022

ABSTRACT

The surveillance and reports on accidents for the current traffic management are carried out by viewing and inspecting CCTV footage. This method is slow to detect incidents and requires a lot of manpower. Since nowadays communication technology and social media play a role in reporting such incidents, this research has been conducted to investigate accidents of interest from text messages reported on social media in Thai. There is currently no research or platform that supports this type of work clearly.

This research presented the development of a neuron network memory-learning model to solve the problem of incidence pattern classification and incidence severity identification from social media messages in Thai. Using deep learning models, such as MLP, CNN, Bi-LSTM, and LSTM+CNN, the study was designed and divided into three experimental patterns. This included examining patterns to identify traffic incidents that may be reported as general news or traffic reporting; examining patterns to indicate the type of incidents, such as traffic accidents, disasters, damaged roads, or others other than those mentioned; and examining patterns to specify the severity levels of the incidents, such as normal, intermediate, traffic lane blocking, or immovable. The experiments demonstrated the capability of CNN+LSTM learning models with the best incidence detection result at 93.44 %. The CNN model gave the best result in identifying a pattern of incidence at 85.29 %, and the LSTM model was best able to determine the severity levels of the incidence at 88.53 %.

Keywords : Neuron Network, Deep Learning, Twitter, Incident, Social Media

กิตติกรรมประกาศ

การศึกษาวิจัยเพื่อการจัดทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความอนุเคราะห์ของบุคคลหลายท่าน ซึ่งไม่อาจจะนำมากล่าวได้ทั้งหมด ผู้วิจัยใคร่ขอกราบขอบพระคุณ ดร.ณัฐวรพล รัชสิริวัชรบุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้ความรู้ คำแนะนำ ตรวจสอบ และแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเอาใจใส่ในทุกขั้นตอน ติดตามความก้าวหน้าในการดำเนินการวิจัย เพื่อให้การเขียนวิทยานิพนธ์ฉบับนี้มีความสมบูรณ์อย่างที่สุด

ขอขอบคุณ เจ้าหน้าที่อำนวยการจราจร ศูนย์ควบคุมจราจรลาดกระบ้ง กองทางหลวงพิเศษระหว่างเมือง กรมทางหลวง และวิศวกร กองวิจัยและพัฒนาวิศวกรรมระบบทางพิเศษ การทางพิเศษแห่งประเทศไทย ที่ให้ความอนุเคราะห์ในการร่วมคัดกรองข้อมูลจราจรสำหรับการเรียนรู้และทดสอบประสิทธิภาพระบบงาน จนเป็นที่ลุล่วงและประสบผลสำเร็จดังผลงานวิจัยที่ปรากฏ

ขอขอบคุณพระคุณบิดา นายเกื้อกุล พวงนาค มารดา นางพรสวรรค์ พวงนาค และครอบครัวของผู้วิจัยที่อยู่เบื้องหลังในความสำเร็จ ที่ได้ให้ความช่วยเหลือสนับสนุน และให้กำลังใจตลอดมา ตลอดจนบุคคลต่าง ๆ ที่ให้ความช่วยเหลืออีกมาก ที่ผู้วิจัยไม่สามารถกล่าวนามได้ทั้งหมดในที่นี้ ผู้วิจัยจึงขอกราบขอบพระคุณเป็นอย่างสูง มา ณ โอกาสนี้

กร พวงนาค



สารบัญ

	หน้า
บทคัดย่อ	(ก)
Abstract	(ข)
กิตติกรรมประกาศ	(ค)
สารบัญ	(ง)
สารบัญตาราง	(ฉ)
สารบัญภาพ	(ช)
1. บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์	2
1.3 ขอบเขตของการศึกษา	3
1.4 สมมติฐาน	3
1.5 กรอบแนวความคิด	4
1.6 ประโยชน์ที่คาดว่าจะได้รับ	7
1.7 นิยามศัพท์	8
1.8 คำสำคัญ	8
2. แนวคิดทฤษฎีที่เกี่ยวข้องและการทบทวนวรรณกรรม	9
2.1 แนวคิดทฤษฎีที่เกี่ยวข้อง	9
2.2 การทบทวนวรรณกรรม	16
3. วิธีดำเนินการ	26
3.1 กรอบแนวคิดการออกแบบ	26
3.2 วิธีการและขั้นตอนการดำเนินงาน	27
3.3 กระบวนการพัฒนา	41
4. กระบวนการทดสอบ	45
4.1 สมมติฐาน	45
4.2 การจัดเตรียมข้อมูล	46
4.3 รูปแบบและขั้นตอนการทดสอบ	48
4.4 การประเมินประสิทธิภาพ	50
5. ผลการวิจัยและพัฒนา	52
5.1 การทดสอบการจำแนกการเกิดอุบัติเหตุการณ์	52
5.2 การทดสอบการจำแนกประเภทของอุบัติเหตุการณ์	56
5.3 การทดสอบการระบุความรุนแรงของอุบัติเหตุการณ์	60

สารบัญ (ต่อ)

	หน้า
6 อภิปรายผล	64
7 สรุปผล	66
7.1 สรุปผล	66
7.2 แนวทางการศึกษาและพัฒนาต่อ	67
8 แผนการนำไปใช้ประโยชน์เชิงพาณิชย์	68
8.1 รูปแบบสำคัญเพื่อดำเนินการทางธุรกิจ	68
8.2 แผนการใช้ประโยชน์ของผลิตภัณฑ์ในเชิงธุรกิจ	72
8.3 การวิเคราะห์ผลตอบแทนทางเศรษฐศาสตร์	73
เอกสารอ้างอิง	75
ภาคผนวก	80
ภาคผนวก ก เอกสารตีพิมพ์	81
ภาคผนวก ก-1 เอกสารตีพิมพ์ ECTI-CIT Transactions - Volume 16 Number 3 - June 2022	82
ภาคผนวก ก-2 เอกสารตอบรับบทความ (Accepted Letter) ECTI-CIT Transactions - Volume 16 - Number 3 - June 2022	83
ภาคผนวก ก-3 เอกสารตีพิมพ์ ECTI-CIT Transactions - Volume 16 Number 3 - August 2022	95
ภาคผนวก ก-4 เอกสารตอบรับบทความ (Accepted Letter) ECTI-CIT Transactions - Volume 16 - Number 3 - August 2022	96
ภาคผนวก ข รายชื่อผู้เชี่ยวชาญ	109
ภาคผนวก ค สถาปัตยกรรมระบบ	110
ประวัติการศึกษาและการทำงาน	112

สารบัญตาราง

ตาราง		หน้า
1.1	การระบุการเกิดอุบัติเหตุการฉ้อโกง	5
1.2	รูปแบบของอุบัติเหตุการฉ้อโกง	6
1.3	ระดับความรุนแรงของอุบัติเหตุการฉ้อโกง	6
1.4	รูปแบบโครงสร้างโครงข่ายประสาทเทียมภายในของการเรียนรู้เชิงลึก	7
3.1	การลำดับความถี่ค่าในการแสดงสถานการณ์	30
3.2	การลำดับความถี่ค่าในการจำแนกสถานการณ์	31
3.3	การลำดับความถี่ค่าที่แสดงความรุนแรงที่ส่งผลกระทบต่อการฉ้อโกง	33
3.4	การลำดับจำนวนค่าที่มีผลต่อการแยกประเภทของข้อความ	36
4.1	ข้อมูลผลการทดสอบการตรวจสอบ	46
4.2	เปรียบเทียบความแม่นยำเมื่อมีการแบ่งข้อมูลต่างกัน	50
6.1	ข้อมูลผลการทดสอบการตรวจสอบอุบัติเหตุการฉ้อโกงด้วยวิธีการต่าง ๆ	64
6.2	ข้อมูลผลการทดสอบเพิ่มประสิทธิภาพของแบบจำลอง	65
8.1	คุณสมบัติของเครื่องแม่ข่าย VPS สำหรับการติดตั้งระบบและให้บริการ	71
8.2	รายละเอียดต้นทุนการพัฒนาและจัดเตรียมโครงสร้างพื้นฐานในการให้บริการระบบต่อปี	73
8.3	การเปรียบเทียบการลงทุนระหว่างการติดตั้งอุปกรณ์แบบเดิมกับการใช้ระบบงานใหม่	74

สารบัญภาพ

ภาพ	หน้า
2.1	การจัดอันดับความนิยมการใช้งานของสื่อสังคมออนไลน์ในประเทศไทย..... 10
2.2	โครงสร้างและองค์ประกอบของโครงข่ายประสาทเทียม 12
2.3	โครงข่ายประสาทเทียมแบบการเรียนรู้เชิงลึก (Deep Neural Networks: DNN).... 13
2.4	สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชัน 15
2.5	โครงข่ายประสาทเทียมแบบรีเคอร์เรนต์ 15
2.6	รูปแบบโนดภายในของโครงข่ายประสาทเทียมแบบรีเคอร์เรนต์..... 16
3.1	กรอบแนวคิดการออกแบบรูปแบบการรวบรวมและจำแนกอุบัติการณ์..... 26
3.2	หน้าจอแสดงผลและองค์ประกอบการแสดงข้อมูลข่าวสาร..... 28
3.3	ตัวอย่างข้อมูลที่จัดเก็บเหตุการณ์ปกติ (การประชาสัมพันธ์ข่าวสาร)..... 28
3.4	ตัวอย่างข้อมูลที่จัดเก็บเหตุการณ์ (อุบัติการณ์)..... 28
3.5	การตัดคำจากข้อความ 30
3.6	การคัดกรองคำก่อนการประมวลผล 30
3.7	กลุ่มข้อความตามความถี่ระหว่างข้อความทั่วไปและข้อความอุบัติการณ์ 31
3.8	กลุ่มข้อความตามความถี่ระหว่างข้อความตามลักษณะ 32
3.9	กลุ่มข้อความตามความถี่ของอุบัติการณ์ที่ไม่มีผลกระทบต่อการจราจร 34
3.10	กลุ่มข้อความตามความถี่ของอุบัติการณ์ที่กระทบต่อการจราจรบางส่วน..... 34
3.11	กลุ่มข้อความตามความถี่ของอุบัติการณ์ที่ปิดกั้นทุกช่องจราจร 34
3.12	จำนวนคำที่ให้ผลลัพธ์การแยกประเภทข้อความที่ดีที่สุด..... 35
3.13	การพัฒนาแบบการเรียนรู้เชิงลึกแบบ MLP 38
3.14	การพัฒนาแบบการเรียนรู้เชิงลึกแบบ CNN..... 39
3.15	การพัฒนาแบบการเรียนรู้เชิงลึกแบบ LSTM..... 40
3.16	รายละเอียดข้อมูล JSON ที่ได้จาก Twitter API..... 43
3.17	ผลลัพธ์การเรียกข้อมูลจาก Twitter API ด้วย Tweepy..... 43
3.18	ผลลัพธ์การตัดคำด้วย PyThaiNLP 43
4.1	ขั้นตอนการทดสอบและบันทึกผล 3 ขั้นตอน 45
4.2	การแบ่งกลุ่มชุดข้อมูลและการกระจายตัวของอุบัติการณ์..... 47
4.3	ขั้นตอนการทดสอบและบันทึกผล 3 ขั้นตอน 47
4.4	ขั้นตอนการทดสอบการจำแนกประเภทข้อความ..... 48
4.5	ขั้นตอนการทดสอบการจำแนกประเภทของอุบัติการณ์..... 49
4.6	ขั้นตอนการทดสอบการระบุความรุนแรง ในการทดสอบและบันทึกผล..... 50

สารบัญภาพ (ต่อ)

ภาพ	หน้า
5.1 ความผิดพลาดการเรียนรู้การจำแนกการเกิดอุบัติเหตุการณ์วิธีการ MPL.....	53
5.2 ความแม่นยำของการเรียนรู้การจำแนกการเกิดอุบัติเหตุการณ์วิธีการ MPL.....	54
5.3 ความผิดพลาดการเรียนรู้การจำแนกการเกิดอุบัติเหตุการณ์วิธีการ CNN.....	54
5.4 ความแม่นยำของการเรียนรู้การจำแนกการเกิดอุบัติเหตุการณ์วิธีการ CNN.....	54
5.5 ความผิดพลาดการเรียนรู้การจำแนกการเกิดอุบัติเหตุการณ์วิธีการ LSTM	55
5.6 ความแม่นยำของการเรียนรู้การจำแนกการเกิดอุบัติเหตุการณ์วิธีการ LSTM	55
5.7 ความแม่นยำของการเรียนรู้การจำแนกการเกิดอุบัติเหตุการณ์วิธีการ LSTM+CNN.....	55
5.8 ความผิดพลาดการเรียนรู้การจำแนกประเภทของอุบัติเหตุการณ์วิธีการ MPL.....	57
5.9 ความแม่นยำของการเรียนรู้การจำแนกประเภทของอุบัติเหตุการณ์วิธีการ MPL.....	57
5.10 ความผิดพลาดการเรียนรู้การจำแนกประเภทของอุบัติเหตุการณ์วิธีการ CNN.....	58
5.11 ความแม่นยำของการเรียนรู้การจำแนกประเภทของอุบัติเหตุการณ์วิธีการ CNN.....	58
5.12 ความผิดพลาดการเรียนรู้การจำแนกประเภทของอุบัติเหตุการณ์วิธีการ LSTM	58
5.13 ความแม่นยำของการเรียนรู้การจำแนกประเภทของอุบัติเหตุการณ์วิธีการ LSTM	59
5.14 ความแม่นยำของการเรียนรู้การจำแนกประเภทของอุบัติเหตุการณ์วิธีการ LSTM+CNN	59
5.15 ความผิดพลาดการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการณ์วิธีการ MPL.....	61
5.16 ความแม่นยำของการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการณ์วิธีการ MPL.....	61
5.17 ความผิดพลาดการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการณ์วิธีการ CNN.....	62
5.18 ความแม่นยำของการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการณ์วิธีการ CNN.....	62
5.19 ความผิดพลาดการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการณ์วิธีการ LSTM.....	62
5.20 ความแม่นยำของการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการณ์วิธีการ LSTM	63
5.21 ความแม่นยำของการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการณ์วิธีการ LSTM+CNN.....	63
6.1 การเปรียบเทียบค่าความถูกต้องของวิธีการในการตรวจสอบอุบัติเหตุการณ์.....	65
6.2 การเปรียบเทียบค่าความผิดพลาดของวิธีการในการตรวจสอบอุบัติเหตุการณ์.....	65
8.1 สถาปัตยกรรมระบบและการดำเนินการทางธุรกิจ.....	68
8.2 กระบวนการให้บริการระบบรวบรวมข่าวอุบัติเหตุการณ์จากช่องทางทวิตเตอร์อัตโนมัติ....	69
8.3 การตรวจสอบสิทธิการเข้าใช้งานโดยการระบุ User Name.....	70
8.4 การรายงานข้อความที่ผ่านการตรวจสอบการตรวจจับอุบัติเหตุการณ์	70
8.5 การรายงานสรุปจำนวนข้อความแต่ละช่องทางและจำนวนเหตุการณ์	71

บทที่ 1

บทนำ (Introduction)

การศึกษาวิจัยการพัฒนาารูปแบบการรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่กรุงเทพมหานครจากข้อมูลทวิตเตอร์วิธีการเรียนรู้เชิงลึก เป็นการวิจัยเชิงคุณภาพที่มุ่งเน้นผลลัพธ์ในการแก้ปัญหาและหาคำตอบที่มีประสิทธิภาพ กระบวนการแก้ปัญหาารูปแบบการเรียนรู้เชิงลึก ซึ่งเป็นวิธีการรูปแบบหนึ่งของปัญญาประดิษฐ์ โดยในเบื้องต้น บทนำนี้จะกล่าวถึง ความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์ ขอบเขตของการศึกษา สมมติฐาน กรอบแนวความคิดและประโยชน์ที่คาดว่าจะได้รับ ดังนี้

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการรายงานอุบัติการณ์ทางการจราจรทางบกของหน่วยงานที่เกี่ยวข้อง และในแต่ละหน่วยงานที่รับผิดชอบมีวิธีการในการรายงานหลากหลายวิธีการ เช่น การรายงานผ่านสถานีวิทยุ (Radio) การรายงานผ่านช่องทางโทรทัศน์ (Television) หรือการรายงานผ่านสื่อสังคมออนไลน์ (Social Media) เป็นต้น ทำให้การลงทุนเกิดความซ้ำซ้อนบนความหลากหลายของช่องทาง ทำให้ประชาชนหรือผู้รับข้อสารไม่สามารถกำหนดหรือเลือกช่องทางที่เหมาะสมเพียงหนึ่งได้ ถึงแม้ว่าในการศึกษาวิจัยในปัจจุบันของ (Theodore, 2015) แสดงให้เห็นว่าเทคโนโลยีด้านการรวบรวมข้อมูลและตรวจจับการเกิดอุบัติการณ์จะมีแนวโน้มไปในทิศทางด้านการรายงานผ่านกลุ่มคนออนไลน์ (Crowdsourcing) ในรูปแบบโมบายแอปพลิเคชัน (Mobile Application) เช่น Waze เป็นต้น หรือแพลตฟอร์มสื่อสังคมออนไลน์ ในรูปแบบการรองรับหลายแอปพลิเคชัน ทั้งเว็บแอปพลิเคชันและโมบายแอปพลิเคชัน เช่น ทวิตเตอร์ (Twitter) เฟซบุ๊ก (Facebook) หรืออินสตาร์แกรม (Instagram) เป็นต้น

ความหลากหลายของช่องทางในการรายงานอุบัติการณ์ดังกล่าว แสดงให้เห็นปัญหาในการเข้าถึงการรับทราบข้อมูลข่าวสารอุบัติการณ์จราจรที่มีทางเลือกหลากหลายช่องทาง นโยบายภาครัฐในการบูรณาการข้อมูลแบบศูนย์กลาง เช่น โครงการพัฒนาศูนย์บูรณาการขนส่งต่อเนื่องหลายรูปแบบแห่งชาติ (National Multimodal Transport Integration Center: NMTIC) ระยะ 1 ถึง 3 หรือระบบรายงานสถานการณ์อุบัติเหตุบนโครงข่ายของกระทรวงคมนาคม (Thailand Road Accident Management Systems: TRAMS) หรือระบบสารสนเทศอุบัติเหตุบนทางหลวง (Highway Accident Information Management System: HAIMS) ของกรมทางหลวง เป็นต้น ดังข้อมูลโครงการดังกล่าวเป็นตัวสะท้อนถึงความพยายามและความต้องการในการรวบรวมข้อมูลของหน่วยงานที่รับผิดชอบหรือเกี่ยวข้อง ซึ่งเป็นมูลเหตุหลักในการศึกษาวิจัยแหล่งที่มาของข้อมูลที่หลากหลายเพื่อการรวบรวมเป็นหนึ่งตามแนวโน้มของเทคโนโลยีปัจจุบันที่กล่าวถึงในข้างต้น ทั้งนี้จากสถิติการใช้งานอินเทอร์เน็ตของดาต้ารีพอร์ตทอล (Datareportal, 2020) ประจำเดือน

มกราคม ปี 2020 จะแสดงให้เห็นว่าสื่อสังคมออนไลน์เป็นช่องทางการรับข้อมูลข่าวสารที่มีผู้ใช้งานจำนวนมาก โดยประชากรส่วนใหญ่ในประเทศไทยใช้อินเทอร์เน็ตเป็นอันดับที่ 5 ของโลก มีค่าเฉลี่ย 9.01 ชั่วโมงต่อวัน บนแพลตฟอร์มสื่อสังคมออนไลน์ 3 อันดับแรกในประเทศไทยประกอบด้วย เฟซบุ๊ก จำนวน 47 ล้านบัญชี อิน스타그램 12 ล้านบัญชี และทวิตเตอร์ 6.55 ล้านบัญชี ทั้งนี้ ทวิตเตอร์เป็นหนึ่งในสื่อสังคมออนไลน์ที่เป็นที่นิยมอยู่ในปัจจุบัน และมีความสามารถในการระบุตำแหน่ง (Geo-Tag) โดยการอ่านข้อมูลผ่านชุดโปรแกรมนักพัฒนา ซึ่งจะทำให้การวิจัยและพัฒนาสามารถระบุตำแหน่งที่มาของข้อมูลได้ในอนาคต

ตั้งข้อมูลสถิติและการศึกษาทบทวนในเบื้องต้น นอกจากปัญหาความหลากหลายของช่องทางการรายงานข้อมูลนั้น เมื่อศึกษาให้ลึกถึงรายละเอียดและขั้นตอนในการวิเคราะห์รวบรวมข้อมูลข่าวสารจะพบว่า การรายงานข้อมูลผ่านสื่อสังคมออนไลน์นั้นมีข้อความข่าวสารอื่นที่ไม่เกี่ยวข้องกับ การรายงานอุบัติการณ์จลาจลอยู่ร่วมจำนวนมาก ซึ่งกระบวนการในการจัดกลุ่มและจำแนก การรายงานอุบัติการณ์จลาจลนี้ จึงเป็นกระบวนการพิจารณาและตัดสินใจที่สำคัญที่จะทำให้ การพัฒนารูปแบบการรวบรวมอุบัติการณ์การจลาจลทางบกสามารถดำเนินการไปได้ วิทยาการ ปัญญาประดิษฐ์ (Artificial Intelligent: AI) ในกลุ่มของจักรกลเรียนรู้ (Machine Learning: ML) นั้น เป็นกระบวนการหนึ่ง ที่สามารถนำมาสนับสนุนการตัดสินใจเพื่อการแก้ปัญหาได้อย่างมีประสิทธิภาพ เทคนิคการเรียนรู้เชิงลึก (Deep Learning: DL) ที่ใช้ได้กับปัญหาการรู้จำและจำแนกผลลัพธ์ แบบอัตโนมัติ (Chen et al., 2019) โดยในการศึกษาวิจัยดังกล่าวนี้ นำเสนอกระบวนการและขั้นตอน ในการพัฒนารูปแบบการรวบรวมอุบัติการณ์การจลาจลทางบกในเขตพื้นที่กรุงเทพมหานครจากข้อมูล ทวิตเตอร์วิธีการเรียนรู้เชิงลึก เพื่อให้ผู้รับข้อมูลข่าวสารสามารถติดตามข่าวสารการรายงาน อุบัติการณ์การจลาจลทางบกทุกเหตุการณ์ที่เกิดขึ้นในเขตพื้นที่ตัวอย่างผ่านช่องทางเดียวที่ออกแบบ และพัฒนาขึ้นในงานวิจัยนี้

1.2 วัตถุประสงค์

- 1.2.1 เพื่อรวบรวมและศึกษาแหล่งที่และอ้างอิงการรายงานอุบัติการณ์ที่มีบนทวิตเตอร์
- 1.2.2 เพื่อศึกษาทบทวนการจัดกลุ่มและจำแนกการรายงานอุบัติการณ์ในรูปแบบภาษาไทย
- 1.2.3 เพื่อออกแบบและพัฒนาวิธีการจำแนกรูปแบบการรายงาน เพื่อการยืนยันการรายงานเหตุ ลักษณะการเกิดเหตุ และระดับความรุนแรง วิธีการโครงข่ายประสาทเทียมแบบการเรียนรู้เชิงลึก
- 1.2.4 เพื่อพัฒนาระบบการรวบรวมอุบัติการณ์การจลาจลทางบกในเขตพื้นที่กรุงเทพมหานคร ศูนย์กลางจากข้อมูลทวิตเตอร์วิธีการเรียนรู้เชิงลึก

1.3 ขอบเขตของการศึกษา

1.3.1 ดำเนินการศึกษาทบทวนงานวิจัยภายในประเทศและต่างประเทศที่เกี่ยวข้องกับการรวบรวมอุบัติการณ์การจราจรทางบกจากสื่อสังคมออนไลน์ที่ผ่านมา โดยไม่จำกัดรูปแบบภาษาหรือประเทศในการรายงานข้อมูลอุบัติการณ์ เพื่อเป็นศึกษากระบวนการ วิธีการ และผลลัพธ์ที่ได้จากการศึกษาวิจัย

1.3.2 ดำเนินการรวบรวมข้อมูลการรายงานอุบัติการณ์ (Abnormal) และเหตุการณ์ปกติ (Normal) บนทวิตเตอร์จากกลุ่มเป้าหมายที่เลือกในระหว่างช่วงวันที่และเวลาเดียวกัน โดยมีจำนวนข้อมูลเหมาะสมต่อการเป็นตัวแทนข้อมูลการรายงานอุบัติการณ์ในระยะเวลา 1 ปี

1.3.2.1 ข้อมูลสำหรับการเรียนรู้ (Training Dataset)

1.3.2.1 ข้อมูลสำหรับการทดสอบประสิทธิภาพ (Testing Dataset)

1.3.3 ดำเนินการออกแบบและพัฒนารูปแบบการรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่กรุงเทพมหานครจากข้อมูลทวิตเตอร์ โดยประกอบไปด้วย

1.3.3.1 กระบวนการรวบรวมและคัดกรองข้อมูลจากแหล่งข้อมูลที่สนใจจากทวิตเตอร์

1.3.3.2 กระบวนการแบ่งคำจากประโยค

1.3.3.3 กระบวนการรู้จำและแยกแยะอุบัติการณ์จราจรกับเหตุการณ์ปกติ

1.3.3.4 กระบวนการรู้จำและจัดกลุ่มรูปแบบของอุบัติการณ์

1.3.3.5 กระบวนการรู้จำและจัดระดับความรุนแรงของอุบัติการณ์

1.3.4 ดำเนินการทดสอบและประเมินประสิทธิภาพของรูปแบบการรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่กรุงเทพมหานครจากข้อมูลทวิตเตอร์ โดยประกอบไปด้วย

1.3.4.1 ความถูกต้องในการแยกแยะอุบัติการณ์กับเหตุการณ์ปกติ

1.3.4.2 ความถูกต้องในการจัดกลุ่มรูปแบบของอุบัติการณ์

1.3.4.3 ความถูกต้องในการจัดระดับความรุนแรงของอุบัติการณ์

1.3.5 ดำเนินการจัดทำระบบรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่กรุงเทพมหานคร ศูนย์กลางจากข้อมูลทวิตเตอร์ต้นแบบ ที่สามารถให้บริการประชาชนและผู้รับข้อมูลข่าวสารแบบสาธารณะ ณ จุดบริการข้อมูลเดียวได้

1.3.6 ดำเนินการออกแบบกระบวนการจัดการระบบรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่กรุงเทพมหานครจากข้อมูลทวิตเตอร์ให้มีความยั่งยืนที่ตอบสนองต่อ สังคม เศรษฐกิจและสิ่งแวดล้อมได้

1.4 สมมติฐาน

การรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่กรุงเทพมหานครจากข้อมูลทวิตเตอร์วิธีการเรียนรู้เชิงลึก สามารถแยกแยะการรายงานอุบัติการณ์และเหตุการณ์ปกติ การระบุรูปแบบ และระดับความรุนแรง จากข้อความภาษาไทย ได้อย่างถูกต้องที่ค่าเฉลี่ยไม่น้อยกว่า 80 % โดยพิจารณาอ้างอิงจากผลลัพธ์การศึกษาวิจัยระดับนานาชาติที่ผ่านมา

1.5 กรอบแนวความคิด

ในการศึกษาวิจัยการพัฒนารูปแบบการรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่กรุงเทพมหานครจากข้อมูลทวิตเตอร์วิธีการเรียนรู้เชิงลึกในเบื้องต้นจะพบว่า ในงานวิจัยที่ผ่านมา Chen et al. (2019) นำเสนอวิธีการและขั้นตอนที่มีลักษณะในการแก้ปัญหาที่คล้ายคลึงกัน โดยในงานวิจัยนี้จะนำกรอบแนวความคิดดังกล่าวมากำหนดขั้นตอนให้สอดคล้องต่อแนวทางการวิจัย 4 ขั้นตอน ดังประกอบไปด้วย

1.5.1 กระบวนการรวบรวมและคัดกรองข้อมูลจากแหล่งข้อมูลที่สนใจจากทวิตเตอร์

1.5.1.1 การพิจารณาเลือกแหล่งที่มาของข้อมูล

การวิจัยการพัฒนารูปแบบการรวบรวมอุบัติการณ์การจราจรทางบกนี้จะพิจารณาเลือกแหล่งที่มาของข้อมูลข่าวสารอุบัติการณ์จากบัญชีทางการ (@ Official ID) ที่มีการประชาสัมพันธ์ข้อมูลข่าวสารในเขตพื้นที่กรุงเทพมหานครอย่างต่อเนื่องจนถึงปัจจุบัน และมีจำนวนข้อความต่อวันมากที่สุดตามลำดับ ทั้งนี้การกำหนดเลือกแหล่งที่มาผู้วิจัยจะไม่จำกัดจำนวนบัญชี แต่จะเลือกใช้บัญชีที่มีการแจ้งข้อมูลข่าวสารอย่างต่อเนื่องไม่น้อยกว่า 10 ข้อความต่อวัน

ทั้งนี้ในการรวบรวมบัญชีของการรายงานข้อมูลอุบัติการณ์จราจรทางบกในเขตพื้นที่กรุงเทพมหานครนั้น ไม่มีการจัดกลุ่มบัญชีให้เรียกใช้งานได้ทันที ผู้วิจัยจะต้องดำเนินการค้นหาบัญชีทางการในการรายงานอุบัติการณ์ผ่านการค้นหาแบบเนื้อความ (Keyword) และการค้นหาจากคำสำคัญ (# Hashtag) โดยใช้คำค้นหาดังนี้ อุบัติเหตุ ของตกล่อน ปิดการจราจร ช่อมถนน การจราจร และสภาพจราจร กับทั้งสองรูปแบบการค้นหา เพื่อเก็บรวบรวมรายชื่อบัญชีทางการและคัดเลือกรายชื่อจากปริมาณข้อความทวิตต่อวันต่อไป

1.5.1.2 การรวบรวมและคัดกรองข้อมูลตัวแทนในการเรียนรู้และทดสอบ

เมื่อการวิจัยมีรายชื่อบัญชีทางการของช่องทางรายงานอุบัติการณ์จราจรทางบกเป็นที่เรียบร้อยแล้ว ผู้วิจัยดำเนินการเก็บรวบรวมข้อมูลตัวแทนในการเรียนรู้ (Training Dataset) และการทดสอบ (Test Dataset) โดยการพิจารณาปริมาณข้อมูลตัวแทน (Sampling Space) จากข้อมูลข้อความทวิตทั้งหมดต่อปี ซึ่งสามารถคำนวณค่าเฉลี่ยข้อความรวมต่อปีได้จากข้อความทวิตทั้งหมดกับระยะเวลาที่บัญชีทางการนั้นถูกเปิดใช้งานจนถึงปัจจุบันสำหรับทุกบัญชีทางการที่ถูกเลือก โดยจำนวนข้อมูลตัวแทน (Cochran, 1963) จะกำหนดให้ค่าความเชื่อมั่นเท่ากับ 95 % เพื่อคำนวณหาจำนวนตัวแทนทั้งหมดและนำไปกำหนดขอบเขตปริมาณข้อมูลที่จะต้องจัดเก็บจากบัญชีทางการที่รวบรวมขึ้นบัญชีไว้ โดยจะต้องจัดเก็บในระหว่างช่วงวันที่และเวลาเดียวกันทุกบัญชี

1.5.2 กระบวนการแบ่งคำจากประโยค (Word Segmentation)

การรับข้อมูลข่าวสารจากทวิตเตอร์จะมีผลลัพธ์ในรูปแบบประโยคหรือกลุ่มคำต่อเนื่อง ซึ่งในการนำไปเข้าสู่กระบวนการเรียนรู้หรือกระบวนการตัดสินใจจะต้องดำเนินการแบ่งคำ (Word) จากประโยคเทคนิค (Conditional Random Field: CRF) CRF Algorithm จำนวน 11 คำที่สามารถต่อเนื่องกันได้สูงสุด หรือเรียกว่า Gram ซึ่งเป็นเทคนิคที่มีประสิทธิภาพและความถูกต้องสูง ร่วมกับ

คลังข้อความภาษาไทยหมวดรายการสารานุกรม (ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, 2565) และการตรวจสอบคลังคำศัพท์ร่วมกับพจนานุกรม (ราชบัณฑิตยสถาน, 2565)

1.5.3 กระบวนการจัดเตรียมข้อมูลและการรู้จำการเรียนรู้เชิงลึก

1.5.3.1 การจัดกลุ่มการรายงานแบบอุบัติการณ์จราจรกับเหตุการณ์ปกติ

ในเบื้องต้นกระบวนการวิจัยต้องสามารถระบุสถานการณ์ของการรายงานอุบัติการณ์ได้ว่าการรายงานข้อมูลดังกล่าวเป็นการรายงานอุบัติการณ์หรือการรายงานข้อมูลข่าวสารทั่วไปอื่น ๆ หรือไม่ (Chen et al., 2019) ทั้งนี้ในขั้นตอนการจัดเก็บข้อมูลต้องมีการบันทึกสถานะของข้อมูลการรายงานว่าเป็นเหตุการณ์ประเภทใด ดังแสดงในตาราง 1.1 โดยในการทดสอบต้องพิจารณาระหว่างความถูกต้องในการแยกแยะระหว่างข้อมูลที่เกี่ยวข้องกับการจราจร (Traffic Relevant) และข้อมูลข่าวสารทั่วไปที่ไม่เกี่ยวกับจราจร (Traffic Irrelevant)

ตาราง 1.1 การระบุการเกิดอุบัติการณ์จราจร

อ้างอิง	ความหมาย	คำอธิบาย
0	การรายงานทั่วไป	ข้อมูลข่าวสารทั่วไปที่ไม่เกี่ยวกับจราจร หรือเหตุการณ์ทางการจราจรที่ไม่เป็นปัจจุบัน เช่น ข่าวประชาสัมพันธ์องค์กร หรือการเชิญชวนร่วมกิจกรรมพิเศษภายในพื้นที่ หรือการรายงานข่าวเหตุการณ์อุบัติการณ์ย่อยหลัง เป็นต้น
1	การรายงานที่เกี่ยวข้องกับการจราจร	ข้อมูลการรายงานที่เกี่ยวข้องกับการจราจรนั้น หมายถึงการรายงานเหตุการณ์จราจรที่กำลังเกิดขึ้น ณ ขณะเวลานั้น ๆ หรือการแจ้งสถานการณ์ สิ้นสุดการจัดการเหตุการณ์ในพื้นที่ เป็นต้น

1.5.3.2 การจัดกลุ่มรูปแบบของอุบัติการณ์

การรายงานอุบัติการณ์จราจรจากหลายแหล่งที่มีจะมีรูปแบบการแจ้งเหตุการณ์ที่หลากหลาย เมื่อศึกษาบทพจนานุกรมวิจัยที่ผ่านมา ผู้วิจัยจะสามารถนำการรายงานอุบัติการณ์ต่าง ๆ จัดกลุ่มตามประเภทของเหตุการณ์ที่เกิดขึ้นได้ 3 รูปแบบ (Agarwal and Toshniwal, 2019) ทั้งนี้เพื่อให้การเก็บข้อมูลมีความครอบคลุมต่อรูปแบบการรายงาน ผู้วิจัยกำหนดให้รูปแบบที่ 0 เท่ากับเหตุการณ์ที่ไม่สามารถระบุรูปแบบได้ชัดเจน ซึ่งในกระบวนการเรียนรู้จะไม่นำข้อมูลดังกล่าวเข้าสู่กระบวนการเรียนรู้ โดยมีรายละเอียดดังแสดงในตาราง 1.2

ตาราง 1.2 รูปแบบของอุบัติเหตุ

รูปแบบ	ความหมาย	คำอธิบาย
0	ไม่สามารถระบุ	อุบัติเหตุที่ไม่มีการระบุรูปแบบได้
1	อุบัติเหตุการณ์ (Incident)	เหตุการณ์ไม่พึงประสงค์ที่ก่อให้เกิดการกีดขวางการจราจรในช่วงระยะเวลาหนึ่ง เช่น รถชน รถเสีย หรือ สิ่งของตกหล่น เป็นต้น
2	ทางชำรุด (Potholes)	เหตุการณ์ที่มีผลกระทบต่ออาการจราจรบนผิวจราจรที่ทำให้ผู้ขับขี่ต้องปรับเปลี่ยนพฤติกรรมหรือความเร็ว เช่น หลุมบนถนน หรือ การปิดกั้นช่องจราจรเพื่อการดำเนินงานบำรุงรักษา เป็นต้น
3	สภาพจราจร (Traffic)	เหตุการณ์รายงานปริมาณจราจร เป็นการแสดงกระแสจราจรที่ส่งผลสืบเนื่องต่อการจัดระดับความรุนแรงของอุบัติเหตุ แต่ไม่ได้เป็นการระบุเหตุการณ์ที่ส่งผลให้เกิดการเปลี่ยนแปลงทางการจราจรในพื้นที่บริเวณดังกล่าว

1.5.3.3 การจัดระดับความรุนแรงของอุบัติเหตุ

ในการศึกษาวิจัยนี้ นิยามความรุนแรง 3 ระดับ ตามแนวทางการศึกษาวิจัยที่ผ่านมา (Zunic et al., 2016; Mulyana et al., 2015) ซึ่งในการเก็บข้อมูล ผู้วิจัยกำหนดให้มีระดับ 0 เพิ่มเติมพิเศษ ในกรณีไม่สามารถระบุรูปแบบความรุนแรงได้ ดังแสดงในตาราง 1.3

ตาราง 1.3 ระดับความรุนแรงของอุบัติเหตุ

ระดับ	ความหมาย	คำอธิบาย
0	ไม่สามารถระบุ	อุบัติเหตุที่ไม่มีการระบุความรุนแรงหรือรูปแบบการเกิดเหตุที่ชัดเจน ซึ่งไม่สามารถกำหนดระดับความรุนแรงในขั้นตอนเก็บรวบรวมข้อมูลได้
1	ระดับปกติ (Normal Traffic / Weak Traffic)	อุบัติเหตุที่มีความรุนแรงต่ำและไม่ส่งผลกระทบต่อสภาพการจราจร
2	ระดับกลาง (Heavy Traffic / Medium Traffic)	อุบัติเหตุที่มีความรุนแรงในระดับกลาง โดยมีผลกระทบต่ออาการจราจร 1-2 ช่องจราจร หรือส่งผลให้เกิดการชะลอตัวของกระแสจราจร
3	ระดับปิดกั้นช่องจราจรหรือหยุดนิ่ง (Standstill Traffic / Traffic Jam)	อุบัติเหตุที่มีความรุนแรงในระดับสูง ซึ่งหมายถึงการปิดกั้นทุกช่องจราจร หรือส่งผลกระทบต่ออาการจราจรในพื้นที่อยู่ในระดับหยุดนิ่ง

1.5.4 กระบวนการแยกแยะการเรียนรู้เชิงลึก

เป็นกระบวนการที่ทำหน้าที่รับข้อมูลค่าประกอบประกอบโยคนำข้อมูลเข้าสู่โครงข่ายประสาทเทียมเพื่อหาคำตอบ (Output) กระบวนการทำงานของโครงข่ายประสาทเทียมสามารถให้ผลลัพธ์ที่มีความถูกต้องที่แตกต่างกันได้อย่างหลากหลายรูปแบบ ทั้งนี้ความถูกต้องแม่นยำจะขึ้นอยู่กับรูปแบบโครงสร้างภายในโครงข่ายประสาทเทียมของวิธีการเรียนรู้เชิงลึก ซึ่งปัจจุบันมีหลากหลายรูปแบบซึ่งในการวิจัยที่ผ่านมาได้มีการนำเสนอผลการเปรียบเทียบประสิทธิภาพการทำงานของโครงสร้างที่แตกต่างกัน 3 รูปแบบ (Chen et al., 2019) โดยผู้วิจัยดำเนินการทดลองกับรูปแบบโครงสร้างโครงข่ายประสาทเทียม 4 รูปแบบ กำหนดให้ทดลองกับโครงสร้าง (Structure) และพารามิเตอร์ (Parameter) ที่แตกต่างกันดังแสดงในตาราง 1.4

ตาราง 1.4 รูปแบบโครงสร้างโครงข่ายประสาทเทียมภายในของการเรียนรู้เชิงลึก

ลำดับ	เทคนิค
1	MLP (Multilayer Perceptron: MLP)
2	CNN (Convolution Neural Network: CNN)
3	LSTM (Long Short-Term Memory: LSTM)
4	LSTM + CNN

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 สามารถพัฒนาแพลตฟอร์มการให้บริการข้อมูลข่าวสารอุบัติการณ์จราจรทางบกแบบสาธารณะในนามหน่วยงานสถาบันการศึกษาได้ จากรูปแบบการรวบรวมอุบัติการณ์การจราจรทางบกในเขตพื้นที่กรุงเทพมหานครจากข้อมูลทวีตเตอร์วิธีการเรียนรู้เชิงลึก

1.6.2 ประชาชนหรือผู้รับข้อมูลข่าวสารอุบัติการณ์จราจรทางบกในเขตพื้นที่กรุงเทพมหานครสามารถรับข้อมูลข่าวสารได้อย่างรวดเร็วและเชื่อถือได้ ผ่านช่องทางการแจ้งประชาสัมพันธ์ช่องทางเดียว

1.6.3 การคมนาคมภายในเขตกรุงเทพมหานครเกิดความคล่องตัวจากการเดินทางแบบกระจายตัวและหลีกเลี่ยงการเข้าพื้นที่ให้เกิดปัญหาการแออัดสะสมของประชาชนและภาคการขนส่ง กรณีเกิดอุบัติการณ์จราจรทางบกที่ตำแหน่งใด ๆ ในเขตพื้นที่

1.6.4 หน่วยงานภาครัฐที่เกี่ยวข้องสามารถลดค่าใช้จ่ายการลงทุน หรือสามารถปรับเปลี่ยนการลงทุนงบประมาณในด้านอื่นทดแทนการพัฒนาาระบบรวบรวมข้อมูลและรายงานอุบัติการณ์จราจรทางบกที่มีความซ้ำซ้อนกันในปัจจุบันได้

1.7 นิยามศัพท์

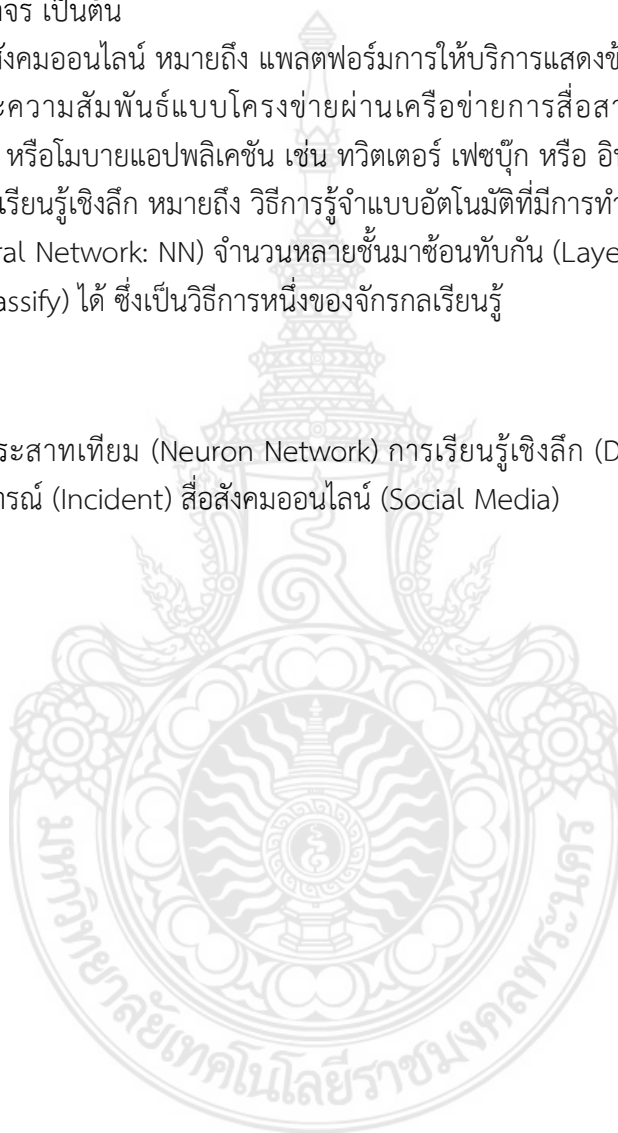
1.7.1 อุบัติการณ์ หมายถึง ทุกเหตุการณ์ที่เกิดขึ้นและมีผลกระทบต่อการจราจร โดยเหตุการณ์ดังกล่าวไม่ต้องเป็นเหตุการณ์ที่เกิดขึ้นบนผิวจราจรเท่านั้น ซึ่งอาจหมายถึงเหตุการณ์ต่าง ๆ ที่เกิดขึ้นในพื้นที่บริเวณข้างเคียงโดยรอบ และส่งผลกระทบต่อการจราจรในพื้นที่ข้างเคียงใด ๆ เช่น เหตุการณ์รถชน เหตุการณ์รถเสีย เหตุการณ์สิ่งของตกหล่น เหตุการณ์หมอกควัน หรือเหตุการณ์งานซ่อมบำรุงรักษาผิวจราจร เป็นต้น

1.7.2 สื่อสังคมออนไลน์ หมายถึง แพลตฟอร์มการให้บริการแสดงข้อมูลในลักษณะการแบ่งปันตามเงื่อนไขและความสัมพันธ์แบบโครงข่ายผ่านเครือข่ายการสื่อสารอินเทอร์เน็ตในรูปแบบเว็บแอปพลิเคชัน หรือโมบายแอปพลิเคชัน เช่น ทวิตเตอร์ เฟซบุ๊ก หรือ อิน스타그램 เป็นต้น

1.7.3 การเรียนรู้เชิงลึก หมายถึง วิธีการรู้จำแบบอัตโนมัติที่มีการทำงานแบบโครงข่ายประสาทของมนุษย์ (Neural Network: NN) จำนวนหลายชั้นมาซ้อนทับกัน (Layer) เพื่อให้มีความสามารถในการแยกแยะ (Classify) ได้ ซึ่งเป็นวิธีการหนึ่งของจักรกลเรียนรู้

1.8 คำสำคัญ

โครงข่ายประสาทเทียม (Neuron Network) การเรียนรู้เชิงลึก (Deep Learning) ทวิตเตอร์ (Twitter) อุบัติการณ์ (Incident) สื่อสังคมออนไลน์ (Social Media)



บทที่ 2

แนวคิดทฤษฎีที่เกี่ยวข้องและการทบทวนวรรณกรรม

การศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง เพื่อให้การรวบรวมข้อมูลและการวิเคราะห์ผลให้ได้ ซึ่งคำตอบตามสมมติฐานงานวิจัย ในบทนี้จะกล่าวถึงแนวคิดของทฤษฎีที่เกี่ยวข้องกับงานวิจัย ดังประกอบไปด้วย สื่อสังคมออนไลน์ การจำแนกข้อความ และโครงข่ายประสาทเทียม และนำเสนอผลการศึกษาทบทวนวรรณกรรมที่เกี่ยวข้อง คำสำคัญที่เกี่ยวข้อง ดังนี้ การตรวจจับอุบัติเหตุการณ์ (Incident Detection) สื่อสังคมออนไลน์ และการเรียนรู้เชิงลึก เพื่อนำผลการศึกษาวิจัยที่ผ่านมา เป็นแนวทางการกำหนดวิธีการวิจัยต่อไป

2.1 แนวคิดทฤษฎีที่เกี่ยวข้อง

ในการศึกษาวิจัยแนวทางการรวบรวมอุบัติเหตุการณ์การจราจรทางบกในเขตพื้นที่กรุงเทพมหานคร จากข้อมูลทวิตเตอร์วิธีการเรียนรู้เชิงลึก ต้องมีการศึกษาถึงองค์ความรู้พื้นฐานที่เกี่ยวข้อง ประกอบไปด้วย สื่อสังคมออนไลน์ เพื่อทราบถึงระบบที่มีให้บริการในปัจจุบัน การใช้งานรายงานอุบัติเหตุการณ์ วิธีการต่อประสาน เป็นต้น การจำแนกข้อความ เพื่อให้การศึกษาวิจัยสามารถนำข้อความหรือรายละเอียดที่มีลักษณะแบบภาษาธรรมชาติ (Natural Language) ที่ได้จากการรายงานอุบัติเหตุการณ์บนสื่อสังคมออนไลน์ และการเรียนรู้เชิงลึก เพื่อการรู้จำและแยกแยะรูปแบบการรายงานอุบัติเหตุการณ์ที่มีความหลากหลายและซับซ้อน โดยมีรายละเอียดดังนี้

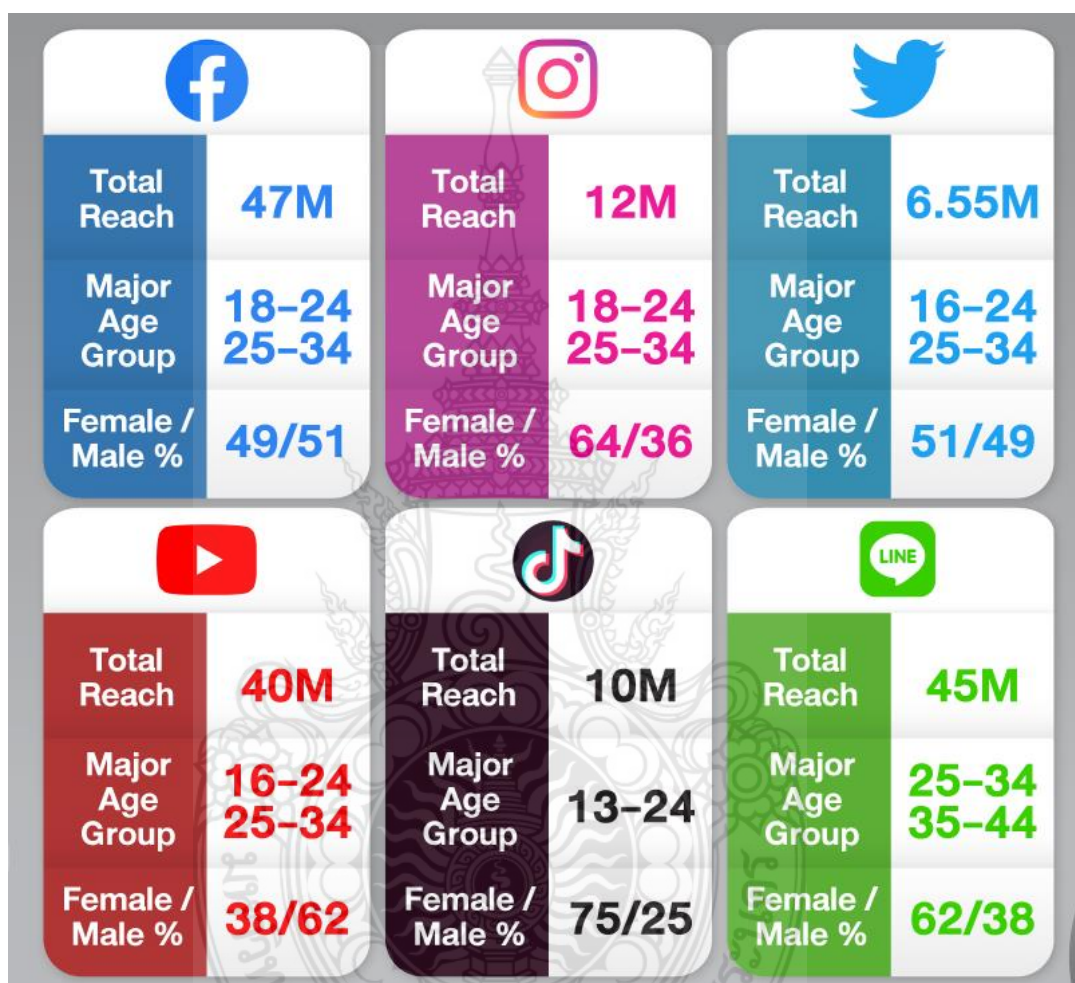
2.1.1 สื่อสังคมออนไลน์

สื่อสังคมออนไลน์ หมายถึง สื่อดิจิทัลที่เป็นเครื่องมือในการปฏิบัติการทางสังคม (Social Tool) เพื่อใช้สื่อสารระหว่างกันในเครือข่ายทางสังคม (Social Network) ผ่านทางเว็บไซต์และโปรแกรมประยุกต์บนสื่อใด ๆ ที่มีการเชื่อมต่อกับอินเทอร์เน็ต โดยเน้นให้ผู้ใช้ทั้งที่เป็นผู้ส่งสารและผู้รับสารมีส่วนร่วม (Collaborative) อย่างสร้างสรรค์ ในการผลิตเนื้อหาขึ้นเอง (User-Generate Content: UGC) ในรูปของข้อมูล ภาพ และเสียง ทั้งนี้ประเภทของสื่อสังคมออนไลน์มีหลายชนิด ขึ้นอยู่กับลักษณะของการนำมาใช้งาน

2.1.1.1 แพลตฟอร์มสื่อสังคมออนไลน์ (Social Media Platform)

จากรายงานพฤติกรรมผู้บริโภคด้านดิจิทัลและสื่อสังคมออนไลน์ของ Global Digital Report ของ We Are Social และ Hootsuite แพลตฟอร์มด้านการบริหารจัดการสื่อสังคมออนไลน์ ที่สำรวจข้อมูลและพฤติกรรมการใช้งานดิจิทัลของผู้คนทั่วโลก ปี 2564 มีการแสดงสถิติให้เห็นว่าปัจจุบันนี้มีผู้ใช้อุปกรณ์พกพาทั่วโลกมากถึง 5.27 พันล้านคน คิดเป็นสัดส่วน 67 % ของประชากรทั้งโลก มีจำนวนผู้ใช้มือถือเพิ่มขึ้น 2.3 % ในปีที่ผ่านมา เพิ่มขึ้น 117 ล้านคน โดยปัจจุบันมีผู้ใช้อินเทอร์เน็ตทั่วโลกมากกว่า 4.80 พันล้านคน คิดเป็น 61 % ของประชากรโลกทั้งหมด (Datareportal, 2020)

ซึ่งเมื่อพิจารณาถึงข้อมูลรายงานของ The Flight 19 Agency ที่จัดทำสรุปข้อมูลผู้ใช้งานแพลตฟอร์มสื่อสังคมออนไลน์หลักของคนไทย จะทำให้เห็นลำดับความนิยมตามช่วงอายุที่แตกต่างกัน ซึ่งจะมี 6 แพลตฟอร์มหลักที่เป็นที่นิยมประกอบไปด้วย เฟซบุ๊ก อิน스타그램 ทวิตเตอร์ ยูทูบ ติกตอก และไลน์ ตามผลการสำรวจประเมิน เมื่อเดือนมีนาคม 2563 (TWF Agency, 2020)



ภาพ 2.1 การจัดอันดับความนิยมการใช้งานของสื่อสังคมออนไลน์ในประเทศไทย (TWF Agency, 2020)

2.1.1.2 ทวิตเตอร์

ทวิตเตอร์ คือ บริการเครือข่ายสังคมออนไลน์ประเภทไมโครบล็อก โดยมีลักษณะการให้บริการในลักษณะการโพสต์ข้อมูลในรูปแบบบล็อกขนาดเล็ก มีเหมือนกับผู้ใช้บริการแบบบล็อกทั่วไป โดยทวิตเตอร์ถูกก่อตั้งเมื่อเดือนมีนาคม ค.ศ. 2006 สหรัฐอเมริกา ซึ่งในปัจจุบันจะมีรูปแบบการให้บริการดังแสดงภาพ 2.1 ทั้งนี้ในการให้บริการของทวิตเตอร์ ระบบสื่อสังคมออนไลน์ดังกล่าวจะมีการแบ่งองค์กรประกอบสำคัญประกอบไปด้วย

- ทวิต (Tweet) เป็นการประกาศข้อมูลข่าวสารโดยไม่มีการกำกับหรือเจาะจงเนื้อหาใด ๆ ทั้งนี้ขึ้นอยู่กับผู้ใช้งานเจ้าของบัญชี ลงสื่อสังคมออนไลน์ ปัจจุบันทวิตเตอร์จำกัดจำนวนการทวิตต่อครั้งได้ไม่เกิน 280 ตัวอักษร
- ไอดี (ID) เป็นการอ้างอิงหรือตรวจสอบการอ้างอิงถึงบัญชีโดยตรง เช่น '@username' จะส่งเป็นการอ้างอิงถึงบัญชีผู้ติดตามไอดีดังกล่าว
- แฮชแท็ก (Hashtag) คือ การเขียนคำตามหลังสัญลักษณ์ '#' จะสร้างแฮชแท็กขึ้นมา แฮชแท็กจะทำให้คำบางคำเป็นคำที่ค้นหาได้ สำหรับข้อมูลเพิ่มเติมของแฮชแท็กเพื่อกำกับประเด็นของการโพสต์ที่เป็นเสมือนคำสำคัญ ซึ่งแฮชแท็กช่วยให้ผู้ใช้งานเข้าสู่ประเด็นสำคัญในการโต้ตอบได้ง่ายขึ้น

2.1.2 การจำแนกข้อความ (Text Classification)

การจำแนกข้อความหรือการแยกคำจากประโยคเพื่อการเริ่มต้นกระบวนการเรียนรู้จำข้อความ ซึ่งจะเป็นที่รู้จักในชื่อของการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) คือ หนึ่งในสาขาของวิทยาศาสตร์คอมพิวเตอร์ ที่เกี่ยวกับปัญญาประดิษฐ์และภาษาศาสตร์คอมพิวเตอร์ (Computational Linguistics) เป็นศาสตร์ที่ศึกษาเกี่ยวกับการทำให้คอมพิวเตอร์สามารถสื่อสารโต้ตอบภาษาของมนุษย์ และทำให้คอมพิวเตอร์เข้าใจภาษามนุษย์มากขึ้น ตัวอย่าง เช่น Siri, Google Assistant และ Alexa เป็นต้น

2.1.3 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม คือกระบวนการคำนวณทางคณิตศาสตร์ที่จำลองการทำงานตามรูปแบบการทำงานพื้นฐานของโครงข่ายประสาทชีวภาพของสมองโดยโครงข่ายประสาทเทียมดังกล่าวจะเรียนรู้การทำงานผ่านข้อมูลตัวอย่างในการเรียนรู้ (Training Dataset) ได้โดยอัตโนมัติ โดยไม่ต้องโปรแกรมหรือกำหนดกฎเกณฑ์การทำงานแบบตายตัว

2.1.3.1 โครงสร้างและองค์ประกอบ (Structure and Component)

โครงสร้างและองค์ประกอบของโครงข่ายประสาทเทียม เริ่มต้นที่หน่วยประมวลผลย่อย ซึ่งจะถูกรเรียกว่า โหนด (Node) โดยโหนดเป็นตัวแทนการจำลองลักษณะการทำงานของเซลล์การส่งสัญญาณระหว่างโหนดที่เชื่อมต่อกันในรูปแบบโครงข่ายดังเช่นการเชื่อมโยงของระบบประสาท โดยโครงสร้างพื้นฐานของโครงข่ายประสาทเทียมจะแบ่งขั้นตอนการประมวลผลไว้ 4 ส่วน ประกอบไปด้วย ชั้นนำเข้าข้อมูล (Input Layer) ชั้นประมวลผลภายใน (Hidden Layer) ชั้นผลลัพธ์ (Output Layer) และ ชั้นการคาดทำนายผลลัพธ์ (Prediction) โดยมีรายละเอียดดังนี้

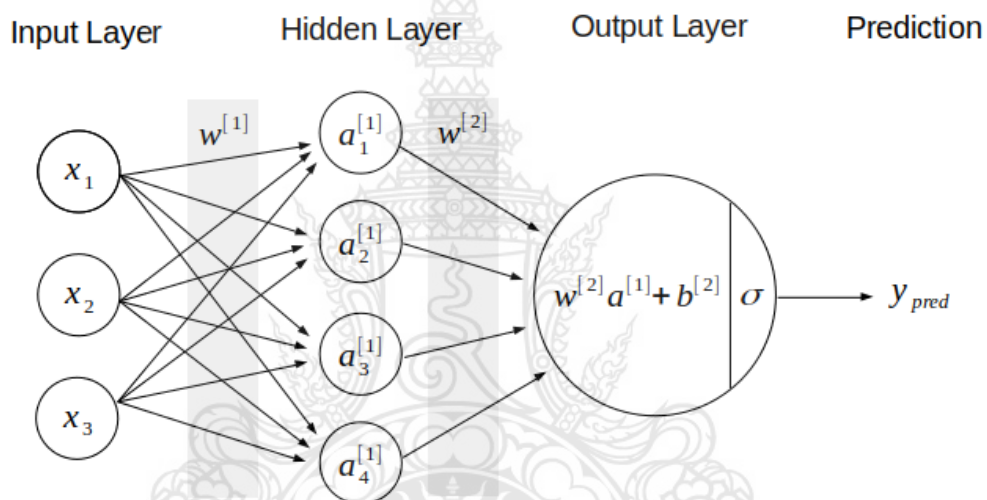
1) ชั้นนำเข้าข้อมูลเข้า คือ ข้อมูลขาเข้า ซึ่งข้อมูล X_n แต่ละโหนดเท่ากับตัวแทนของ Feature ที่นำข้อมูลเข้า หรือหมายถึงคอลัมน์ของข้อมูล เช่น สถานการณ์ สภาพอากาศ ระดับความตึงเครียดทางการจราจร เป็นต้น

2) ชั้นประมวลผลคือ ชั้นที่มีส่วนประมวลผลอยู่ภายในและสามารถมีได้มากกว่าหนึ่งชั้น ซึ่งในแต่ละชั้นจะมีหน่วยประมวลผลที่เรียกว่า Neuron ที่ทำหน้าที่รับข้อมูลจากโหนดของเลเยอร์ก่อนหน้าทุกตัวมาประมวลผล โดยใช้ฟังก์ชันเชิงเส้น (Linear Function) ร่วมกับค่าน้ำหนัก w ของข้อมูล

โนดจากเลเยอร์ก่อนหน้าแต่ละตัว เพื่อคำนวณหาผลลัพธ์ Z และนำเข้าสู่แอคทีฟฟังก์ชัน (Activation Function) ซึ่งมีหลากหลายรูปแบบ เช่น Sigmoid, Tanh หรือ RELU เป็นต้น

3) ชั้นผลลัพธ์ คือชั้นที่ประมวลผลแอคทีฟฟังก์ชันทั้งหมดจากชั้นก่อนหน้า ข้อมูลจากโนดของเลเยอร์ก่อนหน้า a_n ร่วมกับค่าน้ำหนัก $W^{[2]}$ เป็นผลลัพธ์ Z สำหรับการคำนวณในแอคทีฟฟังก์ชัน เช่น Sigmoid Function ได้ผลเป็น $a^{[2]}$ เพื่อนำเข้าสู่ขั้นตอนการคาดการณ์ทำนายผลลัพธ์ต่อไป

4) ชั้นการคาดการณ์ผลลัพธ์ เป็นการนำผลลัพธ์ที่ได้จากแอคทีฟฟังก์ชันของชั้นผลลัพธ์มาตัดสินใจ (Decision Function) เพื่อระบุผลลัพธ์สุดท้ายในรูปแบบเงื่อนไขในลักษณะลอจิก 0 หรือ 1 สำหรับการนำไปใช้งานต่อไป



ภาพ 2.2 โครงสร้างและองค์ประกอบของโครงข่ายประสาทเทียม (ชิตพงษ์ กิตตินราทร, 2563)

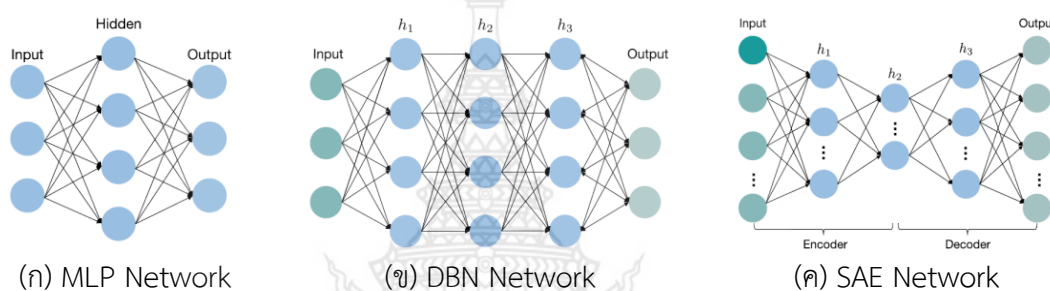
2.1.3.2 การเรียนรู้เชิงลึก

ในปัจจุบันโครงข่ายประสาทเทียมแบบการเรียนรู้เชิงลึก ประสบความสำเร็จ ในด้านความรู้จำ ภาพคอมพิวเตอร์ การรู้จำเสียงพูด และการประมวลผลภาษามนุษย์ และในงานการเรียนรู้และจำแนกที่หลากหลาย ซึ่งมีการพัฒนาประยุกต์ใช้ให้มีความแม่นยำมากขึ้นอย่างต่อเนื่อง ซึ่งต่อมามีการนำรูปแบบการเรียนรู้เชิงลึกมาใช้ในการจำแนกข้อมูลการรายงานอุบัติเหตุการจราจรทางบกแทนวิธีการรูปแบบอื่น ๆ

1) โครงข่ายประสาทเทียมแบบเรียนรู้เชิงลึก (Deep Neural Network: DNN)

โครงข่ายประสาทเทียมสำหรับการเรียนรู้เชิงลึกพื้นฐาน (Min et al., 2016) ที่ใช้ศึกษาเกี่ยวกับโครงข่ายประสาทเทียมเพื่อการเรียนรู้เชิงลึก ได้แสดงรูปแบบของโครงข่ายประสาทเทียมจำนวน 3 รูปแบบประกอบไปด้วย แบบเพอร์เซปตรอนหลายชั้น (Multilayer Perceptron: MLP) แบบเครือข่ายขนาดคงที่ (Deep Belief Network: DBN) และแบบตัวเข้ารหัส (Stacked Auto-Encoder: SAE) ซึ่งโครงสร้างเครือข่ายแบบต่าง ๆ สามารถได้ดังภาพ 2.2 รูปแบบของโครงข่าย

ประสาทเทียมเหล่านี้ประกอบไปด้วย ส่วนการนำเข้าข้อมูลหนึ่งชั้น ส่วนในการคำนวณชั้นประมวลผลภายในจำนวนหนึ่งชั้นหรือมากกว่า และส่วนผลลัพธ์ (Output) สุดท้ายหนึ่งชั้น โดยแต่ละชั้นเวกเตอร์ของส่วนอินพุตจะถูกคูณเมทริกซ์ (Matrix) น้ำหนักทำให้ค่าตัวแปรของภายในถูกปรับเพื่อสร้างเป็นผลรวมของน้ำหนัก จากนั้นผ่านฟังก์ชันไม่เชิงเส้น (Nonlinear Function) หรือแอคทีฟฟังก์ชัน เช่น ฟังก์ชัน Sigmoid, ฟังก์ชัน Hyperbolic Tangent Tanh หรือ ฟังก์ชันเชิงเส้น Rectified ซึ่งจะถูกใช้ในการคำนวณผลลัพธ์ของชั้นประมวลผลภายใน โดยหลักการทำงานของทั้ง 3 รูปแบบจะมีพื้นฐานการคำนวณและกระบวนการทำงานพื้นฐานเช่นเดียวกัน แต่จะมีความแตกต่างระหว่าง MLP, SAE และ DBN คือ การออกแบบชั้นประมวลผลภายในเท่านั้น



ภาพ 2.3 โครงข่ายประสาทเทียมแบบการเรียนรู้เชิงลึก (Wang et al., 2019)

MLP หรือ โครงข่ายประสาทเทียมแบบการเรียนรู้เชิงลึกแบบเพอร์เซ็ปตรอนหลายชั้น ซึ่งทำงานแบบโครงข่ายประสาทเทียมที่ประมวลผลไปเรื่อย ๆ ซึ่งมีอย่างน้อยสามชั้น (หนึ่งชั้นนำเข้า หนึ่งชั้นภายใน หนึ่งชั้นผลลัพธ์) โดยแต่ละโหนดจะเชื่อมต่อไปยังโหนดทั้งหมดของชั้นถัดไป ค่าน้ำหนักของตัวแปรในชั้นภายในจะถูกปรับภายใต้กระบวนการกระจายย้อนกลับ (Back Propagation: BP) จากงานวิจัยที่ผ่านมาพบว่า MLP มีแนวโน้มการให้ผลลัพธ์ที่ดีเมื่อมีจำนวนข้อมูลในการเรียนรู้มากเพียงพอต่อจำนวนชั้นทั้งหมดของโครงสร้างนั้น ๆ

ทั้งนี้โครงสร้างเครือข่ายและกระบวนการปรับค่าของ DBN และ SAE จะแตกต่างกับ MLP โดยชั้นผลลัพธ์จะมีตัวแปรภายในสำหรับการปรับค่าเบื้องต้น เป็นตัวแปรที่สามารถปรับเปลี่ยนได้แบบไม่ต้องการเรียนรู้ ซึ่งเป้าหมายของการค้นหาชุดน้ำหนักเริ่มต้นที่เหมาะสมชุดหนึ่งสำหรับลำดับชั้นผลลัพธ์นั้น ตัวแปรในเลเยอร์ดังกล่าวจะถูกปรับเปลี่ยนเลเยอร์ส่วนบนมีตัวการจำแนก หรือคาดการณ์จากการเรียนรู้ ดังนั้นรูปแบบการทำงานดังกล่าวสามารถช่วยลดจำนวนข้อมูลที่ใช้ในการเรียนรู้ลงได้

ในส่วนของการทำงานแบบ DBN นั้น กำหนดให้กำหนดชั้นเลเยอร์ตามรูปแบบของ RBM (Restricted Boltzmann Machines: RBM) ซึ่งประกอบไปด้วย โครงข่ายประสาทเทียมสองส่วน โดยส่วนที่หนึ่งจะเป็นการเชื่อมโยงระหว่างเลเยอร์ชั้นนำเข้ากับเลเยอร์ชั้นถัดไปที่มีจำนวนโหนดที่น้อยกว่าลงตามลำดับ โดยเชื่อมต่อเข้าสู่โครงข่ายประสาทเทียมส่วนที่สองที่ตำแหน่งเลเยอร์สมมาตร โดยส่วนที่สองจะมีจำนวนโหนดในเลเยอร์เพิ่มขึ้นไปจนถึงสิ้นสุดกับเลเยอร์ผลลัพธ์ รูปแบบ

โครงสร้างดังกล่าวมีจุดประสงค์ เพื่อลดข้อผิดพลาดที่อาจจะเกิดจากโนตสมาชิกภายในให้น้อยที่สุด ในลักษณะการนำเข้าข้อมูลสู่กระบวนการ Encode และ Decode เพื่อการส่งออกเลเยอร์ผลลัพธ์ต่อไป

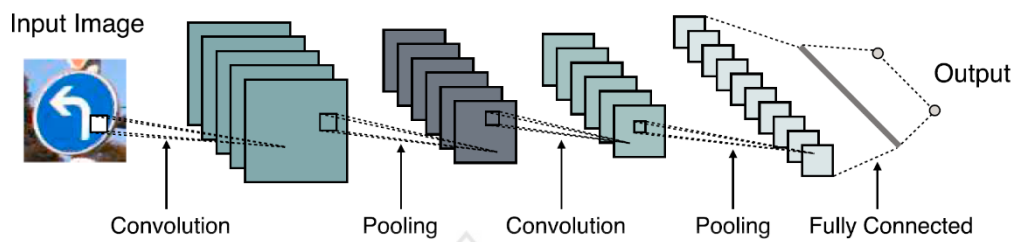
ในงานวิจัยหลายงานที่ผ่านมาแสดงให้เห็นการประยุกต์ใช้ DBN ในงานรู้จำคำพูด (Hinton et al., 2012), การจำแนกเสียง (Lee et al., 2009) และการเข้าใจภาษาธรรมชาติ (Sarikaya et al., 2014), การรู้จำวัตถุ (Maria et al., 2016), การบีบอัดภาพ (Theis et al., 2017) และแก้ไขวีดีโอ (Song et al., 2018) และ RBM ถูกใช้ในการจดจำใบหน้า (Teh et al., 2000) และการค้นหาข้อมูลในสื่อสังคมออนไลน์ (Wang and Li, 2016) ซึ่งจะสังเกตได้ว่ามีงานวิจัยจำนวนมากที่นำ DL มาประยุกต์ใช้ในงานที่หลากหลายได้

2) โครงข่ายประสาทเทียมแบบคอนโวลูชัน

โครงข่ายประสาทแบบคอนโวลูชันเป็นที่รู้จักจากความสำเร็จในการแข่งขันการจำแนกภาพของ ImageNet (Krizhevsky et al., 2012) และภายหลังถูกนำมาใช้อย่างแพร่หลายในการจัดหมวดหมู่วีดีโอ (Karpathy et al., 2014) การรู้จำการเคลื่อนไหวและกระทำ (Wang et al., 2018) และการจำแนกประโยค (Kim, 2014) รูปแบบของ CNN เริ่มจากคอนโวลูชันเลเยอร์ซึ่งมีวัตถุประสงค์ในการดึงรูปแบบสำคัญของข้อมูลสำหรับการเรียนรู้ โดยกระบวนการจะประกอบไปด้วย ตัวกรองหลักหลายชั้นที่ถูกใช้เพื่อเพิ่มความสมบูรณ์ของข้อมูลและการเปลี่ยนค่าของข้อมูลต้นฉบับไปสู่ข้อมูลที่มีรูปแบบพร้อมนำไปใช้งานมากยิ่งขึ้น โดยตัวกรองแต่ละลำดับชั้นจะอยู่ในรูปแบบของเมทริกซ์และในแต่ละเมทริกซ์จะมีค่าประจำตำแหน่งที่เปลี่ยนแปลงได้สำหรับการเรียนรู้ โดยการคอนโวลูชันจะลดขนาดของข้อมูลต้นฉบับตัวกรองที่มีขนาดเท่ากับเคอร์เนล และผลลัพธ์ที่ถูกคำนวณไว้ในแต่ละเคอร์เนลเมทริกซ์ การเลือกทำในรูปแบบเดียวกันทั้งหมดของพื้นที่ข้อมูลต้นฉบับในแต่ละลำดับชั้น

โดยผลลัพธ์ของแอคทิฟฟังก์ชันในเลเยอร์คอนโวลูชันจะถูกส่งเข้าสู่กระบวนการพูลลิงเลเยอร์ โดยมีวัตถุประสงค์ในการรวมข้อมูลและลดการซ้ำซ้อน โดยในระดับพูลลิงเลเยอร์จะทำการคำนวณในรูปแบบเดียวกันบนพื้นที่ทั้งหมดของเมทริกซ์จากเลเยอร์คอนโวลูชันที่อยู่ก่อนหน้าหนึ่งลำดับชั้น การพูลลิงโดยการเลือกค่าสูงสุดบนพื้นที่ขอบเขตการพิจารณา หรือการเลือกใช้ค่าเฉลี่ยของค่าทั้งหมดในขอบเขตพื้นที่ จะถูกเรียกว่าค่าเฉลี่ยพูลลิง เป็นต้น ทั้งนี้ที่ผ่านมามีการคำนวณโดยการพูลลิงค่าสูงสุดเป็นแนวทางการใช้งานที่เป็นที่นิยมที่สุดในปัจจุบัน

การทำงานร่วมกันระหว่างคอนโวลูชันเลเยอร์และพูลลิงเลเยอร์จะสามารถทำซ้ำได้หลายครั้งเพื่อสร้างโครงข่ายประสาทเทียมที่มีจำนวนข้อมูลป้อนเข้าที่ครอบคลุมและมากเพียงพอ ดังแสดงในภาพ 2.4 โดยสถาปัตยกรรมของ CNN จะประกอบไปด้วย คอนโวลูชันเลเยอร์และพูลลิงเลเยอร์มากกว่า 1 ชั้น ดังนั้นเลเยอร์จำนวนมากในโครงข่ายประสาทเทียมแบบ CNN จะทำให้มีโอกาสสูงที่จะสามารถตรวจจับผลลัพธ์ที่ต้องการได้ ในทางกลับกันการเรียกใช้ตัวแปรจำนวนมากนี้อาจเกิดการปรับค่าตัวแปรที่ไม่เหมาะสมได้เช่นกัน ซึ่งโดยทั่วไปการทำงานแบบ CNN จะมีคอนโวลูชันเลเยอร์จำนวนมากกว่า 2 ชั้นขึ้นไป



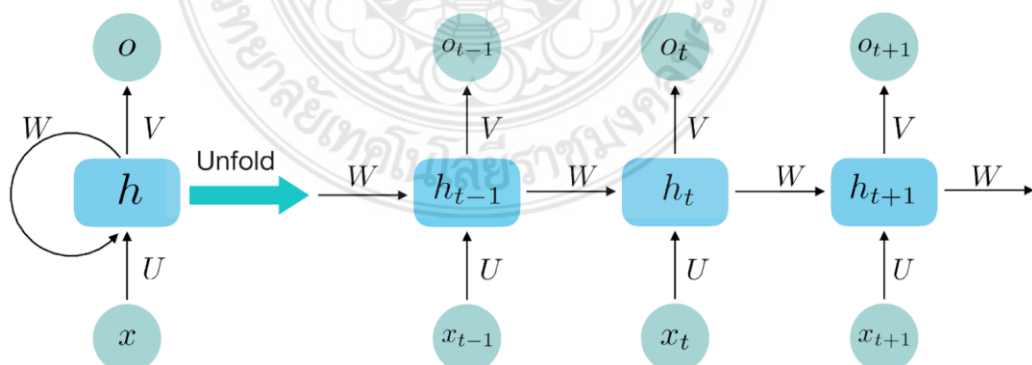
ภาพ 2.4 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Wang et al., 2019)

ในท้ายที่สุดการจับคู่ลักษณะขนาดเล็กจากหลาย ๆ เมทริกซ์เคอร์เนลและแอกทิฟฟังก์ชันจะถูกกำหนดให้มีจำนวนตามข้อมูลป้อนเข้าที่เลเยอร์เชื่อมโยงโครงข่ายประสาทเทียม (Fully Connected) โดยเลเยอร์นี้จะถูกใช้ในขั้นตอนสุดท้ายของโครงข่าย CNN เพื่อใช้สร้างผลลัพธ์ตามจำนวนที่ต้องการในการนำไปใช้งานต่อไป

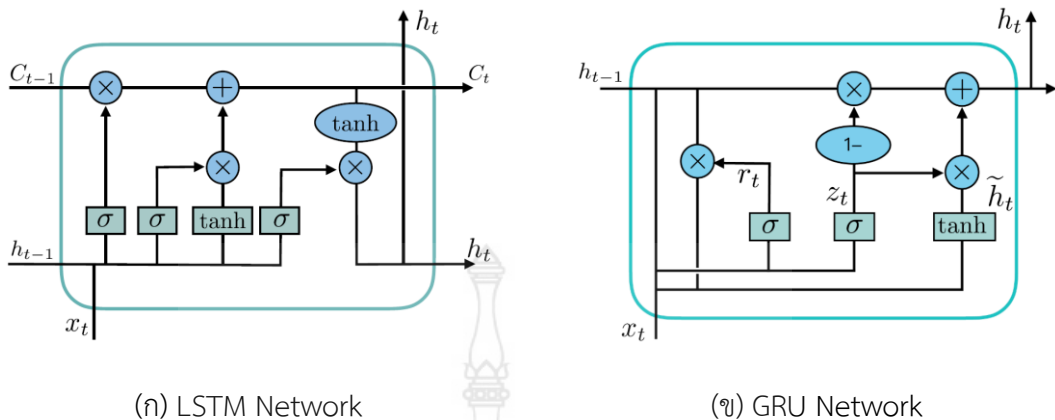
3) โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network: RNN)

โครงข่ายประสาทเทียมแบบวนซ้ำ เป็นโครงข่ายประสาทเทียมที่มีการออกแบบอย่างจำเพาะในลักษณะการใช้งานกับข้อมูลแบบเรียงลำดับ และถูกนำไปประยุกต์ใช้งานอย่างกว้างขวาง เช่น การรู้จำคำพูด (Graves and Jaitly, 2014) การแปลภาษา (Sutskever et al., 2014) การสร้างข้อความ (Sutskever et al., 2011) และการทำคำบรรยายประกอบวิดีโอ (Gao et al., 2017) ดังแสดงในภาพ 2.5 ภายในโครงสร้างแบบ RNN จะมีหน่วยความจำภายในที่เชื่อมโยงและดำเนินการแบบประมวลผลตามข้อมูลอินพุตตามลำดับ ซึ่งจะมีประสิทธิภาพในการเรียนรู้เพื่อการแก้ปัญหาสูงโดยความสามารถในการจับคู่ระหว่างลำดับของอินพุตและเอาต์พุตที่เกิดขึ้นได้โดยตรง

อย่างไรก็ตามรูปแบบ RNN ได้รับการพัฒนาให้โครงสร้างภายในโนดมีรูปแบบความจำระยะยาวและระยะสั้น ตามความเหมาะสมทางเวลาในและเรียนรู้ โดยมีรูปแบบโครงสร้างการทำงานภายใน ดังแสดงในภาพ 2.6 (ก) และ รูปแบบโครงสร้างภายในแบบ GRU ดังแสดงดังภาพ 2.6 (ข)



ภาพ 2.5 โครงข่ายประสาทเทียมแบบรีเคอร์เรนต์



(ก) LSTM Network

(ข) GRU Network

ภาพ 2.6 รูปแบบโน้ตภายในของโครงข่ายประสาทเทียมแบบรีเคอร์เรนต์

2.2 การทบทวนวรรณกรรม

2.2.1 การจัดการข้อมูลจราจรขนาดใหญ่

เมื่อข้อมูลขนาดใหญ่มีการเติบโตถึง Petabytes และเริ่มต้นเข้าสู่ยุคที่เทคโนโลยีระบบการสื่อสารพื้นฐานแบบ 4G ครอบคลุมทั่วโลกและเริ่มต้นนำร่องระบบ 5G เพื่อการเป็นสมาร์ทซิตี (Smart City) เต็มรูปแบบได้ ได้มีงานวิจัยที่แสดงให้เห็นการใช้งานจักรยานไฟฟ้า (E-Bike) ในประเทศจีนกำลังได้รับความนิยมเป็นอย่างมาก เนื่องจากมีราคาถูก ประหยัด และมีความยืดหยุ่นคล่องตัวสูง ซึ่งเหมาะสำหรับผู้ใช้ในเขตเมืองที่มีประชากรหนาแน่น แต่ร้อยละ 30 – 60 ของอุบัติเหตุที่เกิดขึ้นบนท้องถนนมีสาเหตุมาจากจักรยานไฟฟ้าและใน 1 วัน จะมีจักรยานไฟฟ้าถูกขโมยเฉลี่ยถึง 70 คัน ณ เมืองเวินโจว (Wenzhou) รัฐบาลได้ติดตั้งอุปกรณ์สื่อสารที่ใช้พลังงานต่ำที่จักรยานไฟฟ้าและมีอุปกรณ์รับเป็นสถานีสื่อสารต่าง ๆ เพื่อใช้เป็นช่องทางการติดตามรถที่ถูกขโมย โดยผู้วิจัยได้พัฒนาระบบวิเคราะห์และจำลองภาพจากข้อมูลที่ได้จากระบบเก่ามาพัฒนาฟังก์ชันในการวิเคราะห์พฤติกรรมของผู้ใช้ การเฝ้าติดตามความเร็วและการศึกษาการเคลื่อนของจักรยานไฟฟ้า โดยใช้ Mapbox API (Jia et al., 2016)

การประมวลผลข้อมูลขนาดใหญ่ในปี 2016 นั้น ได้มีงานวิจัย แพลตฟอร์ม (Platform) ของระบบวิเคราะห์จัดการการจราจรบนท้องถนนโดยใช้เทคโนโลยีการจัดเก็บข้อมูลแบบกระจาย (Distributed Storage) และเทคนิคการประมวลผลแบบขนาน (Parallel Computing) ซึ่งทำการจัดการเฝ้าติดตามพฤติกรรมของผู้ขับขี่ การตีความสถานะของการจราจรเพื่อวิเคราะห์ข้อมูลสำคัญที่เป็นประโยชน์ การใช้เทคนิคการประมวลผลแบบขนานการทำกลุ่มสมาชิก (Clusters) ช่วยให้ระบบตอบสนองได้เร็วขึ้นเวลาที่ค้นหาหรือวิเคราะห์ข้อมูลจำนวนมาก งานวิจัยใช้ระบบฐานข้อมูลของ Flume และ Hive และใช้ Apache Spark ในการสร้างกลุ่มสมาชิกการประมวลผลแบบขนาน (Peng et al., 2016)

โดยการศึกษาวิจัยเกี่ยวกับลักษณะเฉพาะของการเคลื่อนที่ของผู้ใช้งานโครงข่ายสัญญาณโทรศัพท์ 4G โดยที่มวิจัยกล่าวว่าในปัจจุบันโมเดล (Model) การเคลื่อนที่ของผู้ใช้ สามารถหาได้จาก CDR (Call Detail Records: CDR) หรือ WiFi เป็นหลัก ซึ่งวิธีการของ CDR จะใช้ทำงานได้ในขณะที่ผู้ใช้กำลังทำการโทรศัพท์ หรือ ส่งข้อความ หรือ การใช้ติดตามจาก WiFi ยังไม่ได้ เพื่อให้เข้าใจถึงพฤติกรรมในการเคลื่อนที่ของผู้ใช้งานเครือข่าย (Cellular) ในปริมาณพื้นที่ขนาดใหญ่ได้ ดังนั้นในงานวิจัยนี้จึงศึกษาความแตกต่างของลักษณะเฉพาะในการเคลื่อนที่ที่ได้จากข้อมูลทางเครือข่าย 4G กับ CDR และ 3G โดยการประมวลผลข้อมูลขนาดใหญ่จากเครือข่าย 4G ขนาด 6 TB บนโปรแกรม Hadoop ซึ่งผลลัพธ์ของงานวิจัยเป็นข้อมูลจากเครือข่าย 4G สามารถบอกถึงข้อมูลที่ละเอียดกว่า CDR และ 3G (Sun et al., 2016)

ในขณะที่มีงานวิจัย (Jai et al., 2016) นำเสนอการพัฒนาแพลตฟอร์มการประมวลผลข้อมูลขนาดใหญ่ (Platform Big Data) สำหรับการจราจรอัจฉริยะ (Smart Traffic) โดยศึกษาการประยุกต์ใช้จากงาน การนำข้อมูลจราจรมาทำรูปแบบจำลอง (Visualization) และเครื่องมือสำหรับวิเคราะห์ต่าง ๆ จากงานวิจัยเกี่ยวข้องกับการจราจรอัจฉริยะที่ผ่านมาและนำองค์ความรู้รวบรวมเป็นแพลตฟอร์มการเชื่อมโยงข้อมูล (Integrated Platform) โดยได้ทำการทดสอบกับข้อมูลการจราจรของปักกิ่ง ประเทศจีน ในการประมวลผลข้อมูลขนาดใหญ่ได้ใช้วิธีการ A* Algorithm

Cui et al. (2016) ได้ปรับปรุงพัฒนาระบบ DRIVE Net ซึ่งถูกพัฒนาโดย STAR Lab (US) โดย DRIVE Net คือเครือข่ายประเมินผลและแสดงผลของถนนแบบดิจิทัล (Digital Roadway Interactive Visualization and Evaluation Network) แบบออนไลน์ซึ่งให้บริการข้อมูลเกี่ยวกับการจราจรของถนนต่าง ๆ ซึ่งรุ่นปรับปรุงได้ถูกพัฒนาฟังก์ชันวิเคราะห์เวลาการเดินทาง (Travel Time) จากข้อมูลหลายแหล่งที่มา (Multi-Source) สามารถแสดงผลข้อมูลได้หลากหลายรูปแบบขึ้น โดยนำข้อมูลจาก ข้อมูลขนส่งมวลชน ข้อมูลผู้ใช้รถจักรยานยนต์และที่จอดรถ ข้อมูลจาก Car2go ข้อมูลจากการแหล่งข้อมูลเปิดเผย (Sharing) และข้อมูลอื่น มาทำการวิเคราะห์ข้อมูลเพื่อสร้างเป็นแบบจำลองและระบบการวิเคราะห์ข้อมูล (Data Analysis)

Nandury and Begum (2016) ได้ศึกษาวิธีการรับมือกับการจัดการกับการประมวลผลข้อมูลขนาดใหญ่ จากระบบจราจรในสมาร์ตซิตีที่กล่าวว่าการรับมือกับปริมาณข้อมูลขนาดใหญ่ที่ได้รับจาก อุปกรณ์ ตัวรับรู้และอุปกรณ์อินเทอร์เน็ตกับทุกสรรพสิ่ง (Internet of Thing: IoT) ต่าง ๆ กลายเป็นหัวข้อที่ท้าทาย ในขณะที่วิธีการและความประสิทธิภาพในการรองรับการประมวลผลข้อมูลขนาดใหญ่ ของทั้งสถาปัตยกรรมระบบยังไม่ชัดเจน สถาปัตยกรรม SWIFT ถูกออกแบบเพื่อจัดการการเชื่อมต่อระหว่าง สิ่งของชาญฉลาด (Smart Object) อุปกรณ์ชาญฉลาด (Smart Device) และระบบชาญฉลาด (Smart System) โดยงานวิจัยนี้ได้นำเสนอการทดลองนำสถาปัตยกรรม SWIFT มาทดสอบใช้กับการประมวลผลข้อมูลขนาดใหญ่ของข้อมูลจราจร เช่น การวัดความหนาแน่นของสภาพการจราจร และการควบคุมสัญญาณไฟจราจร การจัดการที่จอดรถ ระบบนำทาง และการตรวจสอบมลพิษจากยานพาหนะ

ในปี ค.ศ. 2012 รัฐบาลไต้หวันได้เปิดตัว "แพลตฟอร์มข้อมูลแบบเปิด" และสนับสนุนให้ประชาชนมีส่วนร่วมในนโยบายสาธารณะและติดตามความโปร่งใสของรัฐบาล โดยในการศึกษาวิจัยนี้ นำเสนอการรวบรวมข้อมูลที่เปิดกว้างของรัฐบาลเมืองเกาหยวนในไต้หวัน 2 แห่ง เพื่อสำรวจความสัมพันธ์ของการรายงานสายด่วนพลเมือง และอุบัติเหตุบนท้องถนนที่เกิดขึ้นจริง เพื่อใช้การวิเคราะห์ความสัมพันธ์ และตรวจสอบความสัมพันธ์ของตัวแปรในชุดข้อมูลทั้งสองนี้ ในผลการวิจัยพบว่า หลุมบ่อบนถนนและสัญญาณไฟจราจรที่ไม่เป็นระบบนั้น เป็นปัญหาที่มีการรายงานมากที่สุด 2 รายการ จากข้อมูลสายด่วนพลเมืองปี 1999 ในเมืองเกาหยวน ผลการวิเคราะห์ชี้ให้เห็นว่าหลุมบ่อบนถนนที่รายงานโดยประชาชน มีความสัมพันธ์อย่างมีนัยสำคัญกับจำนวนอุบัติเหตุบนท้องถนนที่เกิดขึ้นจริงที่เกิดจากหลุมบ่อและการกระแทกบนท้องถนน ซึ่งสามารถสรุปได้ว่าข้อมูลของสายด่วนพลเมืองมีความสำคัญ และสามารถใช้เพื่อคาดการณ์แนวโน้มและสาเหตุต่างๆ ของอุบัติเหตุจราจรได้ (Lin and Tseng, 2017)

การวิเคราะห์ข้อมูลเส้นทางเดินรถของแท็กซี่ ได้มีงานวิจัยนำเสนอการสร้างแพลตฟอร์มการขนส่งอัจฉริยะแบบรวมศูนย์ ซึ่งใช้ข้อมูลจากการประมวลผลบน Hadoop และเทคโนโลยีการประมวลผลแบบขนานขนาดใหญ่ในแพลตฟอร์มการขนส่งอัจฉริยะ มีการเสนอการออกแบบ โมดูลฟังก์ชัน โครงสร้างทางเทคนิค และอัลกอริทึม SGD ซึ่งเมื่อนำมาประมวลผลรวมกับข้อมูลรถแท็กซี่ โดยแพลตฟอร์มดังกล่าวจะวิเคราะห์ข้อมูลจากเส้นทางการเดินทางของรถ สถานะรถเปล่าหรือให้บริการ และผู้โดยสาร แพลตฟอร์มนี้ยังสามารถจัดสรรทรัพยากรของแท็กซี่ได้อย่างมีประสิทธิภาพ และช่วยรัฐบาลในการวางแผนและวิเคราะห์เส้นทางแท็กซี่ที่มีความต้องการสูงได้ (Xie and Luo, 2017)

Zhang et al. (2017) นำเสนออัลกอริทึม Safe Drive เพื่อคอยแจ้งเตือนผู้ขับขี่รถยนต์เมื่อเกิดเหตุผิดปกติขณะขับขี่รถยนต์ โดยอ้างอิงจากอุปกรณ์เซนเซอร์ต่าง ๆ ในรถ ยกตัวอย่างเช่น รอบเครื่องกับความเร็ว ความเร็วกับมุมกระทำบนถนน การเบรคแบบกะทันหัน เพื่อนำมาประมวลผลว่ารถกำลังสูญเสียการควบคุมหรือไม่ ผลลัพธ์ที่ออกมามีความแม่นยำ 83 % ซึ่งนั่นหมายถึง มีค่าความแม่นยำสูงกว่า อัลกอริทึมแบบ Rule Based

ในปี ค.ศ. 2017 การแบ่งปันรถเป็นวิธีกมนาคมที่ได้รับความนิยมในเมืองอัจฉริยะ โดยเฉพาะอย่างยิ่งระบบสนับสนุนที่ช่วยให้ลูกค้าสามารถค้นหารถยนต์ที่พร้อมใช้งานได้ครั้งละหนึ่งคัน จากนั้นเริ่มและหยุดการเช่าตามความต้องการของผู้ใช้งานในพื้นที่ ผ่านโมบายแอปพลิเคชันที่สามารถติดต่อไปยังแบ็กเอนด์บนเว็บแอปพลิเคชันเพื่อแลกเปลี่ยนข้อมูล ซึ่งในงานวิจัยนี้จึงนำเสนอ UMAP ซึ่งเป็นแพลตฟอร์มสำหรับรวบรวมข้อมูลที่เปิดให้ใช้งานบนเว็บแอปพลิเคชันผ่านแบ็กเอนด์เหล่านี้ได้อย่างอิสระ และสามารถแยกแยะรูปแบบพฤติกรรมรถที่ขับขี่ในเมืองต่าง ๆ ได้ ทั้งนี้ UMAP ที่ถูกนำเสนอ มีวัตถุประสงค์เฉพาะสองประการ ส่วนที่หนึ่ง UMAP จะทำหน้าที่ดึงข้อมูลจากแพลตฟอร์มการแชร์รถแบบเรียลไทม์ และส่วนที่สองจะทำหน้าที่ประมวลผลข้อมูล เพื่อดึงข้อมูลขั้นสูงเกี่ยวกับรูปแบบการ

ซับซ้อนและพฤติกรรมของผู้ใช้ ด้วยข้อมูลที่มีอยู่จากแพลตฟอร์มการแชร์รถ ประกอบไปด้วย ข้อมูลแผนที่และทิศทาง ซึ่งข้อมูลนี้ถูกเก็บไว้ในส่วนเก็บข้อมูลส่วนกลาง (Data Lake) จากข้อมูลที่ผ่านมาและนำมาวิเคราะห์ในภายหลัง โดยใช้โมดูลการวิเคราะห์ที่ง่ายต่อการออกแบบและปรับเปลี่ยนรูปแบบบนกรณีศึกษาของเมืองตูริน ด้วยข้อมูลระยะเวลา 50 วัน เพื่อระบุลักษณะทั้งด้านเวลาและด้านพื้นที่ของการเช่า และเพื่อกำหนดลักษณะนิสัยของลูกค้าในการใช้บริการ ซึ่งเราแตกต่างกับทางเลือกในการขนส่งสาธารณะ ผลลัพธ์จะให้ข้อมูลเชิงลึกเกี่ยวกับรูปแบบการซับซ้อนและความต้องการเฉพาะซึ่งเป็นประโยชน์สำหรับนักวางผังเมืองอัจฉริยะ และพิสูจน์ความเป็นไปได้ของแนวทางของเรา (Ciociola et al., 2017)

การขนส่งสินค้าเป็นปัญหาหลักที่ส่งผลให้เกิดความแออัดของการจราจรในเมือง และการปล่อยมลพิษในชั้นบรรยากาศในเมืองหลวงของยุโรป ดังนั้นการทำความเข้าใจรูปแบบการเดินทางที่ไม่มีรูปแบบที่แน่นอนของระบบการขนส่งเชิงพื้นที่ของการขนส่งสินค้าให้ดีขึ้นนั้น ไม่เพียงแต่จะช่วยปรับปรุงประสิทธิภาพของระบบขนส่งสินค้าเท่านั้น แต่จะช่วยลดปัญหาความแออัดของการจราจรในเมืองได้ด้วย งานวิจัยนี้มีการวิเคราะห์ข้อมูลด้วยอุปกรณ์ระบุตัวตนของรถบรรทุก (On-Board Unit: OBU) เพื่อใช้ในการติดตามการเคลื่อนย้ายสินค้าในเขตกรุงบรัสเซลส์ โดยวิธีการที่นำเสนอจะเป็นการจัดคลัสเตอร์เชิงพื้นที่ตามความหนาแน่นด้วยข้อมูล OBU ระยะเวลา 30 วัน ที่จัดเก็บไว้ในแพลตฟอร์มข้อมูลขนาดใหญ่ เพื่อนำข้อมูลดังกล่าวมาประมวลผลโดยอัลกอริทึมการจัดกลุ่มเพื่อระบุข้อมูลเชิงลึกด้วยข้อมูลป้อนเข้า 4 ส่วน ประกอบด้วย ข้อมูลทางเข้า ข้อมูลถนนทางออก ข้อมูลโกดังสินค้า อุตสาหกรรมบรัสเซลส์-แอนต์เวิร์ป และข้อมูลถนนวงแหวนรอบนอกข้ามเขตเมือง นำมาประเมินเชิงปริมาณเพื่อแสดงปัญหาความแออัดตามเส้นทางเพื่อการแก้ไขปัญหาต่อไป (Buroni et al., 2018)

การใช้งานระบบการวิเคราะห์ข้อมูลขนาดใหญ่สำหรับการวิเคราะห์ข้อมูลการโดยสารรถแท็กซี่ในเมืองนิวยอร์ก สำหรับการเปรียบเทียบแพลตฟอร์มสำหรับการคำนวณระหว่าง Amazon Athena, Serverless Microservices ในรูปแบบ SQL หรือ NoSql ด้วย Hadoop และ Spark โดยใช้เทคนิคการปรับแต่งและการเพิ่มประสิทธิภาพที่หลากหลาย เพื่อแก้ปัญหาการวิเคราะห์ข้อมูลที่มีขนาดอินพุต คุณภาพข้อมูล และความซับซ้อนในการคำนวณ ด้วยข้อมูลรถแท็กซี่ของนิวยอร์กตั้งแต่ปี ค.ศ. 2009 ถึง 2017 เพื่อหาจำนวนการโดยสารที่สามารถนำมารวมกันได้ รวมการเรียกแท็กซี่ตามสถานีที่รับ เวลาที่รับ และสถานีที่ส่ง บนเกณฑ์มาตรฐานที่กำหนดไว้เดิมคือการเดินทางภายในห้านาที โดยมีจุดรับและส่งภายในครึ่งกิโลเมตร ที่ต้องประมวลผลการเดินทางมากกว่า 1.2 พันล้านครั้ง โดยใช้ Amazon EMR พร้อม Spark โดยผลการทดสอบการรวมเครื่องมือการประมวลผลประเภทต่าง ๆ และเปรียบเทียบประสิทธิภาพ งานวิจัยนี้สามารถรวมการเดินทางได้ 35 % (Kaplanovich and Yesha, 2018)

การคาดคะเนความเร็วการจราจรบนถนนเป็นปัญหาที่ท้าทายในระบบขนส่งอัจฉริยะและได้รับความสนใจเพิ่มมากขึ้น งานที่มีอยู่ส่วนใหญ่จะอ้างอิงตามข้อมูลการตรวจจับความเร็วที่ได้รับจากเซ็นเซอร์ตรวจวัดสภาพจราจร หรืออุปกรณ์ติดตามตำแหน่งที่ติดตั้งไว้บนยานพาหนะ ซึ่งมีจำนวนน้อยจนไม่เพียงพอต่อการนำไปประยุกต์ใช้งานข้อมูล ด้วยค่าใช้จ่ายในการติดตั้งและบำรุงรักษาที่มีราคาแพง ซึ่งวิธีการตรวจจับสภาพจราจรแบบเดิมเพื่อตรวจจับความเร็วเท่านั้นไม่เพียงพอ โดยเฉพาะเมื่อเกิดเหตุฉุกเฉิน เช่น อุบัติเหตุจราจร โดยบทความนี้ นำเสนอการปรับปรุงการคาดคะเนความเร็วของการจราจรบนถนน โดยการรวมข้อมูลการตรวจจับความเร็วแบบเดิมเข้ากับข้อมูล "การตรวจจับ" แบบใหม่จากแหล่งที่มาอื่นนอกเหนือวิธีการเดิม เช่น ข้อมูลทวิตจากโซเชียลมีเดีย หรือ ข้อมูลเส้นทางจากแผนที่แพลตฟอร์มที่ให้บริการการจราจร ซึ่งการสร้างแบบจำลองข้อมูลจากชุดข้อมูลที่แตกต่างกันนั้นมีความซับซ้อนสูง รวมถึงความไม่แน่นอนของตำแหน่งของข้อมูลความละเอียดต่ำ ภาษาที่คลุมเครือของคำอธิบายการรับส่งข้อมูลในข้อความ และความแตกต่างของข้อมูล เพื่อแก้ปัญหาดังกล่าว งานวิจัยนี้นำเสนอเฟรมเวิร์กวิเคราะห์ความน่าจะเป็นแบบครบวงจรแบบ TEGPAM (Topic-Enhanced Gaussian Process Aggregation Model: TEGPAM) ซึ่งต้องใช้ข้อมูล 3 องค์ประกอบ ประกอบด้วย โมเดลการแยกตำแหน่ง โมเดลหัวข้อการรับส่งข้อมูล และโมเดลกระบวนการเกาส์เซียนความเร็วการรับส่งข้อมูล โดยข้อมูลที่ถูกระบุขึ้นใหม่นี้ทำให้มีข้อมูลประสิทธิภาพมากขึ้นกว่าข้อมูลเดิม ในอนาคตงานวิจัยนี้ จะมีการนำ GP-based และ Distributive GP มาใช้งานร่วมกับการประมวลผล เพื่อที่จะทำให้การทำนายมีความแม่นยำมากขึ้น พร้อมกับสามารถประเมินได้แบบเรียลไทม์ (Lin et al., 2018)

ระบบ ITS (Intelligent Transport System: ITS) ในประเทศอินเดียในปัจจุบันไม่สามารถบริหารจัดการระบบจราจรได้อย่างมีประสิทธิภาพ เนื่องจากลักษณะการขับขี่ รูปแบบถนน และสภาพสิ่งแวดล้อมของประเทศอินเดียนั้น ไม่สามารถนำแพลตฟอร์ม ITS ในประเทศที่มีรูปแบบและโครงสร้างพื้นฐานด้านการจราจรที่เป็นมาตรฐานมาใช้งานได้โดยตรงได้ งานวิจัยนี้นำเสนอการเก็บข้อมูลและวิเคราะห์ลักษณะการจราจรของประเทศอินเดียข้อมูลจริง เพื่อที่จะนำข้อมูลนี้มาพัฒนาระบบ ITS ภายในประเทศ โดยใช้เทคนิคการบริหารจัดการข้อมูลขนาดใหญ่ (Big Data Analytics) เพื่อแก้ปัญหาความแออัดในการจราจร จะต้องมีการปรับวิธีการจัดทำเหมืองข้อมูลที่เหมาะสมซึ่งให้สำหรับยานพาหนะที่มีลำดับความสำคัญ เพื่อแจ้งคำแนะนำที่เหมาะสมต่อการบริหารจัดการจราจรแก่หน่วยงานภาครัฐที่เกี่ยวข้อง ในการใช้ข้อมูลการระบุตำแหน่งแบบ GPS และการวิเคราะห์ข้อมูลการเดินทางขนาดใหญ่ในการใช้งานระบบนี้ได้อย่างมีประสิทธิภาพ (Balamurugan et al., 2018)

การศึกษาเทคโนโลยีอินเทอร์เน็ตกับทุกสรรพสิ่งในการทำระบบจัดการพื้นที่สำหรับจอดรถ ซึ่งพวกเขาเชื่อว่าจะช่วยลดเวลาของผู้ขับขี่ที่จำเป็นจะต้องอยู่บนยานพาหนะ โดยมี 2 เป้าหมายหลัก เป้าหมายระยะยาวคือเพิ่มประสิทธิภาพของระบบจัดการ บริหารพื้นที่จอดรถ เป้าหมายระยะสั้นคือช่วยให้ผู้ขับขี่ใช้เวลาในการหาที่จอดรถน้อยลงโดยการหาที่จอดรถที่เหมาะสม ในงานวิจัยนี้ให้เทคโนโลยีอินเทอร์เน็ตกับทุกสรรพสิ่งในการตรวจสอบพื้นที่ ระยะเวลาการจอดรถในเมืองและทำการวิเคราะห์หาพื้นที่จอดรถบริเวณใกล้ ๆ โดยระบบทำงานโดยไม่ต้องในการส่งข้อมูลเยอะ ไม่สำรองข้อมูลที่อุปกรณ์ตรวจจับ ซึ่งข้อมูลทั้งหมดจากถูกส่งไปยังแม่ข่ายแบบคลาวด์ (Cloud Server) ส่วนการวิเคราะห์ข้อมูลได้เลือกใช้ Hadoop และการลดขนาดแผนที่ (Map Reduce) พร้อมทำงานบน Cluster Computer (Nguyen et al., 2018)

Ren et al. (2019) ดำเนินการศึกษานำเอาเทคโนโลยีบล็อกเชน (Block Chain) มาประยุกต์ใช้เป็นศูนย์กลางในการเก็บ หรือส่งต่อข้อมูล ในระบบ ITS และใช้อุปกรณ์อินเทอร์เน็ตกับทุกสรรพสิ่งติดตั้งเป็นอุปกรณ์ตรวจจับตามที่ตั้งต่าง ๆ โดยได้ทำการเปรียบเทียบกับระบบที่ใช้ศูนย์กลางข้อมูลที่เป็นแม่ข่ายแบบคลาวด์ และได้ออกแบบการส่งข้อมูลระหว่างยานพาหนะเป็นแบบระหว่างสมาชิกโดยตรง (Peer-to-Peer) เพื่อเป็นการอาศัยเทคนิคบล็อกเชนในการกระจายสิทธิการสื่อสารและควบคุม เพื่อให้ยานพาหนะทำงานร่วมกันได้ โดยไม่ต้องผ่านการอนุญาตส่วนประมวลผลส่วนกลาง

Yang and Zhu (2019) ดำเนินการศึกษาวินิจฉัยพัฒนาความสามารถในการจดจำเวลาการเดินทาง ในช่วงของถนนแบบยืดหยุ่นในแพลตฟอร์มให้ข้อมูลจราจร โดยทั้งสองมีความเห็นว่าข้อมูลการประมวลผลข้อมูลขนาดใหญ่จราจรสามารถผลตอบกลับหรือการแนะนำวิธีแก้ปัญหาจราจรติดขัดได้อย่างมีประสิทธิภาพ โดยใช้วิธีการวิเคราะห์เวลาการเดินทางจากการประมวลผลข้อมูลขนาดใหญ่แบบ Dynamic Prediction แบบเวลาจริง (Real Time) ทดสอบกับโครงสร้างของระบบตรวจสอบการจราจรที่ใช้เทคโนโลยี RFID tag และใช้เทคนิคการผสมผสานข้อมูล (Fusion Data) โดยผลการทดสอบในแบบจำลองแสดงให้เห็นว่าวิธีการใหม่ที่ใช้มีความถูกต้องมากขึ้น และสามารถแก้ไขปัญหาจราจรติดขัดและทำให้การจราจรคล่องตัวขึ้น

Zhu et al. (2019) นำเสนองานศึกษาวินิจฉัยเกี่ยวกับการสำรวจระบบวิเคราะห์การประมวลผลข้อมูลขนาดใหญ่ในระบบการขนส่งอัจฉริยะ โดยงานสำรวจวิจัยได้อธิบายถึงประวัติความเป็นมาของลักษณะเฉพาะของทั้งเทคโนโลยีการประมวลผลข้อมูลขนาดใหญ่และ ITS โดยระบบเฟรมเวิร์กที่ใช้ในการดำเนินการวิเคราะห์การประมวลผลข้อมูลขนาดใหญ่ใน ITS วิธีการในการเก็บข้อมูล เครื่องมือในการวิเคราะห์ผล วิธีการวิเคราะห์ผลรวมถึงการวิเคราะห์อุบัติเหตุการจราจรบนท้องถนน การประมาณการความคล่องตัวบนท้องถนน การบริหาร วางแผนและจัดการระบบขนส่งสาธารณะ การวางแผนการเดินทางส่วนบุคคลตลอดจนการวางแผนบำรุงรักษา

2.2.2 การตรวจจับการรายงานอุบัติเหตุบนสื่อสังคมออนไลน์

Chung et al. (2013) นำเสนอการสถาปัตยกรรมในการประมวลผลข้อมูลที่เกี่ยวข้องกับการจราจรขนาดใหญ่ แบบกระจายข้อมูลจำนวนมากแบบทันที (Massive Real-time) ในลักษณะ NoSQL โดยใช้ Hadoop และ HBase ทำหน้าที่เก็บรวบรวมและวิเคราะห์ข้อมูล เพื่อนำข้อมูลดังกล่าวประกอบไปด้วย PeMS (Performance Measurement System: PeMS) ของหน่วยงานทางหลวงแคลิฟอร์เนีย เป็นข้อมูลจากอุปกรณ์ตรวจจับยานพาหนะแบบขดลวดเหนียวนำ (Loop Detector) ที่ให้ข้อมูลความเร็ว (Speed) กระแสจราจร (Flow) และข้อมูลอัตราความหนาแน่น (Occupancy) และ TASAS (Traffic Accident Surveillance and Analysis System: TASAS) เป็นข้อมูลวันและเวลา ลักษณะและความรุนแรง จำนวนยานพาหนะที่เกี่ยวข้อง ที่ถูกรวบรวมย้อนหลัง 10 ปี ขนาด 1 TB มาจัดทำตารางแสดงค่าความน่าจะเป็นการเกิดอุบัติเหตุ จากความเร็วระหว่างความเร็วขาเข้า (Upstream) และขาออก (Downstream) โดยแบ่งระยะช่วงความเร็วในการตรวจจับพบทุก 5 mph ของทั้ง 2 ทิศทาง เพื่อใช้ตรวจสอบแนวโน้มการเกิดอุบัติเหตุต่อไป โดยทำการทดสอบกับข้อมูลจริงสำหรับถนน I-880N รัฐแคลิฟอร์เนีย สหรัฐอเมริกา

Zhang (2015) นำเสนอวิธีการผสมผสานระหว่าง LDA (Latent Dirichlet Allocation: LDA) เทคนิคการเลือกคำหลัก (Topic) ในกลุ่มข้อความหรือหนังสือ และ Document Clustering เทคนิคการจัดกลุ่มคำซ้ำเพื่อแสดงนัยสำคัญในรูปแบบข้อความขนาดใหญ่-เล็ก เพื่อตรวจจับการแจ้งเหตุที่มีในทวิตเตอร์ โดยปัจจุบันมีผู้ใช้งานทวิตเตอร์ที่ยังใช้งานอยู่ 200 ล้านคน และมีการทวิตข้อความเฉลี่ยวันละ 500 ล้านทวิตต่อวัน โดย 3 ข้อความแรกที่จะหมายถึงการเกิดเหตุประกอบไปด้วย Vehicle, Driver และ Tow และลำดับที่สองที่หมายถึงสถานการณ์จะเป็นคำว่า Call, Check, Assist และ Push โดยทดสอบที่เมืองซีแอตเทิล (Seattle) ซึ่งการแจ้งเหตุจะมีรัศมีการแจ้งครอบคลุมประมาณ 600 เมตร จากที่เกิดเหตุ ยิ่งในบริเวณดังกล่าวมีความหนาแน่นสูงการระบุเหตุยิ่งกระจุกตัวใกล้กับพื้นที่เกิดเหตุโดยวิธีการ DBSCAN Algorithm

Georgiou et al. (2015) นำเสนอการรวบรวมทวิตที่เกี่ยวข้องกับจราจรหรือในพื้นที่ที่สนใจและมีข้อความที่เกี่ยวข้อง เช่น “This Traffic” หรือ “On My Way” มาจัดกลุ่มตามช่วงเวลาเรียกว่า ปริมาณการใช้สื่อสังคมออนไลน์ในแต่ละช่วงเวลา (Social Volume) เพื่อนำมาวิเคราะห์ปริมาณจราจรแทน โดยทำการทดลองเปรียบเทียบระหว่างปริมาณจราจรในแต่ละช่วงเวลา (Traffic Volume) และปริมาณการใช้สื่อสังคมออนไลน์ในแต่ละช่วงเวลาของถนนช่วง California Freeway I-405 ซึ่งปริมาณข้อมูลมีความคล้ายคลึงกันและมีความสัมพันธ์เชิงเส้น โดยมีการจัดกลุ่มอัตราส่วนระหว่างปริมาณจราจรในแต่ละช่วงเวลาและปริมาณการใช้สื่อสังคมออนไลน์ในแต่ละช่วงเวลาในแต่ละช่วงเวลา โดยเรียกว่า Shift-Based Model ทำให้สามารถตรวจวัดสภาพจราจรได้ด้วยเงินลงทุนที่น้อยลง

ในขณะนั้นประเทศอินโดนีเซียมีการพัฒนาการรายงานเหตุจากหลากหลายช่องทาง ด้วยการพัฒนาบริการข้อมูลออนไลน์จำนวนมากอย่างรวดเร็วทำให้ผู้ใช้ทางไม่มีเวลาอ่านข้อมูลให้ครบถ้วนก่อนการเดินทาง ซึ่งในงานวิจัยนี้นำเสนอการใช้ข้อมูลทวีตเตอร์ภาษาอินโดนีเซีย มาแยกคำและตีความหมายข้อความจากทวีตเตอร์ ผ่านเครื่องมือติดต่อและค้นหาที่เรียกว่า Twitter API จากนั้นจึงสร้างสรุปเกี่ยวกับการเข้าชม ข้อมูลสรุปสามารถปรับปรุงประสิทธิภาพด้านเวลาและประสิทธิผลในการรับข้อมูลเกี่ยวกับสภาพการจราจร ซึ่งในงานวิจัยดังกล่าวพบปัญหาในด้านโครงสร้างของทวีตเตอร์ ทวีตที่ไม่มีโครงสร้างและรูปแบบที่แน่ชัด ทำให้ยากต่อการประมวลผลข้อมูลคอมพิวเตอร์ ในการแก้ปัญหาเหล่านี้แล้วสรุปโดยใช้การประมวลภาษาธรรมชาติ วิธีการวิเคราะห์ความสัมพันธ์ และความหมายเพื่อปรับปรุงโครงสร้างของคำ และการแยกวิเคราะห์ประโยคเพื่อให้สามารถจำแนกสภาพการจราจรทวีตของประโยคได้อย่างมีประสิทธิภาพ (Aziz et al., 2015)

Zhang et al. (2018) นำเสนอการใช้จักรกลเรียนรู้เชิงลึกในการตรวจสอบอุบัติเหตุจราจรจากข้อมูลสื่อสังคมออนไลน์ ได้ทำการตรวจสอบเนื้อหาทวีตมากกว่า 1 ล้านครั้งในระยะเวลา 1 ปี ในพื้นที่สองเมือง ประกอบไปด้วย เวอร์จิเนียตอนเหนือ และนิวยอร์ก การเลือกคำสำคัญจากประโยค Apriori Algorithm และนำเข้าสู่ ML เพื่อหาคำตอบต่อไป ซึ่งการทดลองจะเปรียบเทียบระหว่างวิธีการ DBN, LMTS, SVMs และ sLDA โดยผลการทดลองจะแสดงให้เห็นว่าการเลือกคำสำคัญที่มีอยู่ในทวีตที่เกี่ยวข้องกับอุบัติเหตุร่วมกับการการตรวจจับอุบัติเหตุจราจรจักรกลเรียนรู้เชิงลึก แบบ DBN จะสามารถตรวจสอบการเกิดอุบัติเหตุได้ถูกต้องถึง 80 % พร้อมการระบุพื้นที่การเกิดอุบัติเหตุได้

Neuhold et al. (2018) นำเสนอการเก็บข้อมูลบนทางด่วนสำหรับการบริหารจัดการจราจรนั้น ขึ้นอยู่กับตำแหน่งการติดตั้งอุปกรณ์ตรวจวัดสภาพจราจรอัตโนมัติ และระบบตรวจสอบสภาพจราจรกล้อง เท่านั้น คณะวิจัยนำเสนอแนวทางในการใช้สื่อสังคมออนไลน์เป็นแหล่งข้อมูลเพิ่มเติม โดย ASFINAG ผู้รับผิดชอบการคมนาคมทางด่วนของประเทศออสเตรีย พัฒนาระบบแสดงข้อมูลความเร็วเฉลี่ยทุกระยะหนึ่งกิโลเมตรในแต่ละช่วงเวลาของวัน ทำให้ผู้ขับขี่ได้รับทราบข้อมูลข่าวสารสถานการณ์โดยรวดเร็วในรูปแบบตารางสี โดยพัฒนาให้สามารถรวบรวมคำอธิบายเหตุการณ์จากเฟซบุ๊กและ RSS (Really Simple Syndication: RSS) ของสถานีวิทยุออสเตรียและข่าวหนังสือพิมพ์ที่เป็นที่นิยม ที่ผ่านการกรองข้อความที่เกี่ยวข้องที่เป็นประโยชน์และประมวลผลเข้ากับข้อมูลการจราจร และนำเสนอเป็นสัญลักษณ์สีแทนสภาพจราจรหรือความเร็วเฉลี่ยบนเส้นทางในขณะนั้น

Jin and Liu (2018) นำเสนอการนำข้อมูลจากสื่อสังคมออนไลน์ที่มีการกล่าวถึงหรือรายงานอุบัติเหตุการณ์ในปัจจุบันมีจำนวนที่มากขึ้น ทำให้สามารถตรวจพบแนวโน้มของการเกิดอุบัติเหตุและเหตุการณ์ที่เกิดบนถนนนั้น ๆ ได้อย่างรวดเร็ว ในงานวิจัยนี้คณะวิจัยได้นำเสนอลักษณะของปัญหา รวมถึงความซับซ้อนของการใช้หลักการของการทำเหมืองข้อมูล (Data Mining) แบบ Spatio-Temporal ในการวิเคราะห์หาการเกิดอุบัติเหตุหรือเหตุการณ์ต่าง ๆ บนท้องถนน

ด้วยอุบัติเหตุบนท้องถนนที่เกิดขึ้นทั่วไป สภาวะอุบัติเหตุที่เกิดขึ้นในแต่ละเหตุการณ์นั้นมีหลายปัจจัยที่ใช้อธิบายอุบัติเหตุทางจราจรได้ เช่น ลักษณะของสภาพแวดล้อมในเวลานั้น ๆ พฤติกรรมการขับขี่ สภาพอากาศ และปัจจัยที่ซับซ้อนอื่น ๆ ที่ไม่แน่นอน ผลลัพธ์การเกิดอุบัติเหตุบนท้องถนนนั้นไม่เป็นรูปแบบเชิงเส้น (Linear) ดังนั้นต้องวิเคราะห์ข้อมูลจากความสัมพันธ์ระหว่างข้อมูลหลายด้านเพื่อนำมาใช้ลดความเสี่ยงการเกิดอุบัติเหตุบนท้องถนนในอนาคต โดยงานวิจัยนี้นำเสนอการนำข้อมูลการแจ้งอุบัติเหตุที่ผ่านมา มาประมวลผลล่วงหน้าและการจัดหมวดหมู่โดยใช้เครื่องมือการทำเหมืองข้อมูล เพื่อหาข้อมูลที่สามารถทำนายสาเหตุของการเกิดอุบัติเหตุได้ ผลลัพธ์จากการวิจัยแสดงให้เห็นว่าวิธีการที่นำเสนอสามารถตรวจสอบและทำนายเหตุการณ์ของอุบัติเหตุการที่มีแนวโน้มที่จะเกิดขึ้นได้ โดยการวิเคราะห์ลักษณะและความสัมพันธ์ของข้อมูล ด้วยแนวคิดของการรายงานเหตุการณ์ผ่านกลุ่มคนออนไลน์ ทั้งนี้ในขั้นตอนการรวบรวมข้อมูลที่เกี่ยวข้องกับอุบัติเหตุเพิ่มเติมจากผู้ใช้งาน ระบบที่เสนอนั้นใช้วิธีการทำเหมืองข้อมูล และเทคนิคการประมวลผลข้อมูลขนาดใหญ่ที่เหมาะสมสำหรับการสำรวจข้อมูลอุบัติเหตุ โดยใช้ Hadoop เพื่อจัดหมวดหมู่อุบัติเหตุบนท้องถนนและตรวจสอบแนวโน้มที่จะเกิดอุบัติเหตุในอนาคต (Abdallaoui et al., 2018)

Chen et al. (2018) นำเสนอการหาเหตุการณ์สำคัญที่เกี่ยวข้องกับการจราจรจากสื่อสังคมออนไลน์ Sina Weibo ซึ่งเป็นแพลตฟอร์มไมโครบล็อกของประเทศจีน โดยใช้เทคนิค Bag of Word Model ในการหาคำสำคัญ ภาษาจีนนำมาเทียบเป็นภาษาอังกฤษเพื่อป้อนเข้าโครงข่ายประสาทเทียมและทดลองการใช้ CNN, LSTM และวิธีการผสมระหว่าง CNN และ LSTM ในการเรียนรู้จำลักษณะคำและเหตุการณ์ เพื่อใช้ในการค้นหาเหตุการณ์ทางจราจร

Yamazaki (2019) นำเสนอการวิเคราะห์การเกิดอุบัติเหตุจากข้อมูลเปิดเผยภาครัฐ (Open Data) คือ ข้อมูลสาธารณะภาครัฐที่สามารถเผยแพร่ได้ ของเมืองนีงาตะ (Niigata) ประเทศญี่ปุ่น ซึ่งข้อมูลที่เป็นประโยชน์ในวิเคราะห์ปัญหาการเกิดอุบัติเหตุดังกล่าวสามารถจำแนกเป็น ข้อมูลการพยากรณ์อากาศ ข้อมูลสถิติเหตุการณ์ของอุบัติเหตุระหว่างยานพาหนะกับยานพาหนะ (Vehicle to Vehicle: V2V) และอุบัติเหตุระหว่างบุคคลกับยานพาหนะ (Human to Vehicle: H2V) และอุบัติเหตุแบบไม่มีคู่กรณีร่วมกับข้อมูลภาพแบบพาโนรามาของสี่แยกไฟแดง จากผลการวิเคราะห์จากข้อมูลทั้งหมดที่กล่าวถึงจะแสดงให้เห็นว่าทางแยกที่ไม่มีไฟจราจรจะเกิดอุบัติเหตุ V2V มากกว่าทางแยกที่มีไฟจราจร

ทั้งนี้ในการศึกษาทบทวนงานวิจัยในหลายปีที่ผ่านมา ด้านการเรียนรู้เชิงลึก มีงานวิจัยที่เผยแพร่ออกมาอย่างมากมายไม่ว่าจะเป็นผลวิจัยการเรียนรู้เชิงลึกด้านภาพเพื่อนำมาสู่การพัฒนาทางด้านคอมพิวเตอร์วิชัน (Computer Vision) ผลการวิจัยทางด้านพยากรณ์ (Predict) เหตุการณ์ต่าง ๆ หรือแม้กระทั่งการทำงานทางด้านการประมวลผลภาษาธรรมชาติ ซึ่งเป็นการทำงานทางด้านการประมวลผลภาษาไม่ว่าจะเป็นการประมวลผลภาษาจากเสียงหรือการประมวลผลภาษาจากตัวอักษรในเอกสารสำหรับการประมวลผลภาษาธรรมชาติจากตัวอักษรในเอกสารได้มีงานวิจัยออกมาอย่างมากมายเช่นกัน เช่น การประมวลผลเพื่อหาความรู้สึกทางด้านเชิงบวก (Positive) หรือเชิงลบ (Negative) ต่อสินค้าที่มีอยู่บนสื่อต่าง ๆ ตลอดจนการประมวลผลเพื่อจำแนกประเภทในรูปแบบอื่น ๆ อย่างไรก็ตามการจำแนกประเภทของข้อความเกี่ยวกับข้อความทางด้านการจราจรนั้นเป็นสิ่งที่น่าสนใจสำหรับภาษาไทยที่มีเอกลักษณ์และมีความซับซ้อนเป็นอย่างมาก และยังไม่มียานวิจัยในการแก้ปัญหาการตรวจจับอุบัติเหตุจากข้อความที่ประกาศหรือเผยแพร่ในสื่อสังคมออนไลน์ ผู้วิจัยได้ศึกษาวิธีการเรียนรู้เชิงลึกจากหลาย ๆ งานวิจัยทั้งในและต่างประเทศจนสรุปวิธีการประมวลผลที่จะใช้ในงานวิจัยครั้งนี้กัน 4 วิธีคือ MLP, CNN, LSTM และ CNN+LSTM ดังรายละเอียดในการนำเสนอการในบทต่อไป

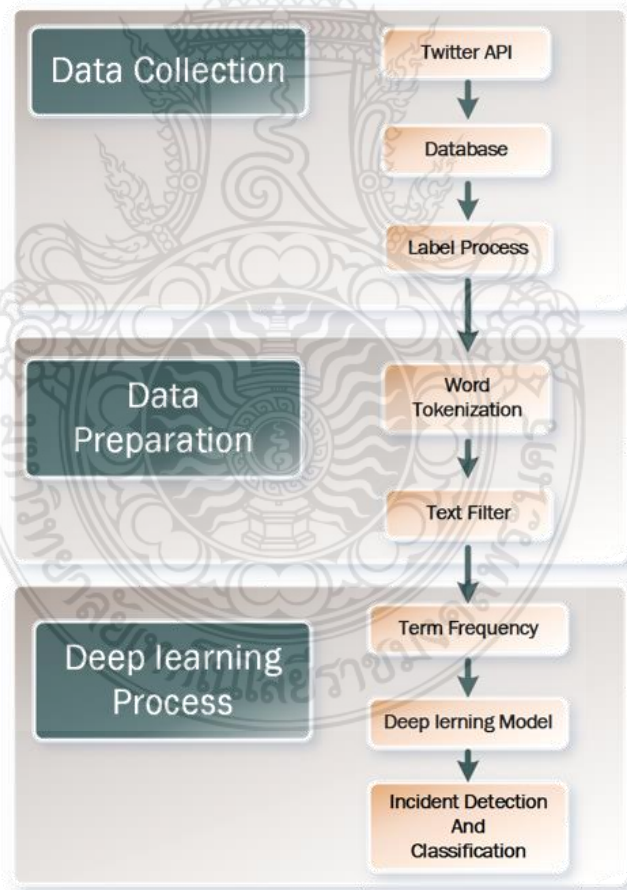


บทที่ 3

การพัฒนาแบบการรวบรวมและจำแนกอุบัติการณ์ (Incident Detection and Classification Development)

3.1 กรอบแนวคิดการออกแบบ (Conceptual Design)

ในการออกแบบระบบการรวบรวมข้อมูลและจำแนกอุบัติการณ์กำหนดเป้าหมายและวัตถุประสงค์ โดยการออกแบบที่สำคัญ คือ การระบุหรือจำแนกข้อความเป้าหมายเป็นข้อความที่เกี่ยวข้องกับอุบัติการณ์หรือไม่ หากข้อความดังกล่าวเป็นอุบัติการณ์ ข้อความดังกล่าวจะมีสถานการณ์ในรูปแบบใด และมีความรุนแรงอยู่ที่ระดับความรุนแรงใด ซึ่งกรอบแนวคิดในการจำแนกอุบัติการณ์วิธีการเรียนรู้เชิงลึก สามารถลำดับขั้นตอนหลักได้ 3 ขั้นตอน ดังแสดงในภาพ 3.1



ภาพ 3.1 กรอบแนวคิดการออกแบบรูปแบบการรวบรวมและจำแนกอุบัติการณ์

1) ขั้นตอนการรวบรวมข้อมูลมาจากแหล่งข้อมูล โดยการวิจัยนี้ พิจารณาเลือกใช้แหล่งข้อมูลจากแพลตฟอร์มสื่อสังคมออนไลน์ในชื่อทวีตเตอร์ที่มี API รองรับการเรียกข้อมูลจากภายนอกซึ่งทำให้สามารถเรียกข้อมูลได้อย่างอัตโนมัติ และสามารถระบุตำแหน่ง (Geo-Location) ได้จากข้อมูล

2) ขั้นตอนการเตรียมข้อมูลสำหรับการเรียนรู้เชิงลึกของระบบโครงข่ายประสาทเทียม ซึ่งข้อมูลที่ได้จากรวบรวมมานั้นจะต้องนำข้อมูลกลุ่มข้อความต้นฉบับ ผ่านกระบวนการคัดกรองและแยกคำไทยให้พร้อมสำหรับการเรียนรู้เชิงลึกของกระบวนการโครงข่ายประสาทเทียมได้

3) ขั้นตอนการเรียนรู้เชิงลึก เป็นขั้นตอนที่มีกระบวนการทำงาน ซึ่งประกอบไปด้วย ส่วนประมวลผลและทดสอบย่อย ๆ เช่น การตั้งค่าการเรียนรู้ (Training) เพื่อสร้างการสร้างโมเดล การทดสอบประสิทธิภาพ พร้อมการปรับตั้งค่าพารามิเตอร์ เพื่อให้วิธีการที่นำเสนอมีความถูกต้องมากขึ้น

3.2 วิธีการและขั้นตอนการดำเนินงาน

3.2.1 กระบวนการรวบรวมและบันทึกข้อความ (Data Collection)

ข้อความที่ได้นำมาศึกษาคือข้อความที่ได้มาจากทวีตเตอร์ โดยภายในข้อความของทวีตเตอร์นั้น มีองค์ประกอบของข้อมูลหลายองค์ประกอบ ซึ่งสามารถนำข้อมูลเหล่านั้นมาใช้งานจำแนกและวิเคราะห์ความหมายได้ โดยส่วนประกอบจากแพลตฟอร์มดังกล่าวจะประกอบไปด้วยองค์ประกอบหลักดังนี้ ชื่อของผู้โพสต์ข้อความ (Name) ข้อความ (Content) เวลา (Date-Time) และแฮชแท็ก ดังแสดงในภาพ 3.2

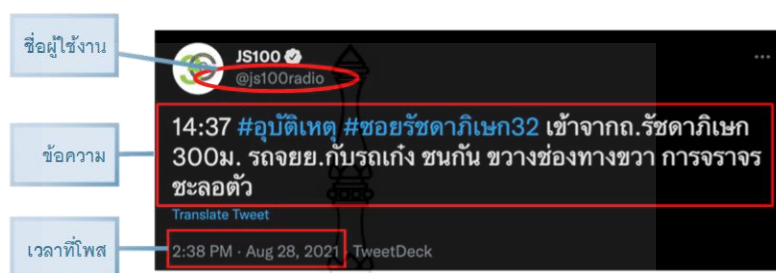
ทวีตเตอร์ที่มีในปัจจุบันนี้มีผู้ใช้งานหลากหลายทั้งที่เป็นทวีตเตอร์รายบุคคล หรือทวีตเตอร์ระดับองค์กร โดยในงานวิจัยพิจารณาเลือกใช้ทวีตเตอร์ในระดับองค์กรที่เกี่ยวข้องกับการรายงานสภาพจราจรสำหรับประเทศไทย จำนวน 3 หน่วยงาน ได้แก่

1) @js100radio ซึ่งเป็นบัญชีทวีตเตอร์ของที่มีรากฐานมาจากการรายงานข่าวทางสื่อวิทยุ ต่อมาได้พัฒนาเพิ่มเติมส่วนของการรายงานข่าวผ่านสื่อสังคมออนไลน์ภายใต้การดูแลของ บริษัท แปซิฟิก คอร์ปอเรชั่น จำกัด มีผู้ติดตามกว่า 3 ล้านบัญชีผู้ใช้งาน

2) @Traffic_1197 เป็นบัญชีทวีตเตอร์ที่อยู่ภายใต้การดูแลของกองบังคับการตำรวจจราจร (บก. 02) ซึ่งเป็นหน่วยงานของราชการมีจำนวนผู้ติดตามกว่า 2 หมื่นบัญชีผู้ใช้งาน

3) @fm91trafficpro เป็นบัญชีทวีตเตอร์ของหน่วยงานราชการคือ สถานีวิทยุพิทักษ์สันติราษฎร์ สวพ. FM91 อยู่ภายใต้การดูแลของกองตำรวจสื่อสาร สำนักงานตำรวจแห่งชาติมีจำนวนผู้ติดตามกว่า 2 ล้านบัญชีผู้ใช้งาน

ส่วนของข้อความที่นำมาวิเคราะห์นั้นมีทั้งสิ้น 3,400 ข้อความ โดยมีการนำข้อความตั้งแต่ วันที่ 5 กุมภาพันธ์ 2564 ถึงวันที่ 23 มีนาคม 2564 มาบันทึกในรูปแบบไฟล์ข้อมูล Excel เพื่อให้สามารถนำไปใช้งานได้สะดวกและยังมีการแยกประเภทของข้อความออก 2 ประเภทคือ ข้อปกติ และ ข้อความแสดงเหตุการณ์



ภาพ 3.2 หน้าจอแสดงผลและองค์ประกอบการแสดงข้อมูลข่าวสาร

ลำดับ	วันที่	เวลา	ข้อความทวีต	สถานะเหตุการณ์
2	5 ก.พ. 64	16:52	ผู้ว่าฯ วีระศักดิ์ อากาศดีขึ้นต่อเนื่อง ทุกอย่างเป็นในเชิงบวก ผลเอกซเรย์ปอดล่าสุดขึ้น 90 % #ข่าวจริงประเทศไทย #FM91 #ผู้ว่าสมุทรสาคร	0
3	5 ก.พ. 64	17:00	เตรียมพบทาวนิโฮม 3 โครงการใหม่ บน 3 ทำเลคุณภาพ . ลงทะเบียนรับส่วนลดสูงสุด 30,000 บาท* ▶ http://bit.ly/35Xw3tc . โกลเด้น ทาวน์ สุขสวัสดิ์ - พระราม 3 http://bit.ly/3448LAV โกลเด้น ทาวน์ ๒ รามอินทรา http://bit.ly/39fgm2U โกลเด้น ทาวน์ สุขุมวิท-ลาซาล http://bit.ly/33U4vDY	0
4	5 ก.พ. 64	17:00	ห้ามพลาด! ข่าวฮอต ประเด็นฮิต เสาร์ที่ 6 กุมภาพันธ์นี้ เวลา 19.10 น. "หนุ่ม กะลา" โพนอิน เชิญชวนคนบ้านเดียวกันส่งมอบพลังงานสะอาดให้โรงพยาบาลชุมชนในแคมเปญ "สู้โควิดด้วย ไฟ จาก ไฟ" พบกัน LIVE ผ่าน 3 ช่องทาง สวท.91 เฟซบุ๊ก ทวิตเตอร์ และยูทูป FM91TRAFFICPRO	0
5	5 ก.พ. 64	17:23	ปาท่องโก๋ การบินไทย ยังได้รับการตอบรับเป็นอย่างดี #ปาท่องโก๋การบินไทย #การบินไทย	0

ภาพ 3.3 ตัวอย่างข้อมูลที่จัดเก็บเหตุการณ์ปกติ (การประชาสัมพันธ์ข่าวสาร)

ลำดับ	วันที่	เวลา	ข้อความทวีต	สถานะเหตุการณ์
1	5 ก.พ. 64	16:48	16.47น. รถจอดเสียอีกคัน !! ทางขึ้นสะพานตากสิน ขาออก รถแท๊กซี่จอดเสีย -เคลื่อนย้ายพ้นแล้ว กลางสะพานตากสิน รถบรรทุก 4ล้อใหญ่จอดเสีย เลนซ้าย -อยู่ระหว่างเคลื่อนย้าย #รายงานจราจร #รถจอดเสีย #FM91	1
8	5 ก.พ. 64	17:45	17.44 น. ถนนเทพรัตน บางนาตราด ขาเข้า ช่วงสะพานวัดศรีเยี่ยมไปหน้ารพ.บางนา 1 รถเยอะชะลอตัวช้า #FM91	1
9	5 ก.พ. 64	17:48	17.45 น. บนต่างระดับสิรินธร ขาเข้า ช่วงทางโค้งมุ่งหน้าเมเจอร์บันเกล้า รถชนกัน 4 คัน กีดขวางเลนขวา #FM91 #รายงานจราจร #อุบัติเหตุ	1
11	5 ก.พ. 64	18:02	18.00 น.อุบัติเหตุ ถนนบางนา-ตราด ขาออก ช่วง กม.15+300 ช่องทางด่วน รถกระบะตู้ทึบพลิกคว่ำขวางเลนซ้าย มีเจ้าหน้าที่แล้ว	1
12	5 ก.พ. 64	18:03	18.03 น. ถนนคลองหลวง ขาออก เชิงทางขึ้นสะพานข้ามคลอง 3 พบรถนั่งส่วนบุคคล ชนกับรถนั่งอนุกรมประสงค์ ขวางเลนขวา ประสานศูนย์วิทยุพทุมธานี ทราบแล้ว ท้ายแถวเริ่มยาวแล้ว	1

ภาพ 3.4 ตัวอย่างข้อมูลที่จัดเก็บเหตุการณ์ที่ก่อให้เกิดผลกระทบต่ออาการจราจร (อุบัติเหตุ)

3.2.2 กระบวนการตัดคำจากข้อความ (Word Tokenization)

ในการประมวลภาษาธรรมชาติจากงานวิจัยที่มีการพิสูจน์มาแล้วมีวิธีการที่จำเป็นอย่างมากคือ การแบ่งประโยคหรือข้อความออกเป็นส่วน ๆ กล่าวคือแบ่งประโยคออกเป็นคำ ๆ ซึ่งมีเทคนิคที่เรียกว่าเมื่อได้ข้อความจากทวิตเตอร์เข้ามาเก็บไว้รูปแบบของข้อมูลรูปแบบที่ต้องการในขั้นตอนนี้ นำเสนอวิธีการตัดคำเพื่อแยกออกมาเป็นส่วน ๆ หรือเป็นหน่วยย่อยที่ไม่ซับซ้อนเรียกว่าการทำโทเคน (Token) โดยเป็นขั้นตอนแรกสุดในการประมวลผลคำทางภาษา ซึ่งสำหรับงานวิจัยนี้ได้ใช้ไพทอน (Python) ในการพัฒนาและได้นำเอาไลบรารี PythaiNLP ที่มีฟังก์ชัน Word Tokenize สามารถใช้งานได้ทันที

3.2.3 กระบวนการคัดกรองคำก่อนการประมวลผล (Text Filter)

เนื่องจากในข้อความที่นำมาประมวลผลยังมีอักขระหรืออักขระที่ไม่มีความหมายในทางการประมวลผลตัวอย่างเช่นมีเครื่องหมาย '#' หรือเครื่องหมาย '/' ซึ่งในทางการประมวลผลแบบภาษาธรรมชาตินั้นไม่มีความหมายเลยต้องดำเนินการคัดกรองคำ (Text Filter) ซึ่งในงานวิจัยนี้ได้ใช้วิธีการค้นหาชนิดปกติ (Regular Expression) จากข้อความที่ผ่านการตัดคำมาแล้ว ตามภาพ 3.6 จะสังเกตว่าข้อความเหลือเพียงแค่ข้อความที่เป็นตัวเลข 0-9 และ ก-๙ เท่านั้น

3.2.4 กระบวนการจัดกลุ่มความถี่ข้อความเพื่อกำหนดจำนวนข้อมูลเข้า

หลักการในการวิเคราะห์ภาษาธรรมชาติมีวิธีการอีกอย่างที่ได้รับคามนิยมในช่วงหลังคือการใช้วิธีการหาความถี่ของคำ (Term Frequency: TF) เมื่อได้จำนวนของคำที่มีความนิยมหรืออีกนัยคือคำที่เกิดขึ้นบ่อยซึ่งการนับความถี่ของคำนี้จะทำให้ทราบว่าข้อความที่นำมาประมวลผลนั้นเป็นข้อความประเภทใด โดยการจัดกลุ่มความถี่ของคำนั้นได้มีการจัดความถี่โดยแยกข้อความออกเป็น 2 ประเภทคือ ข้อความการประชาสัมพันธ์ข่าวหรือข้อมูลทั่วไปกับการรายงานหรือการแจ้งอุบัติเหตุ โดยจะทำการนับจำนวนคำที่เกิดขึ้นทั้งหมดในทุกข้อความมารวมกัน ซึ่งการดำเนินการนับความถี่ของคำเป็นกระบวนการตั้งต้นที่สำคัญสำหรับการเรียนรู้เชิงลึกเนื่องจากการเรียนรู้เชิงลึกนั้นต้องนำเข้าข้อมูลที่เป็นชุดตัวเลข (number) ดังนั้นจึงต้องแปลงค่าให้เป็นตัวเลขโดยจะนำหมายเลขจากอันดับของคำที่เกิดสูงสุดเป็นอันดับที่เท่าไรให้จับคู่กับคำนั้น เช่น คำว่า 'แยก' = 1 , 'ถนน' = 2, 'อุบัติเหตุ' = 3, 'รถ' = 4 และ 'ชน' = 5 เป็นต้น เมื่อเปลี่ยนค่าทั้งประโยคออกมาเป็นตัวเลขจะได้ประโยค ดังเช่น "อุบัติเหตุ", "รถ", "ชน", "กัน", "ที่", "4" และ "แยก" จะได้ [3, 4, 5, 0, 0, 0, 1] ซึ่งเป็นชุดตัวเลขที่สามารถนำไปสู่กระบวนการเรียนรู้ขั้นต่อไปได้ และเพื่อให้การเรียนรู้สามารถทำได้ อย่างเป็นระบบจำเป็นต้องทำการหาความถี่ของคำ (TF) ในแต่ละประเภทของเรียนรู้เชิงลึก โดยผลของการหาความถี่ของคำในข้อความแต่ละประเภทมีรายละเอียดดังนี้

เตรียมพบทวนโฮม 3 โครงการใหม่

เตรียม	พบ	ทวน	โฮม	3	โครงการ	ใหม่
--------	----	-----	-----	---	---------	------

```
txt = excel_data_df.tweet[5]
txt_c = word_tokenize(txt)
print(txt, '\n')
print(txt_c)
```

วันศุกร์แห่งชาติ ! เอ็นวันนีสภาพการจราจรจะเป็นอย่างไร ติดตามได้กับ : เจาะ จุด แจม : 5 ก.พ.64

['วัน', 'ศุกร์', 'แห่งชาติ', '!', 'เอ็น', 'วันนี', 'สภาพ', 'การจราจร', 'จะ', 'เป็น', 'อย่างไร', '!', 'ติดตาม', 'ได้', 'กับ', '!', 'เจาะ', 'จุด', 'แจม', '!', '5', 'ก.พ.', '64']

ภาพ 3.5 การตัดคำจากข้อความ

ข้อความจาก Twitter
มีอักขระที่ไม่มีความหมาย เช่น
“#”, “/”, “!”, “.”

สต.พบผู้ป่วยติดเชื้อใน กทม.ติดเชื้อโควิด 19 จากผู้ดูแลชาวเมียนมา ย่า !! ผู้ดูแลผู้ป่วยต้องสวมหน้ากากตลอดเวลา #FM91 #ผู้ป่วยติดเชื้อ #ติดเชื้อโควิด #ผู้ดูแลชาวเมียนมา #โควิด19

['สธ', 'พบ', 'ผู้ป่วย', 'ติด', 'เชื้อ', 'ใน', '!', 'กม.', 'ติดเชื้อ', 'โควิด', '19', '!', 'จาก', 'ผู้ดูแล', 'ชาว', 'เมียนมา', '!', 'ย่า', '!', '!', 'ผู้ดูแล', 'ผู้ป่วย', 'ต้อง', 'สวมหน้ากาก', 'ตลอดเวลา', '!', '#', 'FM', '91', '!', '#', 'ผู้ป่วย', 'ติด', 'เชื้อ', '!', '#', 'ติดเชื้อ', 'โควิด', '!', '#', 'ผู้ดูแล', 'ชาว', 'เมียนมา', '!', '#', 'โควิด', '19']

10.00 กาญจนภิเษก ช่วง ต่าง ระดับ บาง ร.ร. เทพ ศรินทร์ จุด กลับรถ ได้ สะพาน ช้าม คลอง ปิคอฟ ติด ความสูง ได้ สะพาน การจราจรติดขัด รถติด 10.00 กาญจนภิเษก ช่วง ต่าง ระดับ บาง ร.ร. เทพ ศรินทร์ จุด กลับรถ ได้ สะพาน ช้าม คลอง ปิคอฟ ติด ความสูง ได้ สะพาน การจราจรติดขัด รถติด

ภาพ 3.6 การคัดกรองคำก่อนการประมวลผล

ตาราง 3.1 การลำดับความถี่คำในการแสดงสถานการณ์

อันดับ	การประชาสัมพันธ์ข่าวหรือข้อมูลทั่วไป		การรายงานหรือการแจ้งอุบัติเหตุการณ์	
	คำ	จำนวน	คำ	จำนวน
1	และ	465	แยก	986
2	โควิด	436	รถติด	831
3	ราย	328	อุบัติเหตุ	830
4	100	255	การจราจร	646
5	การ	244	ช่วง	614
6	เขต	234	จราจร	597
7	จาก	218	รายงาน	531
8	เวลา	203	ขวาง	529
9	บาง	202	ช่อง	526
10	100.	197	เคลื่อนตัว	493
11	เพลิง	194	ช้า	479
12	2564	179	กัน	477
13	วันที่	179	สะพาน	476
14	ได้	178	ขาเข้า	379
15	เพลิงไหม้	177	บาง	361
16	ซอย	177	ขาออก	355
17	ข่าว	168	คัน	321
18	แล้ว	161	จาก	289
19	ติดเชื้อ	158	ถึง	288
20	ของ	156	กับ	285

จากกลุ่มตัวอย่างของข้อความที่นำมาหาความถี่ของคำโดยแยกตามแต่ละประเภทของข้อความ เพื่อหาจำนวนของคำซ้ำที่ปรากฏในเอกสารตามตาราง 3.1 แสดงจำนวนคำที่เป็นข้อความอุบัติการณ์ จะมีความถี่ของคำที่แตกต่างกัน เมื่อนำมาแสดงในรูปแบบของ Cloud Text ตามภาพ 3.7 แสดงให้เห็นความแตกต่างได้อย่างชัดเจน โดยมีคำที่เกี่ยวกับข้อความประชาสัมพันธ์ข่าวทั่วไป 3 อันดับแรก คือคำว่า “และ”, “โควิด” และ “ราย” ซึ่งมีจำนวนคำคือ 465, 436 และ 328 ตามลำดับ และคำที่เกี่ยวกับการแจ้งอุบัติการณ์ 3 อันดับแรกคือคำว่า “แยก”, “รถติด” และ “อุบัติเหตุ” ซึ่งมีจำนวนคำคือ 986, 831 และ 830 ตามลำดับ

สำหรับในลำดับถัดไปเป็นการนำเสนอกลุ่มความถี่ของคำที่ในส่วนที่มีความหมายที่บ่งบอกถึงรายละเอียดอุบัติการณ์โดยมีการจำแนกออกเป็น 6 เหตุการณ์ที่มีผลต่อสภาพจราจรดังตาราง 3.2



ภาพ 3.7 กลุ่มข้อความตามความถี่ระหว่างข้อความทั่วไปและข้อความอุบัติการณ์
(ก) การประชาสัมพันธ์ข่าวหรือข้อมูลทั่วไป (ข) การรายงานหรือการแจ้งอุบัติการณ์

ตาราง 3.2 การลำดับความถี่คำในการจำแนกสถานการณ์

อันดับ	ลักษณะที่ 1		ลักษณะที่ 2		ลักษณะที่ 3		ลักษณะที่ 4		ลักษณะที่ 5		ลักษณะที่ 6	
	คำ	จำนวน	คำ	จำนวน	คำ	จำนวน	คำ	จำนวน	คำ	จำนวน	คำ	จำนวน
1	แยก	335	สะพาน	2	อุบัติเหตุ	830	ซึ่ง	4	จราจร	100	ชุมชน	21
2	จราจร	229	ข้าม	2	แยก	589	กม.	4	ปิด	77	ผู้	17
3	รายงาน	216	น้ำท่า	2	รถติด	580	น้ำท่วม	3	แยก	64	การจราจร	14
4	รถติด	215	จีน	2	การจราจร	545	ช่วง	3	รายงาน	53	เส้นทาง	13
5	ขาเข้า	165	พระราม	1	ช่วง	540	พาย	2	ขาเข้า	51	แยก	11
6	สะพาน	147	ขาออก	1	ขวาง	524	ส่งผล	2	และ	50	บข.	8
7	ขาออก	143	ช่วง	1	ช่อง	478	บริเวณ	2	สะพาน	49	กรณี	8
8	ท้าย	140	แยก	1	กัน	466	การจราจร	2	ขาออก	49	การ	7
9	อยู่	130	บ้านแพ้ว	1	ช้า	431	สัญญาณ	2	รถติด	48	และ	7
10	สะสม	125	เชิง	1	เคลื่อนตัว	430	ได้	2	การจราจร	47	รอบ	7
11	มาก	118	ทางลง	1	คัน	316	จาก	2	แจ้ง	39	ส่ง	7
12	ติด	93	ก่อสร้าง	1	สะพาน	291	สอด	2	งาน	38	อาจ	7
13	ท้ายแถว	91	ช่อง	1	กับ	284	145	2	บริเวณ	36	ขาเข้า	6
14	ข้าม	79	ทางซ้าย	1	จราจร	274	น้ำ	2	วันที่	32	กลุ่ม	6
15	บาง	78	ของ	1	รายงาน	267	ป้ายบอกทาง	2	เวลา	32	สะพาน	6
16	ลาดพร้าว	73	ทาง	1	บาง	251	เลน	2	ช่วง	31	ผลกระทบ	6

ตาราง 3.2 การลำดับความถี่คำในการจำแนกสถานการณ์ (ต่อ)

อันดับ	ลักษณะที่ 1		ลักษณะที่ 2		ลักษณะที่ 3		ลักษณะที่ 4		ลักษณะที่ 5		ลักษณะที่ 6	
	คำ	จำนวน	คำ	จำนวน	คำ	จำนวน	คำ	จำนวน	คำ	จำนวน	คำ	จำนวน
17	พระราม	68	หลัก	1	ก่อน	231	ซ้าย	2	ถึง	30	มีนา	6
18	จาก	66	การจรรจร	1	กลาง	225	ขวา	2	ช่องทาง	30	มีอบ	5
19	ได้	55	เคลื่อนตัว	1	จอด	220	ช่องทาง	2	บาง	30	ปิด	5
20	การจรรจร	54	ซ้าย	1	ทางซ้าย	218	มาก	2	22.00	30	100.	5



ภาพ 3.8 กลุ่มข้อความตามความถี่ระหว่างข้อความตามลักษณะรูปแบบของอุบัติเหตุการณ์
 (ก) รายงานสภาพจรรจร (ข) การดำเนินงานของตำรวจ
 (ค) รายงานอุบัติเหตุ (ง) การเกิดภัยพิบัติ
 (จ) รายงานเหตุการณ์ปิดถนน (ฉ) อื่น ๆ

การหาความถี่ของคำที่เกี่ยวข้องกับการเกิดอุบัติเหตุ ซึ่งได้จำแนกออกเป็นรูปแบบต่าง ๆ จำนวน 6 รูปแบบ ประกอบไปด้วย รายงานสภาพจราจร ซึ่งมีความถี่คำสูงสุด 3 อันดับแรกคือ “แยก”, “จราจร” และ “รายงาน” ซึ่งมีจำนวนความถี่ของคำคือ 335, 229 และ 216 ตามลำดับ การดำเนินงานของตำรวจนั้นจากข้อมูลที่ได้รวบรวมมานั้นมีเพียงน้อยนิดทำให้ไม่มีความสำคัญ ในทางการวิเคราะห์ อันดับต่อมาเป็นลักษณะของคำที่เป็นการรายงานอุบัติเหตุ ซึ่งมีค่าที่มีความถี่สูงสุด 3 อันดับ ดังนี้ “อุบัติเหตุ”, “แยก” และ “รถติด” ซึ่งมีจำนวนคำคือ 830, 589 และ 580 ตามลำดับ รายละเอียดตามตาราง 3.2 และภาพ 3.8

ในการจำแนกลักษณะของข้อความรายงานสภาพจราจรหรืออุบัติเหตุ นั้นมีข้อมูลอีกประเภทที่ จำเป็นคือการจำแนกความรุนแรงของอุบัติเหตุ นั้น ๆ โดยในส่วนของ การจำแนกความรุนแรงนั้นมีการรวบรวมและจัดทำอันดับของคำที่เกี่ยวข้องซึ่งมีการแบ่งออกเป็น 3 ระดับความรุนแรง ดังนี้

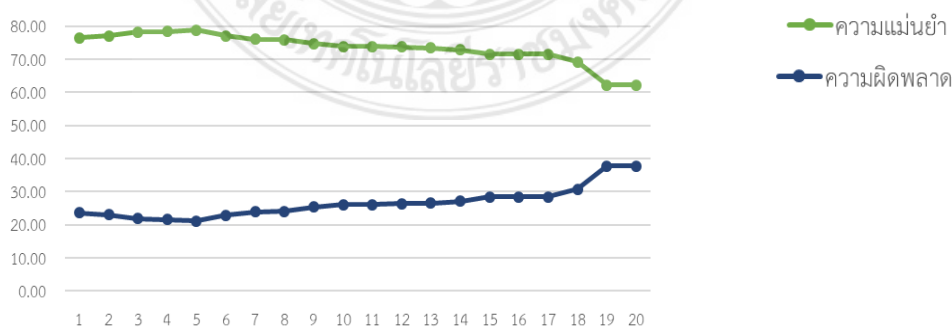
ตาราง 3.3 การลำดับความถี่คำที่แสดงความรุนแรงที่ส่งผลกระทบต่อจราจร

อันดับ	ไม่มีผลกระทบต่อจราจร		กระทบต่อจราจรบางส่วน		ปิดกั้นทุกช่องจราจร	
	คำ	จำนวน	คำ	จำนวน	คำ	จำนวน
1	อุบัติเหตุ	5	อุบัติเหตุ	731	ปิด	51
2	แล้ว	5	รถติด	593	จราจร	47
3	รพ.	4	แยก	582	สะพาน	27
4	ส่ง	3	การจราจร	563	แยก	22
5	เคลื่อนย้าย	3	ช่วง	525	การจราจร	22
6	พัน	3	ขวาง	519	และ	21
7	กีดขวาง	3	ช่อง	499	แจ้ง	18
8	ช่วง	3	กัน	441	รามอินทรา	18
9	ปิดอู่	3	ซ้ำ	431	ขาเข้า	17
10	นครราชสีมา	3	เคลื่อนตัว	429	ช่องทาง	17
11	คืบหน้า	2	คัน	291	รายงาน	16
12	ป่อเต็กตึ๊ง	2	สะพาน	283	รพม.	16
13	ตรวจสอบ	2	จราจร	262	สาย	16
14	รถจักรยานยนต์	2	ก๊ับ	253	22.00	16
15	อยู่	2	บาง	226	ขาออก	15
16	เกษตร	2	รายงาน	243	วันที่	15
17	เวลา	2	ทางซ้าย	226	ข้าม	14
18	ไทเกอร์	2	ถึง	213	ถึง	13
19	ช่วยเหลือ	2	กลาง	211	เวลา	13
20	ก่อน	2	ทางขวา	211	เพื่อ	12

ตามตาราง 3.3 และภาพ 3.9 ถึง 3.11 เป็นการนำเสนอค่าที่แสดงถึงระดับความรุนแรงของการเกิดอุบัติเหตุการซึ่งได้มีการแบ่งไว้ 3 ระดับคือ ไม่มีผลกระทบต่อสภาพจราจรในระดับความรุนแรงนี้จะเห็นว่า ค่าที่มีความสำคัญแท้จริงแล้วคือคำว่า “แล้ว” , “พ้น” และคำว่า “เคลื่อนย้าย” ซึ่งหากดูถึงจำนวนค่าที่พบแล้วอยู่ในลำดับที่ 2, 5 และ 6 ตามลำดับ โดยที่ลำดับของความถี่ค่าในระดับที่ไม่มีผลกระทบต่อสภาพจราจร 3 อันดับแรกประกอบไปด้วยคำว่า “อุบัติเหตุ”, “แล้ว” และ “รพ.” ซึ่งมีจำนวนค่าที่ 5, 5 และ 4 ตามลำดับ และในระดับความรุนแรงที่มีผลกระทบบางส่วนจะเห็นว่า มีคำว่า “อุบัติเหตุ” และคำว่า “รถติดเป็นหลัก” ในขณะที่ 3 อันดับแรกของความถี่ประกอบไปด้วยคำว่า “อุบัติเหตุ”, “รถติด” และ “แยก” ซึ่งมีจำนวนค่าอยู่ที่ 731, 593 และ 582 ตามลำดับ และในระดับความรุนแรงปิดกั้นทุกช่องจราจรมีความถี่ประกอบไปด้วย “ปิด”, “จราจร” และ “สะพาน” ซึ่งมีจำนวนค่าอยู่ที่ 51, 47 และ 27

3.2.5 การทดสอบการแยกแยะสถานการณ์แบบพื้นฐานจากความถี่ค่า

จากการหาความถี่ของค่าจะเห็นว่า มีค่าเป็นจำนวนมากและมีความหลากหลายของการใช้คำในการรายงานสภาพจราจรหรือการแจ้งเหตุการณ์ต่าง ๆ เพื่อให้ทราบถึงจำนวนค่าที่ต้องใช้ในการจำแนกประเภทของข้อความสำหรับกระบวนการเรียนรู้เชิงลึก โดยการหาจำนวนค่าที่เหมาะสมสำหรับกระบวนการเรียนรู้เชิงลึกนั้นได้มีการแบ่งการทดสอบออกเป็น 3 สถานการณ์คือ โดยสถานการณ์แรกเป็นการทดสอบค่าที่มีความถี่สูงสุด 1-20 ค่าโดยทำการค้นหาค่าในข้อความเมื่อมีค่าเพียง 1 ค่าใน 20 ค่าจะนับว่าเป็นอุบัติเหตุ สถานการณ์ที่สองเลือกค่าที่มีความถี่สูงสุด 1-20 ค่า โดยทำการทดสอบเป็นรอบรอบแต่ละรอบจะเพิ่มค่าขึ้นรอบละ 1 ค่าเพื่อค้นหาค่าในข้อความโดยจะต้องเจอค่าทุกคนในแต่ละรอบถึงจะสรุปว่าข้อความนั้นเป็นข้อความอุบัติเหตุ และสถานการณ์ที่สามเลือกค่าที่มีความถี่สูงสุด 1-20 ค่า ทำการค้นหาค่าในข้อความแต่ละรอบโดยในแต่ละรอบจะเพิ่มค่าที่ต้องการขึ้นรอบละ 1 ค่าและทำการค้นหาค่าในข้อความหากเจอค่าในข้อความ ตั้งแต่ 10 % จนถึง 100 % ของแต่ละรอบจึงจะนับว่าข้อความนั้นเป็นข้อความอุบัติเหตุ ซึ่งผลลัพธ์ของการทดสอบเพื่อหาจำนวนที่เหมาะสมที่สุดสำหรับการเรียนรู้เชิงลึกได้ โดยผลลัพธ์ที่ดีที่สุดของการทดสอบคือสถานการณ์แรกคือการค้นหาค่าจากความถี่ค่าที่ 5 ค่าแรกโดยค้นหาค่าในข้อความเจออย่างน้อยหนึ่งค่า รายละเอียดตามตาราง 3.4 และภาพ 3.12



ภาพ 3.12 จำนวนค่าที่ให้ผลลัพธ์การแยกประเภทข้อความที่ดีที่สุด

ตาราง 3.4 การลำดับจำนวนคำที่มีผลต่อการแยกประเภทของข้อความ

จำนวนคำที่ใช้ตรวจสอบ	แจ้งเตือนถูกต้อง (ข้อความ)	แจ้งเตือนผิด (ข้อความ)	แจ้งเตือนทั้งหมด (ข้อความ)	ความถูกต้อง (%)	ความผิดพลาด (%)
1	314	97	411	76.40	23.60
2	368	110	478	76.99	23.01
3	394	110	504	78.17	21.83
4	414	114	528	78.41	21.59
5	432	116	548	78.83	21.17
6	446	133	579	77.03	22.97
7	447	141	588	76.02	23.98
8	456	145	601	75.87	24.13
9	462	157	619	74.64	25.36
10	467	165	632	73.89	26.11
11	469	166	635	73.86	26.14
12	470	168	638	73.67	26.33
13	470	170	640	73.44	26.56
14	473	176	649	72.88	27.12
15	474	189	663	71.49	28.51
16	480	191	671	71.54	28.46
17	480	191	671	71.54	28.46
18	483	215	698	69.20	30.80
19	489	297	786	62.21	37.79
20	490	297	787	62.26	37.74

3.2.6 กระบวนการสร้างรูปแบบการเรียนรู้เชิงลึก (Deep Learning Model)

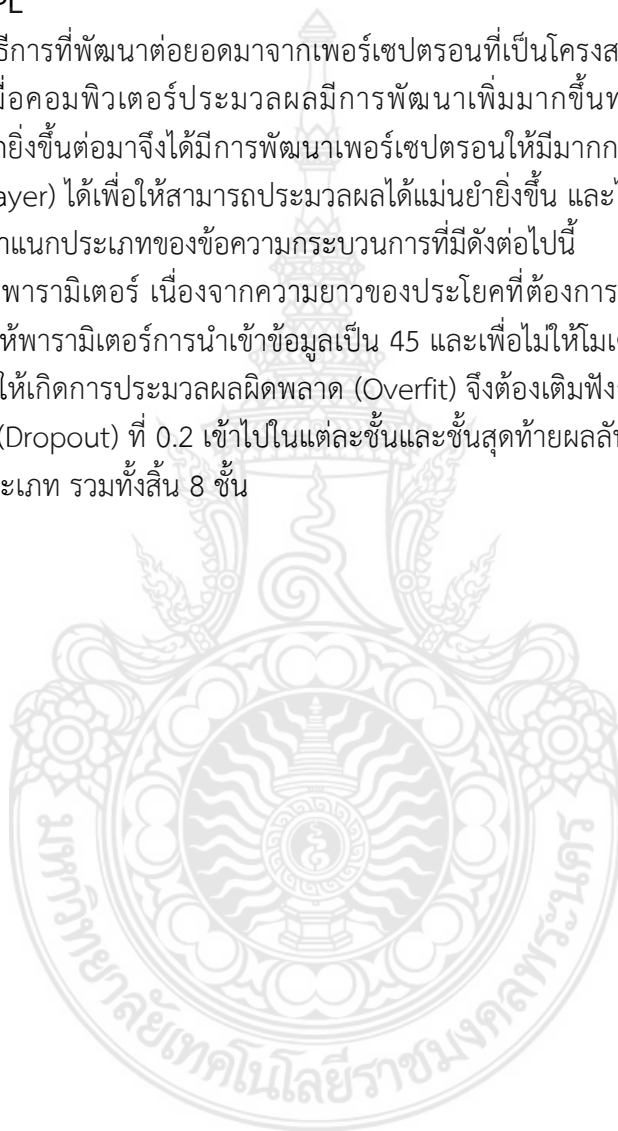
สำหรับงานวิจัยงานวิจัยนี้จะนำเสนอเกี่ยวกับการเพื่อการจำแนกประเภทของข้อความจากสื่อสังคมออนไลน์ทวิตเตอร์ ซึ่งจากงานวิจัยต่าง ๆ นั้นได้นำเอากระบวนการทางด้านการเรียนรู้เชิงลึกเข้ามาทำการประมวลผลข้อความประกอบไปด้วย วิธี MPL ซึ่งเป็นวิธีการเรียนรู้เชิงลึกสำหรับการประมวลผลข้อความเนื่องจากข้อความนั้นเป็นข้อมูลที่อยู่ในรูปแบบการกระจายแบบไม่เป็นเชิงเส้นจึงเป็นวิธีการที่เหมาะสมในการนำมาทดสอบประมวลผลเป็นอย่างดี และในงานวิจัยในครั้งนี้ได้นำเอาวิธีการประมวลผลวิธี CNN เพื่อเป็นการทดสอบผลลัพธ์ที่เกิดขึ้นเนื่องจากเป็นวิธีที่มีความซับซ้อนขึ้นมาอีกระดับ และอีกวิธีนำมาทดสอบคือวิธี LSTM ซึ่งเป็นวิธีการที่เหมาะสมสำหรับการประมวลผลข้อความอีกวิธีที่ได้รับความนิยมเป็นอย่างมากในขณะนี้เนื่องจาก LSTM เป็นการประมวลผลคำที่มีการนำเอาคำที่เกิดก่อนหน้าคำที่จะพิจารณาไปประมวลผลร่วมทำให้การพิจารณาคำปัจจุบันเกิดความแม่นยำขึ้น เช่น หากต้องการทราบความหมายของคำว่า “ชน” ที่หมายถึง “รถชนกัน” ที่ไม่ได้แปลว่า “ชนแก้ว” LSTM จะนำคำก่อนหน้าคือคำว่า “รถ” มาพิจารณาจนทำให้ทราบความหมายของคำว่า “ชน” นั้นคือ “รถชน” ซึ่งเป็นคำที่เกี่ยวกับถนนหรือการจราจรนั่นเอง

เนื่องจากในการประมวลผลการเรียนรู้เชิงลึก นั้นจำเป็นต้องมีการสร้างชุดโปรแกรมเพื่อทำการทดสอบที่เรียกว่าโมเดล กล่าวคือ โมเดลจะเปรียบเสมือนฟังก์ชันที่ทำงานตามคำสั่งเมื่อถูกป้อนพารามิเตอร์เข้าไปและได้ผลลัพธ์ออกมาจากสูตรหรือสมการที่อยู่ในโมเดลนั่นเอง ซึ่งรายละเอียดในการสร้างและทดสอบโมเดลการจำแนกประเภทในแต่ละวิธีการมีรายละเอียด ดังนี้

3.2.6.1 MPL

MPL เป็นวิธีการที่พัฒนาต่อยอดมาจากเพอร์เซปตรอนที่เป็นโครงสร้างที่ไม่ซับซ้อนเนื่องจากมีชั้นเดียวต่อมาเมื่อคอมพิวเตอร์ประมวลผลมีการพัฒนาเพิ่มมากขึ้นทำให้ความสามารถในการประมวลผลมีมากยิ่งขึ้นต่อมาจึงได้มีการพัฒนาเพอร์เซปตรอนให้มีมากกว่า 1 ชั้น หรือเป็นหลาย ๆ ชั้น (Multiple Layer) ได้เพื่อให้สามารถประมวลผลได้แม่นยำยิ่งขึ้น และได้นำเอา MLP มาใช้ในการประมวลผลเพื่อจำแนกประเภทของข้อความกระบวนการที่มีดังต่อไปนี้

การกำหนดพารามิเตอร์ เนื่องจากความยาวของประโยคที่ต้องการทดสอบมีความยาว 45 คำ ดังนั้นจึงกำหนดให้พารามิเตอร์การนำเข้าข้อมูลเป็น 45 และเพื่อไม่ให้โมเดลเกิดการสร้างสมาชิกเกินความจำเป็นและให้เกิดการประมวลผลผิดพลาด (Overfit) จึงต้องเติมฟังก์ชันสุ่มลดจำนวนสมาชิกที่อาจไม่ถูกใช้งาน (Dropout) ที่ 0.2 เข้าไปในแต่ละชั้นและชั้นสุดท้ายผลลัพธ์เป็น 2 เนื่องจากเป็นการจำแนกเป็น 2 ประเภท รวมทั้งสิ้น 8 ชั้น



```

model = tf.keras.Sequential([
    tf.keras.layers.Dense(450, input_shape=(45,), activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(250, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(100, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(50, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(20, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(10, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(5, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(2, activation='softmax'),
])

```

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 450)	20700
dropout (Dropout)	(None, 450)	0
dense_1 (Dense)	(None, 250)	112750
dropout_1 (Dropout)	(None, 250)	0
dense_2 (Dense)	(None, 100)	25100
dropout_2 (Dropout)	(None, 100)	0
dense_3 (Dense)	(None, 50)	5050
dropout_3 (Dropout)	(None, 50)	0
dense_4 (Dense)	(None, 20)	1020
dropout_4 (Dropout)	(None, 20)	0
dense_5 (Dense)	(None, 10)	210
dropout_5 (Dropout)	(None, 10)	0
dense_6 (Dense)	(None, 5)	55
dropout_6 (Dropout)	(None, 5)	0
dense_7 (Dense)	(None, 2)	12

Total params: 164,897
 Trainable params: 164,897
 Non-trainable params: 0

ภาพ 3.13 การพัฒนารูปแบบการเรียนรู้เชิงลึกแบบ MLP

3.2.6.2 CNN

CNN เป็นรูปแบบของวิธีการโครงข่ายประสาทเทียมประเภทหนึ่งที่มีการออกแบบกระบวนการ (Algorithms) เพื่อให้สามารถรองรับชุดตัวเลข เดิม CNN ออกแบบมาเพื่อทำงานกับการประมวลผลรูปภาพซึ่งโดยแท้จริงแล้วคือการแปลงค่าสี RGB ออกมาเป็นชุดตัวเลข 0-255 แล้วจึงนำชุดตัวเลขเข้าโมเดล CNN เพื่อประมวลผล ดังนั้นหาก CNN รองรับการประมวลผลรูปภาพที่แปลงตัวเลขแล้ว โมเดล CNN จะสามารถรองรับการประมวลผลข้อความที่แปลงเป็นตัวเลขแล้วเช่นกันการดำเนินการข้างต้น คือการแปลงข้อความให้เป็นชุดตัวเลขซึ่งจะมีลักษณะคล้ายกับวิธี MLP แต่ CNN จะมีการเพิ่มชั้นของคอนโวลูชันและพูลลิงเข้ามา

```
# Created Embedding (Input) Layer (max_words) --> Convolutional Layer
model.add(Embedding(vocab_size, embedding_dims, input_length=maxlen))
model.add(Dropout(0.2)) # masks various input values
# Create the convolutional layer
model.add(Conv1D(filters, kernel_size, padding='valid', activation='relu'))
# Create the pooling layer
model.add(GlobalMaxPooling1D())
# Create the fully connected layer
model.add(Dense(hidden_dims))
model.add(Dropout(0.2))
model.add(Activation('relu'))
# Create the output layer (num_classes)
model.add(Dense(2))
model.add(Activation('softmax'))
```

Model: "sequential_1"

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 45, 2)	9306
dropout (Dropout)	(None, 45, 2)	0
conv1d (Conv1D)	(None, 41, 250)	2750
global_max_pooling1d (Global	(None, 250)	0
dense (Dense)	(None, 150)	37650
dropout_1 (Dropout)	(None, 150)	0
activation (Activation)	(None, 150)	0
dense_1 (Dense)	(None, 2)	302
activation_1 (Activation)	(None, 2)	0
Total params: 50,008		
Trainable params: 50,008		
Non-trainable params: 0		

ภาพ 3.14 การพัฒนารูปแบบการเรียนรู้เชิงลึกแบบ CNN

3.2.6.3 LSTM

LSTM เป็นวิธีการที่ได้รับความนิยมสำหรับการประมวลผลข้อความหรือการประมวลผลภาษาธรรมชาติเนื่องจากเป็นการใช้เทคนิคที่ถูกพัฒนาต่อยอดมาจาก RNN ซึ่งมีความโดดเด่นในเรื่องการประมวลผลข้อมูลที่เป็นลำดับเวลา (Time Series) กล่าวคือ LSTM จะสามารถจดจำคำที่ผ่านมาให้ในลำดับขั้นก่อนหน้าได้ว่าในอนาคตจะต้องเป็นไปในทางใดได้นานกว่า RNN

LSTM ได้ถูกพัฒนาปรับปรุงประสิทธิภาพให้ดีขึ้น โดยมีชื่อเรียกว่า Bidirectional LSTM หรือ Bi-LSTM จะสามารถจดจำได้จาก Memory Cell ทั้ง Forward และ Backward ได้ ในขณะที่ LSTM จะสามารถจดจำได้แค่ Backward ได้อย่างเดียว และสำหรับงานวิจัยนี้ได้นำเอา Bi-LSTM มาประมวลผลข้อความเพื่อจำแนกประเภทของอุบัติเหตุโดยในขั้นแรกเป็นการสร้างชั้นประมวลผลภายใน เพื่อเป็นผลลัพธ์ ไปสู่ Bi-LSTM ซึ่งจะมีพารามิเตอร์สุ่มลดจำนวนสมาชิกที่อาจไม่ถูกใช้งานเข้ามาช่วยลดการสร้างสมาชิกเกินความจำเป็นและให้เกิดการประมวลผลผิดพลาดของการเทรน สำหรับขั้นถัดมาเป็นขั้นที่สรุปการแบ่งประเภทเพื่อให้ออกมาเป็นคำตอบสุดท้าย

```
model = tf.keras.Sequential([
    tf.keras.layers.Embedding(vocab_size, embedding_dim),
    tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(
        embedding_dim, dropout=0.2,
        recurrent_dropout=0.2)),
    tf.keras.layers.Dense(embedding_dim, input_shape=(2,),
        activation='relu'),
    tf.keras.layers.Dense(2, activation='softmax')
])

model.compile(loss='categorical_crossentropy', optimizer='adam',
    metrics=['accuracy'])
```

Model: "sequential_2"

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, None, 64)	297792
bidirectional_2 (Bidirection	(None, 128)	66048
dense_4 (Dense)	(None, 64)	8256
dense_5 (Dense)	(None, 2)	130
Total params: 372,226		
Trainable params: 372,226		
Non-trainable params: 0		

ภาพ 3.15 การพัฒนารูปแบบการเรียนรู้เชิงลึกแบบ LSTM

3.2.7 กระบวนการตรวจสอบและจำแนกอุบัติการณ์ (Incident Detection and Classification)

เมื่อดำเนินการสร้างโมเดลการจำแนกอุบัติการณ์เป็นที่เรียบร้อยแล้วจากนั้นเป็นการนำข้อมูลเข้าสู่กระบวนการเทรนโดยนำชุดข้อมูล (Dataset) ที่เตรียมไว้ส่งเข้าสู่โมเดลโดยโมเดลจะทำการเทรนตัวเองฟังก์ชัน fit() พารามิเตอร์ดังต่อไปนี้

- x คือข้อความที่ต้องการเทรน
- y คือคำตอบของข้อความ (Label)
- batch size คือ พารามิเตอร์สำหรับระบุจำนวนกลุ่มข้อมูลที่ใช้ในการเทรนแต่ละรอบ
- epochs คือจำนวนรอบที่ใช้ในการเทรนโมเดล
- validation data คือชุดข้อมูลเพื่อให้โมเดลมีการปรับจูนค่าหรือชุดข้อมูลเพื่อการตรวจสอบคำตอบของการทำนาย

และเมื่อโมเดลทำการเทรนจนครบตามจำนวนรอบที่ตั้งค่าไว้กระบวนการต่อมาคือการนำโมเดลที่ได้มาทดสอบกับชุดข้อมูล (Dataset) ที่เตรียมไว้สำหรับการทดสอบซึ่งเป็นชุดข้อมูลที่ไม่เกี่ยวข้องกันจำนวน 1,010 ข้อความ โดยโมเดลที่สามารถจำแนกประเภทของอุบัติการณ์ได้มากที่สุดคือโมเดล Bi-LSTM สามารถจำแนกได้ถูกต้องที่ 957 ข้อความหรือ 94.75 %

3.3 กระบวนการพัฒนา

ในส่วนของการพัฒนาการจำแนกประเภทของอุบัติการณ์ได้นำเอาเครื่องมือต่าง ๆ เข้ามาช่วยในการพัฒนาไม่ว่าจะเป็นระบบฐานข้อมูล Postgresql ซึ่งเป็นโปรแกรมสำหรับจัดการฐานข้อมูลสำหรับการเก็บข้อมูลและจัดทำ Label ก่อนนำไปใช้งานต่อ และเพื่อความสะดวกในการทำ Label จึงได้พัฒนาระบบจัดการข้อมูล Django Framework ซึ่งเป็นระบบจัดการฐานข้อมูลที่พัฒนาอยู่บนพื้นฐานภาษา Python โดยข้อมูลที่น่ามาวิเคราะห์นั้นเป็นข้อมูลที่ได้จากสื่อสังคมออนไลน์ ทวิตเตอร์ที่มี API เพื่อการเชื่อมต่อให้สามารถเรียกข้อมูลแบบอัตโนมัติได้และได้นำเอาเครื่องมือสำหรับจัดการและประมวลผลภาษาธรรมชาติ (NLP) ที่ทำงานบนภาษา Python คือ PyThaiNLP และเครื่องมือหลักในการพัฒนาการโปรแกรมภาษา Python ที่ทำงานอยู่บนเครื่องมือ Anaconda Jupyter โดยมีไลบรารี (Library) และสำหรับการพัฒนางานทางด้านการเรียนรู้เชิงลึก ได้นำเอาไลบรารี Keras ที่ทำงานอยู่บนแพลตฟอร์ม Tensorflow และใช้งานฟังก์ชันงานจากไลบรารีต่าง ๆ บนแพลตฟอร์ม Scikit learn ในการคำนวณ ค่าการประมวลผลในทางคณิตศาสตร์อีกด้วยโดยรายละเอียดแต่ละระบบมีดังต่อไปนี้

3.3.1 Twitter API

ทวิตเตอร์มีการเปิดให้นักพัฒนาสามารถเข้าถึงข้อมูลการทวิตได้จากเครื่องมือที่เรียกว่า Twitter API ซึ่งได้เปิดตัวอย่างเป็นทางการและเปิดให้นักพัฒนาใช้งานมาตั้งแต่ปี 2006 โดยเวอร์ชัน (version) แรกที่เปิดตัวใช้งานคือ Twitter API v1 ในปี 2012 และในปี 2020 ทวิตเตอร์ได้เปิดตัว Twitter API เวอร์ชันล่าสุดคือ Twitter API v2 ซึ่งเป็นเวอร์ชันปัจจุบัน สำหรับเวอร์ชันปัจจุบันมีการเพิ่มคุณสมบัติต่าง ๆ เข้ามาเพื่อจัดการการเข้าถึงของนักพัฒนา โดยมีการแบ่งเป็น 3 ระดับดังนี้

- ระดับมาตรฐาน (Standard) สำหรับระดับมาตรฐานเป็นระดับแรกสุดที่จะให้นักพัฒนาเข้าใช้งาน API ได้เหมาะสำหรับการเริ่มต้นการพัฒนาประกอบไปด้วยกรณีศึกษาที่ดี ซึ่งเหมาะสำหรับการเรียนรู้หรือสอน
- ระดับงานวิจัยทางวิชาการ (Academic Research) สำหรับระดับงานวิจัยทางวิชาการ นักวิจัยสามารถใช้ Twitter API เพื่อทำความเข้าใจข้อความหรือบทสนทนาที่เปิดเผยสู่สาธารณะและในอนาคตจะมีการคัดเลือกนักวิจัยที่มีแนวทางในการสร้างสรรค์หรือปรับปรุงระบบต่าง ๆ และได้เตรียมเครื่องมือหรือคำแนะนำเพื่อให้ง่ายต่อการดำเนินงานในระดับงานวิจัยทางวิชาการ
- ระดับธุรกิจ (Business) นักพัฒนาที่กำลังสร้างสรรค์ธุรกิจบน Twitter API ได้มีการเพิ่มพันธมิตรทางธุรกิจและข้อมูลลูกค้าเข้ามาในข้อมูล API นี้

ในส่วนของหน้ารวมนักพัฒนา (Developer portal) มีการปรับปรุงรูปแบบใหม่เพื่อเป็นเครื่องมือให้นักพัฒนาสามารถสร้างโครงการ (Project) และสามารถเข้าถึงส่วนสนับสนุนได้จากส่วนการจัดการนี้อีกทาง รูปแบบของการเรียกข้อมูลจาก Twitter API สามารถทำได้โดยการระบุรายละเอียดของ Project ที่ถูกสร้างขึ้นภายใน Portal ประกอบไปด้วย

- API Key คือ คีย์เพื่อการอนุญาตให้เข้าถึงแอปพลิเคชันที่สร้างไว้ใน Portal
- API Key Secret คือ คีย์เพื่อบ่งบอกรหัสผ่านและการอนุญาตการเข้าถึงแอปพลิเคชันที่สร้างไว้ใน Portal
- Access Token คือ โทเคนเพื่อแสดงถึงบัญชีของเจ้าของแอปพลิเคชันที่สร้างไว้ใน Portal และช่วยในการเข้าถึงบัญชีทวิตเตอร์ได้
- Access Token Secret คือ รหัสลับของโทเคนเพื่อแสดงถึงบัญชีของเจ้าของแอปพลิเคชันที่สร้างไว้ใน Portal และช่วยในการเข้าถึงบัญชีทวิตเตอร์
- Bearer Token คือโทเคนที่ช่วยให้แอปพลิเคชันสามารถตรวจสอบคำขอร้องเพื่อการตรวจสอบสิทธิ์ความปลอดภัยในระดับ OAuth2.0

เมื่อดำเนินการเกี่ยวกับ Key ต่าง ๆ ถูกต้องแล้วจะได้รับข้อมูลการตอบกลับมาจาก API โดยนักพัฒนาระดับมาตรฐาน (Standard) จะสามารถเรียกข้อความได้ 300 ครั้งต่อ 15 นาที ต่อหนึ่งแอปพลิเคชันที่สร้างไว้และสามารถเรียกข้อความทวิตได้ 500,000 ครั้งต่อเดือน โดยข้อมูลที่ได้จาก API จะมีรูปแบบเป็น JSON ซึ่งข้อมูลที่ได้มาประกอบไปด้วยชั้นของ Object ที่ถูกบรรจุข้อมูลการทวิตของผู้ใช้งานที่นำมาใช้ในงานวิจัย ดังนี้ created_at หมายถึง วันที่ เดือน ปี ของข้อความที่ถูกทวิต id หมายถึง เลขไอดีของข้อความนี้ full_text หมายถึง ข้อความทั้งหมดที่ถูกทวิต entities.hashtags หมายถึง แฮสแท็กของทวิตนี้ user.screen_name หมายถึง ชื่อบัญชีที่โพสต์ทวิตนี้

สำหรับการเรียกข้อมูลจากทวีตเตอร์อัตโนมัติได้ใช้โปรแกรมภาษา Python สำหรับการพัฒนา โดยได้นำเอาไลบรารี Tweepy เพื่ออำนวยความสะดวกในการทำงานโดยมีคำสั่งในการเรียกข้อมูล จาก Twitter API คือฟังก์ชัน “tweepy.Cursor()” และมีการส่งพารามิเตอร์ที่ต้องการคือ “api.user_timeline” หมายถึงประเภทที่ต้องการข้อมูลจากทวีตเตอร์สำหรับงานวิจัยครั้งนี้ต้องการ Timeline ของผู้ใช้งานย้อนหลัง ต่อมาคือพารามิเตอร์ “screen_name” หมายถึงชื่อผู้งานที่ต้องการ ข้อมูล tweet_mode หมายถึงรูปแบบในการเรียกข้อมูลให้ตั้งค่าเป็น “extended” คือการตั้งค่าให้ เรียกข้อมูลมาทั้งหมด และสามารถตั้งค่าจำนวนข้อความที่ต้องการเรียกในแต่ละรอบคือการตั้งค่า ฟังก์ชัน “item()”

```

_json={
  'created_at': 'Wed Mar 03 14: 27: 06 +0000 2021',
  'id': 1367119345948614658,
  'id_str': '1367119345948614658',
  'full_text': '21.26น. ถ.สาทรใต้ มุ่งหน้า แยกสาทร-สุรศักดิ์ ทำอยู่หน้าโรงพยาบาล เซนต์หลุยส์ #รายงานจราจร #รถติด #FM91',
  'truncated': False,
  'display_text_range': [
    0,
    106
  ],
  'entities': {
    'hashtags': [
      {
        'text': 'รายงานจราจร',
      },
    ],
  },
  'user': {
    'screen_name': 'fm91trafficpro',
    'location': '',
    'description': 'สถานีวิทยุข่าวจราจรและความปลอดภัยกสทศ.91 เราฟังได้ 1644 โทรฟรีทั่วประเทศ ตลอด 24 ชั่วโมง',
  },
}

```

ภาพ 3.16 รายละเอียดข้อมูล JSON ที่ได้จาก Twitter API

```

screen_name = js100radio
ID = 1434142955095445520
Geo = None
Message = คนขับมีอาการชักเกร็ง ทำรถบรรทุกป็นข้ามเกาะกลางชนโกดังเก็บของ บาดเจ็บ 2 คน
https://t.co/D32h2RNnuf #JS100 https://t.co/nmXt1Rw1v9
วันที่ = 2021-09-04 20:15:00

```

ภาพ 3.17 ผลลัพธ์การเรียกข้อมูลจาก Twitter API Tweepy

```

>> messages = "คนขับมีอาการชักเกร็ง ทำรถบรรทุกป็นข้ามเกาะกลางชนโกดังเก็บของ บาดเจ็บ 2 คน"
>> textCut = word_tokenize(messages,engine='newmm')
>> print(textCut)
>> ['คนขับ', 'มี', 'อา', 'การชัก', 'เกร็ง', ' ', 'ทำ', 'รถบรรทุก', 'ป็น',
  'ข้าม', 'เกาะ', 'กลาง', 'ชน', 'โกดัง', 'เก็บ', 'ของ', ' ', ' ', 'บาดเจ็บ', ' ',
  ' ', '2', ' ', ' ', 'คน']

```

ภาพ 3.18 ผลลัพธ์การตัดคำ PyThaiNLP

3.3.2 Token

เมื่อได้ข้อความจากทวิตเตอร์แล้วยังไม่สามารถนำข้อมูลมาใช้ได้เนื่องจากข้อความยังคงเป็นประโยคภาษาไทยดังนั้นเพื่อให้สามารถนำคำที่ต้องการมาวิเคราะห์ได้จำเป็นต้องแบ่งประโยคออกมาเป็นคำโดยไพทอนมีไลบรารีที่เป็นโอเพนซอร์ซ (Open Source) สำหรับช่วยให้สามารถพัฒนาระบบต่าง ๆ ภาษาไพทอนนั้น pythainlp เป็นอีกไลบรารีของภาษาไพทอนที่พัฒนาโดยคนไทยเป็นไลบรารีที่มีความสามารถในการแบ่งประโยคภาษาออกมาเป็นคำมีคล้ายกับไลบรารี NLTK (Natural Language Toolkit: NLTK) หากแต่ pythainlp นั้นใช้แยกคำที่เป็นภาษาไทยซึ่งนอกจากตัดคำภาษาไทยแล้ว pythainlp ยังมีความสามารถอื่น ๆ อีกเช่น การตรวจตัวสะกด การวิเคราะห์ชนิดของคำทางไวยากรณ์ การเรียงคำภาษาไทย สำหรับงานวิจัยนี้ได้นำเอาฟังก์ชันการตัดคำมาช่วยในการแบ่งประโยคออกเป็นคำคือฟังก์ชัน “word_tokenize(string,engine)” และมีพารามิเตอร์คือ string หมายถึงข้อความที่ต้องการตัดคำ และ engine หมายถึง วิธีใช้ในการตัดคำ โดยมีผลลัพธ์ ดังนี้

3.3.3 Python

ภาษาเพื่อโปรแกรมสำหรับงานวิจัยนี้คือภาษา Python มาเป็นภาษาหลักในการพัฒนา ซึ่งเป็นภาษาที่ได้รับความนิยมเป็นอย่างมากในการพัฒนาเกี่ยวกับการวิเคราะห์ข้อมูลทั้งยังมีไลบรารีให้สามารถทำงานได้อย่างสะดวกและเป็นโอเพนซอร์ซสามารถใช้งานได้ฟรี สำหรับการโปรแกรมภาษา Python เพื่อการทำงานด้านการวิเคราะห์ข้อมูลได้มีการพัฒนาบนแพลตฟอร์ม Anaconda ซึ่งมีเครื่องมือสำหรับการพัฒนางานด้านการวิเคราะห์ข้อมูลมากมาย และหนึ่งในนั้นคือ Jupyter Notebook ซึ่งเป็นเครื่องมือหลักในการพัฒนางานวิจัยนี้ โดยข้อดีของ Jupyter Notebook คือมีการติดตั้งไลบรารีต่าง ๆ ที่จำเป็นมาให้พร้อมไม่ว่าจะเป็น Numpy, Pandas, Keras, Tensorflow, Scikit-learn, Matplotlib และ Seaborn ซึ่งเป็นไลบรารีหลักที่ใช้ในงานวิเคราะห์ข้อมูล และข้อดีอีกประการของ Jupyter Notebook คือการทำงานที่มีการแบ่งชุดคำสั่งโปรแกรมออกเป็นส่วน โดยจะมีการแบ่งการทำงานออกจากกัน เมื่อรันแต่ละส่วนจะไม่มีผลกระทบต่อส่วนอื่น ๆ เว้นแต่ตัวแปรหรือฟังก์ชันที่ถูกเรียกมาใช้งานซึ่งสามารถเรียกข้ามส่วนได้

3.3.4 Neuron Network Tools

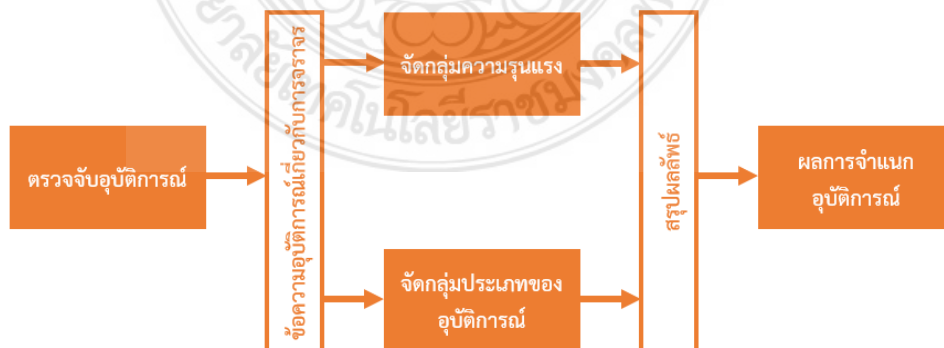
เครื่องมือที่นำมาใช้ในการพัฒนาการเรียนรู้เชิงลึกได้นำเอาเครื่องมือที่มีชื่อว่า Keras ซึ่งเป็นเครื่องมือที่อยู่บนเฟรมเวิร์ค Tensorflow อีกที ข้อดีของ Keras คือสามารถใช้งานได้สะดวกเนื่องจากเป็น API ที่ทำให้นักพัฒนาติดตั้งโปรแกรมและ Import ไลบรารีเข้ามาใช้งานได้ทันที สำหรับงานวิจัยครั้งนี้ได้นำเอาฟังก์ชันของ Keras มาใช้งานเพื่อสร้างโมเดล โดยแต่ละโมเดลจะมีฟังก์ชันที่แตกต่างกันไป สำหรับโมเดล MLP จะมีฟังก์ชันที่สำคัญคือ “Dense()” ซึ่งเป็นฟังก์ชันในการสร้างเลเยอร์และมีพารามิเตอร์คือผลลัพธ์ที่ต้องการให้เลเยอร์นั้นส่งออกมา รวมถึงสามารถเพิ่มฟังก์ชันสำหรับการควบคุมผลลัพธ์ได้ โดยมีโครงสร้างฟังก์ชัน ดังนี้ “Dense(250, activation='relu,')” และสำหรับการสร้างโมเดล CNN จะมีฟังก์ชันในการสร้างโมเดลสองฟังก์ชัน คือ “Conv1D()” และ “GlobalMaxPooling1D()” สำหรับโมเดล Bi-LSTM จะใช้ฟังก์ชัน “Bidirectional(LSTM())”

บทที่ 4

กระบวนการทดสอบ (Testing Process)

4.1 สมมติฐาน

การทดสอบประสิทธิภาพของการตรวจสอบอุบัติการณ์อัตโนมัติหลังจากสร้างโมเดลชุดข้อมูลเทรนเป็นที่เรียบร้อยแล้วโดยโมเดลที่สร้างขึ้นมาประกอบไปด้วย 3 โมเดลคือ โมเดลเพื่อการแบ่งกลุ่มข้อความอุบัติการณ์และข้อความประชาสัมพันธ์ทั่วไป โมเดลเพื่อแบ่งกลุ่มประเภทของอุบัติการณ์ และโมเดลเพื่อแบ่งกลุ่มระดับความรุนแรงของอุบัติการณ์ โดยโมเดลที่ถูกสร้างขึ้นมานั้นได้ใช้ข้อมูลชุดเดียวกันในการเทรน ดังนั้นเพื่อจำกัดความผิดพลาดจึงได้ใช้ชุดข้อมูลสำหรับการทดสอบชุดเดียวกัน โดยจะทำการทดสอบชุดข้อมูลจริงที่เก็บรวบรวมจากการโพสต์ในสื่อสังคมออนไลน์ทวิตเตอร์ (Twitter) โดยจะต้องนำเข้าข้อมูลที่ผ่านขั้นตอนเตรียมข้อมูล (Data Preparation) ซึ่งประกอบไปด้วยการแบ่งประโยคออกเป็นคำ (Word tokenization) การกรองคำที่ไม่ต้องการออก (Data filter) การเปลี่ยนคำเป็นตัวเลข (Text to sequences) และเติมความยาวประโยคให้เท่ากัน (Padding Sequences) หลังจากนั้นนำข้อมูลที่ผ่านกระบวนการตรวจสอบยืนยันว่าข้อความดังกล่าวเป็นอุบัติการณ์จริงหรือไม่ก่อน โดยผลลัพธ์สามารถได้แบ่งเป็น 2 ประเภทคือข้อความที่ได้รับการยืนยันว่าเป็นเหตุการณ์อุบัติการณ์จริงและข้อความที่เป็นการประชาสัมพันธ์ทั่วไปหรือไม่เป็นข้อความอุบัติการณ์ จากนั้นจะนำข้อความประเภทอุบัติการณ์เข้าสู่กระบวนการตรวจสอบประเภทของอุบัติการณ์การจัดกลุ่มประเภทของอุบัติการณ์ซึ่งมีกัน 6 ประเภท โดยมีรูปอินพุตในลักษณะเดียวกันคือเป็นรูปแบบข้อความที่เปลี่ยนเป็นลำดับตัวเลข เมื่อแบ่งประเภทได้สำเร็จจะเข้าสู่กระบวนการเพื่อรายงานผลลัพธ์การตรวจสอบ โดยมีรูปแบบการรายงานร้อยละความถูกต้องของข้อความที่จัดกลุ่มได้ และในขณะเดียวกันคือการนำข้อมูลเดียวกันนี้เข้ากระบวนการแบ่งกลุ่มความรุนแรงซึ่งมีกัน 3 ประเภท เมื่อแบ่งประเภทได้สำเร็จจะเข้าสู่กระบวนการเพื่อรายงานผลลัพธ์การตรวจสอบ โดยมีรูปแบบการรายงานร้อยละความถูกต้องของข้อความที่จัดกลุ่มได้เช่นเดียวกัน ดังแสดงในภาพ 4.1



ภาพ 4.1 ขั้นตอนการการทดสอบและบันทึกผล 3 ขั้นตอน

4.2 การจัดเตรียมข้อมูล

4.2.1 กระบวนการเก็บรวบรวมข้อมูล

ขั้นตอนการจัดเตรียมข้อมูลโดยมีการนำข้อความจากสื่อสังคมออนไลน์ทวิตเตอร์ที่มีการรายงานข้อมูลเกี่ยวกับสภาพจราจรเป็นหลักโดยประกอบไปด้วย @js100radio ซึ่งเป็นบัญชีทวิตเตอร์ของที่มีรากฐานมาจากการรายงานข่าวทางสื่อวิทยุต่อมาได้พัฒนาเพิ่มเติมส่วนของการรายงานข่าวผ่านสื่อสังคมออนไลน์และมีผู้ติดตามกว่า 3 ล้านบัญชีผู้ใช้งาน @Traffic_1197 เป็นบัญชีทวิตเตอร์ที่อยู่ภายใต้การดูแลของกองบังคับการตำรวจจราจร (บก. 02) ซึ่งเป็นหน่วยงานของราชการมีจำนวนผู้ติดตามกว่า 2 หมื่นบัญชีผู้ใช้งาน @fm91trafficpro เป็นบัญชีทวิตเตอร์ของหน่วยงานราชการคือ สถานีวิทยุพิทักษ์สันติราษฎร์ สวพ. FM91 อยู่ภายใต้การดูแลของกองตำรวจสื่อสาร สำนักงานตำรวจแห่งชาติมีจำนวนผู้ติดตามกว่า 2 ล้านบัญชีผู้ใช้งาน โดยมีการนำข้อมูลตั้งแต่วันที่ 5 กุมภาพันธ์ 2564 ถึงวันที่ 23 มีนาคม 2564 มาบันทึกในรูปแบบไฟล์ข้อมูล Excel พร้อมทั้งจัดทำหมวดหมู่ของแต่ละข้อความเจ้าหน้าที่กรมทางหลวงประจำอาคารศูนย์ควบคุมและสั่งการจราจร (CCB ลาดกระบัง) จำนวน 2 คน ซึ่งเป็นบุคลากรที่มีประสบการณ์และมีความเชี่ยวชาญในการ คัดกรองและวิเคราะห์ข้อมูล การจัดการจราจรและเผยแพร่ข้อมูลจราจรเป็นอย่างดี

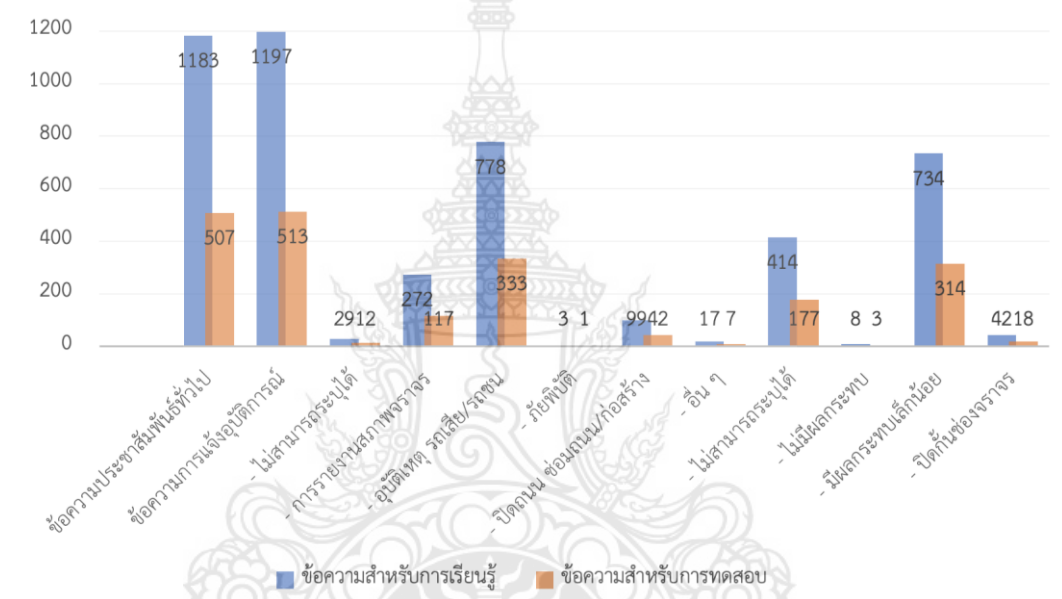
4.2.2 กระบวนการจำแนกชุดข้อมูลเรียนรู้และทดสอบ

ชุดข้อมูล (Dataset) ที่ถูกรวบรวมมาจากสื่อสังคมออนไลน์ทวิตเตอร์โดยมีการรวบรวมข้อมูลข้อความมาทั้งสิ้น 3,400 ข้อความ และดำเนินการจัดกลุ่มของอุบัติการณ์หรือไม่เป็นอุบัติการณ์พร้อมแบ่งแยกประเภทการเกิดอุบัติการณ์และระบุความรุนแรงเป็นที่เรียบร้อย ซึ่งข้อมูลสำหรับการเทรนโมเดลและทดสอบโมเดลนั้นในหลาย ๆ งานวิจัยได้มีการแบ่งข้อมูลออกเป็นสองชุดโดยจะต้องมีการแบ่งแยกกันอย่างชัดเจนระหว่างชุดข้อมูลสำหรับการเทรนและชุดข้อมูลสำหรับการทดสอบซึ่งมีการดำเนินการแบ่งข้อมูลออกเป็น 70 % ของข้อมูลเก็บไว้สำหรับการเทรน และ 30 % ของข้อมูลสำหรับการทดสอบ โดยในส่วนของข้อมูลสำหรับการเทรนนั้นจะมีการแบ่งข้อมูลออกเป็นข้อความประชาสัมพันธ์ทั่วไปและข้อความการแจ้งอุบัติการณ์และในข้อความการแจ้งอุบัติการณ์จะแบ่งออกเป็นข้อความประเภทการเกิดอุบัติการณ์และเป็นข้อความระบุความรุนแรง ตามตาราง 4.1 ภาพ 4.2

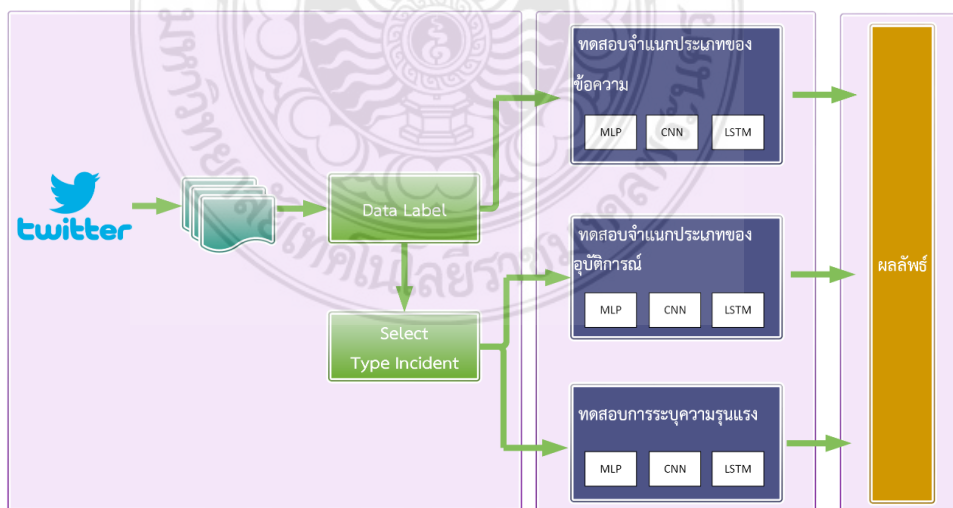
ตาราง 4.1 ข้อมูลผลการทดสอบการตรวจสอบ

ลำดับ	ข้อมูล	ข้อความสำหรับการเรียนรู้		ข้อความสำหรับการทดสอบ	
		จำนวน	สัดส่วน (%)	จำนวน	สัดส่วน (%)
1	ข้อความประชาสัมพันธ์ทั่วไป	1,183	70	507	30
2	ข้อความการแจ้งอุบัติการณ์	1,197	70	503	30
2.1	การจัดรูปแบบอุบัติการณ์				
	- ไม่สามารถระบุได้	29	2.40	12	2.40
	- การรายงานสภาพจราจร	272	22.75	117	22.75
	- อุบัติเหตุ รถเสีย/รถชน	778	64.97	333	64.97
	- ภัยพิบัติ	3	0.23	1	0.23

ลำดับ	ข้อมูล	ข้อความสำหรับการเรียนรู้		ข้อความสำหรับการทดสอบ	
		จำนวน	สัดส่วน (%)	จำนวน	สัดส่วน (%)
	- ปิดถนน ซ่อมถนน/ก่อสร้าง	99	8.25	42	8.25
	- อื่น ๆ	17	1.40	7	1.40
2.2	การระบุความรุนแรง				
	- ไม่สามารถระบุได้	414	34.56	177	34.56
	- ไม่มีผลกระทบ	8	0.64	3	0.64
	- มีผลกระทบเล็กน้อย	734	61.29	314	61.29
	- ปิดกั้นช่องจราจร	42	3.51	18	3.51



ภาพ 4.2 การแบ่งกลุ่มชุดข้อมูลและการกระจายตัวของอุบัติการณ์



ภาพ 4.3 ขั้นตอนการการทดสอบและบันทึกผล 3 ขั้นตอน

4.3 รูปแบบและขั้นตอนการทดสอบ

ในการทดสอบและประเมินผลจะกำหนดให้มีลำดับขั้นตอนในการดำเนินการ จำนวน 3 ขั้นตอน ประกอบไปด้วย ขั้นตอนการทดสอบโมเดลสำหรับการจำแนกประเภทของข้อความ ทำหน้าที่ในการทดสอบความถูกต้องของโมเดลในการจำแนกประเภทของข้อความที่ได้มาจากสื่อสังคมออนไลน์ทวิตเตอร์ เพื่อให้สามารถแยกได้เป็นข้อความประชาสัมพันธ์ทั่วไปหรือเป็นข้อความอุบัติการณ์ ขั้นตอนการทดสอบโมเดลเพื่อจำแนกประเภทของเกิดอุบัติเหตุ ทำหน้าที่จำแนกประเภทข้อความที่ผ่านขั้นตอนการแยกข้อความทั่วไปและข้อความอุบัติการณ์มาแล้วโดยจะทำการทดสอบแบ่งประเภทโมเดลเดี่ยวแยกออกเป็น 6 ประเภทในคราวเดียว และขั้นตอนการจำแนกความรุนแรงของอุบัติเหตุ ทำหน้าที่จำแนกความรุนแรงของอุบัติเหตุจากข้อความที่ผ่านโมเดลก่อนหน้านี้ทั้งสองโมเดลมาเข้าสู่การทดสอบสุดท้ายเพื่อระบุความรุนแรงโดยมีความรุนแรง 4 ระดับซึ่งจะใช้เพียงหนึ่งโมเดลเพื่อระบุความรุนแรงของข้อความอุบัติการณ์ ดังแสดงในภาพ 4.3

4.3.1 ขั้นตอนการทดสอบจำแนกประเภทของข้อความ

ขั้นตอนจำแนกประเภทของข้อความ เริ่มจากนำข้อมูลที่ได้จากการรวบรวมข้อความจากทวิตเตอร์มาผ่านกระบวนการต่าง ๆ เพื่อให้พร้อมเป็นการจัดเตรียมข้อมูลก่อนการทดสอบโดยดำเนินการขั้นตอนประกอบไปด้วย การแยกคำ การกรองคำที่ไม่มีความออก การเปลี่ยนคำเป็นตัวเลขแบบลำดับ และการเพิ่มความยาวข้อความ จากนั้นจึงเข้าสู่กระบวนการทดสอบในขั้นตอนนี้จะดำเนินการผ่านข้อความที่พร้อมสำหรับการทดสอบเข้าสู่โมเดลจำแนกประเภทซึ่งเป็นโมเดลแรก โดยมีวัตถุประสงค์ในการจำแนกข้อความระหว่างข้อความประชาสัมพันธ์ทั่วไปหรือเป็นข้อความการแจ้งอุบัติเหตุ ดังแสดงในภาพ 4.4

4.3.2 ขั้นตอนการทดสอบจำแนกประเภทของอุบัติเหตุ

สำหรับขั้นตอนการทดสอบการจำแนกประเภทของอุบัติเหตุจะดำเนินการเป็นลำดับที่สองถัดจากการจำแนกประเภทข้อความ โดยข้อความที่นำมาทดสอบเป็นข้อความอุบัติการณ์ที่ได้ผ่านการแยกประเภทข้อความมาจากการแบ่งข้อความเจ้าหน้าที่แล้ว ดังนั้น จึงมีการจัดเตรียมข้อมูลก่อนการทดสอบโดยการเก็บเข้าตัวแปรเพื่อผ่านต่อไปยังโมเดลเพื่อจำแนกประเภทอุบัติเหตุ ในขั้นตอนนี้จะดำเนินการทดสอบความถูกต้องของการแยกประเภทของอุบัติเหตุ โดยมีวัตถุประสงค์ในการสรุปคำตอบของข้อความอุบัติการณ์ให้สามารถบอกประเภทของการเกิดอุบัติเหตุได้โดยประเภทของอุบัติเหตุของกลุ่มข้อมูลที่นำมาทดสอบนั้นได้ผ่านขั้นตอนการจำแนกบุคลากรที่เกี่ยวข้องว่าเป็นที่เรียบร้อย โดยมีทั้งสิ้น 6 ประเภทจึงสามารถนำข้อความมาทดสอบได้ทันที ดังแสดงในภาพ 4.5

```
In [10]: correct = accuracy_score(y_true, y_pred, normalize=False)
persen = accuracy_score(y_true, y_pred)*100
print('จำนวนข้อความทดสอบ ',count_test, ' ข้อความ')
print('ทดสอบได้ถูกต้อง ',correct, ' ข้อความ')
print('ทดสอบได้ถูกต้องร้อยละ %.2f' %persen)
```

```
จำนวนข้อความทดสอบ 1010 ข้อความ
ทดสอบได้ถูกต้อง 957 ข้อความ
ทดสอบได้ถูกต้องร้อยละ 94.75
```

ภาพ 4.4 ขั้นตอนการทดสอบการจำแนกประเภทข้อความ

4.3.3 ขั้นตอนการทดสอบการระบุความรุนแรง

ในขั้นตอนการทดสอบประสิทธิภาพของการระบุความรุนแรงมีกระบวนการคือเมื่อได้ข้อความจากทวิตเตอร์ ซึ่งเช่นเดียวกับกระบวนการก่อนหน้าที่จะนำเอาข้อความประเภทอุปติการณ์เท่านั้นมาใช้ในการทดสอบเริ่มกระบวนการโดยนำข้อความที่ผ่านการแยกประเภทเจ้าหน้าที่มาดำเนินการส่วนเตรียมข้อมูล (Preprocess) แล้วจึงนำข้อความที่ผ่านขั้นตอนเตรียมข้อมูลมาเข้าสู่โมเดลเพื่อทดสอบความถูกต้องของโมเดล สำหรับขั้นตอนนี้เป็นารทดสอบโมเดลเพื่อระบุความรุนแรงซึ่งมีอยู่การ 4 ประเภท

In [19]: df_selectrd_test

Out [19]:

	tweet	detail
0	#ทล347 ขาออกมุ่งหน้าอยุธยา การจราจรคล่องตัว แต่...	1.1
1	10:30 #อุบัติเหตุ #ถนนรัชดาภิเษก จากแยกรัชโยธิ...	3.1
2	10:35 #อุบัติเหตุ #ทางยกระดับรามคำแหง จากโอสถส...	3.1
3	รถติดในซอย สุขุมวิท 101 หนักมากคะตติยาวเป็น ก...	1.2
4	ถ.สายเอเชียขาขึ้น ช่วงแยกเข้า อ.ป่าโมก บัด 2 ...	5.4
...
498	14.26น. ถ.ลาดพร้าว ขาเข้า มุ่งหน้า แยกลาดพร้าว...	1.2
499	14.55น. นานกว่านี้ แยะแน่ !! ถ.เทพรัตน (บางนาต...	3.3
500	ชุดแล้ว !! เสี่ยงใช้เส้นทาง 14.58 น. ถนนประเสริ...	5.4
501	15.30น. จุดนี้รถติด ! ถ.กัลปพฤกษ์ มุ่งหน้าไปหน...	3.1
502	"รถเทรลเลอร์ชนกัน" แนะเสี่ยงเส้นทาง... 15.40 น...	3.1

503 rows x 2 columns

```
In [26]: persen = (score*100)/len(result_test)
print('จำนวนข้อความทดสอบ ', len(result_test), ' ข้อความ')
print('ทดสอบได้ถูกต้อง ', score, ' ข้อความ')
print('ทดสอบได้ถูกต้องร้อยละ %.2f' %persen)
```

จำนวนข้อความทดสอบ 503 ข้อความ
 ทดสอบได้ถูกต้อง 392 ข้อความ
 ทดสอบได้ถูกต้องร้อยละ 77.93

ภาพ 4.5 ขั้นตอนการทดสอบการจำแนกประเภทของอุปติการณ์

4.4 การประเมินประสิทธิภาพ

4.4.1 การจัดกลุ่มข้อมูลทดสอบ

เพื่อให้การทดสอบมีความแม่นยำและน่าเชื่อถือ การทดลองจะทำการแบ่งข้อมูลออกเป็น 2 ส่วน สำหรับเพื่อสร้างเป็นข้อมูลสำหรับเรียนรู้ 70 % และข้อมูลสำหรับทดสอบ 30 % (70/30) ซึ่งเป็นอัตราส่วนที่เหมาะสมเนื่องจาก หากมีข้อความสำหรับเรียนรู้มากขึ้นแบบจำลองจะทำงานได้ดีขึ้นแน่นอน (90/10) แต่จะทำให้ข้อมูลทดสอบน้อยเกินไปจนไม่สามารถพิสูจน์ผลได้อย่างแน่ชัด ในทางกลับกันหากลดข้อมูลสำหรับการเรียนรู้เป็น (60/40) จะทำให้แบบจำลองมีความแม่นยำต่ำลงเช่นกัน ผู้วิจัยได้เปรียบเทียบความแม่นยำจากการแบ่งข้อมูลการเรียนรู้และทดสอบแต่ละอัตราดังตาราง 4.2

	tweet	severity
0	#ทล347 ขาออกมุงหน้าอยุธยา การจราจรล่องตัว แต่...	0
1	10:30 #อุบัติเหตุ #ถนนรัชดาภิเษก จากแยกรัชโยธิ...	2
2	10:35 #อุบัติเหตุ #ทางยกระดับรามคำแหง จากโอสถส...	2
3	รถติดในซอย สุขุมวิท 101 หนักมากคะติดยาวเป็น ก...	0
4	ถ.สายเอเชียขาขึ้น ช่วงแยกเข้า อ.ป่าโมก บิด 2 ...	3
...
498	14.26น. ถ.ลาดพร้าว ขาเข้า มุงหน้า แยกลาดพร้าว...	0
499	14.55น. นานกว่านี้ แะแน่ !! ถ.เทพรัตน (บางนาต...	2
500	ขุดแล้ว !! เลี่ยงใช้เส้นทาง 14.58 น. ถนนประเสริ...	0
501	15.30น. จุดนี้รถติด ! ถ.กัลปพฤกษ์ มุงหน้าไปหน...	2
502	"รถเทรลเลอร์ชนกัน" แนะเลี่ยงเส้นทาง... 15.40 น...	3

503 rows × 2 columns

```
In [16]: correct = accuracy_score(y_true, y_pred, normalize=False)
persen = accuracy_score(y_true, y_pred)*100
print('จำนวนข้อความทดสอบ ', len(labelTest), ' ข้อความ')
print('ทดสอบได้ถูกต้อง ', correct, ' ข้อความ')
print('ทดสอบได้ถูกต้องร้อยละ %.2f' %persen)
```

จำนวนข้อความทดสอบ 503 ข้อความ
ทดสอบได้ถูกต้อง 442 ข้อความ
ทดสอบได้ถูกต้องร้อยละ 87.87

ภาพ 4.6 ขั้นตอนการทดสอบการระบุความรุนแรง ในการทดสอบและบันทึกผล

ตาราง 4.2 เปรียบเทียบความแม่นยำเมื่อมีการแบ่งข้อมูลต่างกัน

สัดส่วน	แบบที่ 1	แบบที่ 2	แบบที่ 3	แบบที่ 4	แบบที่ 5	แบบที่ 6
อัตราการเรียนรู้	50	60	70	80	90	95
อัตราข้อความทดสอบ	50	40	30	20	10	5
ความแม่นยำ (%)	81.00	81.34	81.87	80.30	83.93	85.71

จะเห็นว่าการแบ่งข้อมูลออกเป็นอัตรา 90:10 และ 95:5 มีความแม่นยำสูงแต่ไม่มีความน่าเชื่อถือ เนื่องจากข้อความทดสอบมีน้อยเกินไปจนไม่สามารถเป็นตัวแทนของข้อความทั้งหมดได้ ทำการทดสอบด้วยการโปรแกรมภาษา Python ซึ่งสามารถใช้งานไลบรารี Keras ที่มีการให้บริการอยู่บน Tensorflow โดยมีกระบวนการใช้งานด้วยการเรียกใช้ฟังก์ชันสำหรับแต่ละโมเดล เช่น โมเดลด้วยวิธี LSTM จะเรียกใช้งานฟังก์ชัน Bidirectional() แล้วส่ง พารามิเตอร์ LSTM ไปในฟังก์ชันเพื่อเรียกใช้งาน

4.4.2 การทดสอบประสิทธิภาพด้วยวิธีการ K-Fold

หลังจากการปรับแต่งพารามิเตอร์ต่าง ๆ ของโมเดลเพื่อให้พร้อมสำหรับการเรียนรู้เพื่อสร้างโมเดลเป็นที่เรียบร้อยแล้ว เพื่อให้การเรียนรู้สามารถนำข้อมูลที่มีการกระจายตัวมาเรียนรู้นั้นมีกระบวนการคือการแบ่งกลุ่มข้อมูลออกเป็นกลุ่มย่อย (K-Fold) และนำกลุ่มข้อมูลย่อย นั้นมาสลับกันไปมาหรือที่เรียกว่า “การทดสอบแบบไขว้ (Cross Validation)” เพื่อหาค่าเฉลี่ยของประสิทธิภาพของโมเดลที่ผ่านการเรียนรู้ โดยในการทดลองนี้ได้แบ่งกลุ่มข้อมูลออกเป็น 10 กลุ่มหรือ $k=10$ โดยในการแบ่งกลุ่มข้อมูลจะมีการสุ่มค่าในแต่ละกลุ่มเพื่อไม่ให้เกิดค่าที่ซ้ำกันและทำการทดสอบประสิทธิภาพของการเรียนรู้ของโมเดลจำนวน 10 รอบ โดยแต่ละรอบมีการสลับสับเปลี่ยนให้มีการหยิบกลุ่มมาเรียนรู้และทดสอบให้มีกลุ่มสำหรับการเรียนรู้ 70 % และข้อมูลสำหรับการทดสอบ 30 %

4.4.3 การวัดความแม่นยำของวิธีการ

การวัดความแม่นยำการตรวจจับหรือจำแนกอุบัติการณ์ จะสามารถพิจารณาได้จาก 2 พารามิเตอร์ประกอบไปด้วย DR (Detection Rate: DR) เป็นการวัดค่าความถูกต้องของการตรวจจับอุบัติการณ์ มีหน่วยเป็นเปอร์เซ็นต์ สามารถคำนวณได้ดังสมการที่ (1) โดย D_n คือจำนวนครั้งของอุบัติการณ์ที่ตรวจพบโดยวิธีการที่ทดสอบ และ D_t คือจำนวนครั้งของอุบัติการณ์ทั้งหมดที่มีในการทดสอบ

$$DR = \frac{D_n}{D_t} \times 100 \quad (1)$$

ในขณะเดียวกันการวัดความแม่นยำในอีกลักษณะคือการประเมินอัตราการแจ้งเตือนที่ผิดพลาด คือ FAR (False Alarm Rate) มีหน่วยการวัดเป็นค่าเปอร์เซ็นต์ สามารถคำนวณได้ดังสมการที่ (2) โดย N_f คือจำนวนข้อความที่แจ้งเตือนผิดพลาด และ N_t คือจำนวนอุบัติการณ์ทั้งหมดที่มีในการทดสอบ

$$FAR = \frac{N_f}{N_t} \times 100 \quad (2)$$

บทที่ 5

ผลการวิจัยและพัฒนา (Results)

5.1 การทดสอบการตรวจจับการเกิดอุบัติเหตุ

5.1.1 MPL

ในการทดสอบการจำแนกการเกิดอุบัติเหตุวิธีการ MPL กำหนดให้วิธีการมีรายละเอียดในการทดสอบที่มีประสิทธิภาพที่เหมาะสมต่อการแก้ปัญหาที่สุด ดังนี้ จำนวนชั้นประมวลผลภายในเท่ากับ 5 ชั้น สมาชิกในลำดับชั้น 4 ชั้นแรกมีจำนวน 450 โหนดต่อลำดับชั้น รูปแบบฟังก์ชันเชิงเส้นและในชั้นที่ 5 ซึ่งเป็นชั้นสุดท้ายมีจำนวน 2 โหนดเพื่อให้สอดคล้องกับจำนวนกลุ่มข้อมูลที่ต้องการแบ่ง มีรูปแบบฟังก์ชันเป็นซอฟต์แวร์แมกซ์ (Softmax) ใช้ข้อมูลในการทดสอบ 1,010 ข้อมูล โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.1 เป็นผลลัพธ์ค่าความผิดพลาด (Loss) และภาพ 5.2 เป็นผลลัพธ์ทางความแม่นยำ (Accuracy) ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ MPL โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติเหตุเท่ากับ 59.41 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติเหตุผิดพลาดเท่ากับ 40.59 %

5.1.2 CNN

ในการทดสอบการจำแนกการเกิดอุบัติเหตุวิธีการ CNN กำหนดให้วิธีการมีรายละเอียดในการทดสอบที่มีประสิทธิภาพที่เหมาะสมต่อการแก้ปัญหาที่สุด ดังนี้ จำนวนชั้นประมวลผลภายในเท่ากับ 5 ชั้น สมาชิกในแต่ละลำดับชั้นมีจำนวน 45 โหนดในลำดับชั้นแรก ลำดับที่สองเป็นชั้นของคอนโวลูชันเลเยอร์มีจำนวนโหนด 250 โหนด รูปแบบฟังก์ชันเชิงเส้นในชั้นที่สามเป็นชั้นของพูลลิงเลเยอร์มีจำนวนโหนด 250 โหนด และชั้นที่สี่และห้ามีจำนวนโหนด 150 โหนดและ 2 โหนดตามลำดับ โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.3 เป็นผลลัพธ์ค่าความผิดพลาดและภาพ 5.4 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ CNN โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติเหตุเท่ากับ 92.57 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติเหตุผิดพลาดเท่ากับ 7.43 %

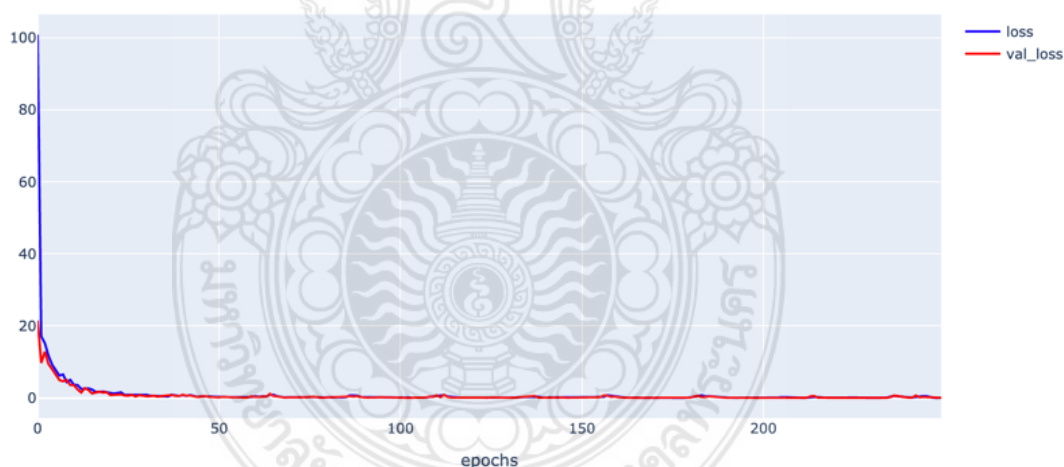
5.1.3 LSTM

ในการทดสอบการจำแนกประเภทของอุบัติเหตุวิธีการ LSTM กำหนดให้วิธีการมีรายละเอียดในการทดสอบที่มีประสิทธิภาพที่เหมาะสมต่อการแก้ปัญหาที่สุด ดังนี้ จำนวนชั้นประมวลผลภายในเท่ากับ 4 ชั้น สมาชิกในแต่ละลำดับชั้นในชั้นแรกมีจำนวน 45 โหนดลำดับชั้นที่สองซึ่งเป็นชั้นของ Bi-LSTM มีจำนวนโหนดทั้งสิ้น 90 โหนด และเพิ่มฟังก์ชันสุ่มลดจำนวนสมาชิกที่อาจไม่ถูกใช้งาน 20 % ในแต่ละชั้น ในลำดับชั้นที่สามมีจำนวนโหนดจำนวน 50 โหนด รูปแบบฟังก์ชันเชิงเส้นและสี่เป็นชั้นเพื่อรวมค่าเป็นผลลัพธ์และ 2 โหนดตามลำดับ โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.5 เป็นผลลัพธ์ค่า

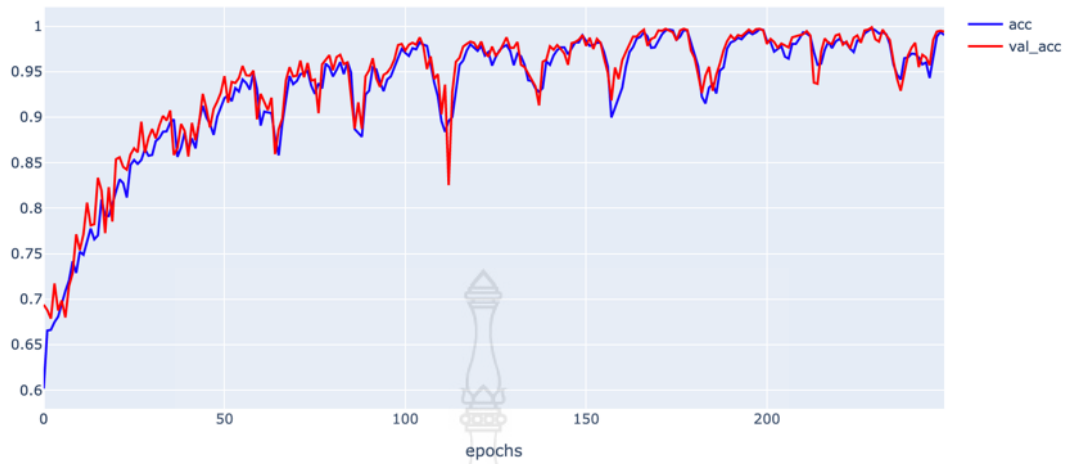
ความผิดพลาดและภาพ 5.6 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ LSTM โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติเหตุเท่ากับ 94.75 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติเหตุผิดพลาดเท่ากับ 5.25 %

5.1.4 LSTM+CNN

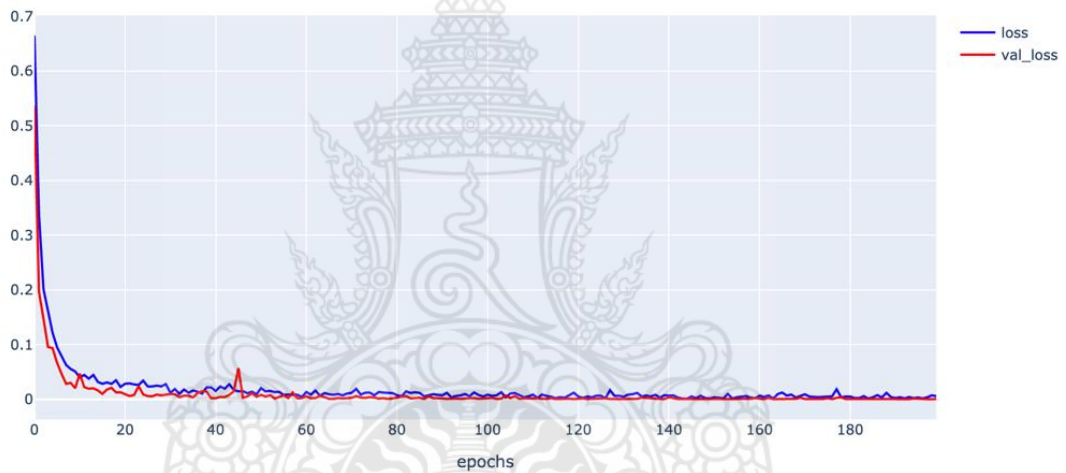
ในการทดสอบการจำแนกประเภทของอุบัติเหตุวิธีการ LSTM+CNN กำหนดให้วิธีการมีการทำงานร่วมกัน โดยการผ่านพารามิเตอร์จาก CNN ที่ได้จากคอนโวลูชันเลเยอร์และนำเข้าสู่ LSTM เลเยอร์และตัดสินใจเลือกผลลัพธ์การทำนายด้วยชั้นซอฟต์แวร์สำหรับส่งพารามิเตอร์เข้าสู่โมเดล โดยวิธีการที่นำเสนอจะมีความแตกต่างจากวิธีการ CNN และ LSTM ก่อนหน้านี้ ในขั้นแรกคือชั้นประมวลผลภายในที่ 32 ลำดับชั้น (Dimension) ในส่วนของคอนโวลูชันเลเยอร์จะมี 250 โหนด และนำผลลัพธ์ส่งเข้าสู่ LSTM เลเยอร์ โดยระบุพารามิเตอร์เข้าเป็น 32 ลำดับชั้น พร้อมกำหนดสุ่มลดจำนวนสมาชิกที่อาจไม่ถูกใช้งานที่ 20 % ในเลเยอร์สุดท้ายจะรวมผลลัพธ์ออกเป็นคำตอบโดยชั้นผลลัพธ์ โดยจะกำหนดผลลัพธ์เป็น 2 สำหรับข้อความการตรวจจับอุบัติเหตุ โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.6 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ LSTM+CNN โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติเหตุเท่ากับ 94.06 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติเหตุผิดพลาดเท่ากับ 5.94 %



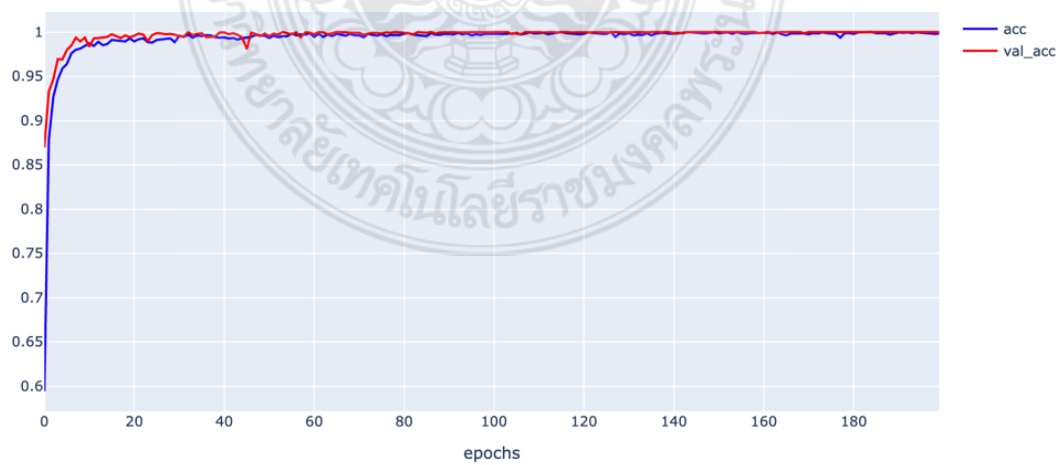
ภาพ 5.1 ความผิดพลาดการเรียนรู้การจำแนกการเกิดอุบัติเหตุวิธีการ MPL



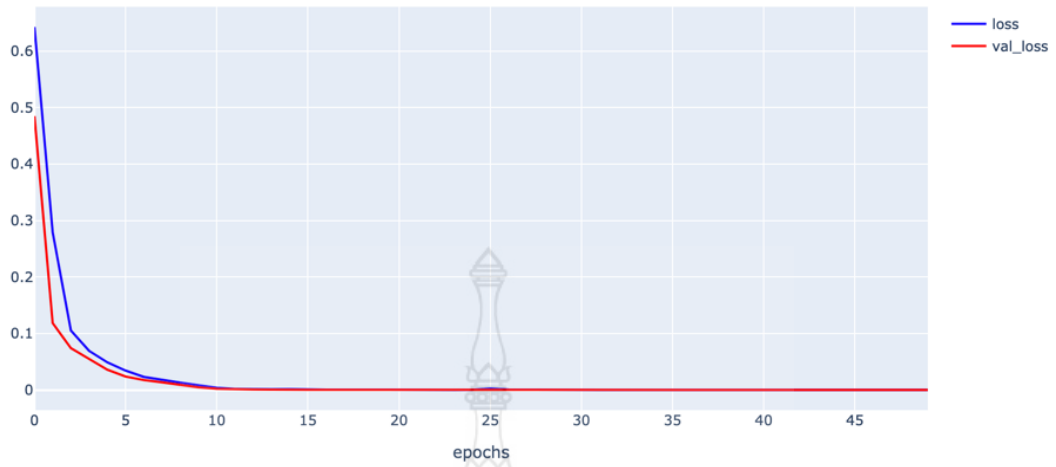
ภาพ 5.2 ความแม่นยำของการเรียนรู้การจำแนกการเกิดอุบัติเหตุการณ์วิธีการ MPL



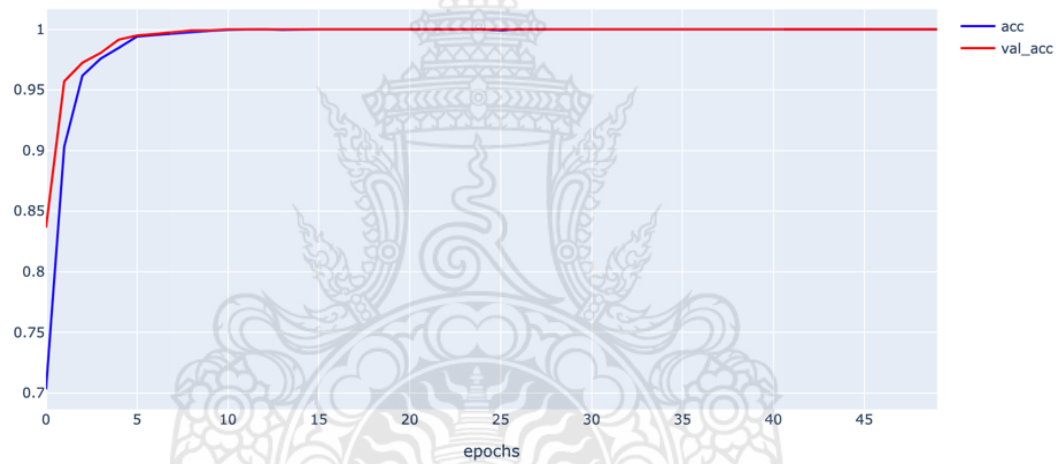
ภาพ 5.3 ความผิดพลาดการเรียนรู้การจำแนกการเกิดอุบัติเหตุการณ์วิธีการ CNN



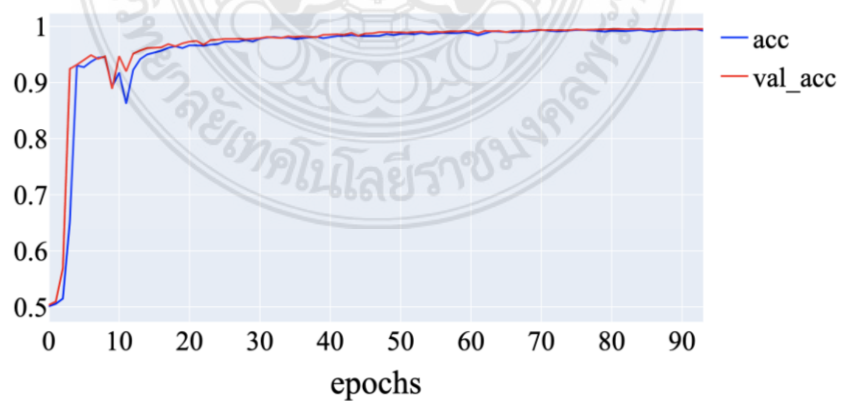
ภาพ 5.4 ความแม่นยำของการเรียนรู้การจำแนกการเกิดอุบัติเหตุการณ์วิธีการ CNN



ภาพ 5.5 ความผิดพลาดการเรียนรู้การจำแนกการเกิดอุบัติเหตุการฉวีวิธีการ LSTM



ภาพ 5.6 ความแม่นยำของการเรียนรู้การจำแนกการเกิดอุบัติเหตุการฉวีวิธีการ LSTM



ภาพ 5.7 ความแม่นยำของการเรียนรู้การจำแนกการเกิดอุบัติเหตุการฉวีวิธีการ LSTM+CNN

5.2 การทดสอบการจำแนกประเภทของอุบัติการณ์ (Incident Classification)

5.2.1 MPL

ในการทดสอบการจำแนกประเภทของอุบัติการณ์วิธีการ MPL กำหนดให้วิธีการมีรายละเอียดในการทดสอบที่มีประสิทธิภาพที่เหมาะสมต่อการแก้ปัญหาที่สุด ดังนี้ จำนวนชั้นประมวลผลภายในเท่ากับ 5 ชั้น สมาชิกในลำดับชั้น 4 ชั้นแรกมีจำนวน 450 โหนดต่อลำดับชั้น รูปแบบฟังก์ชันเชิงเส้นและในชั้นที่ 5 ซึ่งเป็นชั้นสุดท้ายมีจำนวน 4 โหนดเพื่อให้สอดคล้องกับจำนวนกลุ่มข้อมูลที่ต้องการแบ่งประเภท และเพิ่มฟังก์ชันสุ่มลดจำนวนสมาชิกที่อาจไม่ถูกใช้งาน 20 % ในแต่ละชั้นเพื่อลดปัญหาการสร้างสมาชิกเกินความจำเป็นและให้เกิดการประมวลผลผิดพลาด และเมื่อทำการเทรนโมเดลได้มีการเพิ่มฟังก์ชันหยุดการเรียนรู้ (Early Stop) เมื่อค่าความผิดพลาดสูงขึ้น สำหรับข้อมูลที่นำมาทดสอบมีทั้งสิ้น 503 ข้อมูลซึ่งเป็นข้อมูลที่เป็นข้อความอุบัติการณ์เท่านั้น โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.8 เป็นผลลัพธ์ค่าความผิดพลาดและภาพ 5.9 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ MPL โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติการณ์เท่ากับ 67.20 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติการณ์ผิดพลาดเท่ากับ 32.80 %

5.2.2 CNN

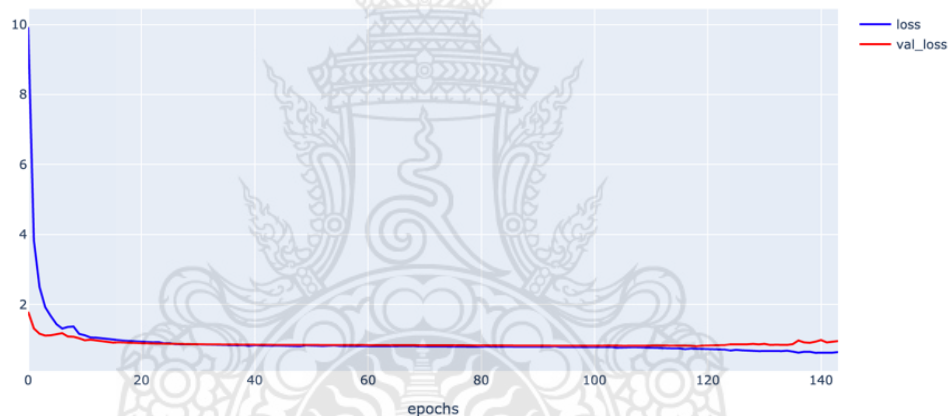
ในการทดสอบการจำแนกประเภทของอุบัติการณ์วิธีการ CNN กำหนดให้วิธีการมีรายละเอียดในการทดสอบที่มีประสิทธิภาพที่เหมาะสมต่อการแก้ปัญหาที่สุด ดังนี้ จำนวนชั้นประมวลผลภายในเท่ากับ 5 ชั้น สมาชิกในแต่ละลำดับชั้น มีจำนวน 45 โหนด ในลำดับชั้นแรก ลำดับที่สองเป็นชั้นของคอนโวลูชันเลเยอร์ มีจำนวนโหนด 250 โหนด รูปแบบฟังก์ชันเชิงเส้นในชั้นที่สามเป็นชั้นของพูลลิงเลเยอร์มีจำนวนโหนด 250 โหนด และชั้นที่สี่และห้ามีจำนวนโหนด 150 โหนดและ 4 โหนดตามลำดับ และเมื่อทำการเรียนรู้โมเดลได้มีการเพิ่มฟังก์ชันเพื่อหยุดการเรียนรู้ เมื่อค่าความผิดพลาดสูงขึ้น โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.10 เป็นผลลัพธ์ค่าความผิดพลาดและภาพ 5.11 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ CNN โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติการณ์เท่ากับ 85.88 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติการณ์ผิดพลาดเท่ากับ 14.12 %

5.2.3 LSTM

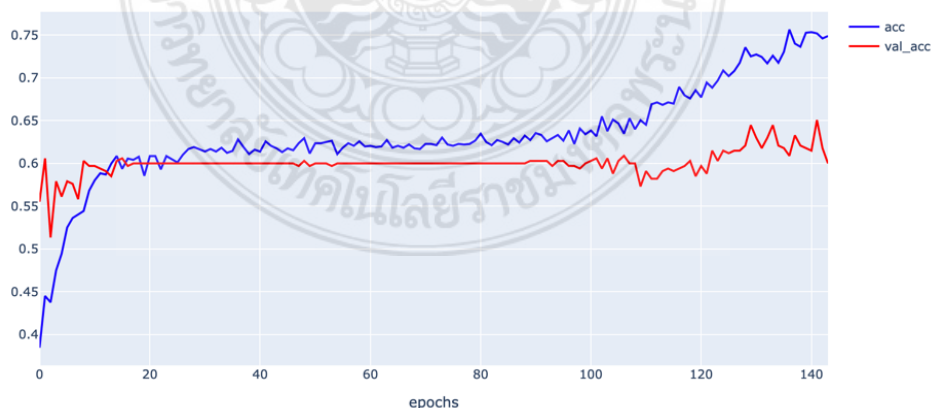
ในการทดสอบการจำแนกประเภทของอุบัติการณ์วิธีการ LSTM กำหนดให้วิธีการมีรายละเอียดในการทดสอบที่มีประสิทธิภาพที่เหมาะสมต่อการแก้ปัญหาที่สุด ดังนี้ จำนวนชั้นประมวลผลภายในในชั้นแรกมีจำนวน 45 โหนดลำดับชั้นที่สองซึ่งเป็นชั้นของ Bi-LSTM มีจำนวนโหนดทั้งสิ้น 90 โหนด และเพิ่มฟังก์ชันสุ่มลดจำนวนสมาชิกที่อาจไม่ถูกใช้งาน 20 % ในแต่ละชั้น ในลำดับชั้นที่สามมีจำนวนโหนดจำนวน 50 โหนด รูปแบบฟังก์ชันเชิงเส้นและสี่เป็นชั้นเพื่อรวมค่าเป็นผลลัพธ์และ 2 โหนดตามลำดับ และเมื่อทำการเรียนรู้โมเดลได้มีการเพิ่มฟังก์ชันเพื่อหยุดการเรียนรู้ เมื่อค่าความผิดพลาดสูงขึ้น โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.12 เป็นผลลัพธ์ค่าความผิดพลาดและภาพ 5.13 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ LSTM โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติการณ์เท่ากับ 87.87 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติการณ์ผิดพลาดเท่ากับ 12.13 %

5.2.4 LSTM+CNN

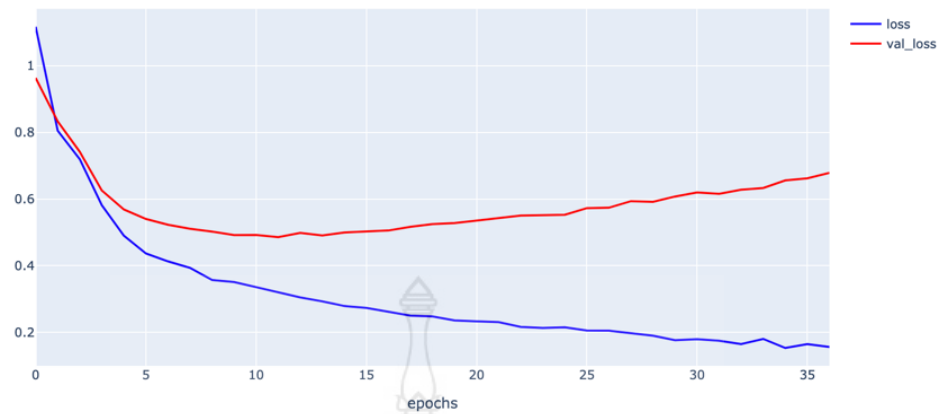
ในการทดสอบการจำแนกประเภทของอุปกรณ์วิธีการ LSTM+CNN กำหนดให้วิธีการมีการทำงานร่วมกัน โดยการผ่านพารามิเตอร์จาก CNN ที่ได้จากคอนโวลูชันเลเยอร์และนำเข้าสู่ LSTM เลเยอร์และตัดสินใจเลือกผลลัพธ์การทำนายด้วยชั้นซอฟต์แวร์แมกซ์ สำหรับการส่งพารามิเตอร์เข้าสู่โมเดล โดยวิธีการที่นำเสนอจะมีความแตกต่างจากวิธีการ CNN และ LSTM ก่อนหน้านี้ ในขั้นแรกคือชั้นประมวลผลภายในที่ 32 ลำดับชั้น ในส่วนของคอนโวลูชันเลเยอร์จะมี 250 โหนด และนำผลลัพธ์ส่งเข้าสู่ LSTM เลเยอร์ โดยระบุพารามิเตอร์เข้าเป็น 32 ลำดับชั้น พร้อมกำหนดสัดส่วนจำนวนสมาชิกที่อาจไม่ถูกใช้งานที่ 20 % ในเลเยอร์สุดท้ายจะรวมผลลัพธ์ออกเป็นค่าตอบโดยชั้นผลลัพธ์ โดยจะกำหนดผลลัพธ์เป็น 5 สำหรับข้อความที่ใช้ในการจำแนกประเภทอุปกรณ์ โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.14 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ LSTM+CNN โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติเหตุเท่ากับ 84.58 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติเหตุผิดพลาดเท่ากับ 15.42 %



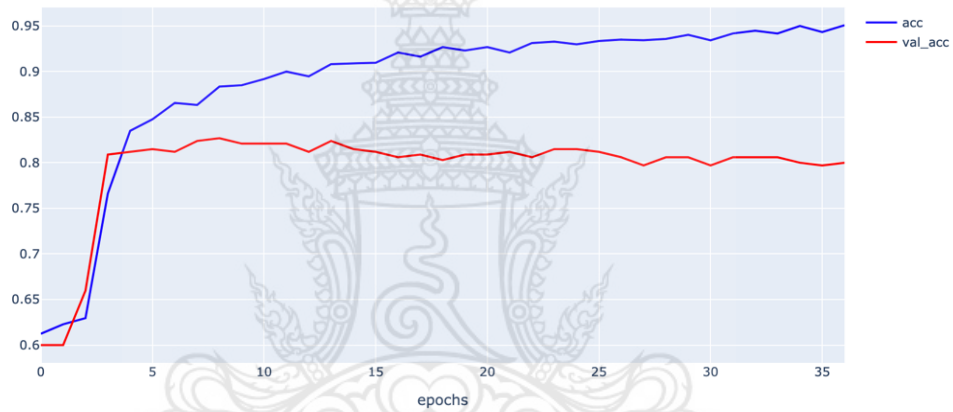
ภาพ 5.8 ความผิดพลาดการเรียนรู้การจำแนกประเภทของอุปกรณ์วิธีการ MPL



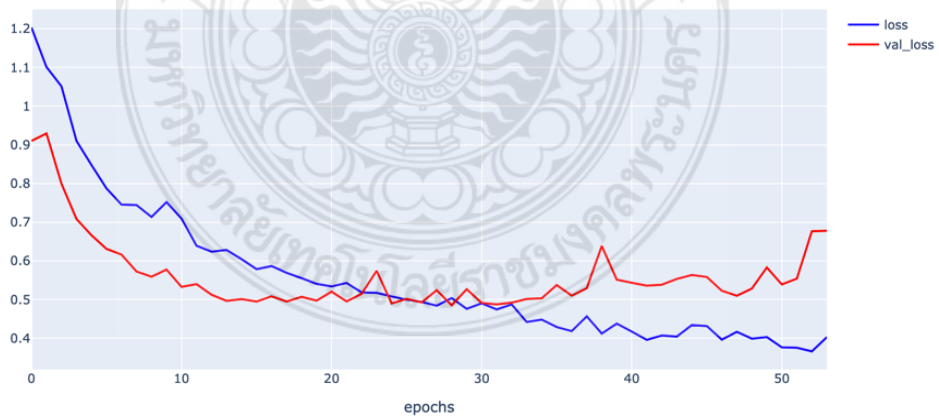
ภาพ 5.9 ความแม่นยำของการเรียนรู้การจำแนกประเภทของอุปกรณ์วิธีการ MPL



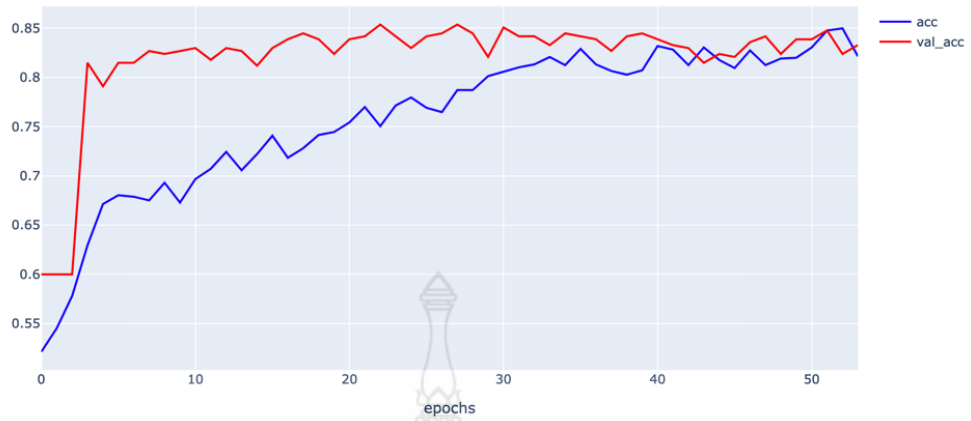
ภาพ 5.10 ความผิดพลาดการเรียนรู้การจำแนกประเภทของอุปกรณ์วิธีการ CNN



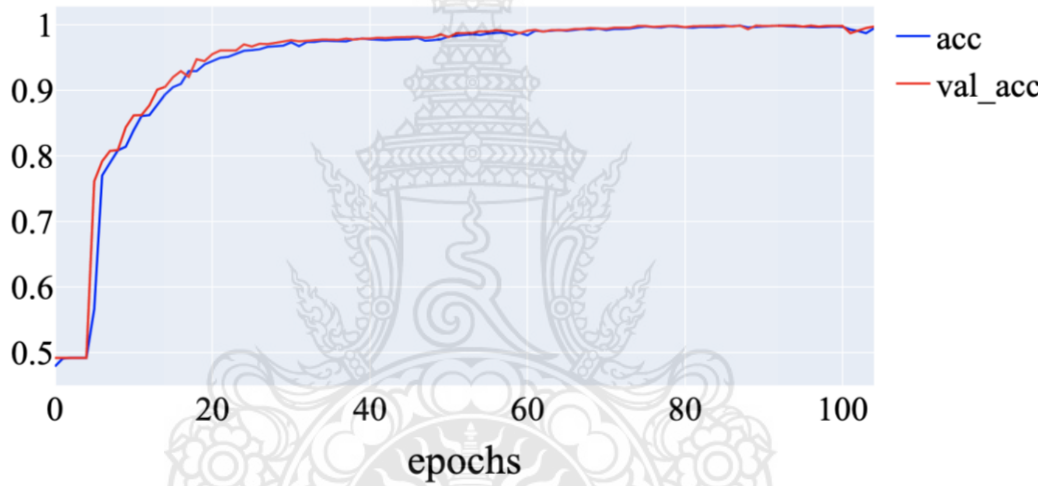
ภาพ 5.11 ความแม่นยำของการเรียนรู้การจำแนกประเภทของอุปกรณ์วิธีการ CNN



ภาพ 5.12 ความผิดพลาดการเรียนรู้การจำแนกประเภทของอุปกรณ์วิธีการ LSTM



ภาพ 5.13 ความแม่นยำของการเรียนรู้การจำแนกประเภทของอุปกรณ์วิธีการ LSTM



ภาพ 5.14 ความแม่นยำของการเรียนรู้การจำแนกประเภทของอุปกรณ์วิธีการ LSTM+CNN

5.3 การทดสอบการระบุความรุนแรงของอุบัติการณ์ (Incident Impact)

5.3.1 MPL

ในการทดสอบการระบุความรุนแรงของอุบัติการณ์วิธีการ MPL กำหนดให้วิธีการมีรายละเอียดในการทดสอบที่มีประสิทธิภาพที่เหมาะสมต่อการแก้ปัญหาที่สุด ดังนี้ จำนวนชั้นประมวลผลภายในเท่ากับ 5 ชั้น สมาชิกในลำดับชั้น 4 ชั้นแรกมีจำนวน 450 โหนดต่อลำดับชั้น รูปแบบฟังก์ชันเชิงเส้นและในชั้นที่ 5 ซึ่งเป็นชั้นสุดท้ายมีจำนวน 6 โหนดเพื่อให้สอดคล้องกับจำนวนกลุ่มข้อมูลที่ต้องการแบ่งประเภทและเพิ่มฟังก์ชันสุ่มลดจำนวนสมาชิกที่อาจไม่ถูกใช้งาน 20 % ในแต่ละชั้นเพื่อลดปัญหาการสร้างสมาชิกเกินความจำเป็นและให้เกิดการประมวลผลผิดพลาด และเมื่อทำการเทรนโมเดลได้มีการเพิ่มฟังก์ชันเพื่อหยุดการเรียนรู้ เมื่อค่าความผิดพลาดสูงขึ้น สำหรับข้อมูลที่นำมาทดสอบมีทั้งสิ้น 503 ข้อมูล ซึ่งเป็นข้อมูลที่เป็นข้อความอุบัติการณ์เท่านั้น โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.15 เป็นผลลัพธ์ค่าความผิดพลาดและภาพ 5.16 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ MPL โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติการณ์เท่ากับ 44.14 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติการณ์ผิดพลาดเท่ากับ 55.86 %

5.3.2 CNN

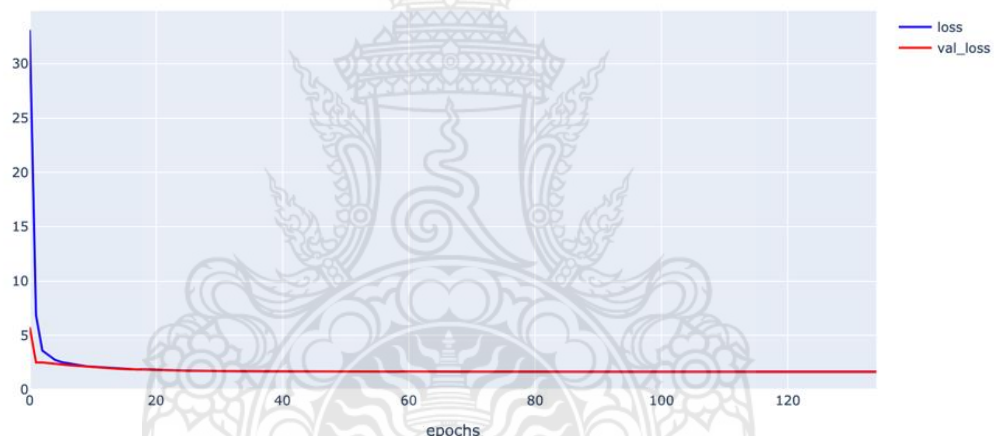
ในการทดสอบการระบุความรุนแรงของอุบัติการณ์วิธีการ CNN กำหนดให้วิธีการมีรายละเอียดในการทดสอบที่มีประสิทธิภาพที่เหมาะสมต่อการแก้ปัญหาที่สุด ดังนี้ จำนวนชั้นประมวลผลภายในมีจำนวน 45 โหนดในลำดับชั้นแรก ลำดับที่สองเป็นชั้นของคอนโวลูชันเลเยอร์มีจำนวนโหนด 250 โหนด รูปแบบฟังก์ชันเชิงเส้นในชั้นที่สามเป็นชั้นของพูลลิงเลเยอร์มีจำนวนโหนด 250 โหนด และชั้นที่สี่และห้ามีจำนวนโหนด 150 โหนดและ 6 โหนดตามลำดับ และเมื่อทำการเรียนรู้โมเดลได้มีการเพิ่มฟังก์ชันเพื่อหยุดการเรียนรู้ เมื่อค่าความผิดพลาดสูงขึ้น โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.17 เป็นผลลัพธ์ค่าความผิดพลาดและภาพ 5.18 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ CNN โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติการณ์เท่ากับ 93.24 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติการณ์ผิดพลาดเท่ากับ 6.76 %

5.3.3 LSTM

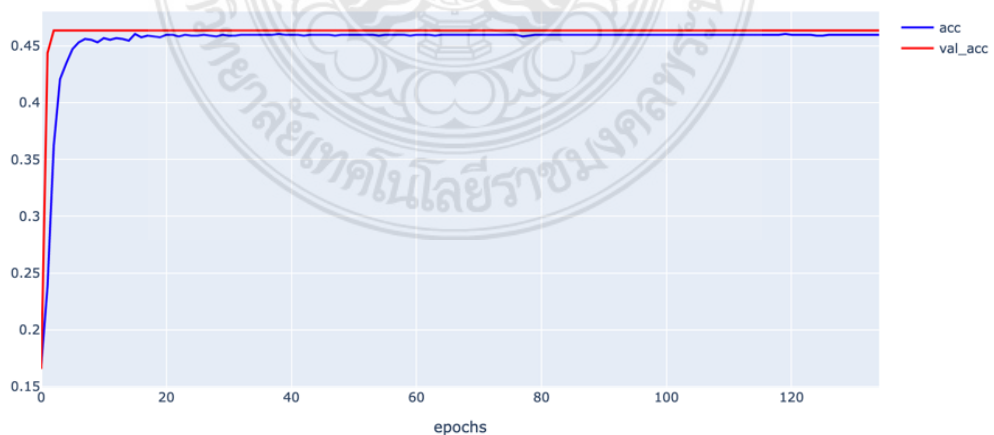
ในการทดสอบการระบุความรุนแรงของอุบัติการณ์วิธีการ LSTM กำหนดให้วิธีการมีรายละเอียดในการทดสอบที่มีประสิทธิภาพที่เหมาะสมต่อการแก้ปัญหาที่สุด ดังนี้ จำนวนชั้นประมวลผลภายในในชั้นแรกมีจำนวน 250 โหนดลำดับชั้นที่สองซึ่งเป็นชั้นของ Bi-LSTM มีจำนวนโหนดทั้งสิ้น 100 โหนด และเพิ่มฟังก์ชันสุ่มลดจำนวนสมาชิกที่อาจไม่ถูกใช้งาน 20 % ในแต่ละชั้น ในลำดับชั้นที่สามมีจำนวนโหนดจำนวน 50 โหนด รูปแบบฟังก์ชันเชิงเส้นและสี่เป็นชั้นเพื่อรวมค่าเป็นผลลัพธ์และ 6 โหนดตามลำดับ และเมื่อทำการเรียนรู้โมเดลได้มีการเพิ่มฟังก์ชันเพื่อหยุดการเรียนรู้ เมื่อค่าความผิดพลาดสูงขึ้น โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.19 เป็นผลลัพธ์ค่าความผิดพลาดและภาพ 5.20 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ LSTM โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติการณ์เท่ากับ 77.92 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติการณ์ผิดพลาดเท่ากับ 22.08 %

5.3.4 LSTM + CNN

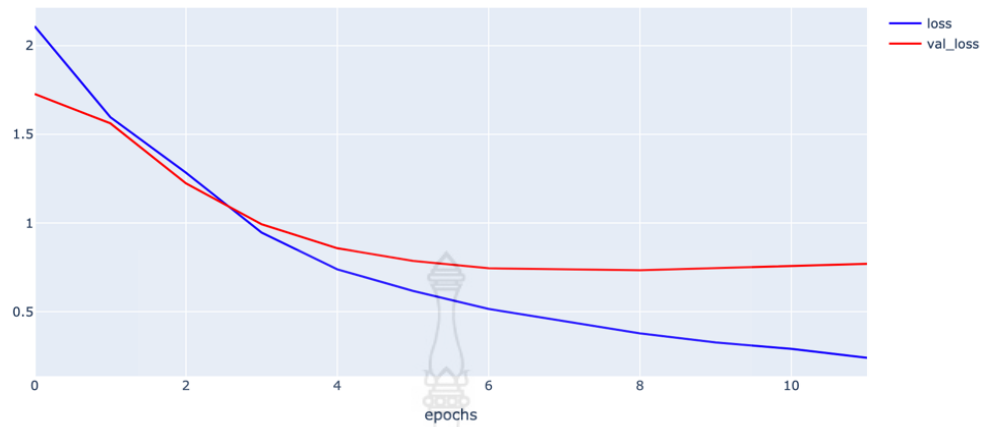
ในการทดสอบการระบุความรุนแรงของอุบัติเหตุการจราจร วิธีการ LSTM+CNN กำหนดให้วิธีการมีการทำงานร่วมกัน โดยการผ่านพารามิเตอร์จาก CNN ที่ได้จากคอนโวลูชันเลเยอร์และนำเข้าสู่ LSTM เลเยอร์ และตัดสินใจเลือกผลลัพธ์การทำนายด้วยชั้นซอฟต์แวร์แมชชีน สำหรับการส่งพารามิเตอร์เข้าสู่โมเดล โดยวิธีการที่นำเสนอจะมีความแตกต่างจากวิธีการ CNN และ LSTM ก่อนหน้านี้ ในขั้นแรกคือชั้นประมวลผลภายในที่ 32 ลำดับชั้น ในส่วนของคอนโวลูชันเลเยอร์จะมี 250 โหนด และนำผลลัพธ์ส่งเข้าสู่ LSTM เลเยอร์ โดยระบุพารามิเตอร์เข้าเป็น 32 ลำดับชั้น พร้อมกำหนดสุ่มลดจำนวนสมาชิกที่อาจไม่ถูกใช้งานที่ 20 % ในเลเยอร์สุดท้ายจะรวมผลลัพธ์ออกเป็นคำตอบโดยชั้นผลลัพธ์ โดยจะกำหนดผลลัพธ์เป็น 3 โหนด สำหรับข้อความความรุนแรง โดยมีผลการเรียนรู้และปรับปรุงประสิทธิภาพดังแสดงในภาพ 5.21 เป็นผลลัพธ์ทางความแม่นยำ ทั้งนี้ในการทดสอบกับข้อมูลจริง วิธีการ LSTM+CNN โดยมีความแม่นยำในการตรวจจับการเกิดอุบัติเหตุเท่ากับ 82.25 % และมีผลลัพธ์ในการตรวจจับการเกิดอุบัติเหตุผิดพลาดเท่ากับ 11.70 %



ภาพ 5.15 ความผิดพลาดการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการจราจรวิธี การ MPL



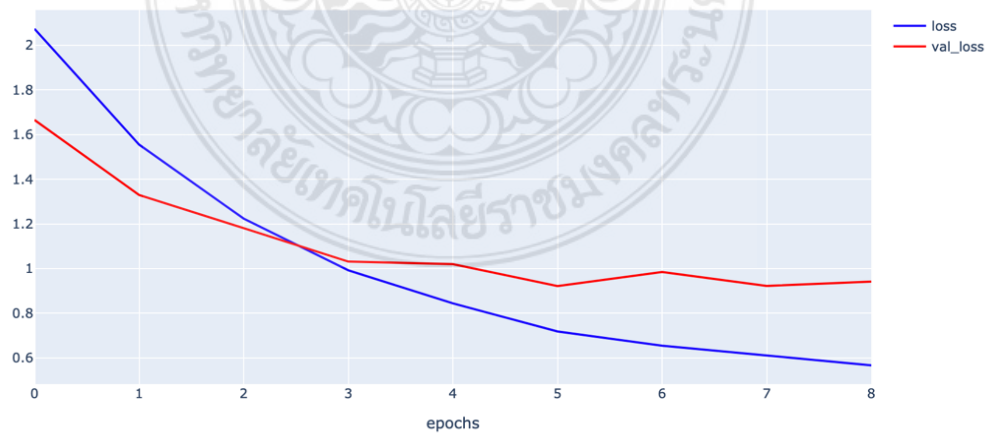
ภาพ 5.16 ความแม่นยำของการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการจราจรวิธี การ MPL



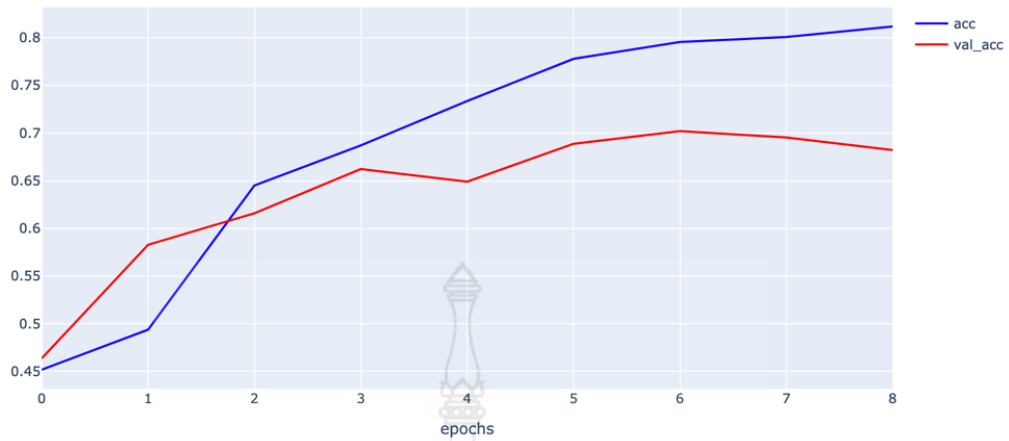
ภาพ 5.17 ความผิดพลาดการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการฉีกการ CNN



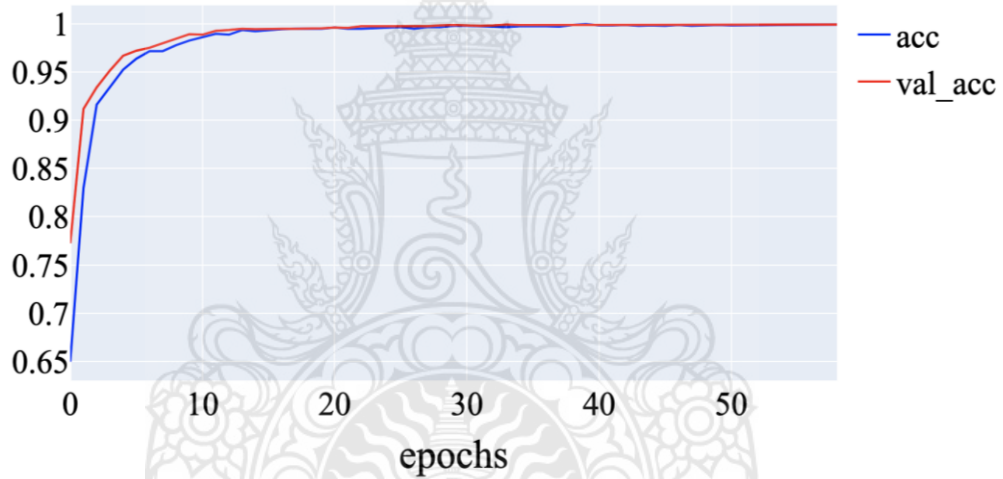
ภาพ 5.18 ความแม่นยำของการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการฉีกการ CNN



ภาพ 5.19 ความผิดพลาดการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการฉีกการ LSTM



ภาพ 5.20 ความแม่นยำของการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการฉีกวิธีการ LSTM



ภาพ 5.21 ความแม่นยำของการเรียนรู้การระบุความรุนแรงของอุบัติเหตุการฉีกวิธีการ LSTM+CNN

บทที่ 6

อภิปรายผล (Discussion)

ในการทดสอบผลจะสามารถบันทึกผลเพื่อทำการเปรียบเทียบประสิทธิภาพของรูปแบบการจำแนกกลุ่มแต่ละประเภทประกอบไปด้วย MLP, CNN, Bi-LSTM ในการจำแนกประเภทอุบัติเหตุ จำแนกประเภท รวมทั้งการระบุความรุนแรงเพื่อให้ทราบถึงความถูกต้องของการทำงานหลังจากการเรียนรู้เชิงลึกของแต่ละระบบและได้ทำการทดสอบบนพื้นฐานข้อความจากทวิตเตอร์โดยสำหรับการจำแนกข้อความการเกิดอุบัติเหตุข้อความจำนวน 1,010 ข้อความ และใน 1,010 ข้อความมีข้อความที่เป็นข้อความอุบัติเหตุ 503 ข้อความเพื่อนำมาทดสอบกับรูปแบบการจำแนกประเภทและรูปแบบการระบุความรุนแรงโดยผลลัพธ์ที่ทดสอบได้แสดงให้เห็นภาพรวมคือรูปแบบการทดสอบโมเดลที่ได้จาก CNN มีค่าเฉลี่ยอัตราความถูกต้องที่สูงที่สุดที่ 93.53 % แต่หากสังเกตจากการจำแนกการเกิดอุบัติเหตุจะเห็นว่ารูปแบบการจำแนกการเกิดอุบัติเหตุที่มีเพียง 2 กลุ่มจากโมเดล CNN+LSTM มีความถูกต้องที่สูงที่สุดคือ 94.06 % แต่เมื่อข้อมูลมีจำนวนกลุ่มที่ต้องจำแนกมากขึ้น (Multi Class) คือการจำแนกประเภทที่มีจำนวน 5 กลุ่ม ข้อมูลและการระบุความรุนแรงที่มี 3 กลุ่มข้อมูล จะสังเกตเห็นผลลัพธ์ของโมเดลที่ได้จาก CNN จะทำได้ดีกว่า อย่างไรก็ตามการเพิ่มประสิทธิภาพของความแม่นยำสามารถทำได้โดยการตัดคำที่ไม่มีความหมายออกและการเพิ่มข้อมูลให้มีความสมดุลกันที่สุดในทุกกลุ่มข้อมูลจะทำให้ความแม่นยำในการจำแนกกลุ่มของข้อความมีมากขึ้นดังตาราง 6.1

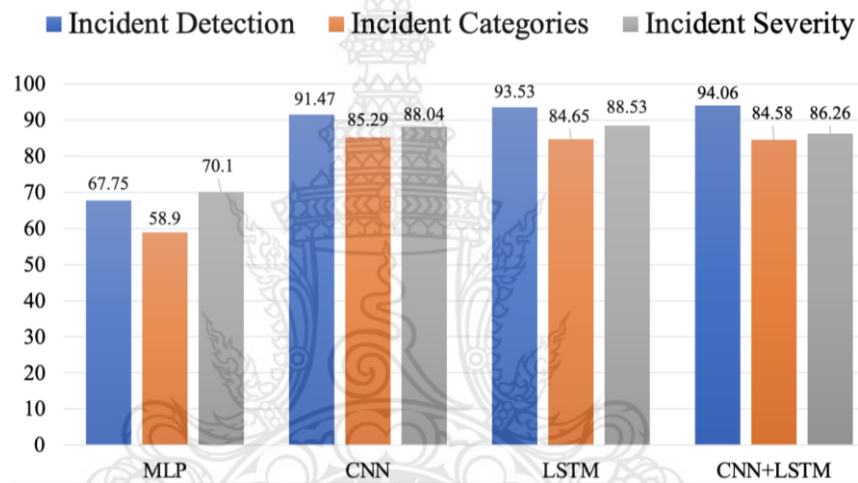
ตั้งข้อมูลผลการทดสอบการตรวจสอบอุบัติเหตุวิธีการต่าง ๆ สามารถแสดงการเปรียบเทียบค่าความถูกต้องของวิธีการในการตรวจสอบอุบัติเหตุลักษณะต่าง ๆ ดังแสดงในภาพ 6.1 ร่วมกับการแสดงค่าความผิดพลาดได้จากภาพ 6.2 ซึ่งในผลการทดสอบจะพบว่าเปรียบเทียบค่าความถูกต้องของวิธีการในการตรวจสอบอุบัติเหตุลักษณะต่าง ๆ ทั้ง 3 รูปแบบ รูปแบบ CNN+LSTM จะมีความสามารถในการจำแนกการเกิดอุบัติเหตุ และการระบุความรุนแรงได้ดีที่สุด และวิธีการ CNN และ MPL มีความถูกต้องลดลงตามลำดับ ในขณะที่เดียวกันวิธีการ CNN มีความสามารถในการจำแนกประเภทได้ดีที่สุดและวิธีการ Bi-LSTM และ MPL มีความถูกต้องลดลงตามลำดับเช่นเดียวกัน

ตาราง 6.1 ข้อมูลผลการทดสอบการตรวจสอบอุบัติเหตุวิธีการต่าง ๆ

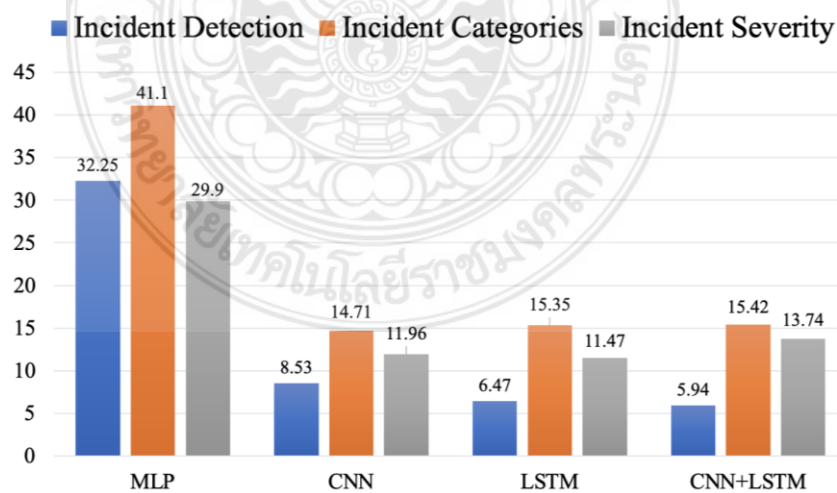
การตรวจสอบอุบัติเหตุ	DR				FAR			
	MLP	CNN	LSTM	CNN+LSTM	MLP	CNN	LSTM	CNN+LSTM
การจำแนกการเกิดอุบัติเหตุ	67.75	91.47	93.53	94.06	32.25	8.53	6.47	5.94
การจำแนกประเภท	58.90	85.29	84.65	84.58	41.10	14.71	15.35	15.42
การระบุความรุนแรง	70.10	88.04	88.53	86.26	29.90	11.96	11.47	13.74
ค่าเฉลี่ย (%)	65.58	88.27	88.90	88.30	34.42	11.73	11.10	11.70

ตาราง 6.2 ข้อมูลผลการทดสอบเพิ่มประสิทธิภาพของแบบจำลอง

รูปแบบแบบจำลอง	ข้อมูลพื้นฐาน	Stop Word	Marge Data	ปรับปรุงสองวิธี
LSTM	ประเภท	87.40	88.35	89.20
	ความรุนแรง	90.07	90.50	90.87
CNN	ประเภท	84.70	85.05	87.89
	ความรุนแรง	91.42	91.76	91.02
MLP	ประเภท	54.34	56.56	57.44
	ความรุนแรง	70.54	71.89	75.37
CNN+LSTM	ประเภท	85.99	86.25	87.20
	ความรุนแรง	86.26	87.28	91.85



ภาพ 6.1 การเปรียบเทียบค่าความถูกต้องของวิธีการในการตรวจสอบอุบัติการณ์



ภาพ 6.2 การเปรียบเทียบค่าความผิดพลาดของวิธีการในการตรวจสอบอุบัติการณ์

บทที่ 7

สรุปผล (Conclusion)

7.1 สรุปผล

การตรวจหาเหตุการณ์การจลาจลบนสื่อสังคมออนไลน์ทวิตเตอร์แพลตฟอร์ม โดยการตรวจจับจากข้อความที่ถูกทวิตจำนวนมาก ผ่าน Twitter API ในงานวิจัยนี้ จะดำเนินการรวบรวมข้อมูลข้อความตัวอย่างดังกล่าวเพื่อนำมาตรวจสอบและระบุรูปแบบพร้อมรายละเอียดต่าง ๆ โดยผู้ที่มีความเชี่ยวชาญในการจัดการจลาจล ร่วมกับวิธีการสำหรับเตรียมข้อมูลด้วย การกรองคำ การตัดคำ และการแปลงคำเป็นตัวเลข ก่อนนำเข้าสู่กระบวนการเรียนรู้เชิงลึก ด้วยข้อความจากสื่อสังคมออนไลน์ โดยการรวบรวมข้อมูลการประกาศของทวิตเตอร์ ซึ่งเป็นสื่อสังคมออนไลน์ที่ได้รับความนิยมเป็นอันดับต้น มีการส่งต่อข่าวสารในรูปแบบข้อความสั้น ผ่าน API ที่รองรับการทำงานแบบอัตโนมัติ เพื่อหาค่าที่มีความเกี่ยวข้องกับการตรวจจับอุบัติการณ์บนสื่อสังคมออนไลน์ แต่ด้วยข้อความบนทวิตเตอร์มีอยู่อย่างมหาศาล จะเป็นการยากในการตรวจหาเนื้อหาที่เกี่ยวข้องหรือที่สนใจอยู่เท่านั้น งานวิจัยนี้จึงนำเสนอการรวบรวมค่าที่เกี่ยวข้องกับข้อความที่ต้องการค้นหา ที่เกี่ยวข้องกับการตรวจจับอุบัติการณ์ บนวิธีการ TF-IDF, Word2Vec และ Markov Chain ด้วยกระบวนการทำงานหลัก 4 ขั้นตอน ผลการทดลองแสดงให้เห็นความสามารถในการประมวลผลตรวจจับการเกิดอุบัติการณ์ของวิธีการที่นำเสนอ พบว่าค่าที่มีความเกี่ยวข้องกับการเกิดอุบัติการณ์และมีประสิทธิภาพความแม่นยำในการตรวจจับข้อความการแจ้งอุบัติการณ์ในรูปแบบภาษาไทยคือ 5 คำแรกจากคำ 20 อันดับแรก โดยมีความแม่นยำของระบบเท่ากับ 85.80 % มีอัตราการตรวจพบ (DR) เท่ากับ 78.83 % และอัตราการแจ้งเตือนผิดพลาด (FAR) เท่ากับ 21.17 %

โดยในการวิจัยการจำแนกเหตุการณ์จะประกอบไปด้วย วิธีการ MLP, CNN, Bi-LSTM และ CNN+LSTM โดยใช้การเปรียบเทียบค่าความถูกต้องของทั้ง 4 วิธีการ สำหรับการตรวจหาเหตุการณ์การจลาจล การระบุประเภทของอุบัติการณ์ การระบุความรุนแรงของเหตุการณ์บน Twitter ภาษาไทย โดยผลการทดลองจะแสดงให้เห็นว่าการพัฒนาระบบตรวจจับการเกิดอุบัติการณ์ สามารถพิจารณาเลือกใช้รูปแบบวิธีการสำหรับการตรวจจับและแยกแยะรูปแบบของข้อมูลต่าง ๆ ได้อย่างมีประสิทธิภาพโดยการประมวลผลร่วมกันระหว่างวิธีการ โดยการเรียนรู้แบบ CNN+LSTM ที่มีผลลัพธ์ในการตรวจจับการเกิดอุบัติการณ์ได้ดีที่สุด ด้วยความแม่นยำ (DR) เท่ากับ 94.06 % และการแจ้งเตือนความผิดพลาด (FAR) เท่ากับ 5.94 % ในขณะที่การระบุรูปแบบของอุบัติการณ์โดยวิธีการ CNN ระบุรูปแบบได้ดีที่สุดถึง ด้วยความแม่นยำ (DR) เท่ากับ 85.29 % และการแจ้งเตือนความผิดพลาด (FAR) เท่ากับ 14.71 % และการระบุระดับความรุนแรงของอุบัติการณ์ โดยวิธีการ Bi-LSTM ระบุความรุนแรงได้ดีที่สุดถึง ด้วยความแม่นยำ (DR) เท่ากับ 88.53 % และการแจ้งเตือนความผิดพลาด (FAR) เท่ากับ 11.47 %

ทั้งนี้ผลสำเร็จของงานวิจัยดังกล่าวนี้ แสดงให้เห็นถึงแนวทางการพัฒนาระบบตรวจจับอุบัติการณ์จำแนกรูปแบบของอุบัติการณ์ที่เกิดขึ้น และระดับความรุนแรงของอุบัติการณ์ จากการใช้ข้อความที่

แบ่งปันผ่านแพลตฟอร์มสื่อสังคมออนไลน์ชื่อ Twitter เพื่อเป็นการทดแทนการลงทุนการติดตั้งอุปกรณ์ตรวจจับสภาพจราจร (Traffic Sensor) หรือกล้องประมวลผลภาพ (Image Processing Camera) เพื่อการตรวจจับการเกิดอุบัติเหตุอัตโนมัติในพื้นที่ที่อยู่ในระยะทำการ ซึ่งจะเป็นการลงทุนงบประมาณจำนวนมากถ้าจะต้องติดตั้งให้ครอบคลุมต่อพื้นที่ผิวจราจรทั้งหมดทั่วเขตพื้นที่กรุงเทพมหานคร โดยการตรวจจับข้อความการแจ้งเหตุอุบัติเหตุในงานวิจัยนี้นำเสนอกระบวนการแบ่งคำจากประโยคที่ได้รับมา และตรวจหาคำเฉพาะที่อยู่ในกลุ่มคำที่คาดว่าจะจะเป็นคำที่ให้ข้อมูลการแจ้งการเกิดอุบัติเหตุและรายละเอียดของเหตุการณ์ จากเทคนิคการตรวจสอบความถี่ของคำในรูปแบบ Word Cloud เพื่อเป็นคำที่ป้อนเข้าสู่กระบวนการตัดสินใจของ กระบวนการประมวลผลที่ได้รับจากรูปแบบการเรียนรู้เชิงลึก แบบผสมผสานร่วมกันระหว่างวิธีการ CNN, LSTM และ CNN+LSTM ดังผลการทดสอบ ซึ่งเป็นเทคนิคและวิธีการผสมผสานใหม่สำหรับการตรวจจับข้อความภาษาไทยในการแจ้งเหตุและแยกแยะรูปแบบการเกิดอุบัติเหตุ สำหรับการนำไปพัฒนาและปรับใช้ในระบบอัตโนมัติในอนาคตต่อไปได้

7.2 แนวทางการศึกษาและพัฒนาต่อ

การศึกษาวิจัยในอนาคต การเลือกรูปแบบของวิธีการเรียนรู้เชิงลึกสำหรับการวิเคราะห์และแยกแยะสถานการณ์ รูปแบบ และความรุนแรง ไม่มีวิธีการใดจะสามารถทำได้ดีที่สุดกับทุกเงื่อนไขที่ต้องการตรวจสอบ แต่ผู้วิจัยต้องเลือกใช้รูปแบบของวิธีการเรียนรู้เชิงลึกที่แตกต่างกันไปเพื่อการแก้ไขปัญหาการทำงานคนละปัญหา ในรูปแบบแยกการประมวลผลการทำงานต่างรูปแบบข้อมูลและสถานการณ์นั้น ๆ ซึ่งการเปลี่ยนแปลงบางเงื่อนไข สามารถส่งผลต่อผลลัพธ์การทำงานได้ทุกความเป็นไป เพื่อให้การศึกษาวิจัยเชิงลึกในโจทย์งานวิจัยนี้มีความแม่นยำและยืดหยุ่นต่อการนำไปพัฒนาใช้งานในอนาคต ผู้วิจัยควรทดสอบกับข้อมูลจากแหล่งที่มา และหลากหลายช่องทางมากขึ้นผลลัพธ์งานวิจัยที่ได้ดังกล่าว จะสามารถชี้แนะแนวทางความเชื่อมั่นได้มากยิ่งขึ้นต่อไป

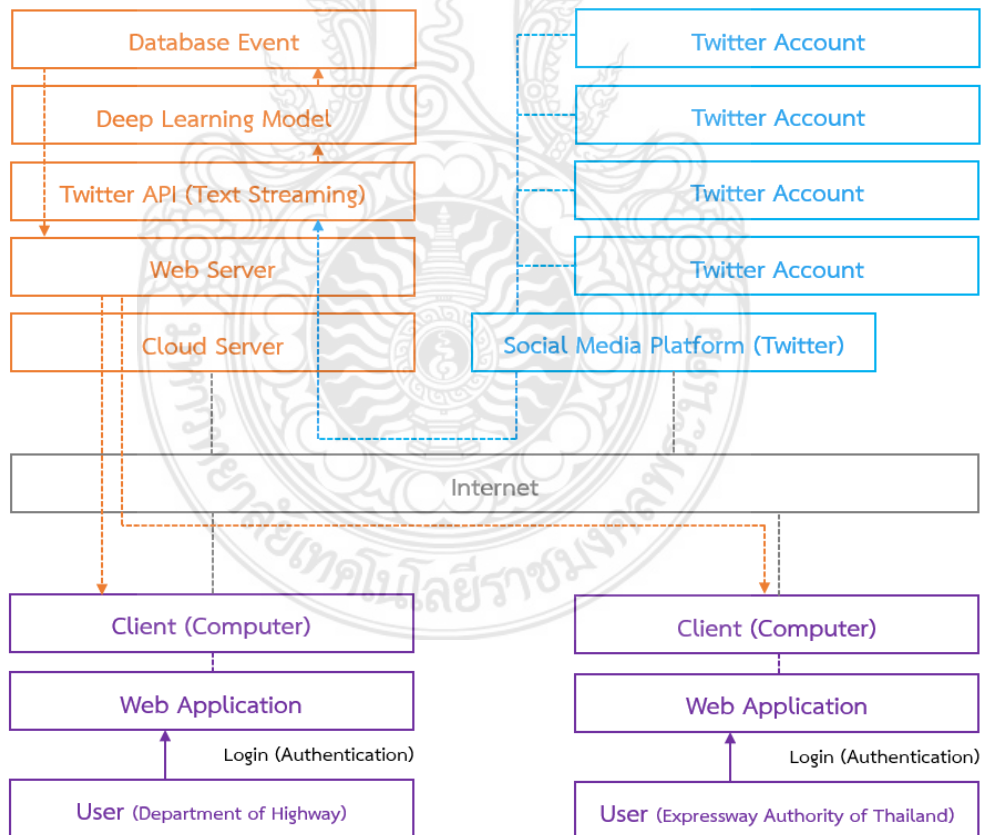
ทั้งนี้งานวิจัยนี้นำเสนอการตรวจจับการเกิดอุบัติเหตุและการจำแนกรูปแบบและความรุนแรงของอุบัติเหตุด้วยการประมวลผลข้อมูลข้อความ (Content) ที่มีการแบ่งปันทวีต ในขณะที่การแบ่งปันดังกล่าว สามารถแนบไฟล์ภาพ (Image) หรือภาพเคลื่อนไหว (Video) ซึ่งข้อมูลแนบดังกล่าวสามารถนำมาใช้ในการตรวจสอบร่วมกันกับการตรวจสอบจากข้อความได้ โดยการประยุกต์ใช้การเรียนรู้เชิงลึกแบบ CNN เข้ามาร่วมเรียนรู้ เป็นต้น

บทที่ 8

แผนการนำไปใช้ประโยชน์เชิงพาณิชย์ (Commercialization Plan)

8.1 รูปแบบสำคัญเพื่อดำเนินการทางธุรกิจ (Business Model)

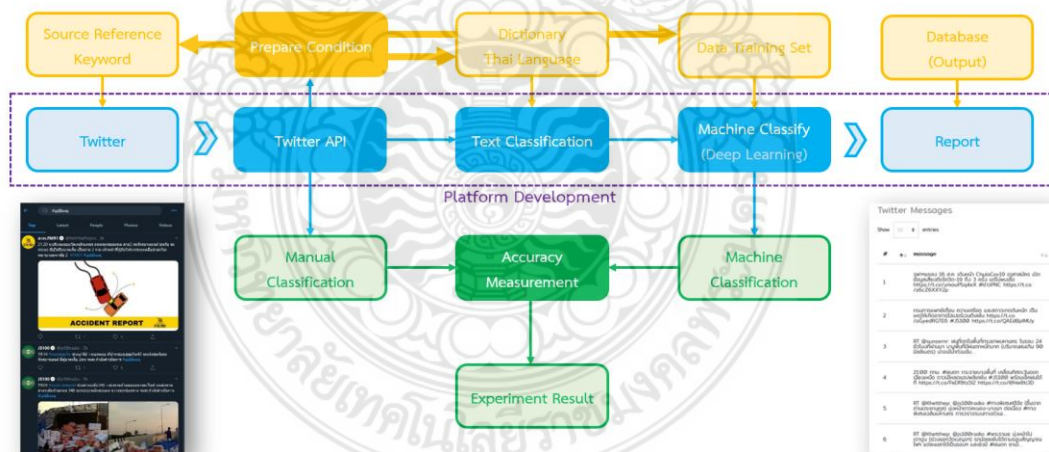
ในการศึกษาวิจัยงานวิจัยนี้ มีวัตถุประสงค์ต้องการศึกษาออกแบบกระบวนการประมวลผลข้อมูลภาษาไทยในระบบทวิตเตอร์เพื่อจำแนกข้อความการแจ้งอุบัติเหตุการณบนถนน สำหรับการส่งข้อความดังกล่าวต่อให้กลับเจ้าหน้าที่หรือผู้ที่มีหน้าที่อำนวยความสะดวกหรือบริหารจัดการจราจรต่อไป เพื่อเป็นการทดแทนการติดตั้งอุปกรณ์ตรวจวัดสภาพจราจรหรือกล้องตรวจจับเหตุการณ์ที่มีระยะและขอบเขตพื้นที่ในการทำงาน โดยกระบวนการที่ถูกออกแบบจะถูกพัฒนาขึ้นอยู่ในรูปแบบเว็บแอปพลิเคชัน บนเครื่องแม่ข่ายที่ผู้วิจัยได้เช่าใช้บริการ และทำการจำหน่ายการให้บริการแบบเก็บค่าบริการใช้งานระบบแบบรายปี (Yearly Subscription) พร้อมการให้สิทธิการเข้าใช้งานผ่านชื่อประจำตัวหน่วยงานต่อไป



ภาพ 8.1 สถาปัตยกรรมระบบและการดำเนินการทางธุรกิจ

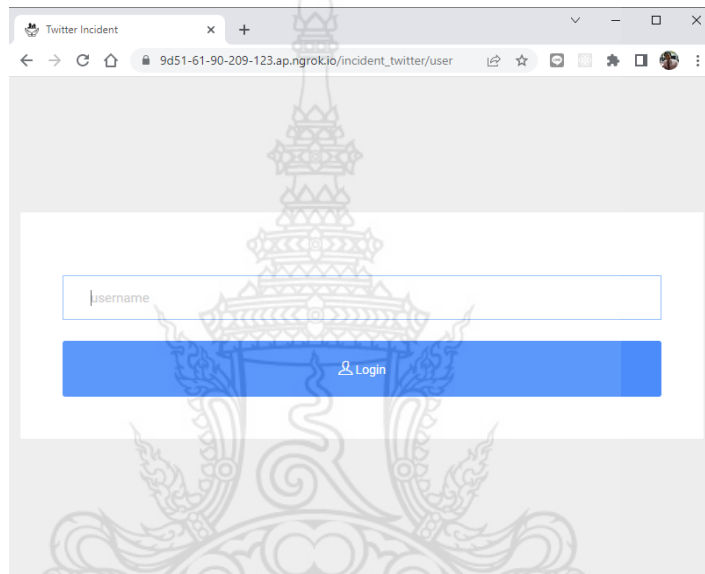
ระบบการให้บริการดังกล่าว จะถูกพัฒนาขึ้นในรูปแบบเว็บแอปพลิเคชัน ทำงานร่วมกับชุดคำสั่งเฉพาะที่มีการทำงานแบบตลอดเวลาบนเครื่องแม่ข่ายให้บริการด้วยภาษา Python ซึ่งจะประกอบไปด้วยการทำงานสำคัญ 3 ขั้นตอน กระบวนการเชื่อมโยงข้อมูลข้อความจากทวิตเตอร์ผ่าน Twitter API ทำหน้าแบบเบื้องหลัง (Background Process) เสมือนมีเจ้าหน้าที่ทำงานดึงข้อความที่ถูกทวิตใหม่เข้าสู่ระบบคัดกรองแบบอัตโนมัติ กระบวนการแยกประโยคเป็นคำ (Text Classification) เพื่อนำเข้าสู่กระบวนการตรวจสอบคำว่าเป็นคำที่อยู่ในกลุ่มคำที่ปรากฏในประโยคแจ้งเหตุอยู่บ่อย ๆ หรือไม่ เพื่อใช้เป็นคำที่บ่งชี้เข้ากระบวนการแยกแยะโดยสมองกล (Machine Classify) โดยระบบการเรียนรู้เชิงลึก ในรูปแบบต่าง ๆ ตามที่ได้รับการเลือกใช้และปรับประสิทธิภาพจนมีประสิทธิภาพสูงสุด ดังแสดงในภาพ 8.2

ในการให้บริการผู้สมัครใช้บริการระบบรวบรวมข่าวอุบัติเหตุจากช่องทางทวิตเตอร์อัตโนมัติ นั้น จะได้อัปเดตการใช้งาน (User) เพื่อใช้ในการยืนยันตัวตนเข้าใช้บริการ ดังแสดงในภาพ 8.3 ผ่านการเรียกใช้บริการบนเว็บแอปพลิเคชันไปยัง URL ที่ระบบแม่ข่ายให้บริการที่ผู้พัฒนาระบบจดทะเบียนไว้ สำหรับการตรวจสอบสิทธิการใช้งานและขอบเขตพื้นที่ที่หน่วยงานต้องการรับทราบและตรวจสอบ ทั้งนี้ระบบดังกล่าวที่ถูกพัฒนาขึ้น มีความสามารถรองรับในการรวบรวมข้อมูลข่าวสารการจราจรได้ทั่วประเทศในประเทศไทย แต่ในงานวิจัยนี้จะกำหนดขอบเขตการรวบรวมข้อมูลเฉพาะในเขตพื้นที่กรุงเทพมหานครและปริมณฑลเพื่อกำหนดให้มีปริมาณข้อมูลที่ต้องประมวลผลอยู่ในระดับที่สามารถประมวลผลและให้บริการได้



ภาพ 8.2 กระบวนการให้บริการระบบรวบรวมข่าวอุบัติเหตุจากช่องทางทวิตเตอร์อัตโนมัติ

หลังจากผู้ใช้งานทำการ Login เข้าใช้งานระบบจากหน้าจอตรวจสอบสิทธิ์ ระบบจะแสดงข้อมูลรายการข้อความที่ผ่านกระบวนการตรวจจับเหตุการณ์ทั้งหมด ทั้งข้อความที่เป็นการประกาศข่าวประชาสัมพันธ์ทั่วไป และการแจ้งเหตุอุบัติการณ์ ซึ่งผู้ใช้งานสามารถตรวจสอบ โดยการคัดกรองให้แสดงผลเฉพาะรายการที่เป็นอุบัติการณ์เท่านั้นได้ โดยเหตุการณ์ที่เป็นอุบัติการณ์จะมีรายละเอียดรูปแบบและความรุนแรงของอุบัติการณ์แนบในส่วนตัวย่อ พร้อมให้เจ้าหน้าที่ทำการยืนยันผลการตรวจสอบว่าข้อความที่รายงานเป็น เหตุการณ์อุบัติการณ์จริงหรือไม่ เพื่อเก็บบันทึกข้อมูลไว้สำหรับการพัฒนาเรียนรู้ได้ในอนาคตและเป็นการรับทราบแจ้งเตือน



ภาพ 8.3 การตรวจสอบสิทธิ์การเข้าใช้งานโดยการระบุ User Name

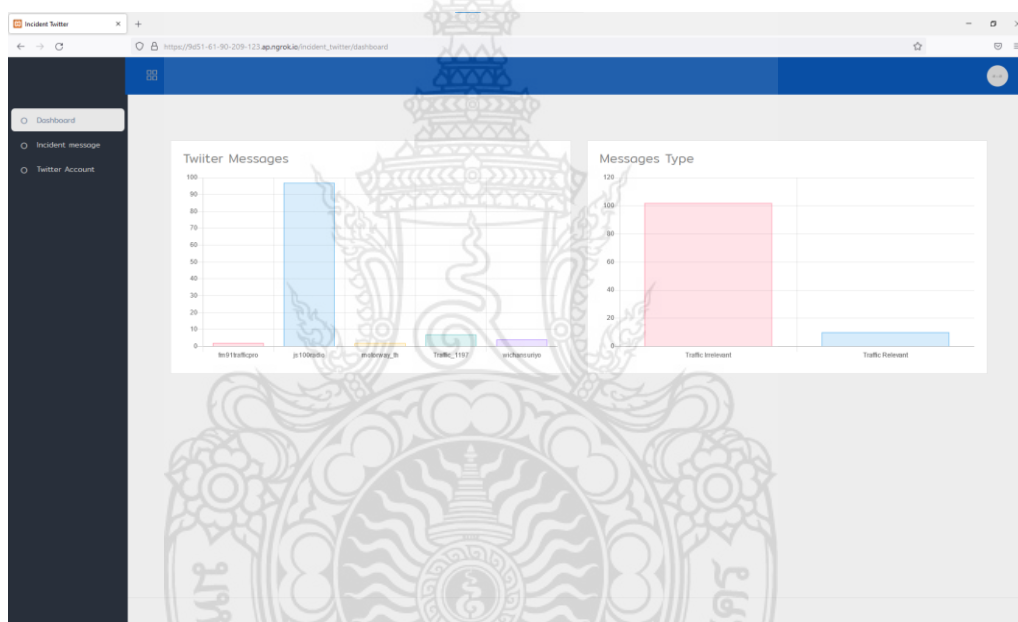
The image shows a dashboard with a sidebar on the left containing 'Dashboard', 'Incident message', and 'Twitter Account'. The main area is titled 'Twitter Messages' and contains a table with 8 rows of incident reports. The table has columns for message content, datetime, screen name, situation, severity, detail, status (Approved), and actions (View).

#	message	datetime	screen_name	situation	severity	detail	Status	Actions
1	แจ้งข่าว 15 ส.ค. เกิดอุบัติเหตุ 19... https://t.co/mn59g4xk... /sCZDjK7Fp	2021-08-12 21:34:23	js100radio	0	0	0	Approved	View
2	กรมการขนส่งทางบก ขอแจ้งด้วย... https://t.co/vv9o295q8y	2021-08-12 21:30:19	js100radio	0	0	0	Approved	View
3	RT @supornsa: แจ้งข่าวดังกล่าว... https://t.co/nW90Uw...	2021-08-12 21:22:22	js100radio	0	0	0	Approved	View
4	21.00 ทน... https://t.co/4e2R8t52...	2021-08-12 21:00:00	js100radio	0	0	0	Approved	View
5	RT @tstetap @js100radio... https://t.co/...	2021-08-12 20:51:10	js100radio	0	0	0	Approved	View
6	RT @tstetap @js100radio... https://t.co/...	2021-08-12 20:49:16	js100radio	1	1	1	Approved	View
7	20.35... https://t.co/...	2021-08-12 20:35:17	js100radio	1	2	2	Approved	View
8	แนวทางการดำเนินงาน... https://t.co/...	2021-08-12 20:30:09	js100radio	0	0	0	Approved	View

ภาพ 8.4 การรายงานข้อความที่ผ่านการตรวจสอบการตรวจจับอุบัติการณ์

ในส่วนสุดท้ายผู้ใช้งานจะสามารถดูข้อมูลสรุปผลการทำงานได้ที่หน้าจอแสดงข้อมูลสรุป (Dashboard) โดยจะประกอบไปด้วยกราฟข้อมูล 2 กราฟ โดยกราฟที่ 1 ดังแสดงข้อมูลด้านซ้ายของภาพที่ 8.5 จะแสดงจำนวนข้อมูลเหตุการณ์การแจ้งเตือนจากช่องทางหรือบัญชีต่าง ๆ ที่มีการเชื่อมโยง เพื่อให้เห็นความเกี่ยวข้องและความเคลื่อนไหวของแหล่งที่มาของข้อมูล และกราฟที่ 2 ดังแสดงข้อมูลด้านขวาของภาพที่ 8.5 จะเป็นการแสดงสรุประหว่างจำนวนข้อความการแจ้งข้อมูลข่าวสารทั่วไปเทียบกับจำนวนข้อความการแจ้งเหตุอุบัติการณ์ทั้งหมดของทุกช่องทางทุกบัญชี

ด้วยระบบงานที่ถูกพัฒนาขึ้นถูกติดตั้งบนเครื่องแม่ข่ายเสมือน (Virtual Private Server: VPS) พร้อมหมายเลข IP แบบ Public IP ทำให้ผู้ใช้บริการภายนอกสามารถเรียกใช้งานระบบได้โดยการอ้างถึงหมายเลขดังกล่าวผ่านเว็บเบราว์เซอร์ (Web Browser) บนระบบเครือข่ายอินเทอร์เน็ตได้ทันที โดยมีคุณสมบัติ ดังแสดงในตาราง 8.1



ภาพ 8.5 การรายงานสรุปจำนวนข้อความแต่ละช่องทางและจำนวนเหตุการณ์

ตาราง 8.1 คุณสมบัติของเครื่องแม่ข่าย VPS สำหรับการติดตั้งระบบและให้บริการ

ลำดับ	รายการ	รายละเอียด
1	หน่วยประมวลผล (CPU)	8 Core 2.6 GHz
2	หน่วยความจำสำรอง (RAM)	32 Gbyte
3	หน่วยความจำหลัก (Disk)	700 Gbyte (SSD)
4	ความเร็วการสื่อสาร (Bandwidth)	Domestic 1 Gbps / International 200 Mbps
5	ระบบปฏิบัติการ (Operating System)	Microsoft Window Server 2019 (STD)
6	ฐานข้อมูล (Database System)	MySQL
7	เว็บเซิร์ฟเวอร์ (Web Server)	IIS

8.2 แผนการใช้ประโยชน์ของผลิตภัณฑ์ในเชิงธุรกิจ (Business Plan)

8.2.1 แผนกลยุทธ์

ดำเนินการจัดทำระบบงานในการตรวจจับอุบัติเหตุทางบก ในรูปแบบเว็บแอปพลิเคชันที่ง่ายต่อการใช้งาน และใช้งบประมาณในการลงทุนที่ต่ำกว่า การจัดตั้งโครงการลงทุนของภาครัฐแบบเดิม ด้วยการอาศัยการเติบโตของข้อมูลในสื่อสังคมออนไลน์ ร่วมกับกระบวนการคิดเชิงนวัตกรรมในรูปแบบระบบง่าย บนการจัดซื้อจัดจ้างแบบ G2G กับหน่วยงานภาครัฐที่เกี่ยวข้อง ตามนโยบายการสนับสนุนธุรกิจด้านนวัตกรรมของคนไทย

8.2.2 เป้าหมาย

- 1) สามารถพัฒนาและเปิดให้ระบบงานที่ทำหน้าที่รวบรวมและตรวจจับข้อความการแจ้งเหตุอุบัติเหตุอัตโนมัติ สำหรับเหตุการณ์ที่เกี่ยวข้องกับการขนส่งและจราจรทางบกได้
- 2) สามารถทดแทนการลงทุนการติดตั้งอุปกรณ์และระบบงานที่ใช้ในการเฝ้าระวังหรือติดตามเหตุอุบัติเหตุที่มีการลงทุนด้วยประมาณที่สูงได้ โดยการให้ระบบงานที่ถูกพัฒนาขึ้นจากงานวิจัยนี้ในรูปแบบการจัดซื้อจัดจ้างระบบงานระหว่างภาครัฐ
- 3) สามารถให้บริการระบบงานได้อย่างต่อเนื่องโดยการรับค่าบำรุงรักษาระบบงานจากผู้ใช้บริการระบบงาน ทดแทนงบประมาณการลงทุนบำรุงรักษาอุปกรณ์ระบบงานเดิมที่มีมูลค่าสูง

8.2.3 การทบทวนเพื่อการใช้ประโยชน์

เมื่อพิจารณาทบทวนการใช้ประโยชน์จากระบบงานที่ถูกพัฒนาขึ้นจากงานวิจัยนี้ จะสามารถดำเนินการจัดทำแผนการใช้ประโยชน์ได้ 3 ระยะ ดังนี้

8.2.3.1 ระยะสั้น เป็นการดำเนินการกระตุ้นและส่งเสริมให้เปลี่ยนแปลงการลงทุนภาครัฐให้มีความเข้าใจและทดลองใช้ เพื่อเริ่มต้นและเปรียบเทียบผลลัพธ์เชิงประสิทธิภาพกับมูลค่าลงทุน ให้เห็นความคุ้มค่ากับทุกหน่วยงานเป้าหมายในช่วง 1 – 2 ปีแรกที่เริ่มให้บริการระบบงาน

8.2.3.2 ระยะกลาง เป็นการพัฒนาปรับปรุงประสิทธิภาพระบบ ในช่วงเวลา 3 – 5 ปีถัดมา โดยผู้พัฒนาระบบต้องมีการเชื่อมโยงบัญชีผู้ใช้งานอย่างหลากหลายและครอบคลุมมากขึ้น เพื่อให้หน่วยงานผู้ใช้งานได้รับข้อมูลอย่างครอบคลุมพื้นที่มากที่สุด

8.2.3.3 ระยะยาว เป็นการดำเนินงานในระยะเวลาหลังการดำเนินงานกิจกรรมโครงการดังกล่าวมาอย่างต่อเนื่องไม่น้อยกว่า 5 ปี ซึ่งจะเป็นช่วงเวลาที่ระบบสามารถทำงานได้อย่างต่อเนื่องและอัตโนมัติได้อย่างถูกต้องและเชื่อถือได้ ผู้ใช้งานระบบสามารถลดต้นทุนโดยการรวมศูนย์ข้อมูล และเจ้าหน้าที่บริหารจัดการลงให้เหลือเพียงศูนย์ประสานงานกลางอย่างยั่งยืนในพื้นที่ส่วนกลางจุดเดียวได้

8.2.4 แผนการใช้งานอย่างยั่งยืน

ด้วยระบบงานนี้เป็นระบบงานที่ต้องอาศัยแหล่งข้อมูลต้นทางจากการแจ้งเหตุโดยผู้ใช้งาน ดังนั้นความยั่งยืนของระบบงานจะสัมพันธ์กับความน่าเชื่อถือ ความรวดเร็ว และความร่วมมือในการให้

ข้อมูลข่าวสารจากประชาชนผู้ใช้งานหรือผู้ประสบเหตุโดยตรง ดังนั้นผู้ใช้งานระบบหรือหน่วยงานต้นสังกัดเจ้าของระบบต้องมีการประชาสัมพันธ์ถึงประโยชน์ที่เกิดขึ้นเพื่อสร้างกระแสและความตระหนักรู้ทางสังคม ประกอบกับการเชิญชวนเข้าร่วมโดยการมอบรางวัลหรือเกียรติบัตรสำหรับผู้มีส่วนร่วมในรูปแบบต่าง ๆ เช่น เป็นผู้แจ้งเหตุที่ถูกต้องมากที่สุดในปีนั้น เป็นผู้แจ้งเหตุได้อย่างรวดเร็วที่สุด เป็นต้น เพื่อให้เกิดความร่วมมือกับภาคประชาชนที่เป็นข้อมูลต้นทางอย่างยั่งยืนต่อไป

8.3 การวิเคราะห์ผลตอบแทนทางเศรษฐศาสตร์

การวิเคราะห์ผลตอบแทนทางเศรษฐศาสตร์ในงานวิจัยนี้จะเน้นจากการสรุปเงินลงทุนเปรียบเทียบกับระหว่างการใช้เทคโนโลยีที่มีในปัจจุบัน กับการลงทุนเข้าใช้บริการจากระบบที่ถูกพัฒนาขึ้น ทั้งนี้การประเมินดังกล่าวจะประกอบไปด้วยการพิจารณา 2 ส่วน คือ ส่วนที่หนึ่ง เป็นส่วนต้นทุนในการพัฒนาและจัดเตรียมโครงสร้างพื้นฐานในการให้บริการระบบ เพื่อใช้ในการประเมินต้นทุนราคาการให้บริการ และส่วนที่สอง เป็นส่วนการประเมินเงินลงทุนโครงการตามกรอบแนวทางการพัฒนาและติดตั้งระบบในปัจจุบัน เปรียบเทียบกับการลงทุนเข้าใช้บริการระบบงานที่ถูกพัฒนาขึ้น โดยมีข้อมูลและรายละเอียดดังนี้

ตาราง 8.2 รายละเอียดต้นทุนการพัฒนาและจัดเตรียมโครงสร้างพื้นฐานในการให้บริการระบบต่อปี

ลำดับ	รายการ	ต่อหน่วย	จำนวน	หน่วย	มูลค่ารวม
1	ต้นทุนในการลงทุนครั้งแรก (CAPEX)				1,320,000
1.1	ค่านักวิเคราะห์และออกแบบระบบ (SA) (1 คน คนละ 60,000 บาท ทำงาน 8 ชม.)	60,000	6	เดือน	360,000
1.2	ค่านักพัฒนาปรับปรุงระบบ (Programmer) (2 คน คนละ 40,000 บาท ทำงาน 8 ชม.)	80,000	6	เดือน	960,000
2	ต้นทุนดำเนินงาน (OPEX)				227,190
2.1	ค่าเช่าใช้บริการคอมพิวเตอร์แม่ข่าย (VPS)	3,000	12	เดือน	32,400
2.2	ค่าลิขสิทธิ์ระบบปฏิบัติการ (OS)	900	12	เดือน	10,800
2.3	ค่าบริการชื่อโดเมนเว็บ (Domain Name)	390	1	ปี	390
2.4	ค่าลิขสิทธิ์ฐานข้อมูล (DBMS: MySQL)	-	1	ปี	-
2.5	ค่าลิขสิทธิ์เว็บเซิร์ฟเวอร์ (Web Server: IIS)	-	1	ปี	-
2.6	ค่าเจ้าหน้าที่บำรุงรักษาระบบ	15,000	12	เดือน	180,000
	ราคาจำหน่ายระบบ ในปีที่ 1 (1) + (2)	1,547,190	1.5	เท่า	2,320,785
	ราคาจำหน่ายในปีที่ 1 รวมภาษี	2,320,785	7	%	2,483,240
	ราคาบำรุงรักษาในปีที่ 2 เป็นต้นไป	227,190	1.5	เท่า	340,785
	ราคาบำรุงรักษาในปีที่ 2 รวมภาษี	340,785	7	%	364,640

ตาราง 8.3 การเปรียบเทียบการลงทุนระหว่างการติดตั้งอุปกรณ์แบบเดิมกับการเข้าใช้ระบบงานใหม่

ลำดับ	รายการ	งบประมาณ (บาท)	
		ระบบงานเดิม	ระบบงานนำเสนอ
1	งบดำเนินงานโครงการ ระยะทาง 20 กม. *	15,555,232	2,483,240
2	งบบำรุงรักษา (เฉลี่ยต่อปีจากค่าใช้จ่าย 5 ปี) **	1,306,639	364,640

หมายเหตุ * อ้างอิงโครงการจัดซื้อจัดจ้างภาครัฐ (หน่วยงานด้านการคมนาคมขนส่ง) ที่เทียบเคียงในปีงบประมาณ 2564
 ** ประเมินงบประมาณตามมาตรฐาน ICT ค่าใช้จ่ายบำรุงรักษา อัตราก้าวหน้า 7 %, 8 %, 9 % ต่อปีจากงบประมาณ

จากตาราง 8.2 และ 8.3 จะแสดงให้เห็นที่มาของราคาต้นทุนและราคาในการให้บริการระบบงานเพื่อนำมาเปรียบเทียบกับการลงทุนโครงการในลักษณะจัดหาและพัฒนาอุปกรณ์มาติดตั้งและบำรุงรักษาระบบงานตามรูปแบบการดำเนินงานในปัจจุบันกับการเข้าใช้บริการจากระบบงานที่ถูกพัฒนาขึ้นจากงานวิจัยดังกล่าว แสดงให้เห็นว่า การเข้าใช้ระบบงานที่นำเสนอ หน่วยงานสามารถลดงบประมาณการลงทุนเป็นเงินจำนวน 13,071,992 บาท หรือคิดเป็น 84.04 % จากวงเงินงบประมาณการลงทุนเดิม และสามารถลดงบประมาณบำรุงรักษาได้เฉลี่ยปีงบประมาณละ 942,000 บาท หรือคิดเป็น 72.09 % จากวงเงินงบประมาณเดิม ซึ่งสามารถสรุปได้ว่า ระบบงานที่ถูกพัฒนาขึ้นจากงานวิจัยนี้ จะสามารถช่วยลดงบประมาณการลงทุนโครงการได้



เอกสารอ้างอิง

- ชิตพงษ์ กิตตินราดร. 2563. **Neural Network Algorithm**. [ออนไลน์] เข้าถึงได้จาก : <https://guopai.github.io/ml-blog/14.html>, 9 กันยายน 2564.
- ราชบัณฑิตยสถาน. 2565. **คลังคำศัพท์ร่วมกับพจนานุกรม**. [ออนไลน์] เข้าถึงได้จาก : <https://dictionary.orst.go.th/>, 9 กันยายน 2564.
- ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. 2565. **คลังข้อความภาษาไทยหมวดรายการสารานุกรม**. [ออนไลน์] เข้าถึงได้จาก : <https://www.nectec.or.th/corpus/>, 9 กันยายน 2564.
- Abdallaoui, H. A., Fazziki, A., Ennaji, F. Z. and Sadgal, M. 2018. “Decision Support System for the Analysis of Traffic Accident Big Data.” **14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)**. Las Palmas de Gran Canaria, Spain : 514 – 521.
- Agarwal, A. and Toshniwal, D. 2019. “Face off: Travel Habits, Road Conditions and Traffic City Characteristics Bared Using Twitter.” **IEEE Access**. 7 (May) : 66536 – 66552.
- Aziz, M. V. G., Prihatmanto, A. S., Henriyan, D. and Wijaya, R. 2015. “Design and Implementation of Natural Language Processing with Syntax and Semantic Analysis for Extract Traffic Conditions from Social Media Data.” **5th IEEE International Conference on System Engineering and Technology (ICSET)**. Shah Alam, Malaysia : 43 – 48.
- Balamurugan, S. A. A., Lilian, J. S. and Sasikala, S. 2018. “The Future of India Creeping up in Building a Smart City: Intelligent Traffic Analysis Platform.” **Second International Conference on Inventive Communication and Computational Technologies (ICICCT)**. Coimbatore, India : 518 – 522.
- Buroni, G., Borgne, Y. L., Borgne, Y. L. and Determe, K. 2018. “Cluster Analysis of On-Board-Unit Truck Big Data from the Brussels Capital Region.” **21st International Conference on Intelligent Transportation Systems (ITSC)**. Maui, HI., USA : 2074 – 2079.
- Chen, Y., Lv, Y., Wang, X., Li, L. and Wang, F. 2019. “Detecting Traffic Information from Social Media Texts with Deep Learning Approaches.” **IEEE Transactions on Intelligent Transportation Systems**. 20, 8 (August) : 3049 – 3058.
- Chung, D., Rui, X., Min, D. and Yeo, H. 2013. “Road Traffic Big Data Collision Analysis Processing Framework.” **7th International Conference on Application of Information and Communication Technologies**. Baku, Azerbaijan : 1 – 4.

- Ciociola, A., Cocca, M., Giordano, D., Mellia, M., Morichetta, A., Putina, A. and Salutari, F. 2017. "UMAP: Urban Mobility Analysis Platform to Harvest Car Sharing Data." **IEEE SmartWorld, Ubiquitous Intelligence & Computing**, Francisco. CA. USA : 1 - 8.
- Cochran, W. G. 1963. **Sampling Technique**. 2nd Edition. John Wiley and Sons Inc., New York.
- Cui, Z., Zhang, S., Henrickson, K. C. and Wang, Y. 2016. "New Progress of DRIVE Net: An E-Science Transportation Platform for Data Sharing, Visualization, Modeling, and Analysis." **IEEE International Smart Cities Conference (ISC2)**. Trento. Italy : 1 – 2.
- Datareportal. 2020. **Digital 2020: Global Digital Overview**. [Online] Available : <https://datareportal.com/reports/digital-2020-global-digital-overview/>, September 9, 2021.
- Gao, L., Guo, Z., Zhang, H., Xu, X. and Shen, H. T. 2017. "Video captioning with attention-based LSTM and semantic consistency." **IEEE Transactions on Multimedia**. 19, 9 (July) : 2045 – 2055.
- Georgiou, T., El-Abadi, A., Yan, X. and George, J. 2015. "Mining Complaints for Traffic-Jam Estimation: A Social Sensor Application." **Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining**. Paris, France : 330 - 335.
- Graves, A. and Jaitly, N. 2014. "Towards end-to-end Speech Recognition with Recurrent Neural Networks." **Proceedings of the 31st International Conference on International Conference on Machine Learning**. 32, 2 : 1764 – 1772.
- Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. and Kingsbury, B. 2012. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups." **IEEE Signal Processing Magazine**. 29, 6 (October) : 82–97.
- Jia, X., Cheng, P. and Chen, J. 2016. "A Data Analysis and Visualization System for Large-Scale E-Bike Data." **IEEE International Conference on Big Data (Big Data)**. Washington, DC. USA : 3998 - 4000.
- Jin, F. and Liu, H. 2018. "Detect Hidden Road Hazards Combining Multiple Social Media Data." **IEEE International Conference on Big Data**. Seattle. WA. USA : 4559 - 4560
- Kaplunovich, A. and Yesha, Y. 2018. "Consolidating Billions of Taxi Rides with AWS EMR and Spark in the Cloud: Tuning, Analytics and Best Practices." **IEEE International Conference on Big Data**. Seattle, WA,. USA : 4501 - 4507.

- Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Li, F. 2014. "Large-Scale Video Classification with Convolutional Neural Networks." **2014 IEEE Conference on Computer Vision and Pattern Recognition**. Columbus, USA : 1725 - 1732.
- Kim, Y. 2014. "Convolutional Neural Networks for Sentence Classification." **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. Doha, Qatar : 1746–1751.
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." **Communications of the ACM**. 60, (June) : 84 – 90.
- Lee, H., Largman, Y., Pham, P. and Andrew Y. Ng, 2009. "Unsupervised Feature Learning for Audio Classification using Convolutional Deep Belief Networks." **23rd Annual Conference on Neural Information Processing Systems (NIPS)**. (December) : 1096–1104.
- Lin, B. and Tseng, S. 2017. "A Predictive Analysis of Citizen Hotlines 1999 and Traffic Accidents: A Case Study of Taoyuan City." **IEEE International Conference on Big Data and Smart Computing (BigComp)**. Jeju, Korea (South) : 374 - 376.
- Lin, L., Li, J., Chen, F., Ye, J. and Huai, J. 2018. "Road Traffic Speed Prediction: A Probabilistic Model Fusing Multi-Source Data." **IEEE Transactions on Knowledge and Data Engineering**. 30, 7 (June) : 1310-1323.
- Maria, J., Amaro, J., Falcao, G. and Alexandre, L. A., 2016. "Stacked Autoencoders using Low-power Accelerated Architectures for Object Recognition in Autonomous Systems." **Neural Process Lett**. 43, 12 (May) : 445 – 458.
- Min, S., Lee, B. and Yoon, S. 2016. "Deep Learning in Bioinformatics." **Briefings in Bioinformatics**. 18, 5 (July) : 851 – 869.
- Mulyana, A., Hindersah, H. and Prihatmanto, A. S. 2015. "Gamification Design of Traffic Data Collection Through Social Reporting." **4th International Conference on Interactive Digital Media (ICIDM)**. Bandung, Indonesia : 1 – 4.
- Nandury, S. V. and Begum, B. A. 2016. "Strategies to Handle Big Data for Traffic Management in Smart Cities." **International Conference on Advances in Computing, Communications and Informatics (ICACCI)**. Jaipur, India : 356 - 364.
- Neuhold, R., Gursch, H., Kern, R. and Cik, M. 2018. "Driver's Dashboard – Using Social Media Data as Additional Information for Motorway Operators." **IET Intelligent Transport Systems**. 12, 9 (September) : 1116 - 1122.

- Nguyen, S., Salcic, Z. and Zhang, X. 2018. "Big Data Processing in Fog - Smart Parking Case Study." **2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)**. Melbourne, VIC, Australia : 127 - 134.
- Peng, W., Li, B. and Zhu, X. 2016. "An Analysis Platform of Road Traffic Management System Log Data Based on Distributed Storage and Parallel Computing Techniques." **IEEE International Conferences on Big Data and Cloud Computing (BDCloud)**. Atlanta, GA., USA : 585 - 589.
- Ren, Q., Man, K. L., Li, M. and Gao, B. 2019. "Using Blockchain to Enhance and Optimize IoT-based Intelligent Traffic System." **International Conference on Platform Technology and Service (PlatCon)**. Jeju, Korea (South) : 1 - 4.
- Sarikaya, R., Hinton, G. E. and Deoras, A. 2014. "Application of Deep Belief Networks for Natural Language Understanding." **IEEE/ACM Transaction Audio, Speech and Language Processing**. 22, 4 (April) : 778 - 784.
- Song, J., Zhang, H., Li, X., Gao, L., Wang, M. and Hong, R. 2018. "Self-Supervised Video Hashing With Hierarchical Binary Auto-Encoder." **IEEE Transactions on Image Processing**. 27, 7 (July) : 3210 - 3221.
- Sun, W., Miao, D., Qin, X. and Wei, G. 2016. "Characterizing User Mobility from the View of 4G Cellular Network." **17th IEEE International Conference on Mobile Data Management (MDM)**. Porto, Portugal : 34 - 39.
- Sutskever, I., Martens, J. and Hinton, G. E. 2011. "Generating Text with Recurrent Neural Networks." **Proceedings of the 28th International Conference on Machine Learning**. Bellevue, WA., USA : 1017 - 1024.
- Sutskever, I., Vinyals, O. and Le, Q. V., 2014. "Sequence to Sequence Learning with Neural Networks." **Advances in Neural Information Processing Systems**. Montreal, Canada : 1 - 9.
- Teh, Y. W. and Hinton, G. E. 2000. "Rate-coded Restricted Boltzmann Machines for Face Recognition." **Proceedings of the 13th International Conference on Neural Information Processing Systems (NIPS)**. (January) : 872 - 878.
- Theis, L., Shi, W., Cunningham, A. and Huszár, F. 2017. "Lossy Image Compression with Compressive Autoencoders. Cornell University." **5th International Conference on Learning Representations (ICLR)**. Toulon, France : 1 - 19.
- TWF Agency. 2020. **Digital 2020 Global Digital Overview**. [Online] Available : <https://www.twfdigital.com/blog/2020/04/summary-of-social-network-in-thailand-march2020/>, September 9, 2021.

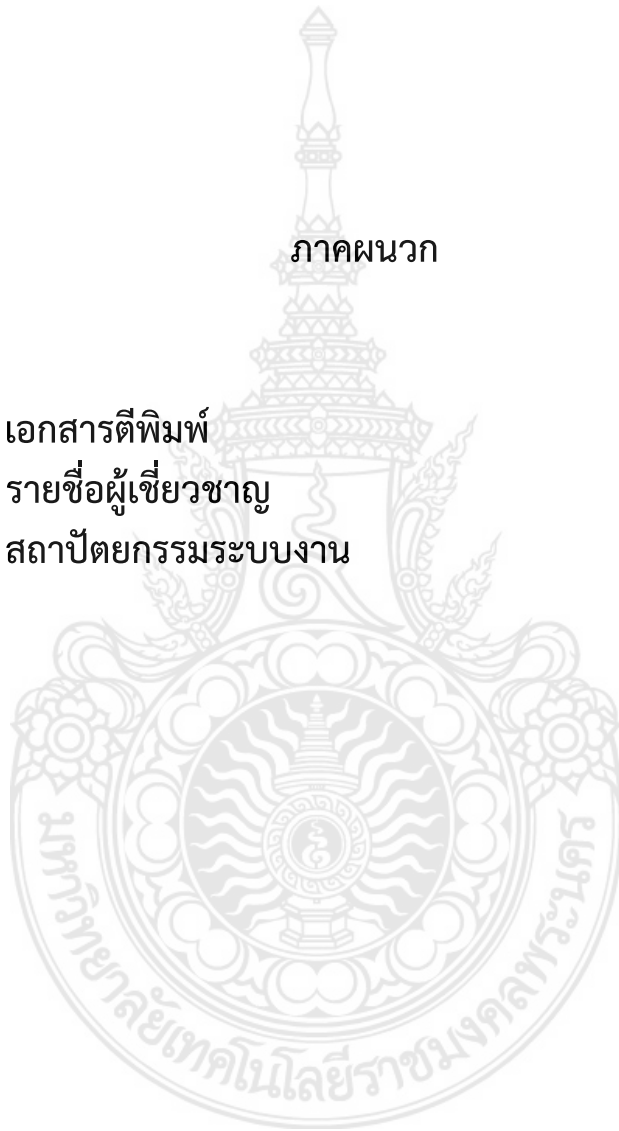
- Wang, X. and Li, Z. 2016. "Integrated Platform for Smart Traffic Big Data." **International Conference on Logistics, Informatics and Service Sciences (LISS)**. Sydney, Australia : 1 – 6.
- Wang, X., Gao, L., Wang, P., Sun, X. and Liu, X. 2018. "Two-Stream 3-D Convnet Fusion for Action Recognition in Videos with Arbitrary Size and Length." **IEEE Transactions on Multimedia**. 20, 3 (September) : 634 – 644.
- Wang, Y., Zhang, D., Liu, Y., Dai, B., and Lee, L. H. 2019. "Enhancing Transportation Systems via Deep Learning: A Survey." **Transportation Research Part C: Emerging Technologies**. 99, (February) : 144 - 163.
- Xie, J. and Luo, J. 2017. "Construction for the City Taxi Trajectory Data Analysis System by Hadoop Platform." **IEEE 2nd International Conference on Big Data Analysis (ICBDA)**. Beijing, China : 527 - 531.
- Yamazaki, T. 2019. "Analysis of Traffic Accident Occurrence in Niigata Prefecture of Japan using Open Data." **IEEE International Symposium on Circuits and Systems (ISCAS)**. Sapporo, Japan : 1 - 4.
- Yang, Z. and Zhu, M. 2019. "A Dynamic Prediction Model of Real-Time Link Travel Time Based on Traffic Big Data." **International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)**. Changsha, China : 330 – 333.
- Zhang, M., Chen, C., Wo, T., Xie, T., Bhuiyan, Z. A. and Lin, X. 2017. "SafeDrive: Online Driving Anomaly Detection from Large-Scale Vehicle Data." **IEEE Transactions on Industrial Informatics**. 13, 4 (February) : 2087-2096.
- Zhang, S. 2015. "Using Twitter to Enhance Traffic Incident Awareness." **IEEE 18th International Conference on Intelligent Transportation Systems**. Gran Canaria. Spain.
- Zhang, Z., He, Q. and Gao, J. 2018. "A Deep Learning Approach for Detecting Traffic Accidents from Social Media Data." **Transportation Research Part C Emerging Technologies**. 86, (January) : 580 - 596.
- Zhu, L., Yu, F. R., Wang, Y. and Ning, B. 2019. "Big Data Analytics in Intelligent Transportation Systems: A Survey." **IEEE Transactions on Intelligent Transportation Systems**. 20, 1 (April) : 383 – 398.
- Zunic, E., Djekovic, A. and Donko, D. 2016. "Application of Big Data and Text Mining Methods and Technologies in Modern Business Analyzing Social Networks Data About Traffic Tracking." **9th International Symposium on Telecommunications (BIHTEL)**. Sarajevo, Bosnia : 1 - 6.

ภาคผนวก

ภาคผนวก ก เอกสารตีพิมพ์

ภาคผนวก ข รายชื่อผู้เชี่ยวชาญ

ภาคผนวก ค สถาปัตยกรรมระบบงาน




ภาคผนวก ก
เอกสารตีพิมพ์

1. K. Puangnak and N. Rachsirivatcharabul, “The Collection of Road Traffic Incidents in Bangkok from Twitter Data based on Deep Learning Algorithm”, ECTI-CIT Transactions, vol. 16, no. 3, pp. 267–276, Jun. 2022.
2. K. Puangnak and N. Rachsirivatcharabul, “Incident Detection Techniques for the Thai language on Twitter”, ECTI-CIT Transactions, vol. 16, no. 3, pp. 313–323, Aug. 2022.



ภาพ ก-1 หน้าปกเอกสารตีพิมพ์ ECTI-CIT Transactions - Volume. 16 – Number 3 - June 2022
<https://ph01.tci-thaijo.org/index.php/ecticit/article/view/248535>

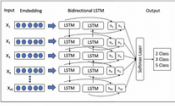
[Register](#) [Login](#)



[About the Journal](#)
[Current](#)
[Articles in Press](#)
[Archives](#)
[Editorial Team](#)
[Announcements](#)
[Author Guideline](#)
[Others ▾](#)

Home / Archives / Vol. 16 No. 3 (2022): ECTI Transactions on CIT (Sep 2022) / Research Article

The Collection of Road Traffic Incidents in Bangkok from Twitter Data based on Deep Learning Algorithm



PDF

Published: Jun 18, 2022

DOI: <https://doi.org/10.37936/ecti-cit.2022163.248535>

Keywords:

Incident Detection, Deep Learning Algorithm

Korn Puangnak

Rajamangala University of Technology Phra Nakhon, Thailand

Natworapol Rachsirivatcharabul

Rajamangala University of Technology Phra Nakhon, Thailand

Abstract

Text processing technology from Twitter to report notification formats that are known in many countries with verification on different languages. This research presents the development of a neural network memory learning model. To solve the problem of classifying incidence patterns and identifying severity of incidents from Thai social media messages. For gathering incident data and reporting incidents externally from a single reporting platform by using deep learning models like MLP, CNN and LSTM which is designed by dividing the study into 3 types, including examination traffic incidence identification pattern that can identify the report as general news or traffic reporting Incident Identification Patterns. These include traffic conditions, accidents, disasters, damaged roads, or other than the aforementioned patterns, and the pattern indicating the severity of the incidence consists of normal level, medium level and lane blocking or stationary levels. The results demonstrated the ability of LSTM learning with the best results in incidence detection and incidence pattern identification at 93.44% and 87.40%, respectively, and the CNN method was able to State the severity of the incidence at best, reaching 91.42%.

How to Cite

K. Puangnak and N. Rachsirivatcharabul, "The Collection of Road Traffic Incidents in Bangkok from Twitter Data based on Deep Learning Algorithm", *ECTI-CIT Transactions*, vol. 16, no. 3, pp. 267–276, Jun. 2022.


More Citation Formats ▾
Download Citation ▾

Issue

Vol. 16 No. 3 (2022): ECTI Transactions on CIT (Sep 2022)

Section

Research Article



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](#).

Make a Submission


Information

[For Readers](#)






[For Authors](#)

[For Librarians](#)

Indexed In



Scopus[®]

Citescore by Scopus

0.1

2022 CiteScore

Powered by Scopus

ภาพ ก-2 เอกสารตอบรับบทความ (Accepted Letter) ECTI-CIT Transactions - Volume. 16 –
Number 3 - June 2022

Your Submission - [EMID:bbOf10dd871c9d2b]

ภายนอก

กล่องจดหมาย x



ECTI-CIT <em@editorialmanager.com>

ส. 4 มิ.ย. 10:44



ถึง ฉัน ▾

Ref.: Ms. No. ECTI-CIT-D-22-00011R1

The Collection of Road Traffic Incidents in Bangkok from Twitter Data based on Deep Learning Algorithm
ECTI Transactions on Computer and Information Technology

Dear Asst.Prof. PUANGNAK,

I am pleased to tell you that your work has now been accepted for publication in ECTI Transactions on Computer and Information Technology.

It was accepted on 2022-05-21 00:16:08

Comments from the Editor and Reviewers can be found below.

Thank you for submitting your work to this journal.

With kind regards

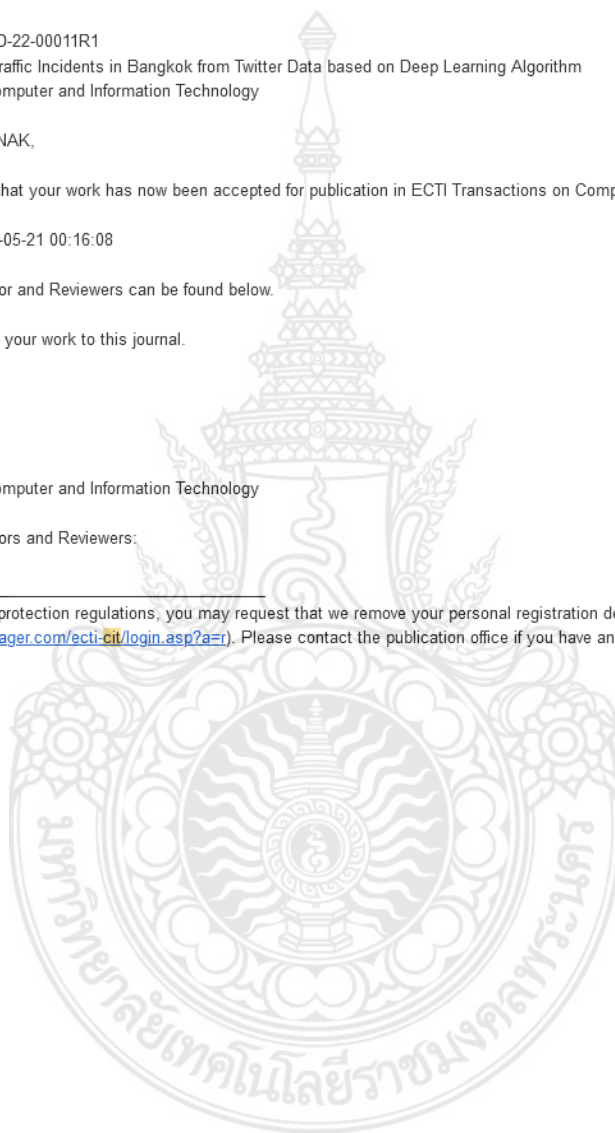
Krisana Chinnasarn

Editor

ECTI Transactions on Computer and Information Technology

Comments from the Editors and Reviewers:

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: <https://www.editorialmanager.com/ecti-cit/login.asp?a=r>). Please contact the publication office if you have any questions.





ISSN 2286 – 9131

Transactions

on Computer and Information Technology
Volume 16, No. 3, September 2022

Research Article

Two-Dimensional Error Correction Code for Spin-Transfer Torque Magnetic Random-Access Memory (STT-MRAM) Caches.....	<i>C. Duangthong, P. Supnithi and W. Phakphisut</i>	237
The Implementation of Discrete-Event Simulation and Demand Forecasting Using Temporal Fusion Transformers to Validate Spare Parts Inventory Policy for The Petrochemicals Industry.....	<i>S. Immuphat and M. Toahchoodee</i>	247
A Mathematical Model and PI Controller Design Based on Indirect Vector Control for Permanent Magnet Synchronous Motor.....	<i>N. Kongchoo, P. Santiprapan and N. Jindapetch</i>	259
The Collection of Road Traffic Incidents in Bangkok from Twitter Data based on Deep Learning Algorithm.....	<i>K. Puangnak and N. Rachsirivatcharabul</i>	267
Capturing Spatial Relationship Mapping Patterns between GPS Coordinates and Road Network Using Machine Learning and Partitioning Techniques.....	<i>T. Boonchoo</i>	277
Privacy Threats and Privacy Preservation Techniques for Farmer Data Collection based on Data Shuffling.....	<i>S. Riyana, N. Ito, T. Chaïya, U. Sriwichai, N. Dussadee, T. Chaichana, R. Assawarachan, T. Maneechukate, S. Tantikul and N. Riyana</i>	289
Optimized CNN Model with Derived Kemels for Apnea Classification.....	<i>S. A and S. M</i>	302
Incident Detection Techniques for the Thai language on Twitter.....	<i>K. Puangnak and N. Rachsirivatcharabul</i>	313
Hybrid Visual and Optimal Elliptic Curve Cryptography for Medical Image Security in Iot.....	<i>L. A. Kumar, S. Srivastava, S. R. Balaji, F. H. Shajin and P. Rajesh</i>	324
High-speed Firewall Rule Verification Technique Improves Throughput Performance for IP Version 6.....	<i>S. Khummanee, U. Saisangchan, K. Somsuk and S. Atsawaraungsuk</i>	338

Submission Guideline

In your manuscript, one author has been designated as the corresponding author with contact details:

- E-mail address
- Affiliation

Further considerations before your submission:

- Manuscript has been 'spell checked' and 'grammar checked'
- All references quoted in the Reference Section are cited in the body of the manuscript
- Permission for use of copyrighted materials from other sources must be declared
- The manuscript length must be at least 8 pages and no longer than 12 pages with two (2) columns.

Form more information visit: <https://ph01.tci-thajjo.org/index.php/ecticit/author-guideline>

<https://ph01.tci-thajjo.org/index.php/ecticit/>

ECTI
E
C
T
I



The Collection of Road Traffic Incidents in Bangkok from Twitter Data based on Deep Learning Algorithm

Korn Puangnak¹ and Natworapol Rachsirivatcharabul²

ABSTRACT

Surveillance and accident reports for current traffic management. Surveillance by monitoring the situation by using CCTV. This detecting method need to many operators to observe real time. At present communication technology and social media participate in report traffic incidents. Also, the accident was report by people who were facing the accident. Therefore, we decide to create software to investigate accidents from text reported via social media in Thai language. At present, there is no research or work platform that clearly supports this form of work. Many countries use analysis technology to detect the information from Twitter to report incident notification with verification on different languages. The purpose of this research is "to develop the deep learning technology and to solve the problem of classifying incidence patterns and identifying severity of incidents from social media in Thai's massage." For collect incident data and reporting incidents externally from a single reporting platform. Using the deep learning model MLP, CNN, Bi-LSTM, and LSTM+CNN. We can identify the twitter message as general news or traffic reporting. The traffic conditions such as traffic information, accidents, disasters, damaged roads, or other than those mentioned above. And the incidence severity level was identified as normal, lane-blocking or lane-closure. The examination demonstrated the capability of CNN+LSTM learning with the best results in incidence detection and incidence patterns at 93.44%, and CNN results in image identification. The incidence model was best at 85.29%, respectively, and the LSTM method best rated the severity of the incidence, reaching 88.53%.

Article information:

Keywords: Incident Detection, Deep Learning Algorithm

Article history:

Received: January 1, 2022

Revised: May 14, 2022

Accepted: May 21, 2022

Published: June 18, 2022

(Online)

DOI: 10.37936/ecti-cit.2022162.248535

1. INTRODUCTION

Reporting of road traffic incidences by relevant agencies at present, each responsible agency has different methods for detecting incidences including detecting the occurrence of the incident by surveillance of the traffic control room staff through CCTV. By relying on the display of random images or lap times of image showing in each camera on the highway or automatic Image Processing and decided by the processor chip on the surveillance CCTV or central computer processing by image processing software or detecting the occurrence of an incident with a sensor type detector such as Microwave Radar by taking traffic changes data to calculate the incidence forecast, etc. All of the technologies mentioned are those

that require investment and the installation of a large number of related tools and equipments. As the technology of mobile devices or smartphones is developing by leaps and bounds, it plays a role in many areas. On the daily life of the people who use more roads including notifying news via social media channels on various platforms. Twitter is one of the most popular social media platforms and supports the development of programs to bring information to use, including being able to detect incidents by processing tweets. Notify the news immediately.

In the past research on solving the problem of automatic incidence detection with a large amount of data, beginning with a 2017 study by Bo-Huei Lin, Shu-Fen Tseng [1] analyzed data that the government allows the public to participate in collecting data by

^{1,2}The authors are with Faculty of Engineering, Rajamangala University of Technology Phra Nakhon, E-mail: Korn.p@rmutp.ac.th and Natworapol.b@rmutp.ac.th

informing them via hotline to report on problems encountered on the way to compare with road accidents by the results in the same year's large-scale. Computing research, Lin Cao [2] presented a big data platform to collect data for public traffic services. In the matter of informing the bus time calling a taxi and providing real-time traffic information. There is also mixed research by Jianhua Xie and Jiahui Luo [3] present data collection and analysis route waiting area for passengers, time interval and the number of taxis in the city to develop a smart traffic system with the government. On the other hand, there are different types of traffic research for example: Mingming Zhang [4] introduces the SafeDrive algorithm to alert motorists in the event of a malfunction while driving. By referring to various sensor devices in the car, for example machine rpm and speed, speed and action angles on the road sudden braking to process whether the car is losing control or not. The result is 83% accurate, which means have higher accuracy Rule Based Algorithm at the end of the year, Alessandro Cioccola [5] presented a car sharing platform, UMAP, to compare each company's data. What types of people have the characteristics of choosing a car rental company? distance and time analysis, pickup point and destination.

In 2018, by Giovanni Buroni et al. [6], a study of transport routes coupled with the time of trucks for bringing in a new management process improvement to reduce transportation costs and pollution problems and to help solving the problem of traffic in the city. This article was collected and tested in the Brussels capital region. Belgium, the same year, by Alex Kaplunovich and Yelena Yesha [7] collecting data from taxis in New York City from the beginning of the pickup at the station during the day that the driving style parking in any area. Time of start-up and distance from parking until returning the car at the station to analyze the pollution of taxis in New York City. This year, there are various researches in applying big data to transportation work such as the research of Lu Lin [8] improves road traffic speed predictions by fusing traditional speed detection data with new data types. By bringing data from many aspects to create more efficiency of the data in the future, this research will include GP-based and Distributive GP. In order to make the predictions more accurate along with being able to assess in real time meanwhile, a case study of research for large-scale Intelligent Transport System (ITS) is presented by S Appavu [9]. It is a presentation of the unique model of ITS in India. Due to the driving style, road layout and environmental conditions of India. The ITS Platform in developed countries cannot be used directly. The publisher therefore has to collect and analyze the traffic characteristics of India with real data. In order to use this information to develop ITS systems in the country, another example of research in bringing in-

ternet of things (IoT) to help manage by Sang Nguyen [10] the researchers studied IoT technology to create a parking management system that they believe will reduce the time drivers need to be in vehicles with two main goals. The long-term goal is to optimize the management system. Manage parking spaces the short-term goal is help drivers spend less time finding parking spaces by finding suitable parking spaces. In this research, The IoT technology is provided to monitor the area. The duration of parking throughout the city and analysis of nearby parking spaces. The system works without sending a lot of data. Does not back up the data detected by the device. All the data from the cloud server is sent to the cloud server. The analytics uses Hadoop and Map Reduce and is ready to run on the cluster computer.

In the year 2019, Qilei Ren [11] studied the application of Block Chain technology as the center of storage or forwarding data in an ITS and using IoT devices installed as a detection device in various places. It has been compared with a cloud server data centered system and directly designed peer-to-peer data transmission between vehicles. There is research to solve traffic problems by Yang Zhao-xia and ZHU Ming-hua [12] researched and developed the ability to estimate Travel Time in a flexible range of roads in the Traffic Information Platform. They both commented that Big Data traffic can effectively provide feedback or suggest a solution to traffic congestion. Using Travel Time analysis method from Big Data in dynamic prediction in real time tested on traffic monitoring system structure using RFID tag technology and using fusion data. The test results in the model show that the new method used is more accurate and can solve traffic congestion problems and make traffic smoother. Meanwhile, Li Zhu [13] presented a research study exploring Big Data analytics systems in the intelligent transportation system. The survey describes the history of the unique characteristics of both Big Data technologies and the ITS, a framework used to perform Big Data analysis in ITS results analysis tool. Methods for analyzing results include analysis of road traffic accidents. Road mobility estimation, administration, planning and management of the public transport system. Personal travel planning as well as maintenance planning.

Currently, the research trend of new technology device which is IoT has a priority to develop the Auto detect incident system. In addition, the research the study about analyses data from any interested group with detect incident from social media messages in 2016, by Peerapon Vateekul et al. [14]. The author presents the method to get people feeling and emotion from Thai Tweeter message by using Deep Learning Method with Short Term Memory (LSTM) model and Dynamic Convolutional Neural Network (DCNN) model then compare with other method such as Naïve Bayes and SVM. The result indicate that

these two methods were get more high accuracy especially DCNN with 75.35%. Besides, Supon Klaitin et al. [15] present categorize message Thai Twitter by naive Bayes classifiers which is the one of machine learning for categorize detail of event and summary data with separate Tweet for two group that is Information of Road with average correction 88.42% and Classification with average correction 76.40%. In 2020, Tanatorn Tanantong et al. [16] present method for get keyword from Thai message in Twitter with N-gram technique for classified the word that does not in dictionary and increase accuracy of partial word in Twitter by using message regarding to Thai university for classify keyword and trend analysis. The result indicates that they got 70% accuracy.

On the other hand, in a group of researches to solve such problems. Modern methods based on technology, such as pseudo-neural network methods or deep learning, have been introduced into the analysis and processing by Zhenhua Zhang [17]. In 2018, it presented the use of Machine Learning to detect traffic accidents from social media data. More than 1 million tweets were reviewed in one year in two cities: Northern Virginia and New York by selecting key words from sentences with Apriori Algorithm and importing them into ML for further answers. The experiment compares DBN, LMTS, SVMs, and sLDA methods with identifying the area of incidence or according to research by Robert Neuhold [18] said data collection on expressways for traffic management depends on the installation of automatic traffic and traffic monitoring systems with cameras only. The research team presented guidelines for using social media as an additional source of information by ASFINAG, the Austrian expressway transport authority. Developed a system to display the average speed of each kilometer at different times of the day. This allows the driver to quickly get information on the situation in a color chart. It has evolved to be able to collect event descriptions from Facebook and RSS of Austrian radio stations and popular newspaper news which is filtered through useful relevant messages and processed into traffic data and presented as a color symbol for traffic conditions or average speed on the route at that time. Meanwhile, Fang Jin[19] said: More information from social media that has been mentioned or reported on the current incidence. This makes it possible to detect the tendency of accidents and incidents on the road. quickly In this paper, the research team presents the nature of the problem as well as the complexity of applying Spatio-Temporal Data Mining principles. To analyze accidents or incidents on the road Hasna El Alaoui [20] said road accidents are still common. Safety has become a major concern problem. There are many factors that explain a traffic accident such as the nature of the environment at that time, behavior, weather conditions and other complex factors that are uncertain. The out-

come of road accidents is not linear. Therefore, it is imperative to explore the relationship between data from many aspects to reduce the risk. The research team has used the data of past accident reports come pre-process and classification by using Data Mining tools to find information that can predict the cause of the accident. The results obtained can validate the data and can help predict new events in the future with similar data. The goal of this research is to analyze the data and select the most accurate data obtained for validation by analyzing the nature and relationships of the data. The research team proposes a decision system for analyzing traffic accident data to extract information related to road risk prevention and also introduces the concept of crowdsourcing. In the process of collecting additional accident-related information from road users the proposed system uses data mining methods and big data techniques suitable for surveying accident data and uses Hadoop to categorize road accidents and determine what is causing the problem. Until now research on social gathering data is ongoing and it is a replacement tool for traditional technologies such as Tatsuya Yamazaki [21] that presents Open Data Accident analysis. Publicly available public information of the city of Niigata, Japan which useful information in the analysis of such incidence problems can be classified as weather forecast information, event statistics of vehicle-to-vehicle accidents (V2V) and Human-to-Vehicle accidents (H2V) and accidents without parties together with panoramic images of red light intersections from the analysis results of all the information mentioned will show that traffic junctions without traffic lights are more V2V incidence than junctions with traffic lights, ultimately Yuan Yuan Chen [22]. Presenting traffic-related milestones from Sina Weibo social media, China's Microblogging Platform, using the Bag of Word Model technique to find keywords. Chinese is compared to English for input into the neural network (NN) and to experiment with the use of Convolutional Neural Networks (CNN), Long Short-term Memory (LSTM), and a combination of CNN and LSTM methods to learn to recognize words and events to be used to number and find traffic incidents.

2. DATA PREPARATION

Data preparation to create practice and test data sets consisted of 3 models: a model for grouping incident messages and general publicity messages, a model to categorize the types of incidences and a model to classify the severity of the incidence. The model that was created used the same set of data for training, so to limit errors the same test data set was used. It will test the actual data set collected from social media posts on Twitter. The data must be imported through the Data Preparation process which consists of dividing sentences into words, filtering un-

wanted words out, changing words into numbers and adding equal sentence lengths. that has come to make labels of text in the next step. The process step of this method was shown in Fig.1

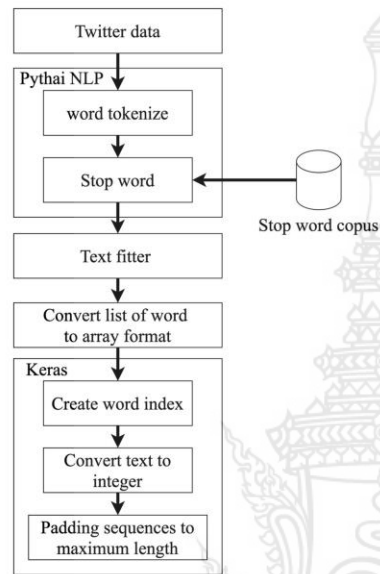


Fig.1: Data Preparation Process.

Fig.1 shown the Tweet message was separated to each word with remove word by Stop word method, text filter and convert to array format then increase the word length by using maximum word length

2.1 Identifying Traffic Incidents

Steps to classify messages it starts with taking the information gathered from Twitter messages through various processes in order to prepare the data before the test by carrying out the process of extracting words. filtering out words converting words to numbers in sequence and increasing the length of the text then enters the testing process. In this step, the test-ready text is passed into the classification model, which is the first model. Its purpose is to classify the message between a general press release or an incident notification message.

2.2 Identifying Patterns of Incidence

For the incidence classification testing procedure, it is performed second after the text classification. The text to be tested is an incidence message that has been categorized from the message division of the

staff. Therefore, pre-test data is prepared by storing it in the tolerance variable and passing it on to the model to classify the incidence. At this stage, a test for the accuracy of the incidence classification is carried out. The objective is to summarize the answers to the incidence messages to be able to tell the type of incidence by the incidence type of the test data group that has passed the professional personnel classification process. There are total five types, so the message can be tested immediately.

2.3 Identifying the Severity of the Incident

In the process of testing the violence isolation when receiving a message from Twitter. As in the previous process, the incident-only message was used in the test, the process was initiated by the operator-classified message used for pre-processing. Then the text that has gone through the data preparation process is brought into the model to test the validity of the model. For this step, the model is tested to identify three types of profound severity.

2.4 Converting Words into Numbers

Neural network behavior is the process of importing a set of numbers to calculate mathematical calculations according to each learning method. All words from a data set with word indexing process. It is generated by word frequency by giving word indexes in order of repetition of that word. For example, as shown in table 1.

Table 1: Word Index.

Index	Word
1	แยก
2	อุบัติเหตุ
3	รถติด
4	จราจร
5	รายงาน
6	การจราจร
7	ช่วง
8	และ
9	ช่อง
10	กัน
11	ขวาง

2.5 Incident of "Stop word"

Word that always used in many sentences but it does not mean in the main clause. However, some repeated word was the key word of the clause. Many research of NLP was encounter on this issue that is "the, a, an" word for English and, "และ, กับ, ที่, ซึ่ง" Word for Thai. These words were called "Stop word". For the preparation data Stop word was cleaned for get

the main clause to decrease complexity of learning simulation.

In this research used Stop word list of “pythainlp” which is Thai language. The main concept was cleaning Stop word that is in pythainlp from sentence. e.g. “15.56 เริ่ม คิดค้น 100% แล้ว วางงานจราจร สุทธิ” will clean Stop word to “15.56 เริ่ม คิดค้น วางงานจราจร สุทธิ” which is remain the meaningful of sentence. After cleaning Stop word, the Copus will decrease 7% of total word.

2.6 Data Augmentation

The detect word and sentence of incident tweet in Twitter was clarified to 5 types and, several of incident was clarified in 3 types. For compare number of messages, indicate some type was a majority and minority that shown in Fig 2 and Fig 3 respectively. The imbalance data of classification could decrease the data reliability in training simulation. To solve this problem, we recommend the Data Augmentation to add and separate imbalance data that shown in pic x (b) and pic x (b)

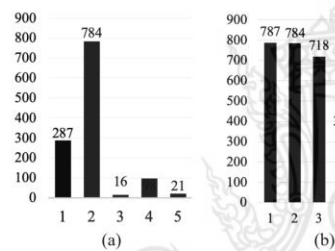


Fig.2: Multi-Layer Perceptron Model.

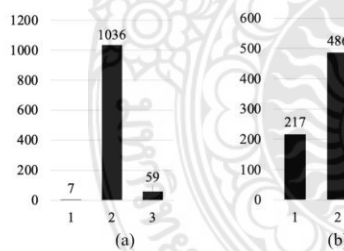


Fig.3: Multi-Layer Perceptron Model.

3. MODEL CONFIGURATION INCIDENT DETECTION WITH IN DEEP LEARNING METHODS

3.1 Deep Learning Process with MLP Method

Multi-Layer Perceptron (MLP) is a method developed from perceptron which is a simple structure because it has a single layer. Later, when computing has evolved more and more, the ability to process has been developed later. Perceptron to have more than one layer or as multiple layers to be able to process more precisely and have adopted MLP in processing to classify the following types of process messages.

Parameterization is length of the sentence to be tested is 45 words long. Therefore, the input parameter is set to 45 and in order not to overfit the model, it has to add a dropout function of 0.2 to the number of layers equal to 5 layers, the first 4 node members of the hierarchy have 450 nodes per node hierarchy. Linear function format and in the fifth layer, the last layer has each layer and the last layer output is 2 for specifying the traffic incidence, output is 5 for specifying the pattern of functions. Incidence and output are 3 for determining the severity of the incidence to match the number of data segments to be divided It has a function format as the Softmax.

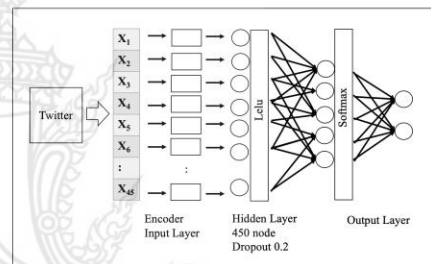


Fig.4: Multi-Layer Perceptron Model.

3.2 Deep Learning Process with CNN Method

Convolution Neural Network (CNN) is a form of neural network method, a type of process design algorithms in order to be able to support a set of numbers. CNNs were originally designed to work with image processing, which is actually converting RGB color values to a set of numbers 0-255 and then taking the set of numbers into the CNN model for processing. Therefore, if a CNN supports processing numerically converted images, then the CNN model can support processing numerically converted text as well as above is converts a message into a series of numbers, which is similar to the MLP method, but CNN adds layers of Convolution and Pooling to filter words by 4654×45 dimension of filter (n × d) with n is number of word in corpus and d is the length of filter that same with dimension of characterize the text to bring the result to the last layer and output it as result Number of layers equal to 5 layers, members in each node number 45 nodes in the first hierarchy. The second is the class

of the convolutional layer has 250 nodes, the Linear Function format is the third pooling layer, 250 nodes, and the fourth and fifth layers have 2. The output is 5 nodes for specifying the pattern of incidence, and the output is 3 nodes for indicating the severity of the incidence, respectively.

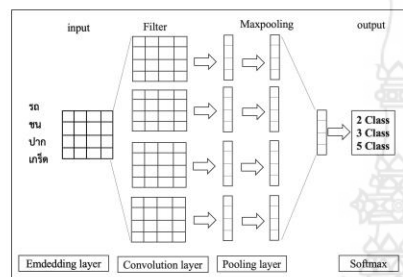


Fig.5: Convolution Neural Network Model.

3.3 Bi-LSTM Deep Learning Process

Long/Short Term Memory, or LSTM, is a popular method for text processing or natural language actions because it employs a technique developed from Recurrent Neural Networks: RNN, which is unique in that its time series, that is, LSTMs are able to remember past words in a previous hierarchy longer than RNNs in the future.

The LSTM has been developed further. Bidirectional LSTM, unlike Bi-LSTM, can recognize both forward and backward memory cells, whereas LSTM can recognize only backward and for this research, Bi-LSTM is used to process messages to classify the type of incidence. The first layer is to create an Embedding Layer as an output to Bi-LSTM, which has dropout parameters to reduce the overfit of Trends for the next level is the class that summarizes the classification to come out as a final answer. The number of layers is equal to 4 layers, each member of the node in the first layer has 45 nodes of the second hierarchy which is the layer of Bidirectional LSTM has a total of 90 nodes and adds 20% dropout functionality to each layer. In the third hierarchy there are 50 nodes. Linear Function layout and four nodes to sum values as result and 2 nodes for incidence identification, output 5 nodes for incidence pattern identification, and output to 3 nodes for incidence severity level identification number.

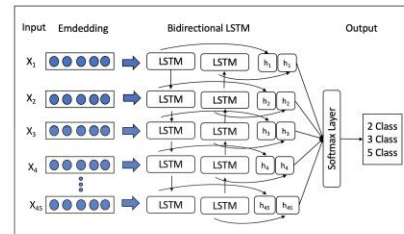


Fig.6: Bi-LSTM Model.

3.4 Bi-LSTM Deep Learning Process

Past research has been processed by combining CNN and LSTM. By passing parameters from the CNN obtained from the convolution layer and importing them into the LSTM layer and deciding the prediction results with the Softmax layer for passing the parameters to the model. The method presented here differs from the previous CNN and LSTM methods. The first layer is an embedding layer with 32 dimensions. Convolution layer has 250 nodes and the result is sent to the LSTM layer by specifying parameters. Enter a 32 dimension with a dropout set to 20%. In the final layer, the results are aggregated into responses by a dense layer. Output is set to 2 for incident detection messages, output to 3 for incident classification messages, and output to 5 nodes for severity messages.

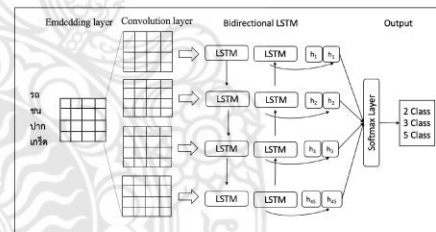


Fig.7: Hybrid CNN-LSTM Model.

4. EXPERIMENT

4.1 Assumption and Test cases

The experiment starts with data mining from Twitter account that relate to traffic report which are @js100radio, @Traffic_1197 and @fm91trafficpro. The total data is 3363 messages from 5 February 2021 to 23 March 2021 then we classified data to 3 type which is "Incident identification" with consist of Normal message and Traffic message, "Accident case" with consist of Traffic, Incident, Disaster, Potholes and Other and "Severity identification" with consist

of Normal, Heavy and Standstill. We select Train-Test split by 70:30 for machine learning which is the best ratio of our raw data that does not overfitting with high of train data that shown in Table 2.

Table 2: Comparison of DR with different ratio.

Ratio	1	2	3	4	5	6
Learning Ratio	50	60	70	80	90	95
Testing Ratio	50	40	30	20	10	5
DR (%)	81.00	81.34	81.87	80.30	83.93	85.71

The table 2 indicate that the ratio 90:10 and 95:5 is high accuracy, but it does not reflect the whole message due to the less test data.

The experiment was setting by using Python with Keras library in Tensorflow that processes to call function for each model for example, modeling with LSTM method by send LSTM parameter to Bidirectional() function.

4.2 Performance Testing by K-Fold Method

After adjusting the parameters of the model to make it ready for learning to model it. In order for learning to bring distributed data to learn, the process is to divide the data into subgroups (k-fold) and bring the subgroups then come back and forth, also known as "Cross validation" to find the mean of the efficiency of the learned model, in this experiment, the data groups were divided into 10 groups, or k=10. In each group, the values were randomly assigned to avoid possible differences. Iterated and tested the model's learning performance for 10 cycles, with each alternating cycle having a learning group picked up and tested, with 70% learning group and 30% testing data. With result in table 2 so, we select that ratio for separate data.

4.3 Accuracy Measurement

To measure accuracy of incident detection, it can be considered with two parameter which is DR (Detection Rate) that for measure correction of incident detection (%). The calculation was shown in equation (1) with D_n is number of incidents that detect by this method and D_t is number of incidents of this experiment.

$$DR = \frac{D_n}{D_t} \times 100 \quad (1)$$

In addition, the measure accuracy is FAR (%) (False Alarm Rate). The equation was shown in equation (2) with N_f is number of messages that false alert and N_t is number of incidents in this experiment.

$$FAR = \frac{N_f}{N_t} \times 100 \quad (2)$$

5. RESULTS

5.1 Results of the identification of traffic incidents

Required to use the data in the test, 1010 data that have already passed the data preparation process.

MLP in the MPL incidence classification test. In the test with real data, the incidence detection accuracy was 67.75% and the incidence detection error was 32.25%.

CNN in the CNN method incidence classification test, the incidence detection accuracy was 91.47% and the incidence detection error was 8.58%.

Bi-LSTM in the LSTM method incidence classification test with an incidence detection accuracy of 93.53% and an incidence detection result of 6.47%.

CNN+LSTM in the combining CNN method and LSTM method incidence classification test with an incidence detection accuracy of 94.06% and an incidence detection result of 5.94%.

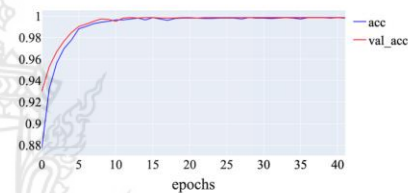


Fig.8: Incident Detection Accuracy of LSTM.

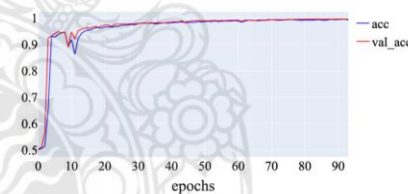


Fig.9: Incident Detection Accuracy of CNN+LSTM.

5.2 Categories of incidents

For the data tested, there were 503 data, which were only incidence text data.

MLP in the MPL method incidence classification test. In the test with real data, the incidence detection accuracy was 58.90% and the error incidence detection result was 41.10%.

CNN in the CNN incidence classification test. In the test with real data, the incidence detection accuracy was 85.29% and the error incidence detection result was 14.71%.

Bi-LSTM in the Bi-LSTM method incidence classification test. The incidence detection accuracy was 84.65% and the incidence detection accuracy was 15.35%.

CNN+LSTM in the combining CNN method and LSTM method incidence classification test. The incidence detection accuracy was 84.58% and the incidence detection accuracy was 15.42%.

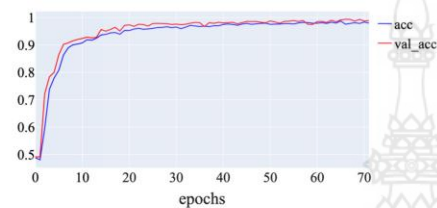


Fig.10: Incident Classification Accuracy of CNN.

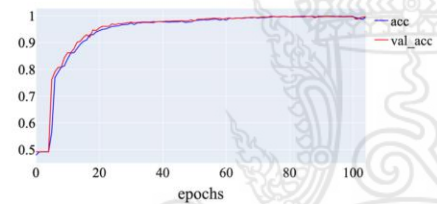


Fig.11: Incident Classification Accuracy of CNN+LSTM.

5.3 Incident Severity Determination Results

For the data tested, there were 503 data, which were only incidence text data.

MLP in the MPL method incidence determination test. The MPL method has an incidence severity accuracy of 44.14% and the error was 55.86%.

CNN in the CNN method incidence determination test. The CNN method has an incidence severity accuracy of 93.24% and the error was 6.76%.

Bi-LSTM in the incidence severity determination test the Bi-LSTM method, with an incidence severity accuracy of 77.92% and the error rate was 22.08%.

CNN+LSTM in the incidence severity determination test the CNN+LSTM method, with an incidence severity accuracy of 86.25% and an incidence severity outcome the error rate was 11.70%.

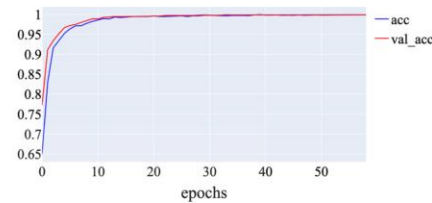


Fig.12: Incident Severity Accuracy of CNN.

5.4 Analyze the results of the experiment

In the test, results can be recorded to compare the effectiveness of each classification model including MLP, CNN, Bi-LSTM in incidence classification and severity identification to determine their validity. Of the work after deep learning of each system and a text-based test from Twitter for the classification of 1,010 messages, incidence messages and in 1,010 there were 503 incidence messages to be tested against classification and strength-identification patterns. The results of the tests showed the overall picture: the CNN-derived model test pattern had the highest average accuracy rate of 93.53%. However, when observed in the incidence classification, the incidence classification model of only two groups from the CNN-LSTM model had the highest accuracy of 94.06%. Need to Multi Class is a classification with 5 data groups and specifying severity with 3 data groups. It can be seen that the model results obtained from CNN are better, as detailed in the table 3.

However, we can use Data Augmentation to increase accuracy by remove no meaning word and adding to balance message. The result will show in Table 4.

Table 3: DR and FAR of Deep Learning Method.

Activity	DR				FAR			
	MLP	CNN	LSTM	CNN+LSTM	MLP	CNN	LSTM	CNN+LSTM
Detection	67.75	91.47	93.53	91.06	32.25	8.53	6.47	5.94
Categories	58.9	85.29	84.65	81.58	41.1	14.71	15.35	15.42
Severity	70.1	88.04	88.53	86.26	29.9	11.96	11.47	13.74
AVG	65.58	88.27	88.90	88.30	34.42	11.73	11.10	11.70

Table 4: Increase Efficiency of Learning Simulation.

LSTM	Patterns	NI	SW	MD	SWMD
		87.40	88.35	88.41	89.20
LSTM	Severity	90.07	90.50	90.78	90.87
	Patterns	84.70	85.05	87.22	87.89
CNN	Severity	91.42	91.76	91.56	91.02
	Patterns	54.34	56.56	58.12	57.44
MLP	Severity	70.54	71.89	72.45	75.37
	Patterns	85.99	86.25	86.89	87.20
CNN+LSTM	Severity	86.26	87.28	88.20	91.85

As the data of the incidence verification test data, the various methods can be shown to compare the accuracy of the methods in the incidence examination of various characteristics as shown in Fig.13, together with the visual representation of the error values in Fig.14. In the test results, it was found that by comparing the validity values of the three different incidence detection methods, the Bi-LSTM model was able to classify the occurrence of incidence. And the best identification of violence and the CNN and MPL methods have decreased accuracy respectively. At the same time, the CNN method has the best classification capability, the Bi-LSTM and MPL methods are respectively decreasing in accuracy.

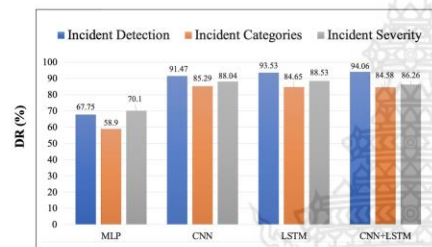


Fig.13: Detection Rate Comparative

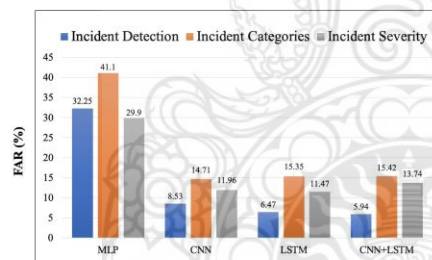


Fig.14: False Alarm Rate Comparative.

6. CONCLUSIONS

Traffic incident detection on Twitter social media platform by detecting from a large number of tweets via the Twitter API in this research the sample text will be collected for review and detailed formatting. By people who have expertise in traffic management along with methods for preparing data with word filtering, word wrapping, and word-to-numeric conversion before being introduced into the deep learning process. It includes MLP, CNN, and Bi-LSTM and CNN+LSTM methods, using a three-method validity comparison for traffic incident detection, Identifying

the type of incidence and identifying the severity of incidents on Twitter in Thai.

The research will show that the development of an incidence detection system. It is possible to choose a method for detecting and discriminating various forms of data effectively by co-processing between methods. By learning CNN+LSTM with the best incidence detection results with DR 94.06%, FAR 5.94%, while the CNN model of incidence was best identified by with DR 85.29%, FAR 14.71% and Bi-LSTM that well process on this pattern is the best for severity of the incident was identified by DR 88.53% and FAR 11.47%.

The experiment results indicate that for each method of deep learning was proper with different issues of incident detection therefore, the selection method for solve the problem should be consideration for work processing to find solution.

References

- [1] B.-H. Lin and S.-F. Tseng, "A predictive analysis of citizen hotlines 1999 and traffic accidents: A case study of Taoyuan city," *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp.374-376, 2017.
- [2] L. Cao, F. Zhu, X. Dong, Z. Shen, J. Yu, B. Hu and G. Xiong, "Big data platform & typical APP services for urban public transportation," *2017 Chinese Automation Congress (CAC)*, pp. 7565-7570, 2017.
- [3] J. Xie and J. Luo, "Construction for the city taxi trajectory data analysis system by Hadoop platform," *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pp. 527-531, 2017.
- [4] M. Zhang, C. Chen, T. Wo, T. Xie, M. Z. A. Bhuiyan and X. Lin, "SafeDrive, Online Driving Anomaly Detection from Large-Scale Vehicle Data," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2087-2096, 2017.
- [5] A. Ciociola, M. Cocca, D. Giordano, M. Melia, A. Morichetta, A. Putina and F. Salutari, "UMAP, Urban mobility analysis platform to harvest car sharing data," *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*, pp. 1-8, 2017.
- [6] G. Buroni, Y. Le Borgne, G. Bontempi and K. Determe, "Cluster Analysis of On-Board-Unit Truck Big Data from the Brussels Capital Region," *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2074-2079, 2018.
- [7] A. Kaplunovich and Y. Yesha, "Consolidating billions of Taxi rides with AWS EMR and Spark in the Cloud Tuning, Analytics and Best Prac-

- tices," *2018 IEEE International Conference on Big Data (Big Data)*, pp. 4501-4507, 2018.
- [8] L. Lin, J. Li, F. Chen, J. Ye and J. Huai, "Road Traffic Speed Prediction: A Probabilistic Model Fusing Multi-Source Data," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1310-1323, 2018.
- [9] S. A. A. Balamurugan, J. F. Lilian and S. Sasikala, "The Future of India Creeping up in Building a Smart City Intelligent Traffic Analysis Platform," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 518-523, 2018.
- [10] S. Nguyen, Z. Salcic and X. Zhang, "Big Data Processing in Fog - Smart Parking Case Study," *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)*, pp. 127-134, 2018.
- [11] Q. Ren, K. L. Man, M. Li and B. Gao, "Using Blockchain to Enhance and Optimize IoT-based Intelligent Traffic System," *2019 International Conference on Platform Technology and Service (PlatCon)*, pp. 1-4, 2019.
- [12] Z. -X. Yang and M. -H. Zhu, "A Dynamic Prediction Model of Real-Time Link Travel Time Based on Traffic Big Data," *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pp. 330-333, 2019.
- [13] L. Zhu, F. R. Yu, Y. Wang, B. Ning and T. Tang, "Big Data Analytics in Intelligent Transportation Systems A Survey," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383-398, 2019.
- [14] P. Vateekul and T. Koomsubha, "A study of sentiment analysis using deep learning techniques on Thai Twitter data," *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1-6, 2016.
- [15] S. Klaithin and C. Haruechaiyasak "Traffic information extraction and classification from Thai Twitter," *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1-6, 2016.
- [16] R. Bishnoi, M. Ebrahimi, F. Oboril and M. B. Tahoori, "Improving Write Performance for STT-MRAM," in *IEEE Transactions on Magnetics*, vol. 52, no. 8, pp. 1-11, Aug. 2016.
- [17] Z. Zhang, Q. He, J. Gao and M. Ni, "A deep learning approach for detecting traffic accidents from social media data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 580-596, 2018.
- [18] R. Neuhold, H. Gursch, R. Kern and M. Cik, "Driver's dashboard - using social media data as additional information for motorway operators," *IET Intelligent Transport Systems*, vol. 12, no. 9, pp. 1116-1122, 2018.
- [19] F. Jin and H. Liu, "Detect Hidden Road Hazards combining Multiple Social Media Data," *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018.
- [20] H. El Alaoui El Abdallaoui, A. El Fazziki, F. Z. Ennaji and M. Sadgal, "Decision Support System for the Analysis of Traffic Accident Big Data," *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 514-521, 2018.
- [21] T. Yamazaki, "Analysis of Traffic Accident Occurrence in Niigata Prefecture of Japan using Open Data," *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1-4, 2019.
- [22] Y. Chen, Y. Lv, X. Wang, L. Li and F.-Y. Wang, "Detecting Traffic Information from Social Media Texts With Deep Learning Approaches," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 3049-3058, 2019.



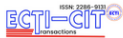
transport system).

Korn Puangnak received his Master degree from Department of Computer Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 2011. Currently, he is an assistant professor at Department of Engineering, Rajamangala University of Technology Phra Nakhon, Thailand. His research falls in the domain of machine learning, and ITS (intelligent



Natworapol Rachsirivatcharabul received his Ph.D. degree from Engineering Management, University of Missouri, USA, in 1999. Currently, he is the President at Rajamangala University of Technology Phra Nakhon, Thailand. His research falls in the domain of energy management, transportation, smart city, and sustainable industrial management.

ภาพ ก-3 หน้าปกเอกสารตีพิมพ์ ECTI-CIT Transactions - Volume. 16 – Number 3 - August 2022
<https://ph01.tci-thaijo.org/index.php/ecticit/article/view/247328>


[About the Journal](#)
[Current](#)
[Archives](#)
[Editorial Team](#)
[Announcements](#)
[Author Guideline](#)
[Others ▾](#)

[Home](#) / [Archives](#) / [Vol. 16 No. 3 \(2022\): ECTI Transactions on CIT \(Sep 2022\)](#)
 / Research Article

[Make a Submission](#)

Incident Detection Techniques for the Thai language on Twitter



PDF

Published: Aug 6, 2022

DOI: <https://doi.org/10.37936/ecti-cit.2022163.247328>

Keywords:
 Twitter Incident Detection
 Word Cloud

KORN PUANGNAK

Rajamangala University of Technology Phra Nakhon, Thailand

Natworapol Rachsirivatcharabul

Rajamangala University of Technology Phra Nakhon, Thailand

Abstract

Nowadays, the rate of road incidents is continuously increasing as a result of elevated capability of vehicle acceleration that increases the risk of driver's mistakes. Such road incidents directly impact the flow of traffic in such area and affect directly and indirectly to the economy, society, and environment. Incident monitoring and detection in Thailand is currently done by the responsible authority through CCTV and the traffic flow data from traffic flow measurement, both means of monitoring and detection have high operation costs. Online communication, on the other hand, has seen significant growth in the present days resulting in a fast growth of online social media use for various characteristic of communication replacing telephone calls. This article will present forms of incident detection from social media posts that have been data-mined from Twitter with autonomous API designed to screen for messages related to incident detection consisting of 4 steps. The experiment demonstrated the ability of the proposed method to detect incidents in Thai language with the accuracy of 85.80%, the detection rate (DR) of 78.83%, and false alarm rate (FAR) of 21.17%, based on the top 5 ranked keywords, out of 20 first keywords.

How to Cite

K. PUANGNAK and N. Rachsirivatcharabul, "Incident Detection Techniques for the Thai language on Twitter", *ECTI-CIT Transactions*, vol. 16, no. 3, pp. 313–323, Aug. 2022.

[More Citation Formats ▾](#)

[Download Citation ▾](#)

Issue

[Vol. 16 No. 3 \(2022\): ECTI Transactions on CIT \(Sep 2022\)](#)

Section

Research Article



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](#).

Information

[For Readers](#)

[For Authors](#)

[For Librarians](#)

Indexed In

ECTI Transactions on
 Computer and Information..



Scopus®



DOAJ | DIRECTORY OF OPEN ACCESS JOURNALS



Citescore by Scopus

0.1 2022 CiteScore

ภาพ ก-4 เอกสารตอบรับบทความ (Accepted Letter) ECTI-CIT Transactions - Volume. 16 –
Number 3 - August 2022

Your Submission - [EMID:fc84fccf89518c77] ภายนอก กล่องจดหมาย 📧 🔗

ECTI-CIT <em@editorialmanager.com> ถึงฉัน ส. 25 มิ.ย. 14:08 ☆ ← ⋮

🌐 อังกฤษ > ไทย แปลข้อความ เปิดสำหรับ: อังกฤษ

Ref.: Ms. No. ECTI-CIT-D-22-00012R1
Incident Detection Techniques for the Thai language on Twitter
ECTI Transactions on Computer and Information Technology

Dear Asst.Prof. PUANGNAK,

I am pleased to tell you that your work has now been accepted for publication in ECTI Transactions on Computer and Information Technology.

It was accepted on 2022-06-25 03:05:23

Comments from the Editor and Reviewers can be found below.

Thank you for submitting your work to this journal.


With kind regards

Krisana Chinnasarn
Editor
ECTI Transactions on Computer and Information Technology

Comments from the Editors and Reviewers:

Reviewer #1: The authors have addressed all the comments and I have no further comments.

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: <https://www.editorialmanager.com/ecti-cit/login.asp?a=r>). Please contact the publication office if you have any questions.





ISSN 2286 – 9131

Transactions

on Computer and Information Technology
Volume 16, No. 3, September 2022

Research Article

Two-Dimensional Error Correction Code for Spin-Transfer Torque Magnetic Random-Access Memory (STT-MRAM) Caches.....	<i>C. Duangthong, P. Supnithi and W. Phakphisut</i>	237
The Implementation of Discrete-Event Simulation and Demand Forecasting Using Temporal Fusion Transformers to Validate Spare Parts Inventory Policy for The Petrochemicals Industry.....	<i>S. Immuphat and M. Toahchoodee</i>	247
A Mathematical Model and PI Controller Design Based on Indirect Vector Control for Permanent Magnet Synchronous Motor.....	<i>N. Kongchoo, P. Santiprapan and N. Jindapetch</i>	259
The Collection of Road Traffic Incidents in Bangkok from Twitter Data based on Deep Learning Algorithm.....	<i>K. Puangnak and N. Rachsirivatcharabul</i>	267
Capturing Spatial Relationship Mapping Patterns between GPS Coordinates and Road Network Using Machine Learning and Partitioning Techniques.....	<i>T. Boonchoo</i>	277
Privacy Threats and Privacy Preservation Techniques for Farmer Data Collection based on Data Shuffling.....	<i>S. Riyana, N. Ito, T. Chaïya, U. Sriwichai, N. Dussadee, T. Chaichana, R. Assawarachan, T. Maneechukate, S. Tantikul and N. Riyana</i>	289
Optimized CNN Model with Derived Kemels for Apnea Classification.....	<i>S. A and S. M</i>	302
Incident Detection Techniques for the Thai language on Twitter.....	<i>K. Puangnak and N. Rachsirivatcharabul</i>	313
Hybrid Visual and Optimal Elliptic Curve Cryptography for Medical Image Security in Iot.....	<i>L. A. Kumar, S. Srivastava, S. R. Balaji, F. H. Shajin and P. Rajesh</i>	324
High-speed Firewall Rule Verification Technique Improves Throughput Performance for IP Version 6.....	<i>S. Khummanee, U. Saisangchan, K. Somsuk and S. Atsawaraungsuk</i>	338

Submission Guideline

In your manuscript, one author has been designated as the corresponding author with contact details:

- E-mail address
- Affiliation

Further considerations before your submission:

- Manuscript has been 'spell checked' and 'grammar checked'
- All references quoted in the Reference Section are cited in the body of the manuscript
- Permission for use of copyrighted materials from other sources must be declared
- The manuscript length must be at least 8 pages and no longer than 12 pages with two (2) columns.

Form more information visit: <https://ph01.tci-thaijo.org/index.php/ecticit/author-guideline>

<https://ph01.tci-thaijo.org/index.php/ecticit/>

ECTI
E
C
T
I



Incident Detection Techniques for the Thai language on Twitter

Korn Puangnak¹ and Natworapol Rachsirivatcharabul²

ABSTRACT

Nowadays, the rate of road incidents is continuously increasing as a result of the elevated capability of vehicle acceleration that increases the risk of driver's mistakes. Such road incidents directly impact the flow of traffic in an area and affect the economy directly and indirectly. These incidents also have an impact on society and the environment. Incident monitoring and detection in Thailand is currently done by the responsible authority through CCTV and traffic flow data from traffic flow measurement. Both monitoring and detection have high operation costs. Online communication, on the other hand, has seen significant growth in the present day resulting in a fast growth of online social media use for various characteristic of communication, replacing telephone calls.

This article will present forms of incident detection from social media posts that have been data-mined from Twitter with an autonomous API designed to screen for messages related to incident detection. But there are an enormous number of messages on Twitter. It will be difficult to detect only content that is relevant or of interest. Therefore, this research presents a collection of search terms that are related to incidence detection for the TF-IDF, Word2Vec, and Markov Chain methods. This incident detection method consists of 4 steps. The experiment demonstrated the ability of the proposed method to detect incidents in Thai language with a detection rate of 81.71%, and a false alarm rate of 10.85%, based on the top 5 ranked keywords out of a list of the 20 first keywords.

Article information:

Keywords: Twitter, Incident Detection, Word Cloud

Article history:

Received: December 31, 2021

Revised: June 4, 2022

Accepted: June 25, 2022

Published: August 6, 2022

(Online)

DOI: 10.37936/ecti-cit.2022163.247328

1. INTRODUCTION

Previous literature reviews related to incident management by traffic information analysis to determine the location or site of incident, indicates that most of the research focuses on application of various existing traffic information formats. The practice of such research started back in 2016 due to the rise of big data to the Petabyte-level, combined with global 4G infrastructure development at the same time along with 5G pioneering projects moving toward the fully functional Smart City concept. One of the research projects showed the rising popularity of E-bike usage in China due to its accessibility, low running cost, flexibility, and mobility, making it suitable for densely populated cities. However, 30 to 60 percent of road incidents involve E-bikes. Approximately 70 E-bikes were stolen on a daily basis in Wenzhou, so the gov-

ernment had to install low-energy transmitters on E-bikes that work with stations to track the stolen vehicles. Xiaoxia Jia, Peng Cheng, and Jiming Chen [1] developed an analysis and visualization system that aims for E-bike monitoring, mobility analysis in a web application framework, user behavior analysis, speed monitoring, and E-bike movement using the Mapbox API. At the same time, Wei Peng et al. [2] researched platforms for traffic management using distributed storage and parallel computing to monitor drivers' behaviors and interpret traffic status for relevant information. Parallel computing and clustering helped improve responsiveness in search and analysis of the large amount of data. The parallel computing clusters utilized Flume and Hive's database system and Apache Spark, similar to some other research. Weijian Sun et al. [3] researched specific mobility characteristics of 4G cellular network users and stated that

^{1,2}The authors are with Faculty of Engineering, Rajamangala University of Technology Phra Nakhon, Thailand, E-mail: Korn.p@rmutp.ac.th and Natworapol.b@rmutp.ac.th

at present, the mobility of users could be tracked from call detail records (CDR) or Wi-Fi while the users make calls or text messaging or use the Wi-Fi tracking function. However, it does not allow the researcher to access or understand the mobility of large-scale cellular network users, so the research aimed to specifically study the difference in mobility data obtained from a 4G network compared to those from CDR and 3G networks through processing of 6 Terabytes (TB) of big data from a 4G network. The result produced was more detailed than data obtained from CDR and 3G networks in similar research. Xiaoxia Wang and Zhanqiang Li [4] compiled visualization and analysis tools developed from previous research on smart traffic systems into an integrated platform and tested them with traffic data of Beijing, China, for assessing capability of big data processing. The test used the A* Algorithm for analysis. Moreover, research regarding traffic data was also developed into other uses by Zhiyng Cui et al. [5]. They improved the DRIVE Net developed by STAR Lab (US), an online digital roadway interactive visualization and evaluation network that provided data about traffic information on various roads. The improvement included "multi-sourced travel time analysis" with capability to display results in more formats. Mass transit data, motorcycle drivers' data, parking lot data, data from Car2go, shared data from various sources, and other data was collected, analyzed, and visualized through data analysis within the same year. Satyanarayana Nandury and Beneyaz Begum [6] investigated traffic system data handling in smart cities. They stated that handling big data from instruments, sensors, and various IoTs was becoming a challenge as the overall architecture of the system remained unclear. The SWIFT architecture was designed to serve as a connection between smart objects, smart devices, and smart systems. The research presented an experiment where the SWIFT architecture was tested with a large collection of traffic information data to determine various parameters such as traffic density, traffic signalling, parking management, navigation, and vehicle pollution monitoring.

On the other hand, other research about utilizing social media to navigate traffic information big data started back in 2013. Duckwon Chung et al. [7] presented an architecture used in massive real-time traffic information big data distributed complex event processing (CEP). The processing used a NoSQL database utilizing Hadoop and HBase to collect and analyze data from the California Highway Commission. The assessed data came from the performance measurement system (PeMS), collected from loop detectors, including speed, flow, occupancy, and from the TASAS (traffic accident surveillance and analysis system) which provided the coordinated date, time, characteristics, severity, and number of vehicles involved in incidents during the past 10 years. The 1

TB of data was used to summarize the probability of occurrence according to the upstream and downstream velocities, divided into intervals of 5 mph. The probability would then be tested in predicting future occurrence in I-880N highway, California. In the following year, the same process was repeated in Korea by In Jung Lee [8]. In 2015, Shen Zhan et al. [9] presented a methodology that combines Latent Dirichlet Allocation (LDA) with topic wording in text strings and document clustering to visualize the significance of each specific word of varying sizes to categorize incident reports in Twitter. Currently, there are approximately 200 million active users of Twitter and about 500 daily tweets. The first 3 words that would imply than an incident occurred were "vehicle", "driver", and "tow". The second words that would imply the situations were "call", "check", "assist", and "push". This research was tested in Seattle with DBSCAN algorithm, and the reports covered an area within a radius of approximately 600 meters from the site of incident. It was also found that the denser the area is, the closer the reports would be related to the site of incident. Theodore Georgiou et al. [10] summarized a collection of tweets related to traffic in areas of interest that contained relevant keywords such as "this traffic" or "on my way" in a time-based grouping called "Social Volume". The social volume was then compared to the actual traffic volume of California freeway I-405 and it was found that the interpretation was similar and contained a linear relationship. The ratio between traffic volume and social volume was grouped based on time period in a model called the "shift-based model" that allowed the traffic status to be measured with lower investment. Mochamad Vicky et al. [11] presented a design and development of natural language processing for Indonesian to distinguish words and interpret texts from Twitter with a tool called Twitter API. This was an important origin to the trend of design and development of incident detection from social media without installation or investment in additional instruments to obtain optimal benefit out of the significantly expanded social reporting volume.

2. DEFINITION AND CLUSTERING OF DATA

2.1 Clustering of News and Incident Reports

As a prerequisite, the data must be categorized and clustered into 2 groups: news reports and incident reports [12]. The process of data collection must also include the status of the report to indicate whether it falls into either of the clusters as described in Table 1. The test must also account for the accuracy in distinguishing between news and incident data.

Table 1: Traffic Relevance Determination.

Index	Interpretation	Description
0	News Report	General updates that are not relevant to traffic or are outdated such as public statements of organizations, local activity announcement, or posts about old incidents.
1	Incident Report	Reports about traffic-relevant incidents that are current or reports about resolution of such incidents.

2.2 Clustering of Occurrences

Reports from different sources usually have different formats. From past research, it was found that the researcher could categorize the reports into 3 clusters [13]. In this study, the author has defined reference “0” as undefined incidents to be excluded from the learning process. A further description is provided in Table 2.

Table 2: Categories of Occurrences.

Index	Interpretation	Description
0	Undefined	Occurrences unable to be categorized.
1	Traffic	Traffic volume or flow reports that affect severity clustering but do not describe the cause of the traffic volume or flow observed.
2	Incident	Undesired events resulting in temporary obstruction such as car crashes, breakdowns, or fallen objects.
3	Disaster	Natural occurrences that slow down or obstruct traffic flow such as floods, fallen timber, or mudslides.
4	Potholes	Occurrences that affect road surfaces requiring the drivers to adapt or decelerate such as holes and maintenance blockage.
5	Other	Other occurrences that could not be otherwise be categorized such as demonstrations, building collapses, and roadside fires.

2.3 Clustering of Severity

This study categorizes occurrences into 3 levels according to other past research [14-15] with an addi-

tional 0th level for an unidentifiable level of severity. A further description is given in Table 3.

Table 3: Severity of Occurrences.

Index	Interpretation	Description
0	Uncategorized	Occurrences without severity level or without clear event descriptions, making them unable to be categorized in the data collection process.
1	Normal	Low severity occurrences with no effect on traffic.
2	Intermediate	Occurrences with moderate severity that affect 1-2 traffic lanes or cause traffic flow slowdown.
3	Lane closure	Occurrences with high severity that result in total blockage or an area locked in a standstill.

3. INCIDENT DETECTION PROCESS FROM THAI LANGUAGE

The incident detection process from Thai language utilizes natural language processing like that which has already used in various languages around the world, but adapted for use with Thai language. There are multiple instances of NLP research for Thai language and more research being conducted continuously. Thai language is considered a continuous script without spaces or sentence terminators. Data is mined from messages on the Twitter platform. One of the ubiquitously used methods is to create a vocabulary from the frequency of words related to traffic. It was expected that there would be a large number of words involved and that a relevance checking method would be needed.

3.1 Message Collection and Recording

Data is collected from the Twitter accounts that regularly and mainly report about traffic situations such as @js100radio (an online extension from radio traffic reports with over 3 million followers), @Traffic_1197 (a twitter account under supervision of a traffic police department with over 20,000 followers), and @fm91trafficpro (a twitter account of the communications police department with over 2 million followers). The data was collected during the period from 5 February 2021 to 23 March 2021 with Twitter’s support by letting the developer access tweets through tools using the Twitter API. The data was received in the form of JSON-Objects comprising 3,400 messages. The data was then clustered according to each message’s nature using features such as relevance, categories, and degree of severity.

The messages for processing were analyzed by experts working on traffic surveillance in Thai governance such as Inter - City Motorway Division, Department of Highways (DOH), and Expressway Authority of Thailand (EXAT). To see how messages were categorized, the message groups before starting processing are shown in Table 4. Table 4 shows the example message in the first column, traffic relevance determination in the R column, categories of occurrences in the C column, and severity of occurrences in the S column, respectively.

Table 4: Example Grouping of Messages.

Message	R	C	S
16:52 ผู้ว่าฯ วีระศักดิ์ อำนวยการสืบค้นต่อเนื่อง ทุกอย่างเป็นไปเรียบร้อย ขณะรถบรรทุกปิดตัวผู้ขับขี่ขึ้น 90% ท้าวจิวประเทศไทย #FM91	0	0	0
13:56 อนุสรณ์สถาน ถ.พหลโยธิน มีขากออก ไปสี่แยกเมือง รด เคลื่อนตัวช้า สาเหตุปรับผิวจราจร "ทางขนาน" ปากทางเมือง เอก สี่แยกศรีนครินทร์ #รายงานจราจร #FM91	1	1	1
18:04 ถนนบรมราชชนนีขาเข้าเอื้องฯ สน.คลองจั่น รถกระบะชนกับ รถนั่งส่วนบุคคล ช่องทางด่วน เลนซ้าย มีเจ้าหน้าที่แล้ว	1	2	2
10:35 นำทีมฝนลาดพร้าว ปากซอยมหาดไทย	1	3	2
22:50 ถนนรามอินทรา ขาออก ผังตรงข้าม รพ.สินแพทย์ #เจ้าหน้าที่กำลังปฏิบัติงาน แขวงทางหลวงรังสิตไฟฟ้า #ปิดการจราจรฝั่งขาออก แล้วเบี่ยงให้ใช้เลนส่วนในฝั่งขาเข้าการจราจร ชลลลล // #FM91 #รายงานจราจร	1	4	3
01:18 หลังยุติการชุมนุมหน้า สน.ปทุมวัน สร้างความเสียหายกับรถทางราชการตำรวจ 8 คัน // #FM91	1	5	0

3.2 Word Tokenization Process

The raw data could not be used instantly because it existed in long continuous sentences. In order for the machine to be able to analyze the data, it needed to be tokenized. The word tokenization was done by a Python open-source library called PyThaiNLP which worked similarly to the natural language toolkit (NLTK), but the PyThaiNLP could be used for Thai language. Apart from tokenization, PyThaiNLP could also complete other tasks such as word correction, word categorization, and ordering. This research mainly utilizes the tokenization function or "word_tokenize(string.engine)". The parameters are "string" (the data to be tokenized), and "engine" (the method to be used in tokenization). A sample result is shown in Figure 1.

```
>>> message = "คนขับมีอาการชักเกร็ง ชีวรถบรรทุกในช่องทางกลางของสี่แยกของ บาดเจ็บ 2 คน"
>>> textCut = word_tokenize(message, engine='newmm')
>>> print(textCut)
>>> ['คนขับ', 'มี', 'อาการ', 'ชักเกร็ง', 'ชีวิ', 'รถบรรทุก', 'ใน', 'ช่องทาง', 'กลาง', 'ของ', 'สี่แยก', 'ของ', 'บาดเจ็บ', ':', '2', ':', 'คน']
```

Fig.1: Word Tokenization by PyThaiNLP.

3.3 Text Filtering Process

A string consists of various letters and symbols that do not affect the processing such as "#", "!", and various URLs. Such components are not needed in NLP and must be filtered out. This research uses regular expression search on the filtered text. It can be seen from Figure 2 that only text that included numbers 0-9 and Thai letters of ก-ฮ and อ-ค remained. The third section is a grouping of word cloud and term frequency.

ส.พ.ผู้ว่าฯ วีระศักดิ์ใน ทม.คลองจั่น 19 จากผู้ดูแลชาวเมียนมา นำ !! ผู้ดูแลผู้ว่าฯต้อง
มหน้าภาคคลองเวลา #FM91 #ผู้ว่าฯวีระศักดิ์ #คลองจั่น 19

['ส', 'พ', 'ผู้', 'ว่า', 'ฯ', 'วี', 'ระ', 'ศ', 'ก', 'ดี', 'ใน', 'ท', 'ม', 'ค', 'ล', 'อง', 'จ', 'ั่น', '1', '9', 'จาก', 'ผู้', 'ดูแล', 'ชาว', 'เม', 'ียน', 'มา', 'นำ', '!!', 'ผู้', 'ดูแล', 'ผู้', 'ว่า', 'ฯ', 'ต้อง', 'ม', 'หน้า', 'ภาค', 'คลอง', 'เวลา', '#', 'F', 'M', '9', '1', 'ส', 'พ', 'ผู้', 'ว่า', 'ฯ', 'วี', 'ระ', 'ศ', 'ก', 'ดี', 'ใน', 'ท', 'ม', 'ค', 'ล', 'อง', 'จ', 'ั่น', '1', '9', 'ผู้', 'ดูแล', 'ผู้', 'ว่า', 'ฯ', 'เม', 'ียน', 'มา', 'นำ', '!!', 'ผู้', 'ว่า', 'ฯ', '1', '9']

10.00 ภายหลังเลิกเขต ช่วง ช่วง บาง ร.ร. เพศ ศิรินทร์ จุฑ กลับรถ ได้ สะพานข้าม
คลอง 10.00 ภายหลังเลิกเขต ช่วง ช่วง บาง ร.ร. เพศ ศิรินทร์ จุฑ กลับรถ ได้ สะพานข้าม
คลอง 10.00 ภายหลังเลิกเขต ช่วง ช่วง บาง ร.ร. เพศ ศิรินทร์ จุฑ กลับรถ ได้ สะพานข้าม
คลอง 10.00 ภายหลังเลิกเขต ช่วง ช่วง บาง ร.ร. เพศ ศิรินทร์ จุฑ กลับรถ ได้ สะพานข้าม

Fig.2: Text Filtering for Pre-processing.

3.4 Term Frequency Determination

Term frequency determination was used to find the most frequently used words. This research focuses on word combinations that imply road incidents. 70% of the data was sampled to determine the term frequency related to the aspects of interest. Table 3 shows the data clusters with words and their relevance. Table 4 shows the data clusters with words indicating categories of the occurrences. Table 5 shows the data clusters with words indicating degree of severity of each occurrence.

Table 5: Word Relevance.

No	News Report		Incident Report	
	Word	Count	Word	Count
1	และ	465	แยก	986
2	หัว	436	รถ	831
3	ราย	328	อุบัติเหตุ	830
4	100	255	การจราจร	646
5	การ	244	ช่วง	614
6	เขต	234	จราจร	597
7	จาก	218	รายงาน	531
8	เวลา	203	ช่วง	529
9	บาง	202	ของ	526
10	100	197	เคลื่อนตัว	493
11	หลัง	194	เข้า	479
12	2564	179	คัน	477
13	วันที่	179	สะพาน	476
14	ได้	178	ขาเข้า	379
15	เพลิงไหม้	177	บาง	361
16	ชอย	177	ขาออก	355
17	ข่าว	168	คัน	321
18	แล้ว	161	จาก	289
19	ติดขัด	158	ถึง	288
20	ของ	156	กับ	285

Table 6: Word Categories (3 of 5 Categories).

No	Traffic		Incident		Disaster	
	Word	Count	Word	Count	Word	Count
1	แยก	335	อุบัติเหตุ	830	ซึ่ง	4
2	จราจร	229	แยก	589	กม.	4
3	รายงาน	216	รถติด	580	น้ำท่วม	3
4	รถติด	215	การจราจร	545	ช่วง	3
5	ขาเข้า	165	ช่วง	540	พายุ	2
6	สะพาน	147	ขวาง	524	ส่งผล	2
7	ขาออก	143	ช่อง	478	บริเวณ	2
8	ท้าย	140	กัน	466	การจราจร	2
9	อยู่	130	ช้า	431	สัญญาณ	2
10	สะสม	125	เคลื่อนตัว	430	ได้	2
11	มาก	118	คัน	316	จาก	2
12	ติด	93	สะพาน	291	สอดคล้อง	2
13	ท้ายแถว	91	กับ	284	145	2
14	ข้าม	79	จราจร	274	น้ำ	2
15	บาง	78	รายงาน	267	ป้าย	2
16	ลาดพร้าว	73	บาง	251	เลน	2
17	พระราม	68	ก่อน	231	ซ้าย	2
18	จาก	66	กลาง	225	ขวา	2
19	ได้	55	จอด	220	ช่องทาง	2
20	การจราจร	54	ทางซ้าย	218	มาก	2



Fig.5: Word Cloud of Disaster Categories.

Table 7: Severity levels of words.

No	Traffic		Incident		Disaster	
	Word	Count	Word	Count	Word	Count
1	แยก	335	อุบัติเหตุ	830	ซึ่ง	4
2	จราจร	229	แยก	589	กม.	4
3	รายงาน	216	รถติด	580	น้ำท่วม	3
4	รถติด	215	การจราจร	545	ช่วง	3
5	ขาเข้า	165	ช่วง	540	พายุ	2
6	สะพาน	147	ขวาง	524	ส่งผล	2
7	ขาออก	143	ช่อง	478	บริเวณ	2
8	ท้าย	140	กัน	466	การจราจร	2
9	อยู่	130	ช้า	431	สัญญาณ	2
10	สะสม	125	เคลื่อนตัว	430	ได้	2
11	มาก	118	คัน	316	จาก	2
12	ติด	93	สะพาน	291	สอดคล้อง	2
13	ท้ายแถว	91	กับ	284	145	2
14	ข้าม	79	จราจร	274	น้ำ	2
15	บาง	78	รายงาน	267	ป้าย	2
16	ลาดพร้าว	73	บาง	251	เลน	2
17	พระราม	68	ก่อน	231	ซ้าย	2
18	จาก	66	กลาง	225	ขวา	2
19	ได้	55	จอด	220	ช่องทาง	2
20	การจราจร	54	ทางซ้าย	218	มาก	2



Fig.3: Word Cloud of Traffic Categories.



Fig.4: Word Cloud of Incident Categories.



Fig.6: Word Cloud of Normal Level.



Fig.7: Word Cloud of Intermediate Level.

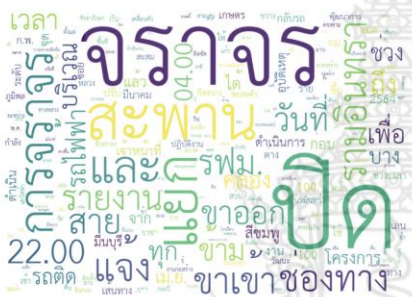


Fig.8: Word Cloud of Lane Closure Level.

3.5 Term Weight Determination

Term weight determination was used to group words by word weight to find semantic representations of text by converting the word or text of interest to a mathematical form, such as a vector of words in text. Using the word2vec techniques, similar vectors mean that the words have similar meanings. For example, accident, parking, and traffic jam have similar vector values, as shown in Figure 9.



Fig.9: Word Vector Technical.

3.6 Term Probabilities Determination

Term probability determination was used to find the probabilities of link words with the Markov Chain

method. The Markov Chain method finds the probability of the next word by matching the next word using the previous word. Which can be described as $S = \{S_1, S_2, S_3, \dots, S_n\}$ where S is probability of continuous group word and S_n is sequence word as shown in Figure 10. This will find the number of most frequently used words from Table 8 showing the number of most frequently used words.

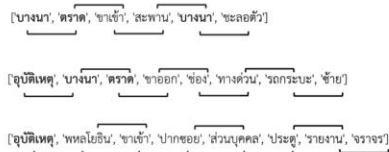


Fig.10: Match to Find the Probability of the Next Word.

Table 8: Severity levels of words.

S	S ₁	S ₂	S ₃	Frequently
บางนา	ตราด	ชะลอตัว	ตราด	3
ตราด	ขาเข้า	ขาออก	-	2
สะพาน	บางนา	-	-	1
อุบัติเหตุ	บางนา	พหลโยธิน	-	2
ขาออก	ช่อง	-	-	1
ทางด่วน	รถกระบะ	-	-	1
รถกระบะ	ซ้าย	-	-	1
พหลโยธิน	ขาเข้า	-	-	1
ขาเข้า	ปากซอย	-	-	1
ปากซอย	ส่วนบุคคล	-	-	1
ประตู	รายงาน	-	-	1
รายงาน	จราจร	-	-	-

3.7 The Process for Selecting the Best Search Terms

The process of determining the best search words uses the best word test results to detect incidence and to distinguish patterns. To choose the best word, this research compares the grouping of words using three different methods. That consist of the method as below.

The TF-IDF method can generate the best possible event to detection term in 6 formats consist of traffic relevant, traffic incident, disaster, potholes, etc.,

The Word2Vec method can generate the best severity level of an incident. consists of Normal and Lane Closure.

The Markov Chain method can generate the best severity level of an incident for one pattern is Intermediate.

The best terms are selected for further use in the process of detecting and distinguishing incidence patterns as detailed in Table 9.

Table 9: Word Detection Technique Comparison.

Detection Group	Word Detection Technique (%)		
	TF	Word2vec	Markov Chain
News	85.50	85.20	80.34
Traffic	52.80	52.40	51.00
Incident	72.16	72.00	71.12
Disaster	88.82	87.53	54.53
Potholes	82.10	75.54	44.85
Etc.	57.05	55.96	54.67
Normal	55.80	61.43	48.90
Intermediate	67.00	64.21	68.00
Lane Closure	52.40	53.47	50.54

4. EXPERIMENT

4.1 Hypothesis

The term frequency is tested to determine the optimum word number to be used in Thai language for data clustering according to the data type. This experiment uses 70% of the data for term frequency determination and the remaining 30% of the data for testing. Twenty words from each cluster are then used to generate the processes for word search. The experiment is divided into 3 levels: 1. Determining whether a keyword exists in the string; 2. Keyword selection from 1 to 20 keywords per set. All keywords must be used in a completed round of test; and 3. Keyword selection from 1 to 20 keywords per set to determine condition that provide the most accuracy in each round of test. The accuracy ranges from 10 to 100 percent. The test is done in 3 clusters. Cluster 1 is used to determine the relevance of occurrences. Cluster 2 is used for categorization of occurrences. Cluster 3 is used to determine the degree of severity for occurrences.

4.2 Performance Benchmark

In this experiment, the performance of each method is evaluated by using two parameters which come from the Detection Rate (DR) that measures correction of incident detection. The calculation is shown in equation (1), and the corresponding False Alarm Rate (FAR) is shown in equation (2). D_n is number of incidents detected by this method, D_t is the number of incidents of this experiment and, N_f is the number of incidents in this experiment.

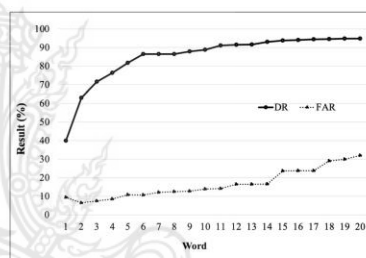
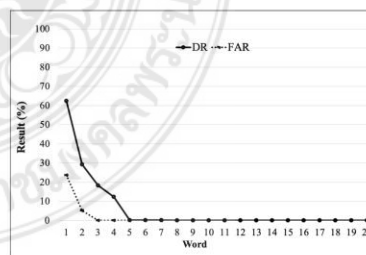
$$DR = \frac{D_n}{D_t} \times 100 \quad (1)$$

$$FAR = \frac{N_f}{D_t} \times 100 \quad (2)$$

4.3 Results

The test is done in 3 clusters where cluster 1 is the determination of the relevance of occurrences. Using the three test methods, Table 9 shows the de-

tection group of relevant occurrences, categories of occurrences, and severity of occurrences. To find the message group that is proper for detecting an incident with a condition, the test collects groups of messages of 20 words for each detection. For the detection process, it does not need to test with all words. To detect and categorize incidents into a message group, this research tests by increasing word every testing round starting with 1 word until 20 words are used in the last testing round. Each test uses the main three methods which are the "some keywords method" (detect the message with 1 word in sentences), "all keywords method" (detect the message with all word in sentence), and "ratio keywords method" (detect the message with word ratio in sentence such as 50% detection in 10 target word in sentence that means 5 words should be detected in this sentence). The evaluation in this research considers all three methods for indicating the highest performance processing auto detection.

**Fig.11:** Incident Detection Result (Some Keywords).**Fig.12:** Incident Detection Result (All Keywords).

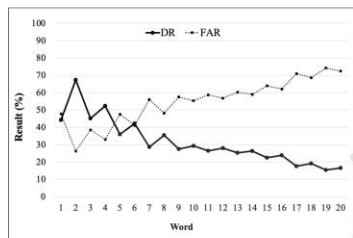


Fig.13: Incident Detection Result (50% Ratio).

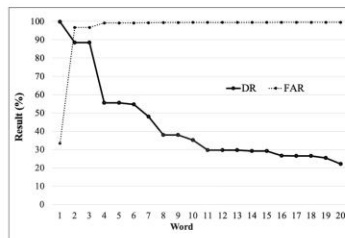


Fig.16: Categories: Disaster Report (Some Keywords).

Cluster 2 is the categorization of occurrences based on 5 main characteristics. The incident detection results from searches where at least 1 term is present are displayed in Figure 11. The number of keywords that result in the most accurate output is 5 as demonstrated in Figure 14. Figures 15-17 show that the optimum term frequency is 5. Figure 18 shows that incident detection results from searches where all terms must be present are significantly less accurate. The traffic status report accuracy is optimal at 3 keywords without the ability to detect or categorize the text string. Figure 19 displays the result of detection with ratio. The optimal result comes from 3 search terms at the ratio of 30%, as show in Figure 20.

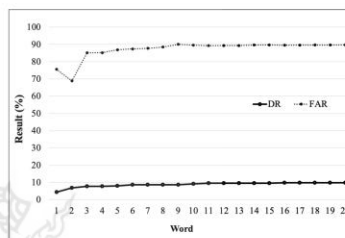


Fig.17: Categories: Potholes Report (Some Keywords).

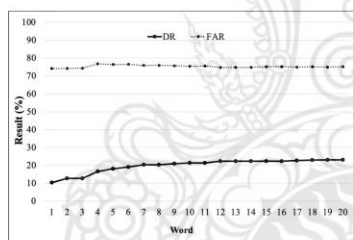


Fig.14: Categories: Traffic Report (Some Keywords).

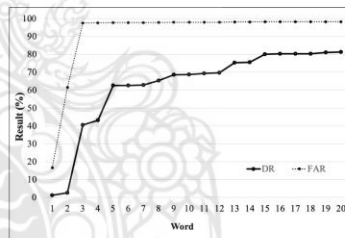


Fig.18: Categories: Etc (Some Keywords).

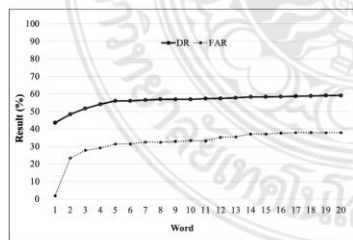


Fig.15: Categories: Incident Report (Some Keywords).

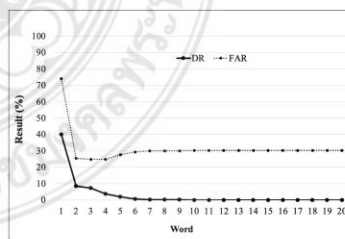


Fig.19: Incident Categories (All Keywords).

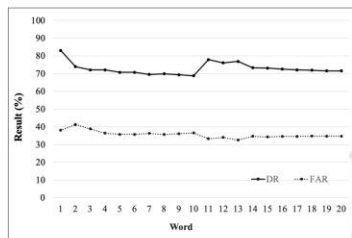


Fig.20: Incident Categories (30% Ratio Keywords).

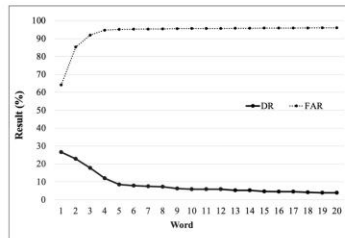


Fig.23: Incident Severity: Lane Closure (Some Keywords).

Cluster 3 is the result of the detection of text strings that indicate the severity of occurrences, divided into 3 degrees. Figure 21 shows the result of detection of occurrences that do not impact the traffic with the optimum of 5 keywords. Figure 22 shows the result of detection of occurrences that moderately impact the traffic with the optimum of 5 keywords. Figure 23 shows the result of detection of occurrences that result in total blockage of traffic with the optimum of 5 keywords. Figure 24 shows that the optimum for searches where all terms must be present is also 5 keywords. From the test, the optimal result comes from 8 search terms at the ratio of 30%, as shown in Figure 25.

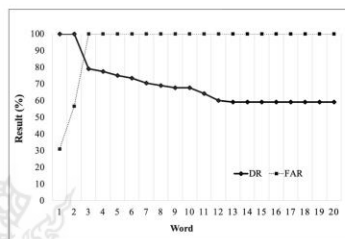


Fig.24: Incident Severity (All Keywords).

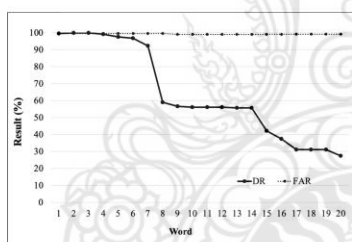


Fig.21: Incident Severity: Normal (Some Keywords).

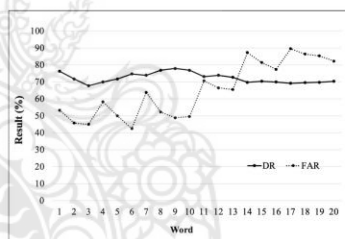


Fig.25: Incident Severity (30% Ratio Keywords).

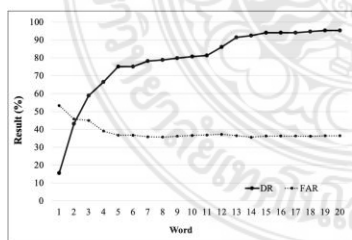


Fig.22: Incident Severity: Intermediate (Some Keywords).

Table 10: Incident Classification Results.

Amount of Word	Incident Classification (%)		
	Categories	DR (%)	FAR (%)
1	Traffic Relevant	10.34	74.13
	Traffic	43.54	1.79
	Incident	99.80	33.33
	Disaster	4.37	75.56
	Potholes	1.19	16.67
5	Traffic Relevant	18.09	76.36
	Traffic	56.06	31.39
	Incident	55.47	99.12
	Disaster	7.95	86.89
	Potholes	62.62	97.78

Table 11: Incident Detection Results.

Amount of Word	Incident Detection Event	
	DR (%)	FAR (%)
1	39.96	9.46
2	63.02	6.49
3	71.57	7.46
4	76.34	8.57
5	81.71	10.85

Table 12: Incident Severity Results.

Categories	Incident Severity		
	Amount of Word	DR (%)	FAR (%)
Normal	1	99.55	99.55
	2	100.00	99.59
	3	100.00	99.60
	4	99.21	99.60
	5	97.65	99.61
Intermediate	1	15.57	53.36
	2	43.10	45.81
	3	58.92	45.00
	4	66.52	39.06
	5	75.06	36.74
Lane Closure	1	26.56	64.10
	2	22.78	85.32
	3	17.78	91.96
	4	11.94	94.74
	5	8.44	95.12

Table 13: First Five Words for Incident Detection.

Detection Group	Word				
Traffic Relevant	แยก	รถติด	อุบัติเหตุ	การจราจร	ช่วง
Traffic	แยก	จราจร	รายงาน	รถติด	ข้างหน้า
Incident	อุบัติเหตุ	แยก	รถติด	การจราจร	ช่วง
Disaster	ช่วง	ถนน	น้ำท่วม	ช่วง	พายุ
Potholes	จราจร	ปิด	แยก	รายงาน	ข้างหน้า
Etc.	ชุมชน	ผู้	การจราจร	เส้นทาง	แยก
Normal	อุบัติเหตุ	แล้ว	รถ	ส่ง	ย้าย
Intermediate	อุบัติเหตุ	รถติด	แยก	การจราจร	ช่วง
Lane Closure	ปิด	จราจร	สะพาน	แยก	การจราจร

4.4 Analysis of the Results

In the experiment, the 20 most frequent words were taken from a grouping process using various methods. Each method was fed a specified group of words, and produced a specified pattern incidence, a classification of incidence patterns, and the severity of the incident most accurate was employed to test the efficiency of detecting and distinguishing different types of incidence patterns. This experiment confirmed that the messages received from Twitter as incident messages can be used to determine kinds of event and how severity. This is done by checking whether the message contains words that come from a given word group or not. It starts with a test on the

condition that if there is only one word in the text, it is considered an event or event pattern according to that word grouping. And then gradually the words checked increases to 2 and 3 words until all words are processed. The test results are shown in Figures 11 to 25. The test results show the effectiveness of the method. Using a wide variety of different number of words. In Tables 10 to 12, from the above test results it can be concluded that using five words from a group of words ranked by frequency can best identify and distinguish the patterns and severity of the incidences, as shown in Table 13.

5. CONCLUSION

This research proposes incident detection through the method searching text messages from Twitter social media by selecting the words used in the search based on word frequencies, using TF-IDF, Word2Vec, and Markov Chain methods. This gives appropriate word groups to find and distinguish events to lead to incidence detection. Using four processes that consist of: 1. Data collection and clustering; 2. Data preparation by filters; 3. Compilation and ordering of keywords related to incidents; and 4. Determination of the appropriate number of search terms for incident detection through social media posts. The search results can be used for analysis and determination of road incidents. The method determines whether they are relevant or not, are of what category (such as traffic slowdown, incidents, disasters, and potholes), and how severe they are based on 3 degrees of severity: normal, intermediate, and lane closure. The determination could be done by determining the frequency of keywords and grouping them in Word Clouds together with comparison of the ratio of appearing keywords. The most accurate setting will result in analysis that can be used to report incidents through a single reporting platform to the public.

The results from the experiment show that the proposed method can use the collected key terms to detect and categorize the terms with a DR of 81.71% and a FAR of 10.85%. This research contributes to the development of natural language processing, especially in regard to road incidents and traffic conditions. It will serve as an important steppingstone for future research that might need to use word clouds for language processing. Fast and effective data mining about road incidents on Twitter will help reduce time for data preparation and analysis for future research.

References

- [1] X. Jia, P. Cheng and J. Chen, "A data analysis and visualization system for large-scale e-bike data," *2016 IEEE International Conference on Big Data (Big Data)*, pp. 3998-4000, 2016.
- [2] W. Peng, Y. Li, B. Li and X. Zhu, "An Analysis Platform of Road Traffic Management

- System Log Data Based on Distributed Storage and Parallel Computing Techniques.” *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*, pp. 585-589, 2016.
- [3] W. Sun, D. Miao, X. Qin and G. Wei, “Characterizing User Mobility from the View of 4G Cellular Network,” *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, pp. 34-39, 2016.
- [4] X. Wang and Z. Li, “Integrated platform for smart traffic big data,” *2016 International Conference on Logistics, Informatics and Service Sciences (LISS)*, pp. 1-6, 2016.
- [5] Z. Cui, S. Zhang, K. C. Henrickson and Y. Wang, “New progress of DRIVE Net: An E-science transportation platform for data sharing, visualization, modeling, and analysis,” *2016 IEEE International Smart Cities Conference (ISC2)*, pp. 1-2, 2016.
- [6] S. V. Nandury and B.A. Begum, “Strategies to handle big data for traffic management in smart cities,” *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 356-364, 2016.
- [7] D. Chung, X. Rui, D. Min and H. Yeo, “Road traffic big data collision analysis processing framework,” *2013 7th International Conference on Application of Information and Communication Technologies*, pp. 1-4, 2013.
- [8] I. J. Lee, “Big data processing framework of road traffic collision using distributed CEP,” *The 16th Asia-Pacific Network Operations and Management Symposium*, pp. 1-4, 2014.
- [9] S. Zhang, “Using Twitter to Enhance Traffic Incident Awareness,” *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pp. 2941-2946, 2015.
- [10] T. Georgiou, A.E. Abbadi, X. Yan and J. George, “Mining complaints for traffic-jam estimation: A social sensor application,” *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 330-335, 2015.
- [11] M.V. G. Aziz, A.S. Prihatmanto, D. Henriyan and R. Wijaya, “Design and implementation of natural language processing with syntax and semantic analysis for extract traffic conditions from social media data,” *2015 5th IEEE International Conference on System Engineering and Technology (ICSET)*, pp. 43-48, 2015.
- [12] Y. Chen, Y. Lv, X. Wang, L. Li and F. -Y. Wang, “Detecting Traffic Information from Social Media Texts with Deep Learning Approaches,” in *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 3049-3058, 2019.
- [13] A. Agarwal and D. Toshniwal, “Face off: Travel Habits, Road Conditions and Traffic City Characteristics Bared Using Twitter,” in *IEEE Access*, vol. 7, pp. 66536-66552, 2019.
- [14] E. Žunić, A. Djedović and D. Donko, “Application of Big Data and text mining methods and technologies in modern business analyzing social networks data about traffic tracking,” *2016 XI International Symposium on Telecommunications (BIHTEL)*, Sarajevo, pp. 1-6, 2016.
- [15] A. Mulyana, H. Hindersah and A. S. Prihatmanto, “Gamification design of traffic data collection through social reporting,” *2015 4th International Conference on Interactive Digital Media (ICIDM)*, pp. 1-4, 2015.



transport system).

Korn Puangnak received his Master degree from Department of Computer Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. in 2011. Currently, he is an assistant professor at Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Phra Nakhon, Thailand. His research falls in the domain of machine learning, and ITS (intelligent

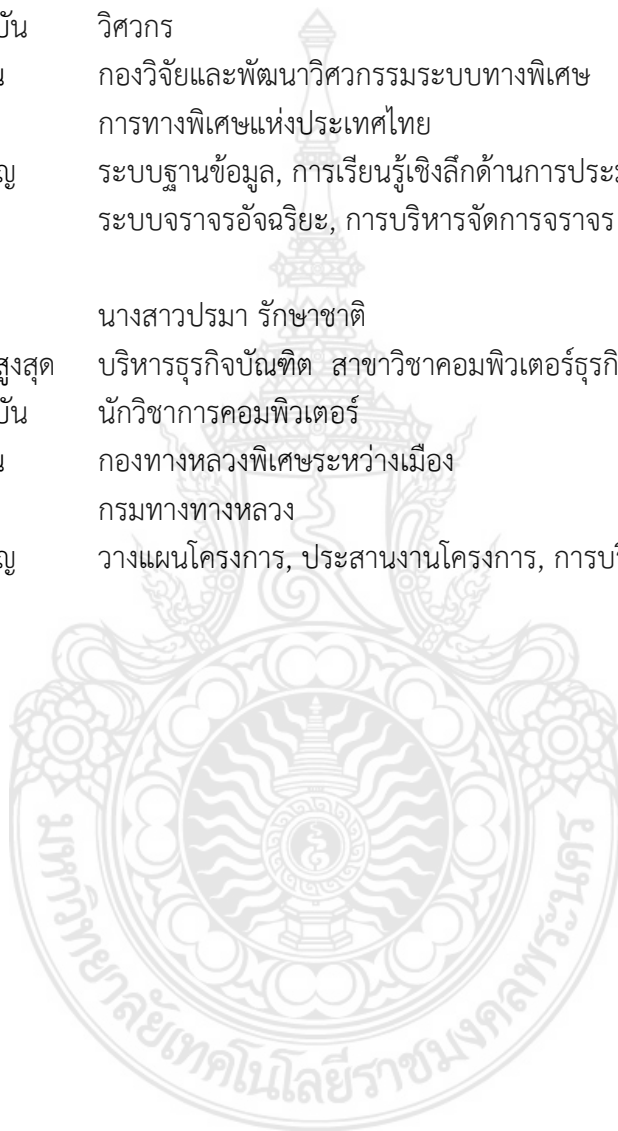


Natworapol Rachsiriwatcharabul received his Ph.D. degree from Engineering Management, University of Missouri, USA. in 1999. Currently, he is the President at Rajamangala University of Technology Phra Nakhon, Thailand. His research falls in the domain of energy management, transportation, smart city, and sustainable industrial management.

ภาคผนวก ข

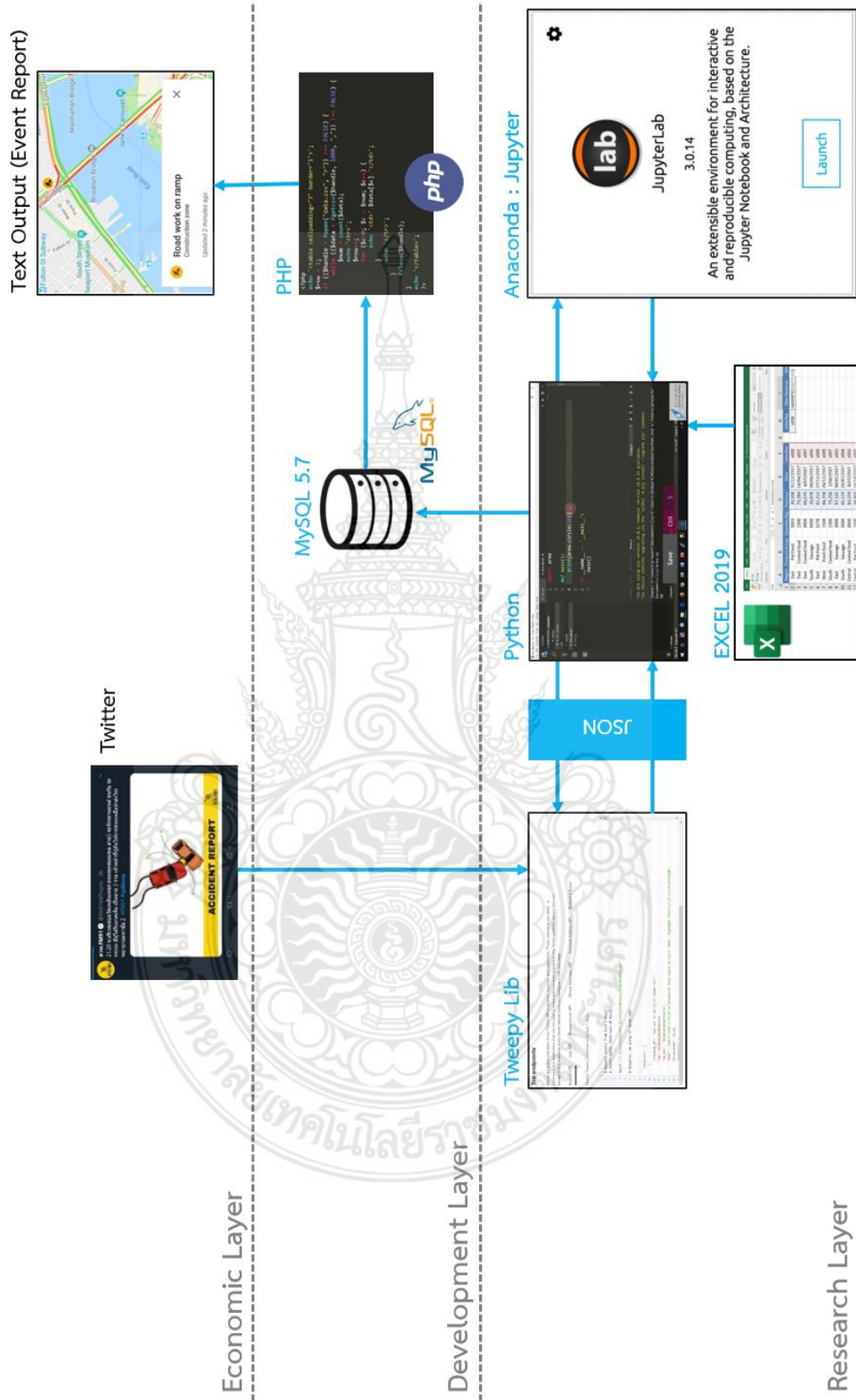
รายชื่อผู้เชี่ยวชาญ

- | | |
|--------------------|---|
| 1. ชื่อ นามสกุล | นายวัชชัย รักษาชาติ |
| วุฒิการศึกษาสูงสุด | วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ |
| ตำแหน่งปัจจุบัน | วิศวกร |
| สถานที่ทำงาน | กองวิจัยและพัฒนาวิศวกรรมระบบทางพิเศษ
การทางพิเศษแห่งประเทศไทย |
| ความเชี่ยวชาญ | ระบบฐานข้อมูล, การเรียนรู้เชิงลึกด้านการประมวลผลภาษาธรรมชาติ
ระบบจราจรอัจฉริยะ, การบริหารจัดการจราจร |
| | |
| 2. ชื่อ นามสกุล | นางสาวปรมา รักษาชาติ |
| วุฒิการศึกษาสูงสุด | บริหารธุรกิจบัณฑิต สาขาวิชาคอมพิวเตอร์ธุรกิจ |
| ตำแหน่งปัจจุบัน | นักวิชาการคอมพิวเตอร์ |
| สถานที่ทำงาน | กองทางหลวงพิเศษระหว่างเมือง
กรมทางหลวง |
| ความเชี่ยวชาญ | วางแผนโครงการ, ประสานงานโครงการ, การบริหารจัดการจราจร |



ภาคผนวก ค
สถาปัตยกรรมระบบงาน





Economic Layer

Development Layer

Research Layer

ประวัติการศึกษาและการทำงาน

ชื่อ นามสกุล นาย กร พวงนาค
วัน เดือน ปีเกิด 13 ตุลาคม 2526
ภูมิลำเนา กรุงเทพมหานคร

ประวัติการศึกษา

วุฒิการศึกษา	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วิศวกรรมศาสตรมหาบัณฑิต (วิศวกรรมคอมพิวเตอร์)	สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง (สจล.)	2554
วิศวกรรมศาสตรบัณฑิต (วิศวกรรมคอมพิวเตอร์)	สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง (สจล.)	2549

ตำแหน่งและสถานที่ทำงานปัจจุบัน

รองอธิการบดีฝ่ายพัฒนาดิจิทัล
มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
1381 ถนนประชากรราษฎร์ 1 แขวงวงศ์สว่าง เขตบางซื่อ กทม. 10800
โทรศัพท์ 02-836-3000

