ระบบทำนายความเคลื่อนไหวของกลุ่มคนจากสื่อสังคมออนไลน์

CARAVAN: CROWD FLOW PREDICTION FROM SOCIAL MEDIA SYSTEM

พิทยารัตน์ แซ่โง้ว
ศกลวรรณ นาวาเจริญ

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2559

ระบบทำนายความเคลื่อนไหวของกลุ่มคนจากสื่อสังคมออนไลน์

CARAVAN: CROWD FLOW PREDICTION FROM SOCIAL MEDIA SYSTEM

พิทยารัตน์ แซ่โง้ว

ศกลวรรณ นาวาเจริญ

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2559

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2559

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง ระบบทำนายความเคลื่อนไหวของกลุ่มคนจากสื่อสังคมออนไลน์

CARAVAN: CROWD FLOW PREDICTION FROM SOCIAL MEDIA SYSTEM

ผู้จัดทำ

1. นางสาวพิทยารัตน์ แซ่โง้ว รหัสนักศึกษา 56010858
2. นางสาวศกลวรรณ นาวาเจริญ รหัสนักศึกษา 56011171

_____อาจารย์ที่ปรึกษา

(ดร.อรทัย สังข์เพ็ชร)

_____อาจารย์ที่ปรึกษา

(ดร.อักฤทธิ์ สังข์เพ็ชร)

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบทำนายความเคลื่อนไหวของกลุ่มคนจากสื่อสังคมออนไลน์

นางสาวพิทยารัตน์	แซ่โง้ว	56010858
นางสาวศกลวรรณ	นาวาเจริญ	56011171
ดร.อรทัย	สังข์เพ็ชร	อาจารย์ที่ปรึกษา
ดร.อักฤทธิ์	สังข์เพ็ชร	อาจารย์ที่ปรึกษาร่วม

ปีการศึกษา 2559

บทคัดย่อ

ปัญหาหนึ่งของการเดินทาง เช่น การใช้บริการขนส่งมวลชน ผู้ใช้ทราบเพียงระยะเวลาที่ใช้เดินทางจากสถานีต้นทางไปยังสถานีปลายทาง แต่ไม่ทราบจำนวนผู้ใช้บริการในแต่ละสถานี ทำให้เสียเวลาในการรอคิว หนึ่งในข้อมูลสำคัญที่ผู้เดินทางควรทราบเพื่อวางแผนการเดินทาง คือ สถานที่ที่ต้องการจะไปนั้นมีความหนาแน่นของกลุ่มคนมากน้อยเพียงใด และมีทิศทางการเคลื่อนไหวอย่างไร ซึ่งในปัจจุบันยังไม่มีระบบที่สามารถนำเสนอข้อมูลเหล่านี้ได้ ถ้าผู้ใช้ทราบข้อมูลเหล่านี้ ก็จะช่วยในการวางแผนการเดินทางได้มีประสิทธิภาพขึ้น คณะผู้จัดทำจึงสร้างระบบทำนายความเคลื่อนไหวของกลุ่มคนขึ้นมา โดยระบบนี้ใช้ข้อมูลจากสื่อสังคมออนไลน์ที่ได้รับความนิยมอย่างมากในปัจจุบัน ร่วมกับกระบวนการ วิเคราะห์ และประมวลผล เพื่อทำให้ผู้ใช้เหล่านี้สามารถนำข้อมูลการเคลื่อนไหวของกลุ่มคนที่ได้นี้ไปใช้ เพื่อวางแผนการเดินทางให้ตรงกับความต้องการได้

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

CARAVAN: CROWD FLOW PREDICTION FROM SOCIAL MEDIA SYSTEM

Ms. Pittayarat	Saengow	56010858
Ms. Sakonwan	Navacharoen	56011171
Dr. Orathai	Sangpetch	Advisor
Dr. Akkarit	Sangpetch	Co-Advisor

Academic Year 2016

ABSTRACT

One of challenges in using public transportation is that users do not know how busy the transportation will be when they travel. Users only know the distance and the amount of time that takes to travel from the origin to the destination. During certain periods of time, users may have to wait for a long time due to a massive number of people. As a result, users may be late for important work or waste their time in a long waiting line. If users know how crowd flows at any moment, they can take this information into their plan, making it more efficient in travelling. Nowadays, there is not such system. In this project, we plan to create a model to predict the crowd flow by using the data from social media to analyze the number of people at a certain place and the direction they are likely to move to. This project should help users to commute more efficiently.

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดีด้วยความช่วยเหลือจากหลายฝ่ายทั้งในทางตรงและทางอ้อม โดยจะสำเร็จลงไม่ได้หากปราศจากความช่วยเหลือของบุคคลเหล่านี้

อาจารย์ที่ปรึกษาทั้งสองท่านคือ ดร.อรทัย สังข์เพ็ชร และ ดร.อักฤทธิ์ สังข์เพ็ชร ซึ่งท่านเป็นผู้ที่ให้คำแนะนำ คำปรึกษา ข้อคิดเห็นอันเป็นประโยชน์ต่อการพัฒนาผลงานรวมถึงคอยติดตามดูแลการดำเนินงานอย่างใกล้ชิดเพื่อที่จะทำให้การทำงานต่าง ๆ เป็นไปอย่างราบรื่นและออกมามีประสิทธิภาพสูงสุด

อาจารย์และบุคลากรในภาควิชาวิศวกรรมคอมพิวเตอร์ที่ได้ให้คำแนะนำและถ่ายทอดความรู้มาตลอด

ในท้ายที่สุดนี้ขอขอบคุณบิดา มารดา และครอบครัวที่ได้อบรมเลี้ยงดูและสั่งสอน พร้อมทั้งให้โอกาสในการศึกษาและให้กำลังใจเสมอมา

พิทยารัตน์ แซ่โง้ว

ศกลวรรณ นาวาเจริญ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

สารบัญ IV

สารบัญตาราง VI

สารบัญรูป VII

บทที่ 1 บทนำ 1

1.1 ความเป็นมาของปัญหา 1

1.2 วัตถุประสงค์ของโครงงาน 2

1.3 ประโยชน์ที่คาดว่าจะได้รับ 2

1.4 ขอบเขตของโครงงาน 2

1.5 สมมติฐานของโครงงาน 3

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง 4

2.1 ทฤษฎีที่เกี่ยวข้องด้านการจัดการข้อมูล 4

2.2 งานวิจัยที่เกี่ยวข้อง 8

บทที่ 3 การวิเคราะห์และออกแบบระบบ 10

3.1 โครงสร้างของระบบ 10

บทที่ 4 การทดลองและผลการทดลอง 25

4.1 การทดสอบการดึงข้อมูลจาก Twitter 25

4.2 การทดสอบการดึงข้อมูลจาก Foursquare 27

4.3 การทดสอบการดึงข้อมูลจาก Facebook 33

4.4 การทดลองหาความสัมพันธ์ของข้อมูลที่ได้จากสื่อสังคมออนไลน์ 36

4.5 การทดลองหาจำนวนวันที่ใช้เพื่อเป็นข้อมูลต้นในการทำนาย 40

4.6 การทดลองหาค่าความแม่นยำของการทำนายความหนาแน่นในช่วง 5 นาทีถัดไปของแต่ละสถานที่ 44

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.7 การทดลองหาค่าความแม่นยำของการทำนายความหนาแน่นในช่วง 15 นาทีถัดไปของแต่ละสถานที่ 46
4.8 การทดลองการทำนายความเคลื่อนไหวของกลุ่มคน 48

บทที่ 5 บทสรุปและข้อเสนอแนะ 49
5.1 บทสรุป 49
5.2 ปัญหา และแนวทางแก้ไข 49
5.3 แนวทางการพัฒนาต่อ 50

บรรณานุกรม 51

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตาราง หน้า

2.1 ตารางแสดงการแบ่งช่วงเวลาของแต่ละสัปดาห์ 8

3.1 ตารางแสดง ENTITY ของข้อมูลจาก TWITTER 13

3.2 ตารางแสดง ENTITY ของข้อมูลจาก FOURSQUARE 13

3.3 ตารางแสดง ENTITY ของข้อมูลจาก FACEBOOK 13

3.4 ตัวอย่างตารางที่ได้จากขั้นตอน PERIOD SLICING 15

3.5 ตัวอย่างตารางข้อมูลที่ได้จากขั้นตอน FREQUENCY COUNTING 15

3.6 ตัวอย่างกระบวนการทำนายจำนวนคน 16

3.7 ตัวอย่างกระบวนการทำนายจำนวนคนเมื่อพบค่าที่ตรงกันหรือใกล้เคียงมากกว่า 1 ค่า ในกรณีพบค่าที่ตรงกันหรือใกล้เคียงมากกว่า 1 ค่าของช่วงเวลา 5 นาทีถัดไป 17

3.8 ตัวอย่างกระบวนการทำนายจำนวนคนเมื่อพบค่าที่ตรงกันหรือใกล้เคียงมากกว่า 1 ค่า ในกรณีพบค่าที่แตกต่างกัน ค่าละ 1 ครั้งของช่วงเวลา 5 นาทีถัดไป 18

3.9 ตัวอย่างตารางแสดงช่วงความหนาแน่น 19

4.1 จำนวนการเช็คอินในช่วงวันและเวลาต่างๆที่ สยามพารากอน ของ FOURSQUARE 40

4.2 ช่วงความหนาแน่นของการเช็คอินในช่วงวันและเวลาต่างๆ ที่ สยามพารากอน ของ FOURSQUARE 41

4.3 ผลการทำนายในช่วงวันและเวลาต่างๆ ที่ สยามพารากอน ของ FOURSQUARE 41

4.4 ค่าความแม่นยำเมื่อใช้จำนวนวันเป็นข้อมูลต้นแตกต่างกัน 42

4.6 เปอร์เซ็นต์ความแม่นยำของการทำนายความหนาแน่นในช่วง 5 นาทีถัดไป ในแต่ละสถานที่.. 44

4.7 เปอร์เซ็นต์ความแม่นยำของการทำนายความหนาแน่นในช่วง 15 นาทีถัดไป ในแต่ละสถานที่ 46

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูป หน้า

2.1 APACHE SPARK SYSTEM 5

2.2 รูปจำลองการเก็บข้อมูลแบบแยกตามแถว และคอลัมน์ 7

3.1 โครงสร้างของระบบ 10

3.2 ENTITY RELATIONSHIP DIAGRAM ของฐานข้อมูลจากสื่อสังคมออนไลน์ 12

3.3 กระบวนการในทำนายความหนาแน่น และความเคลื่อนไหวของกลุ่มคน 14

3.4 แสดงการทำนายความเคลื่อนไหวของกลุ่มคน ขั้นตอนที่ 1 20

3.5 แสดงการทำนายความเคลื่อนไหวของกลุ่มคน ขั้นตอนที่ 2 21

3.6 แสดงการทำนายความเคลื่อนไหวของกลุ่มคน ขั้นตอนที่ 3 21

3.7 แสดงการทำนายความเคลื่อนไหวของกลุ่มคน ขั้นตอนที่ 4 22

3.8 แสดงการทำนายความเคลื่อนไหวของกลุ่มคน ขั้นตอนที่ 5 22

3.9 แอปพลิเคชันแสดงความหนาแน่นของกลุ่มคน 23

3.10 แอปพลิเคชันแสดงทิศทางการเคลื่อนไหวของกลุ่มคน 24

3.11 แอปพลิเคชันแสดงการเลือกช่วงเวลาในการทำนาย 24

4.1 กราฟแสดงจำนวนคนที่เช็คอินที่ สยามพารากอน ในแต่ละช่วงเวลาของแต่ละวันจาก FOURSQUARE 36

4.3 กราฟแสดงจำนวนรูปที่เช็คอินที่ สยามพารากอน ในแต่ละช่วงเวลาของแต่ละวันจาก FOURSQUARE 38

4.4 กราฟแสดงจำนวนทวิตที่กล่าวถึง สยามพารากอน ในแต่ละช่วงเวลาของแต่ละวันจาก FOURSQUARE 39

4.5 กราฟแสดงเปอร์เซ็นต์ความแม่นยำตามจำนวนวันที่ใช้ 43

4.6 กราฟแสดงเปอร์เซ็นต์ความแม่นยำของการทำนายความหนาแน่นในช่วง 5 นาทีถัดไป ในแต่ละสถานที่ 45

4.7 กราฟแสดงเปอร์เซ็นต์ความแม่นยำของการทำนายความหนาแน่นในช่วง 10 นาทีถัดไป ในแต่ละสถานที่ 47

4.8 กราฟแสดงจำนวนครั้งที่ทำนายว่ามีการเคลื่อนไหวของกลุ่มคนในช่วงเวลาต่างๆ ของสยามพารากอน 48

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

ปัญหาหนึ่งของการเดินทาง เช่น การใช้บริการขนส่งมวลชน ผู้ใช้ทราบเพียงระยะเวลาที่ใช้เดินทางจากสถานีต้นทางไปยังสถานีปลายทาง แต่ไม่ทราบจำนวนผู้ใช้บริการในแต่ละสถานี ทำให้เสียเวลาในการรอคิว นอกจากนี้ การไปในสถานที่ต่าง ๆ ที่มีผู้คนจำนวนมาก อาจทำให้เกิดปัญหาได้ ไม่ว่าจะเป็น การไม่ได้รับความสะดวกในการดำเนินการต่าง ๆ ทำให้เสียเวลามากกว่าปกติ ทำให้รู้สึกเสียอารมณ์ได้ หรือมีที่พักไม่เพียงพอ

หนึ่งในข้อมูลสำคัญที่ผู้เดินทางควรทราบเพื่อวางแผนการเดินทาง คือ สถานที่ที่ต้องการจะไปนั้นมีความหนาแน่นของกลุ่มคนมากน้อยเพียงใด และมีทิศทางการเคลื่อนไหวของกลุ่มคนจากสถานที่นั้นเป็นอย่างไร ด้วยเหตุนี้ทางคณะผู้จัดทำเล็งเห็นว่าในปัจจุบันยังไม่มีระบบที่สามารถนำเสนอข้อมูลเหล่านี้ได้ ผู้พัฒนาจึงสร้างระบบทำนายความเคลื่อนไหวของกลุ่มคนขึ้นมา

ระบบทำนายความเคลื่อนไหวของกลุ่มคนนี้ใช้ข้อมูลจากสื่อสังคมออนไลน์ที่ได้รับความนิยมอย่างมากในปัจจุบัน คือ Facebook, Twitter และ Foursquare ซึ่งมีข้อมูล เช่น ข้อความสถานะ การระบุตำแหน่ง รูปภาพของผู้ใช้ ฯลฯ เนื่องจากบัญชีผู้ใช้ในสื่อสังคมออนไลน์สามารถเป็นตัวแทนของประชากรในชีวิตจริงได้ และข้อมูลที่ได้จากสื่อสังคมออนไลน์นี้มีจำนวนมาก มีความเคลื่อนไหวตลอดเวลา ทำให้ได้ข้อมูลที่มีขนาดใหญ่ เพื่อนำมาเข้ากระบวนการ วิเคราะห์ และประมวลผล เพื่อใช้ในการทำนายความเคลื่อนไหวของกลุ่มคน

ข้อมูลที่ได้จากระบบทำนายความเคลื่อนไหวของกลุ่มคนนี้ ทำให้ผู้ใช้ระบบสามารถทราบถึงข้อมูลความหนาแน่นของกลุ่มคน และ สามารถนำข้อมูลการเคลื่อนไหวของกลุ่มคนที่ได้นี้ไปใช้ เพื่อวางแผนการเดินทางให้ตรงกับความต้องการได้

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 วัตถุประสงค์ของโครงงาน

1) เพื่อให้ผู้ใช้ทราบว่าแต่ละสถาน ณ ขณะนั้นมีปริมาณคนมากหรือน้อย
2) เพื่อให้ผู้ใช้ทราบว่า ณ สถานที่นั้นจะมีปริมาณผู้คนเพิ่มขึ้นหรือน้อยลง
3) เพื่อให้ผู้ใช้ทราบว่ากลุ่มคน ณ สถานที่ปัจจุบันจะมีแนวโน้มไปสถานที่ใดต่อไป
4) เพื่อให้ผู้ใช้หลีกเลี่ยงสถานที่ที่มีผู้คนเป็นจำนวนมากได้
5) เพื่อให้ผู้ใช้สามารถนำข้อมูลการทำนายจากระบบมาใช้วางแผนกิจกรรมต่าง ๆ เช่น การเดินทาง การท่องเที่ยว เป็นต้น
6) เพื่อประยุกต์ใช้ข้อมูลจากสื่อสังคมออนไลน์

1.3 ประโยชน์ที่คาดว่าจะได้รับ

ปัญหาหนึ่งของการเดินทาง เช่น การใช้บริการขนส่งมวลชน ผู้ใช้ทราบแต่ระยะเวลาที่ใช้เดินทางจากสถานีต้นทางไปยังสถานีปลายทาง แต่ไม่ทราบจำนวนผู้ใช้บริการในแต่ละสถานี ทำให้เสียเวลาในการรอคิว นอกจากนี้ การไปในสถานที่ต่าง ๆ ที่มีผู้คนจำนวนมาก อาจทำให้เกิดปัญหาได้ ไม่ว่าจะเป็น การไม่ได้รับความสะดวกในการดำเนินการต่าง ๆ ทำให้เสียเวลามากกว่าปกติ ทำให้รู้สึกเสียอารมณ์ได้ หรือมีที่พักไม่เพียงพอ

เมื่อเราสามารถรู้ว่าคนที่อยู่ในสถานที่ที่เราอยู่ตอนนี้มีแนวโน้มไปที่ใดต่อ และรู้ว่าสถานที่ที่เราจะไปต่อมีแนวโน้มที่คนจะเยอะหรือน้อย เราก็จะสามารถหลีกเลี่ยงสถานที่นั้นได้ และสามารถนำข้อมูลไปวางแผนการเดินทางได้

1.4 ขอบเขตของโครงงาน

1) ระบบสามารถแสดงระดับความหนาแน่นของผู้คนในสถานที่ต่าง ๆ ณ เวลาปัจจุบัน
2) ระบบสามารถทำนายระดับความหนาแน่นของผู้คนในสถานที่ต่าง ๆ ได้
3) ระบบสามารถทำนายความเคลื่อนไหวของกลุ่มคนได้
4) ระบบสามารถแสดงผลผ่านแอปพลิเคชั่นบนโทรศัพท์มือถือได้
5) ระบบนี้ทำการทดลองในสถานที่บริเวณสยาม ซึ่งมีสถานที่ ได้แก่ สยามพารากอน, สยามดิสคัฟเวอรี่, สยามเซ็นเตอร์, เอ็ม บี เค เซ็นเตอร์, เซ็นทรัลเวิลด์, รถไฟฟ้าบีทีเอส สถานีสยาม และรถไฟฟ้าบีทีเอส สถานีสนามกีฬาแห่งชาติ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5 สมมติฐานของโครงงาน

1) สมมติให้การโพสต์ข้อความที่ระบุชื่อสถานที่ใน Twitter หมายถึงผู้โพสต์กำลังอยู่ในสถานที่นั้น ในเวลาที่โพสต์ข้อความ
2) สมมติให้การโพสต์ข้อความ และ ความคิดเห็น บนเพจใน Facebook ที่เกี่ยวข้องกับสถานที่ใดๆ หมายถึงการที่ผู้โพสต์อยู่ในสถานที่นั้น ในเวลาที่โพสต์ข้อความ
3) สมมติให้การโพสต์ข้อความแนะนำเกี่ยวกับสถานที่ และการโพสต์รูปภาพ บนสถานที่ใน Foursquare หมายถึงการที่ผู้โพสต์อยู่ในสถานที่นั้น ในเวลาที่โพสต์ข้อความ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้องด้านการจัดการข้อมูล

ในโครงงานนี้ การจัดการข้อมูลเป็นสิ่งที่สำคัญ เนื่องจาก ระบบต้องการใช้ข้อมูลจำนวนมากจากสื่อสังคมออนไลน์ ได้แก่ Foursquare, Twitter, Facebook เพื่อมาใช้วัดจำนวนคนในแต่ละสถานที่ในเวลาต่างๆ ทำให้มีการเก็บข้อมูลที่มีการกล่าวถึงสถานที่และเก็บข้อมูลนั้นๆอย่างสม่ำเสมอ ซึ่งเมื่อมีข้อมูลจำนวนมาก ระบบต้องสามารถทำงานได้เร็ว จึงจะมีประสิทธิภาพ เครื่องมือที่เลือกใช้ในการจัดการข้อมูล มีดังนี้

2.1.1 Apache Hadoop

Hadoop [2,3] เป็นซอฟต์แวร์ ถูกออกแบบมาเพื่อสำหรับไว้จัดเก็บ และ ประมวลผลข้อมูลขนาดใหญ่ โดย Hadoop เป็น cluster computing [9] ซึ่งจะประมวลผลแบบกระจายผ่านเครื่องคอมพิวเตอร์จำนวนมากที่ถูกเชื่อมต่อเข้าเป็นกลุ่มเดียวกันภายใต้ระบบเครือข่ายความเร็วสูง และสนับสนุนการทำงานแบบขนาน มีความสามารถในการกระจายงานที่ทำไปยังเครื่องภายในระบบเพื่อให้การประมวลผลมีประสิทธิภาพสูงขึ้นและเพื่อให้สามารถรองรับการขยายระบบ โดยมีชุดคำสั่ง หรือ API เพื่อช่วยอำนวยความสะดวก โดยสามารถรองรับการขยายตัวของข้อมูลได้ มีความน่าเชื่อถือสูง

Hadoop ประกอบไปด้วยส่วนเก็บข้อมูล ที่เรียกว่า Hadoop Distributed File System (HDFS) และส่วนประมวลผล ที่เรียกว่า MapReduce

Hadoop มีด้วยกัน 4 โมดูลได้แก่

1) Hadoop Common
   เป็นโปรแกรมพื้นฐานที่สนับสนุนโมดูลของ Hadoop อื่นๆ
2) Hadoop Distributed File System (HDFS)
   เป็น distributed file system ซึ่งเป็นแอปพลิเคชั่นแบบ client/serverที่ทำให้เข้าถึงและประมวลผลข้อมูลที่เก็บบนเซิฟเวอร์ ได้อย่างรวดเร็ว และช่วยจัดการไฟล์และไดเรคทอรีของเซิฟเวอร์
3) Hadoop YARN เป็น framework สำหรับการทำ scheduling และ cluster management
4) Hadoop MapReduce เป็นระบบ YARN-based ที่ใช้สำหรับการประมวลผลแบบขนานในข้อมูลที่มีขนาดใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

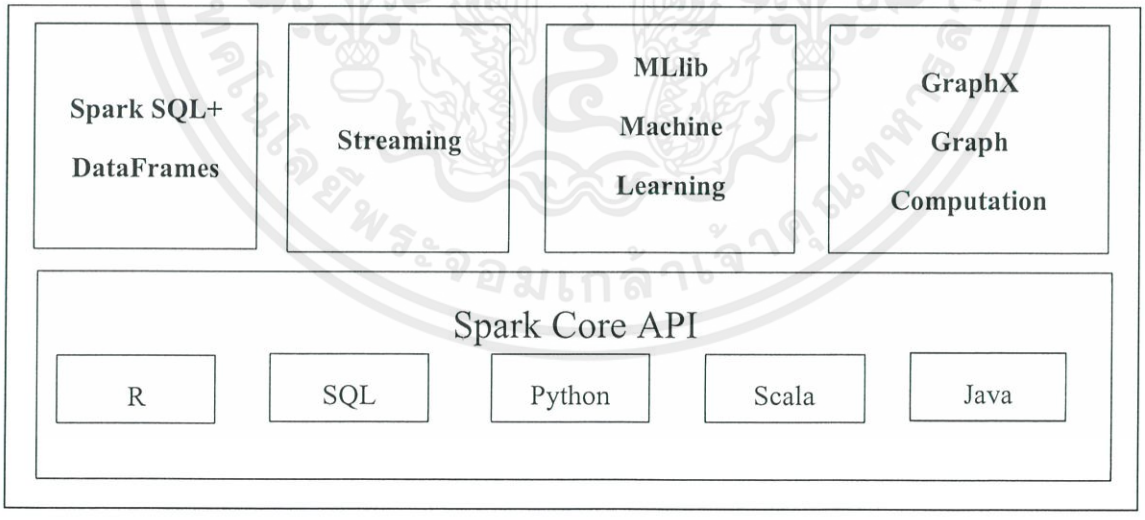
ซึ่งในโครงงานนี้จะใช้ Hadoop Distributed File System (HDFS) เพราะว่าระบบต้องใช้และวิเคราะห์ข้อมูลเป็นจำนวนมาก เนื่องจาก สื่อสังคมออนไลน์มีผู้ใช้ตลอดเวลา จึงมีข้อมูลเพิ่มขึ้นเรื่อยๆ

2.1.2 Apache Spark

Spark [6,7] เป็น cluster computing framework [9] สำหรับประมวลผลข้อมูลขนาดใหญ่ให้สามารถทำงานได้อย่างรวดเร็ว ใช้งานง่าย และ วิเคราะห์ข้อมูลที่ซับซ้อน ซึ่ง Spark ถูกพัฒนามาจาก Hadoop ในเรื่องของ MapReduce โดยทำให้มีประสิทธิภาพมากยิ่งขึ้น

โดยในโครงงานนี้ นำมาใช้ Spark SQL เพื่อให้สามารถใช้ภาษา SQL ในการเก็บไฟล์ลง Spark ได้

Spark SQL เป็นโมดูลของ Apache Spark สำหรับทำงานกับข้อมูลที่มีโครงสร้าง ที่ทำงานบน Spark Core ผ่าน data abstraction ที่เรียกว่า DataFrames ซึ่งสนับสนุน ข้อมูลที่สามารถใช้ได้กับฐานข้อมูล relational ที่มีแถวและคอมลัมน์ (structured data) และ ข้อมูลที่มีรูปแบบของโครงสร้างที่ไม่สอดคล้องกับฐานข้อมูล Relational หรือ ฐานข้อมูลแบบอื่นๆ แต่จะมีเครื่องหมายเพื่อไว้แบ่งความหมายและลำดับแทน(semi-structured data) โดยมีภาษาที่ใช้เฉพาะด้าน หรือ domain-specific language (DSL) เพื่อจัดการกับ DataFrames ใน Scala, Java หรือ Python ได้ และยังสามารถทำเหมือนกับ distributed SQL query engine คือสามารถประมวลผล SQL queries ใน Spark ได้อีกด้วย



รูป 2.1 Apache Spark System

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3 ข้อมูลจาก Social Media

ในโครงงานนี้ใช้ข้อมูลจาก Social Media เพื่อมาใช้ในการทำนายการเคลื่อนไหวของกลุ่มคนโดยเลือกข้อมูลที่เกี่ยวข้องกับแต่ละสถานที่ในเวลาต่างๆ โดย Social Media ที่เลือกมามีดังนี้

2.1.3.1 Twitter

Twitter เป็นบริการเครือข่ายสังคมออนไลน์จำพวกไมโครบล็อก โดยผู้ใช้สามารถส่งข้อความยาวไม่เกิน 140 ตัวอักษร โดยเรียกการส่งข้อความนี้ว่า ทวีต ให้บริการผ่านเวปไซต์ และแอปพลิเคชันมือถือ โดยปัจจุบันมีผู้ใช้มากกว่า 313 ล้านบัญชี [16,17]

ในโครงงานนี้คาดว่าจะได้ข้อมูลสถานที่ที่มีคนกล่าวถึงในทวีตเพื่อนำไปใช้วัดจำนวนคนที่อยู่ในสถานที่นั้นในเวลาต่างๆ

2.1.3.2 Foursquare

Foursquare เป็นบริษัทด้านเทคโนโลยีที่ใช้ความสามารถในด้านการกำหนดรู้สถานที่ตั้ง โดยให้บริการ 2 แอพพลิเคชันมือถือ คือ Foursquare ที่ให้บริการค้นหาสถานที่ พร้อมคำแนะนำจากผู้ใช้ และ Swarm เป็นเกมที่ใช้การเช็คอินในสถานที่ต่างๆเพื่อทำคะแนน โดยในแต่ละเดือนมีผู้คนมากกว่า 50 ล้านคนทั่วโลกที่ใช้ Foursquare และ Swarm [15]

ในโครงงานนี้คาดว่าจะได้ข้อมูลจำนวนคนที่เช็คอิน ณ สถานที่ต่างๆ ในขณะนั้น , ข้อมูลของชื่อเรียกในการโพสต์แนะนำเกี่ยวกับสถานที่ของ Foursquare (Tip) และรูปภาพที่โพสต์บนสถานที่นั้น เพื่อนำไปใช้วัดจำนวนคนที่อยู่ในสถานที่นั้นในเวลาต่างๆ

2.1.3.3 Facebook

Facebook เป็นบริการเครือข่ายสังคมออนไลน์ขนาดใหญ่ ผู้ใช้สามารถสร้างหน้าประวัติของตนเอง สามารถเพิ่มเพื่อน แลกเปลี่ยนข้อความ โพสสถานะ รูปภาพ วีดีโอ และ ลิ้งค์ นอกจากนั้นผู้ใช้ยังสามารถเข้าร่วมกลุ่มที่มีความสนใจเดียวกันในหมวดหมู่ต่างๆได้ โดยมีผู้ใช้มากกว่า 1.65 พันล้านบัญชี จากการสำรวจในวันที่ 31 มีนาคม 2559 [10]

ในโครงงานนี้คาดว่าจะได้ข้อมูลโพสต์และความคิดเห็นบนเพจที่เกี่ยวข้องกับสถานที่นั้นๆ เพื่อนำไปใช้วัดจำนวนคนที่อยู่ในสถานที่ที่เกี่ยวข้องกับเพจนั้นในเวลาต่างๆ

2.1.4 Location Based Service

เทคโนโลยี Location Based Service (LBS) [11] หมายถึง การบริการการบอกตำแหน่งทางภูมิศาสตร์ โดยใช้อุปกรณ์พกพา เช่น โทรศัพท์มือถือ PDA หรืออุปกรณ์ต่างๆผ่านสัญญาณเครือข่ายของผู้ให้บริการต่างๆ การให้บริการตำแหน่งที่อยู่นั้น ต้องอาศัยอุปกรณ์เฉพาะในการเชื่อมต่อกับดาวเทียม เช่น เครื่องรับสัญญาณ GPS ซึ่ง Location Based Service (LBS) ได้รับความนิยมขึ้นเรื่อยๆ และการสื่อสารไร้สายยังเพิ่มขึ้นมาจนถึงทุกวันนี้ Location Based Service (LBS)

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นบริการที่ใช้งานอยู่บนเทคโนโลยีไร้สาย ทำให้บุคคล หรือองค์กร สามารถระบุตำแหน่งที่อยู่ของผู้ใช้อุปกรณ์ไร้สายได้อย่างแม่นยำอีกด้วย

ใน [1] นั้นมีการกล่าวถึงการใช้ Location Based Service (LBS) เพื่อนำข้อมูลเกี่ยวกับการเช็คอินซึ่งมีการบอกตำแหน่งอยู่มาช่วยในการทำนาย

2.1.5 Apache Parquet

Apache Parquet เป็นรูปแบบของระบบจัดการฐานข้อมูลที่เก็บตารางข้อมูลเป็นคอลัมน์ (columnar database) ที่มีข้อได้เปรียบคือมีการบีบอัดข้อมูล และมีประสิทธิภาพ ให้บริการแก่โปรเจคที่ใช้ระบบ Hadoop โดยสามารถใช้ร่วมกับทุก framework รูปแบบข้อมูล หรือภาษาคอมพิวเตอร์ [13]

ระบบจัดการฐานข้อมูลที่เก็บตารางข้อมูลเป็นคอลัมน์ (columnar database หรือ column-oriented DBMS) เก็บตารางฐานข้อมูลแยกตามคอลัมน์ พร้อมกับ attribute values ของคอลัมน์เดียวกันเก็บไว้ติดกัน บีบอัด และเก็บรวมกัน เมื่อเทียบกับ ระบบฐานข้อมูลดั้งเดิมที่เก็บเป็นแถวต่อกัน[14] แสดงได้ดังรูป 2.2

รูป 2.2 รูปจำลองการเก็บข้อมูลแบบแยกตามแถว และคอลัมน์

ระบบนี้ใช้ฐานข้อมูลที่เก็บตารางข้อมูลเป็นคอลัมน์เนื่องจาก ในการประมวลผลส่วนใหญ่มีการ query ข้อมูลที่ต้องการข้อมูลจากคอลัมน์เดียวจำนวนมากมาใช้เพื่อประมวลผล การจัดเก็บตามคอลัมน์จะทำได้มีประสิทธิภาพมากกว่าการจัดเก็บแบบแถวที่ในการอ่านข้อมูลจะต้องดึงข้อมูลมาทั้งแถวแต่ใช้แค่คอลัมน์เดียว ทำให้เปลืองพื้นที่ในการเข้าถึงข้อมูลในแต่ละครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 Where will I go next?: Predicting Future Categorical Check-ins in Location Based Social Networks [1]

งานวิจัยนี้ศึกษาแนวทางการทำนายประเภทของสถานที่ของผู้ใช้ในอนาคตโดยการใช้ Location Based Social Networks(LBSN) ซึ่งงานวิจัยนี้จะมีผลลัพธ์ 2 อย่าง คือ ประเภทของสถานที่ที่มีแนวโน้มจะไปต่อ เช่น ร้านอาหาร ห้างสรรพสินค้า และ ช่วงเวลาที่คาดว่าจะไปซึ่งเป็นช่วง 6 ช่วง คือ Early Hours, Morning, Noon, Afternoon, Evening, Night โดยการทำนายนั้นเกิดจากการแบ่งประเภทของกิจกรรมของผู้ใช้จากการเช็คอิน ซึ่งให้แต่ละการเช็คอินในเวลาที่ถูกระบุไว้เป็นรูปแบบเฉพาะอันหนึ่ง ซึ่งจะใช้ข้อมูลจริงจาก Foursquare โดยนำข้อมูลมาทำกระบวนการทางสถิติ

งานวิจัยนี้มีกระบวนการทำงาน 3 ขั้นตอน ดังนี้

1) **Period slicing**

แบ่งสัปดาห์เป็น 7 วัน คือ Mon, Tue, Wed, Thu, Fri, Sat, Sun และแบ่งในแต่ละวันเป็น 6 ช่วงเวลา ได้แก่ Early Hours, Morning, Noon, Afternoon, Evening, Night ดังนั้นจะแบ่งได้ทั้งหมด 42 ช่วงเวลาใน 1 สัปดาห์

2) **Frequency counting**

สร้างฮิสโตแกรม หรือ การนับความถี่ในการเช็คอินแต่ละประเภทของสถานที่ในแต่ละช่วงเวลาของสัปดาห์ ซึ่งแสดงได้ดังตารางที่ 2.1

ตาราง 2.1 ตารางแสดงการแบ่งช่วงเวลาของแต่ละสัปดาห์

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Early Hours							
Morning							
Noon							
Afternoon							
Evening							
Night							

3) Prediction

อัลกอริทึมนี้จะใช้ข้อมูลในย้อนหลังเป็นเวลา 14 วัน เพื่อคำนวณความน่าจะเป็นของจำนวนเช็คอินในแต่ละช่วงเวลา เพื่อทำนายประเภทของสถานที่ ที่ผู้คนจะเดินทางไปต่อ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

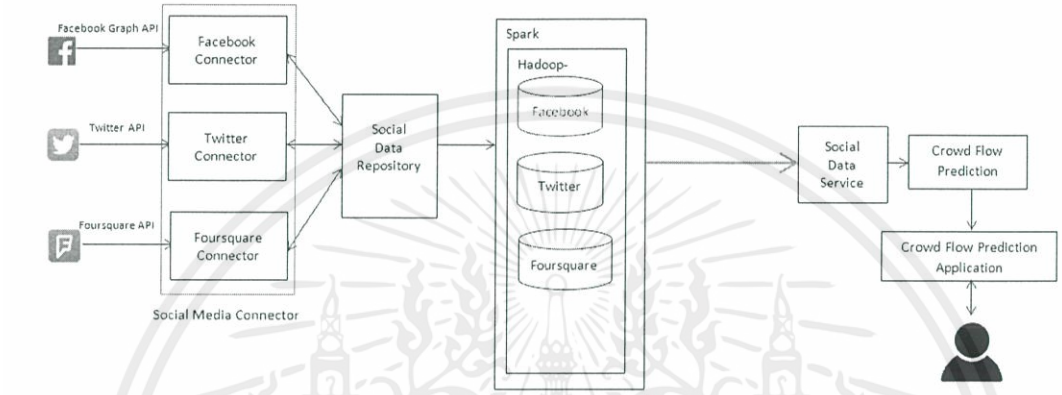
ในงานวิจัยนี้พบปัญหาที่ไม่ตรงกับการแก้ปัญหาของโครงงานดังต่อไปนี้

- เป็นการทำนายประเภทของสถานที่ แต่โครงงานนี้ต้องการทำนายสถานที่แบบเฉพาะเจาะจง
- แบ่งช่วงเวลากว้างมากเกินไป ไม่เหมาะกับการทำนายความเคลื่อนไหวของกลุ่มคนในช่วงเวลาเร่งด่วน จึงแก้ปัญหาโดยการแบ่งช่วงเวลาละ 5 นาที มีทั้งหมด 288 ช่วงเวลา

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การวิเคราะห์และออกแบบระบบ

3.1 โครงสร้างของระบบ



รูป 3.1 โครงสร้างของระบบ

จากรูปที่ 3.1 โครงสร้างของระบบแบ่งออกเป็น 5 ส่วนดังนี้

3.1.1 ส่วนจัดการข้อมูลจากสื่อสังคมออนไลน์ (Social Media Connector)

เป็นส่วนที่ทำหน้าที่ติดต่อกับ API ของแต่ละสังคมออนไลน์ ได้แก่ Facebook, Twitter และ Foursquare ประกอบไปด้วย 3 ส่วน ได้แก่

3.1.1.1 Facebook connector

นำข้อมูลมาจาก Facebook Graph API โดยเลือกใช้ endpoint ต่อไปนี้

1) ค้นหาข้อมูลเพจ
(graph.facebook.com/{page-id})
2) ค้นหาข้อมูลโพสต์ในเพจ
(graph.facebook.com/{page-id}/feed)
3) ค้นหาข้อมูลความคิดเห็นในเพจ
(graph.facebook.com/{page-id}/comment)

3.1.1.2 Twitter connector

นำข้อมูลมาจาก Twitter ผ่านทาง REST APIs [8] โดยเลือกใช้ API ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) ค้นหา tweet (https://api.twitter.com/1.1/search/)

3.1.1.3 Foursquare connector

นำข้อมูลมาจาก Foursquare API [5] โดยเลือกใช้ API ต่อไปนี้

2) ค้นหาสถานที่

(https://api.foursquare.com/v2/venues/search)

3) ค้นหาข้อมูลเกี่ยวกับสถานที่

(https://api.foursquare.com/v2/venues/VENUE_ID)

4) ค้นหาจำนวนคนที่อยู่ขณะนี้

(https://api.foursquare.com/v2/venues/VENUE_ID/herenow)

5) ค้นหาคำแนะนำเกี่ยวกับสถานที่

(https://api.foursquare.com/v2/venues/VENUE_ID/tips)

6) ค้นหารูปภาพที่เกี่ยวกับสถานที่

(https://api.foursquare.com/v2/venues/VENUE_ID/photos)

3.1.2 ส่วนจัดการเก็บข้อมูลลงในฐานข้อมูล (Social Data Repository)

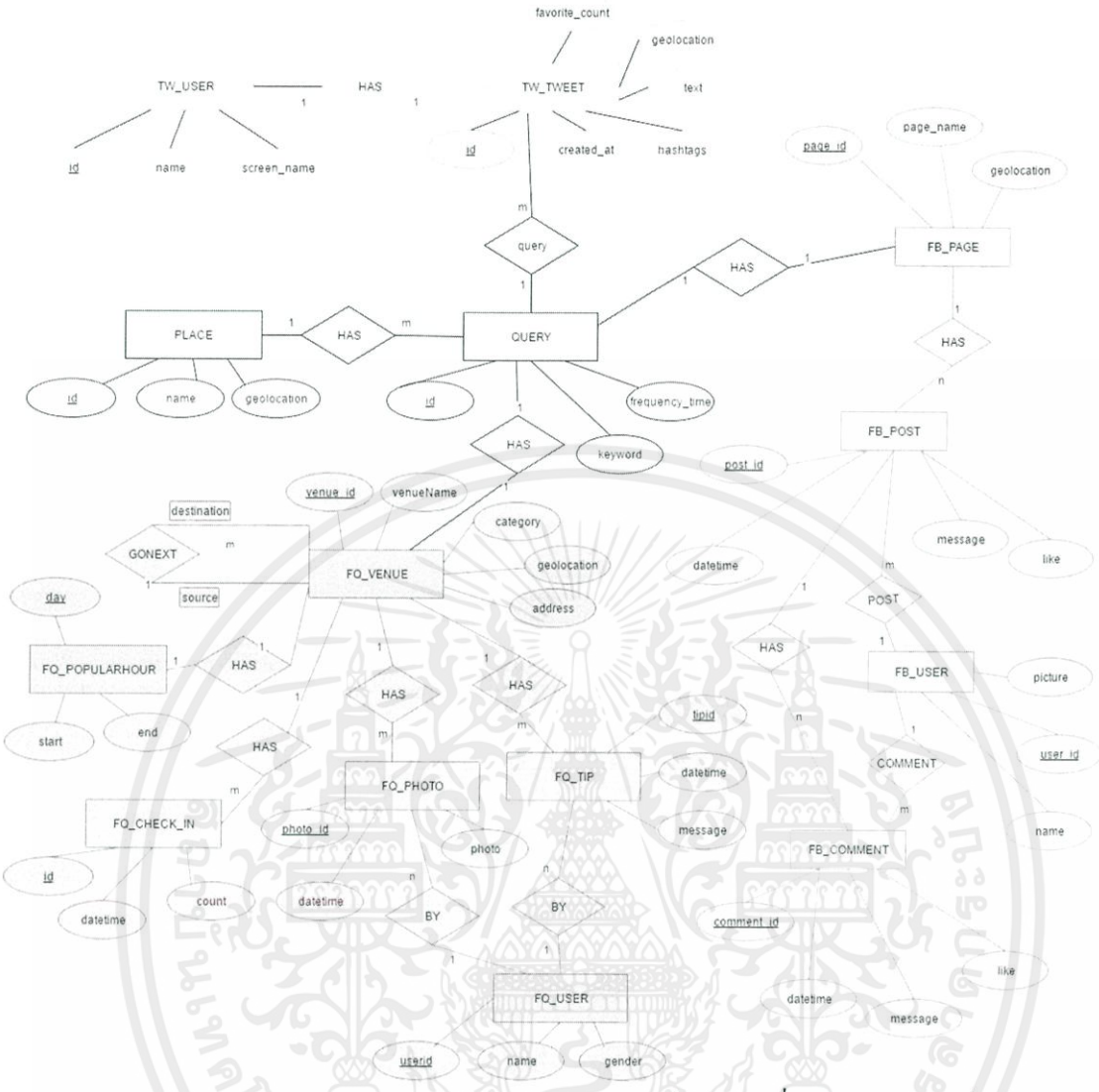
เป็นส่วนที่ทำหน้าที่จัดการนำข้อมูลที่ได้จากสื่อสังคมออนไลน์ประกอบด้วย Twitter, Foursquare และ Facebook ที่ได้จากส่วน Social Media Connector มาลงในตารางที่สัมพันธ์กับข้อมูลที่ได้รับในฐานข้อมูล เพื่อให้อยู่ในรูปแบบที่พร้อมจะนำไปประมวลผลต่อไป โดยใช้ SparkSQL เป็นเครื่องมือที่ช่วยให้ใช้ภาษา SQL เพื่อ query ข้อมูลผ่าน Spark ที่มีความสามารถในการประมวลผลแบบกลุ่มได้

3.1.3 ส่วนฐานข้อมูล (Database)

เป็นส่วนที่ทำหน้าที่เก็บข้อมูลต่างๆ เช่น ข้อมูลการเช็คอินสถานที่ และเวลา ที่ได้มาจากสื่อสังคมออนไลน์ ประกอบด้วย Twitter, Facebook และ Foursquare เก็บข้อมูลลงใน Hadoop ที่ช่วยจัดเก็บ และ ประมวลผลข้อมูลขนาดใหญ่ สามารถรองรับการขยายตัวของข้อมูลได้ เก็บข้อมูลโดยใช้ Apache Parquet ซึ่งเก็บข้อมูลแบ่งตามคอลัมน์

ซึ่งมีโครงสร้างและความสัมพันธ์ของข้อมูล (Relationship) แสดงได้ดังแบบจำลองอธิบายโครงสร้างของฐานของฐานข้อมูล Entity Relationship Diagram ดังรูปที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 3.2 Entity Relationship Diagram ของฐานข้อมูลจากสื่อสังคมออนไลน์

จากรูป 3.2 ฐานข้อมูลประกอบด้วยข้อมูลจาก Twitter, Foursquare และ Facebook ซึ่งมี Entity แสดงสถานที่คือ 'QUERY', 'FQ_VENUE' และ 'FB_POST' ตามลำดับ เชื่อมต่อสถานที่ของแต่ละสื่อสังคมออนไลน์เข้าด้วยกันด้วย 'PLACE' และมี 'QUERY' ไว้สำหรับเก็บคำค้นหาสำหรับสื่อสังคมต่าง ๆเข้าด้วยกัน โดยที่ข้อมูลสถานที่ที่ได้จากสื่อสังคมออนไลน์แต่ละบริการก็จะให้ข้อมูลแตกต่างกัน ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3.1 Twitter

ตาราง 3.1 ตารางแสดง Entity ของข้อมูลจาก Twitter

Entity	คำอธิบาย
TW_TWEET	ทวีตที่ผู้ใช้โพสต์ลง Twitter
TW_USER	บัญชีผู้ใช้ Twitter

3.1.3.2 Foursquare

ตาราง 3.2 ตารางแสดง Entity ของข้อมูลจาก Foursquare

Entity	คำอธิบาย
FQ_VENUE	สถานที่ใน Foursquare
FQ_CHECK_IN	จำนวนคนที่อยู่ในแต่ละสถานที่
FQ_POPULARHOUR	ช่วงเวลายอดนิยมในแต่ละสถานที่
FQ_PHOTO	รูปภาพที่ผู้ใช้ถ่ายในแต่ละสถานที่
FQ_TIP	คำแนะนำที่ผู้ใช้แนะนำเกี่ยวกับสถานที่
FQ_USER	บัญชีผู้ใช้ Foursquare

3.1.3.3 Facebook

ตาราง 3.3 ตารางแสดง Entity ของข้อมูลจาก Facebook

Entity	คำอธิบาย
FB_PAGE	เพจใน Facebook
FB_POST	โพสต์ในหน้าเพจ
FB_COMMENT	ความคิดเห็นในแต่ละโพสต์
FB_USER	บัญชีผู้ใช้ หรือ บัญชีเพจ Facebook

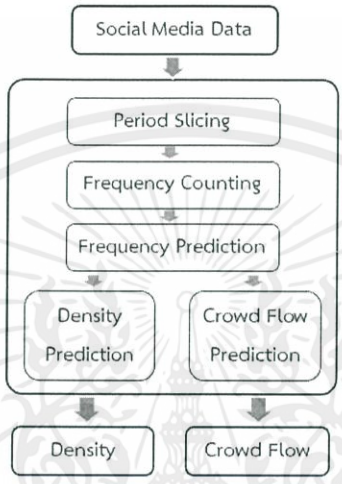
3.1.4 ส่วนจัดการข้อมูล (Social Data Service)

เป็นส่วนที่ทำหน้าที่นำข้อมูลจากฐานข้อมูลไปให้ระบบทำนายการเคลื่อนไหวของกลุ่มคนใช้ เพื่อวิเคราะห์และทำนายความเคลื่อนไหวของกลุ่มคน

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.5 ส่วนการวิเคราะห์ความหนาแน่นประชากร และ การทำนายเคลื่อนไหวของกลุ่มคน (Crowd flow prediction System)

เป็นส่วนที่ทำการประมวลผล วิเคราะห์ ข้อมูลจากส่วนจัดการเก็บข้อมูล เพื่อนำมาทำนายการเคลื่อนไหวของกลุ่มคน ความหนาแน่นประชากรของสถานที่ต่าง ๆ และแนวโน้มที่การเปลี่ยนแปลงความหนาแน่นประชากรของสถานที่นั้น ๆ ซึ่งมีการประยุกต์ใช้วิธีการมาจาก [1]



รูป 3.3 กระบวนการในทำนายความหนาแน่น และความเคลื่อนไหวของกลุ่มคน

3.1.5.1 การเก็บข้อมูลจากสื่อสังคมออนไลน์ (Social Media Data)

นำข้อมูลมาจากสื่อสังคมออนไลน์มาวิเคราะห์ ได้แก่

1) จำนวนข้อความที่มีคนกล่าวถึงสถานที่ต่างๆ จาก Twitter
2) จำนวนของคนที่เช็คอิน ณ สถานที่ต่างๆในเวลาที่แตกต่างกัน, ข้อมูลการโพสต์แนะนำสถานที่(Tip) และรูปภาพที่โพสต์บนสถานที่นั้น จาก Foursquare
3) จำนวนการโพสต์และความคิดเห็นบนเพจที่เกี่ยวข้องกับสถานที่นั้นๆ จาก Facebook

3.1.5.2 Period Slicing

แบ่งเวลาออกเป็นช่วงๆ โดยจะแบ่งสัปดาห์เป็น 7 วัน คือ Mon, Tue, Wed, Thu, Fri, Sat, Sun และแบ่งแต่ละวันเป็นช่วงๆ ในแต่ละวันเป็น 288 ช่วงเวลา ช่วงเวลาละ 5 นาที

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 3.4 ตัวอย่างตารางที่ได้จากขั้นตอน Period Slicing

เวลา	วัน							
	1/3/2017	*2/3/2017*	*3/3/2017*	*4/3/2017*	*5/3/2017*	*6/3/2017*	*7/3/2017*	...
00:00								
...								
11.45								
11.50								
11.55								
12.00								
12.05								
...								
23:50								
23:55								

3.1.5.3 Frequency Counting

นับความถี่ของจำนวนคนที่อยู่ในสถานที่นั้นๆ (การเช็คอินหรือการกล่าวถึงสถานที่นั้นๆ) ในแต่ละช่วงเวลา

ตาราง 3.5 ตัวอย่างตารางข้อมูลที่ได้จากขั้นตอน Frequency Counting

เวลา	วัน							
	1/3/2017	*2/3/2017*	*3/3/2017*	*4/3/2017*	*5/3/2017*	*6/3/2017*	*7/3/2017*	...
00:00	4	3	2	5	1	1	2	4
...								
11.45	15	19	12	11	20	25	22	15
11.50	11	23	23	22	16	11	10	11
11.55	20	33	12	22	19	22	33	20
12.00	24	21	10	15	12	14	30	24
12.05	22	40	33	21	23	22	18	22
...								
23:50	2	1	5	6	8	9	7	2
23:55	5	6	10	12	12	9	3	5

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.5.4 Frequency Prediction

ทำนายโดยใช้จำนวนคนในช่องของช่วงเวลาปัจจุบัน เพื่อทำนายช่องของช่วงเวลาที่ต้องการทำนาย โดยดูจำนวนคนในช่องเวลาปัจจุบันว่ามีค่าเท่ากันหรือใกล้เคียงกับวันใดในช่วงเวลาเดียวกัน จากนั้นจึงทำนายว่าช่วงเวลาที่ต้องการทำนายของวันปัจจุบันมีค่าเท่ากับเดียวค่าใดของวันที่มีค่าใกล้เคียง โดยใช้ข้อมูลย้อนหลัง 14 วัน

แสดงได้ดังตัวอย่างการทำนายจำนวนคนของช่วงเวลา 5 นาทีถัดไป ได้ดังนี้

ตาราง 3.6 ตัวอย่างกระบวนการทำนายจำนวนคน

เวลา	วัน						
	1/3/2017	*2/3/2017*	*3/3/2017*	...	*13/3/2017*	*14/3/2017*	*วันที่ต้องการทำนาย*
00:00	4	3	2		5	1 1	2 0
...							
11.45	15	19	12		11	20	11 (ช่วงเวลาปัจจุบัน)
11.50	11	23	22		22	16	22 (ช่วงเวลาทำนาย)
11.55	20	33	12		22	19	
12.00	24	21	10		15	12	
12.05	22	40	33		21	23	
...							
23:50	2	1	5		6	8	
23:55	5	6	10		12	12	

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- หากพบค่าที่ตรงกันหรือใกล้เคียงมากกว่า 1 ค่าของช่วงเวลา 5 นาทีถัดไป จะเลือกค่าช่วงเวลา 5 นาทีถัดไปที่พบมากที่สุดเพื่อมาเป็นผลลัพธ์ของการทำนายในช่วงเวลา 5 นาทีถัดไป

ตาราง 3.7 ตัวอย่างกระบวนการทำนายจำนวนคนเมื่อพบค่าที่ตรงกันหรือใกล้เคียงมากกว่า 1 ค่า ในกรณีพบค่าที่ตรงกันหรือใกล้เคียงมากกว่า 1 ค่าของช่วงเวลา 5 นาทีถัดไป

เวลา	วัน						
	1/3/2017	*2/3/2017*	*3/3/2017*	...	*13/3/2017*	*14/3/2017*	*วันที่ต้องการทำนาย*
00:00	4	3	2		5	1 1	20
...							
11.45	15	19	12		11	20	11
11.50	11	23	22		22	22	22 (ช่วงเวลาปัจจุบัน)
11.55	20	33	12		21	21	21 (ช่วงเวลาทำนาย)
12.00	24	21	10		15	12	
12.05	22	40	33		21	23	
...							
23:50	2	1	5		6	8	
23:55	5	6	10		12	12	

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- หากพบค่าที่แตกต่างกัน ค่าละ 1 ครั้งของช่วงเวลา 5 นาทีถัดไป จะเลือกค่าช่วงเวลา 5 นาทีถัดไปที่พบมากที่สุดเพื่อมาเป็นผลลัพธ์ของการทำนายในช่วงเวลา 5 นาทีถัดไป

ตาราง 3.8 ตัวอย่างกระบวนการทำนายจำนวนคนเมื่อพบค่าที่ตรงกันหรือใกล้เคียงมากกว่า 1 ค่า ในกรณีพบค่าที่แตกต่างกัน ค่าละ 1 ครั้งของช่วงเวลา 5 นาทีถัดไป

เวลา	วัน						
	1/3/2017	*2/3/2017*	*3/3/2017*	*...*	*13/3/2017*	*14/3/2017*	*วันที่ต้องการทำนาย*
00:00	4	3	2		5	1 1	2 0
...							
11.45	15	19	12		11	20	11
11.50	11	22	22		22	22	22 (ช่วงเวลาปัจจุบัน)
11.55	20	33	12		23	19	23 (ช่วงเวลาทำนาย)
12.00	24	21	10		15	12	
12.05	22	40	33		21	23	
...							
23:50	2	1	5		6	8	
23:55	5	6	10		12	12	

จากตารางด้านบน คอลัมน์วันสุดท้าย หมายถึง วันที่ต้องการทำนายหรือวัน ณ ปัจจุบัน โดยช่องเวลาสีเหลือง หมายถึง ช่วงเวลา ณ ปัจจุบัน

3.1.5.5 Density Prediction

นำค่าจำนวนคนที่มากที่สุด (max) ของแต่ละสถานที่มาแบ่งเป็น 3 ช่วง เพื่อทำนายความหนาแน่นในแต่ละสถานที่ โดยคำนวณความถี่ในแต่ละช่วงได้จาก

$$\text{จำนวนความถี่ในแต่ละช่วง}(\square\square\square\square\square) = \frac{[\square\square\square+1]}{3} \qquad (3.1)$$

ช่วงที่ 1 ความถี่ในช่วงเวลานั้น < range หมายถึงมีจำนวนคน น้อย
ช่วงที่ 2 ความถี่ในช่วงเวลานั้น < 2*range หมายถึงมีจำนวนคน ปานกลาง
ช่วงที่ 3 ความถี่ในช่วงเวลานั้น > 2*range หมายถึงมีจำนวนคน มาก

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 3.9 ตัวอย่างตารางแสดงช่วงความหนาแน่น

เวลา	วัน						
	1/3/2017	*2/3/2017*	*3/3/2017*	...	*13/3/2017*	*14/3/2017*	*15/3/2017*
00:00	น้อย	น้อย	น้อย		น้อย	น้อย	น้อย
...							
11.45	ปานกลาง	ปานกลาง	น้อย		น้อย	ปานกลาง	น้อย
11.50	น้อย	ปานกลาง	ปานกลาง		ปานกลาง	ปานกลาง	ปานกลาง
11.55	ปานกลาง	มาก	น้อย		ปานกลาง	ปานกลาง	
12.00	ปานกลาง	ปานกลาง	น้อย		ปานกลาง	น้อย	
12.05	ปานกลาง	มาก	มาก		ปานกลาง	ปานกลาง	
...							
23:50	น้อย	น้อย	น้อย		น้อย	น้อย	
23:55	น้อย	น้อย	น้อย		น้อย	น้อย	

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

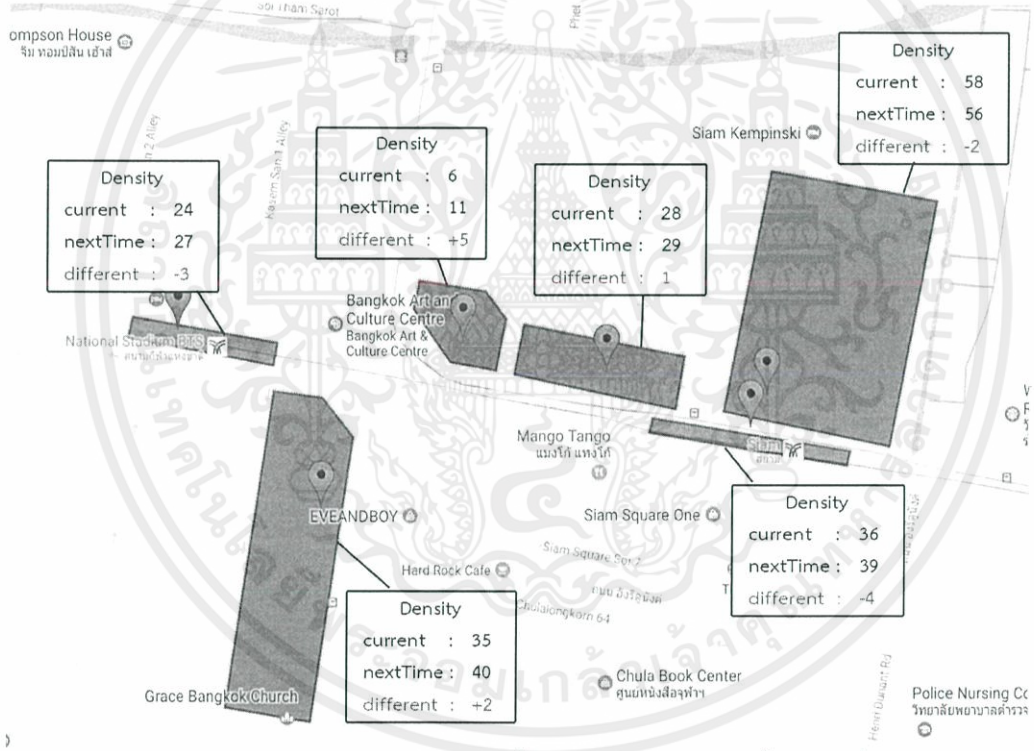
3.1.5.6 Crowd Flow Prediction

เปรียบเทียบจำนวนคนของสถานที่ปัจจุบันว่ามีการเปลี่ยนแปลงตรงข้ามมีค่าใกล้เคียงกับการเปลี่ยนแปลงของสถานที่ใกล้เคียงใดมากที่สุด เพื่อทำนายความเคลื่อนไหวของคนในช่วงเวลาถัดไป

สมมติให้ ระยะเวลา 5 นาที สามารถเคลื่อนย้ายไปได้ไม่เกิน 400 เมตร

โดยขั้นตอนการทำนายเป็นดังนี้

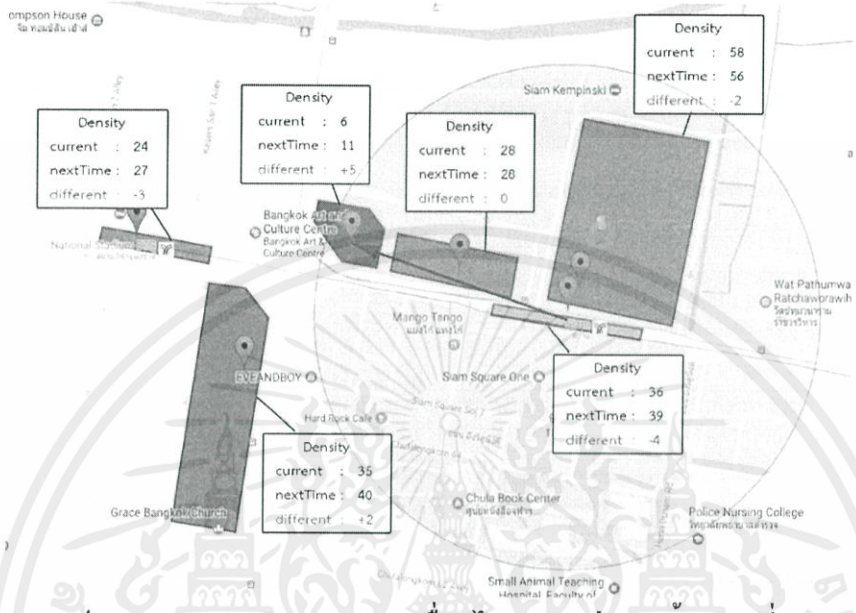
1. คำนวณการเปลี่ยนแปลงความหนาแน่นจากปัจจุบันกับอีก 5 นาทีข้างหน้าที่ได้จากตารางความหนาแน่น ของแต่ละสถานที่
    1. ถ้าจำนวนคนลดลง จะแทนด้วยเครื่องหมาย –
    2. ถ้าจำนวนคนเพิ่มขึ้น จะแทนด้วยเครื่องหมาย +



รูป 3.4 แสดงการทำนายความเคลื่อนไหวของกลุ่มคน ขั้นตอนที่ 1

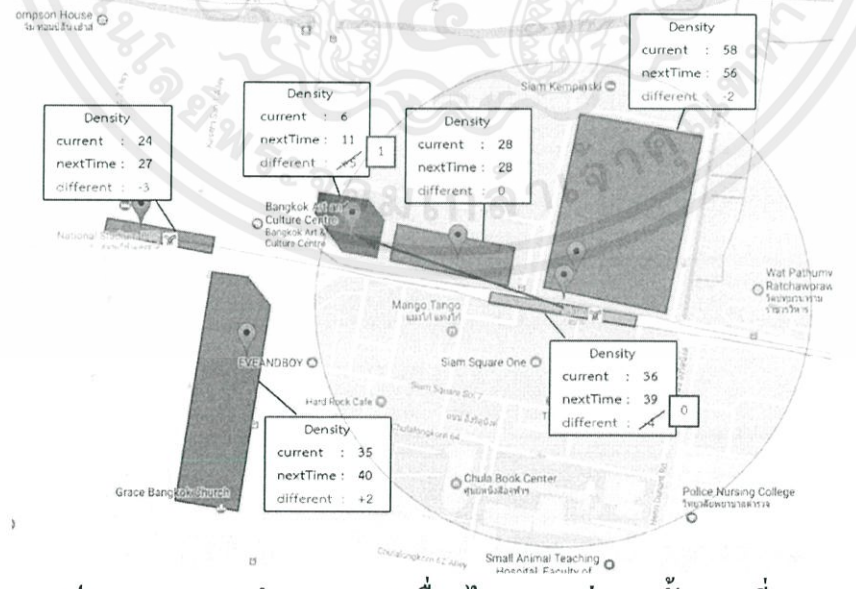
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. จากนั้นเริ่มจากสถานที่ที่ค่าลบที่มากที่สุดก่อน และ หาสถานที่ในรัศมี 400เมตรที่เป็นค่าบวกที่ค่าใกล้เคียงกับค่าสัมบูรณ์ของค่าลบมากที่สุด ให้ทำนายว่าผู้คนจะเดินทางไปสถานที่นั้น



รูป 3.5 แสดงการทำนายความเคลื่อนไหวของกลุ่มคน ขั้นตอนที่ 2

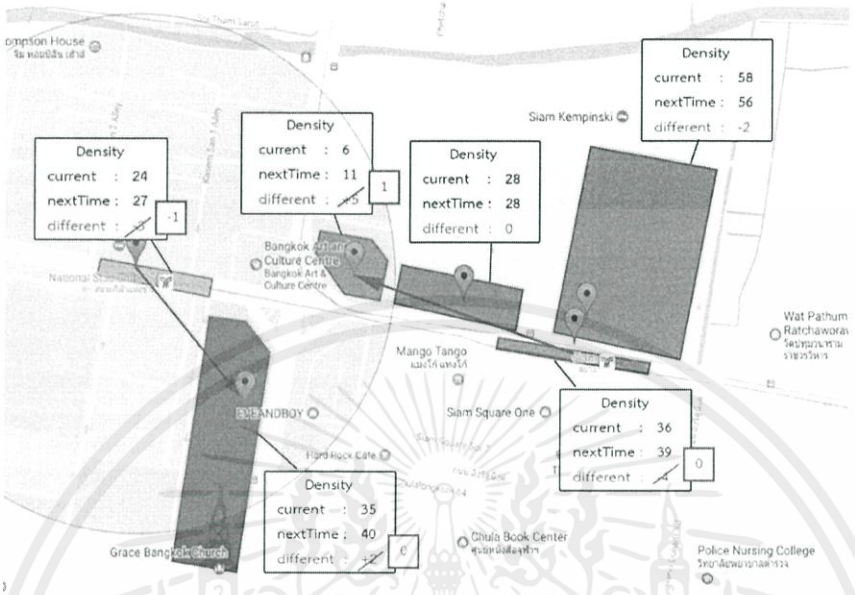
3. ทำการหักจำนวนคนคงเหลือ ของแต่ละสถานที่ด้วยผลต่างของจำนวนคนทั้ง 2 สถานที่



รูป 3.6 แสดงการทำนายความเคลื่อนไหวของกลุ่มคน ขั้นตอนที่ 3

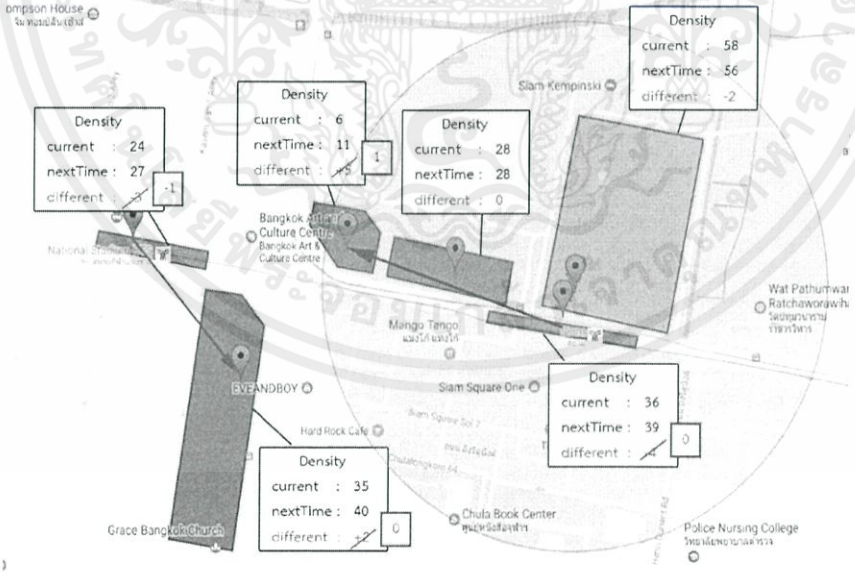
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. จากนั้นไปทำนายสถานที่ที่มีค่าลบมากขึ้นเป็นลำดับถัดไป โดยใช้วิธีเดิมตาม ข้อ 2-3



รูป 3.7 แสดงการทำนายความเคลื่อนไหวของกลุ่มคน ขั้นตอนที่ 4

5. ทำนายสถานที่ลบลำดับถัดไปเรื่อยๆจนหมด

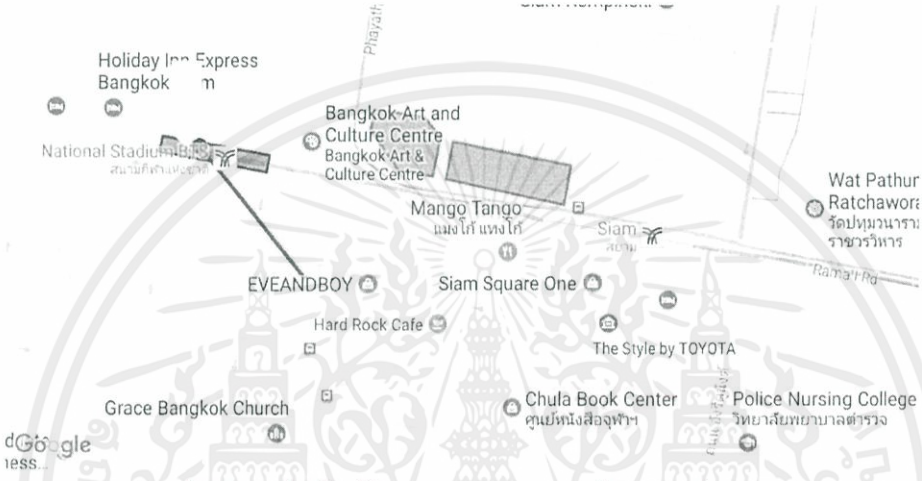


รูป 3.8 แสดงการทำนายความเคลื่อนไหวของกลุ่มคน ขั้นตอนที่ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.6 แอปพลิเคชันแสดงผลสำหรับผู้ใช้งาน (Crowd flow Prediction Application)

แอปพลิเคชันบนมือถือสำหรับแสดงผลการทำนายความหนาแน่น และทิศทางการเคลื่อนไหวของกลุ่มคนให้แก่ผู้ใช้งาน สร้างแอปพลิเคชันด้วยภาษา JavaScript และใช้ React Native ซึ่งเป็น library ภาษา JavaScript ที่ใช้สร้างส่วนแสดงผลผู้ใช้งาน ที่สามารถใช้ได้ง่าย และจะ render ในส่วนแสดงผลเฉพาะส่วนที่มีการเปลี่ยนแปลงของข้อมูลเท่านั้น ทำให้ตอบสนองการใช้งานได้รวดเร็ว[4] โดยแอปพลิเคชันแสดงได้ดังรูปที่ 3.9 ถึง 3.11



รูป 3.9 แอปพลิเคชันแสดงความหนาแน่นของกลุ่มคน

จากรูปที่ 3.9 ระบบจะแสดงความหนาแน่นของประชากร ณ เวลาปัจจุบัน โดยแสดงสีบนพื้นที่ของสถานที่นั้นๆ แบ่งเป็น 3 ระดับ คือ

1) ระดับสีเขียว หมายถึง มีความหนาแน่นประชากรต่ำ
2) ระดับสีเหลือง หมายถึง มีความหนาแน่นประชากรปานกลาง
3) ระดับสีแดง หมายถึง มีความหนาแน่นประชากรสูง

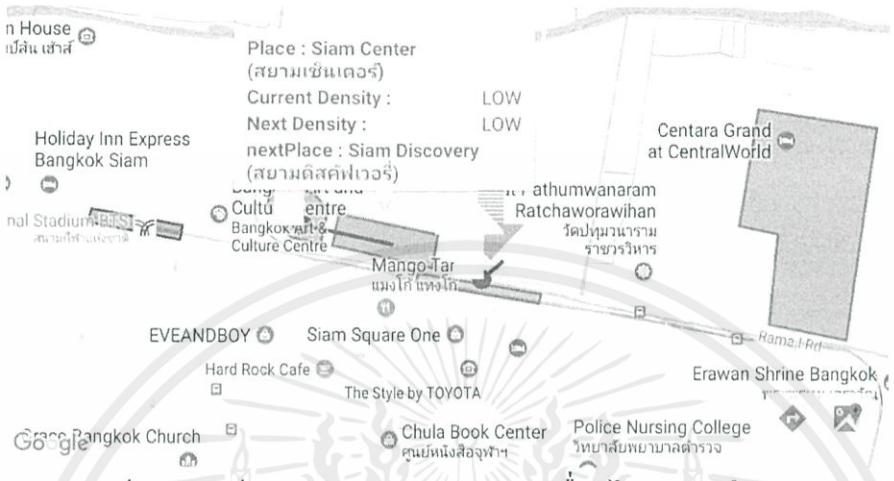
นอกจากนั้นระบบจะแสดงการเปลี่ยนแปลงของประชากรในเวลาถัดไป โดยแสดงเป็นลูกศรปรากฏบนพื้นที่ของสถานที่นั้นๆ แบ่งเป็น 4 แบบ คือ

1) ลูกศรขึ้นสีแดง หมายถึง มีความหนาแน่นประชากรเพิ่มขึ้นเป็นระดับสูงในเวลาถัดไป
2) ลูกศรขึ้นสีเหลือง หมายถึง มีความหนาแน่นประชากรเพิ่มขึ้นเป็นระดับปานกลางในเวลาถัดไป
3) ลูกศรลงสีเหลือง หมายถึง มีความหนาแน่นประชากรลดลงเป็นระดับปานกลางในเวลาถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

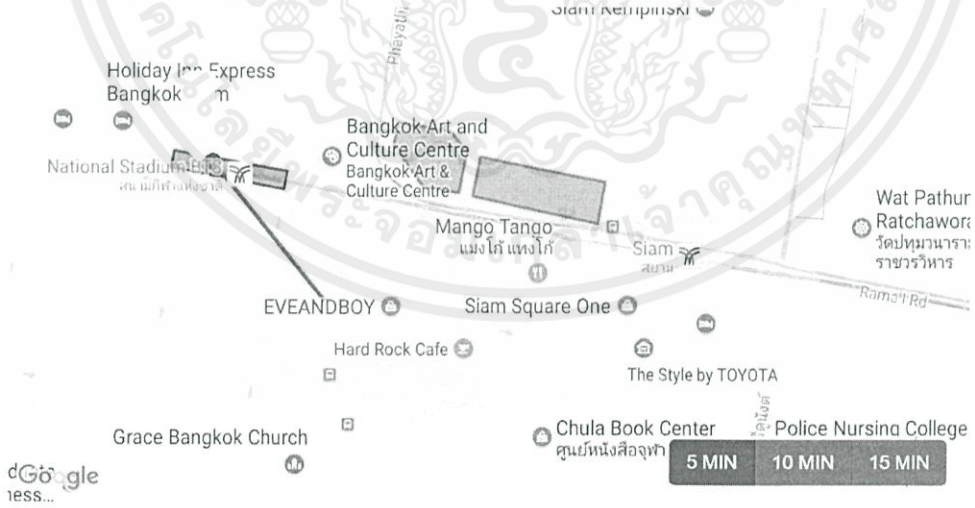
4) ลูกศรลงสีเขียว หมายถึง มีความหนาแน่นประชากรลดลงเป็นระดับต่ำในเวลาถัดไป

หากสถานที่ใดไม่มีลูกศร หมายถึง ความหนาแน่นเท่าเดิมในเวลาถัดไป



รูป 3.10 แอปพลิเคชันแสดงทิศทางการเคลื่อนไหวของกลุ่มคน

จากรูปที่ 3.10 ระบบจะแสดงสถานที่ที่ผู้คนจากสถานที่ต้นทางที่เลือก จะเดินทางต่อในเวลา5 นาทีถัดไป โดยปรากฏข้อความแสดงสถานที่ที่มีแนวโน้มว่าผู้คนจะเดินทางไปต่อมาก พร้อมกับแสดงเส้นสีม่วงจากสถานที่ดังกล่าวไปยังสถานที่ที่ผู้คนจะเดินทางต่อไปในเวลา 5 นาทีถัดไป โดยด้านวงกลม หมายถึง สถานที่ปลายทาง



รูป 3.11 แอปพลิเคชันแสดงการเลือกช่วงเวลาในการทำนาย

จากรูปที่ 3.11 แอปพลิเคชันสามารถเลือกช่วงเวลาในการทำนายได้ คือ 5 นาที, 10 นาที และ 15 นาที

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การทดลองและผลการทดลอง

4.1 การทดสอบการดึงข้อมูลจาก Twitter

4.1.1 วัตถุประสงค์

เพื่อดึงข้อมูลโพสต์ที่เกี่ยวกับสถานที่ที่ต้องการค้นหาจาก Twitter Rest API และตรวจสอบรูปแบบของข้อมูลที่ได้รับ

4.1.2 วิธีการทดสอบ

สมัคร Developer account ของ Twitter เพื่อทำการขอ API Key API Secret Consumer Key และ Consumer Secret เพื่อใช้ในการดึงข้อมูล โดยการดึงข้อมูลจะใช้ twit ซึ่งเป็นไลบรารี่ของ Node.js ที่ใช้เชื่อมต่อกับ Twitter Rest API และใช้คำว่า kmitl เป็นคำค้นหา

โปรแกรม 4.1 TwitterService

```
let T =new Twit({
  consumer_key:'eFroY0pfuCmbSf4OrwFckHDUY',
  consumer_secret:
'xCJRkrUHuPGAGs7ZCAdKOjqDfIcghzMGjfrS6xLXHZUKwJ430h',
  access_token:'3854752459-
6KWXoTPeBbxUdydKsAOMohUru2DBioZ9jmealFm',
  access_token_secret:
'9qow1TnZqKaWVgZ6V22qcNeqEf5IoFroVUTGs846N6hfQ',
  timeout_ms:60 *1000
})
export function searchTweet(q){
  return new Promise((resolve, reject)=> {
    T.get('search/tweets', { q:q, count:100}, (err, data,
response)=> {
   if(err){ reject(err)}
       else { resolve(data)}
    })
  })
}
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3 ผลการทดลอง

จากการทดลองพบว่าโปรแกรม TwitterService สามารถดึงข้อมูลจาก Twitter ได้โดยมีรูปแบบข้อมูลดังนี้

ตัวอย่าง 4.1 ผลลัพธ์ที่ได้จากการดึงข้อมูล Twitter ด้วยคำว่า kmitl


```
{
 "created_at":"Tue Nov 08 04:14:33 +0000 2016",
 "id":795841917997240300,
 "id_str":"795841917997240320",
 "text":"I'm at ตึกภาควิชาวิศวกรรมเครื่องกล KMITL in กรุงเทพ
https://t.co/dkVC2vIxhR",
 "truncated":false,
 "entities":{…},
 "metadata":{…},
 "source":"<a href=\"http://foursquare.com\"
rel=\"nofollow\">Foursquare</a>",
 "in_reply_to_status_id":null,
 "in_reply_to_status_id_str":null,
 "in_reply_to_user_id":null,
 "in_reply_to_user_id_str":null,
 "in_reply_to_screen_name":null,
 "user":{…},
 "geo":{…},
 "coordinates":{…},
 "place":{…},
 "contributors":null,
 "is_quote_status":false,
 "retweet_count":0,
 "favorite_count":0,
 "favorited":false,
 "retweeted":false,
 "possibly_sensitive":false,
 "lang":"th"
}
```


เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่าง 4.1 ผลลัพธ์ที่ได้จากการดึงข้อมูล Twitter ด้วยคำว่า kmitl (ต่อ)

```
  "favorite_count": 0,
  "favorited": false,
  "retweeted": false,
  "possibly_sensitive": false,
  "lang": "th"
}
```

4.2 การทดสอบการดึงข้อมูลจาก Foursquare

4.2.1 วัตถุประสงค์

เพื่อดึงข้อมูลรายละเอียดต่างๆ คำแนะนำ รูปภาพ จำนวนคนในเวลาปัจจุบัน ช่วงเวลาที่ได้รับความนิยม ในสถานที่ที่ต้องการค้นหา และค้นหาสถานที่ถัดไปที่ผู้คนนิยมเดินทางไปต่อจากสถานที่ปัจจุบัน จาก Foursquare API และตรวจสอบรูปแบบของข้อมูลที่ได้รับ

4.2.2 วิธีการทดสอบ

สมัคร account ของ Foursquare แล้วสร้าง App เพื่อทำการขอ Client ID และ Client Secret เพื่อใช้ในการดึงข้อมูลจาก Foursquare API โดยใช้สถานที่ “King Mongkut's Institute of Technology Ladkrabang (KMITL)” ที่มี VENUE_ID คือ “4c034d0cf56c2d7fa6c71c66” เป็นคำค้นหา

โปรแกรม 4.2 FoursquareService

```
import request from 'request'
let params ={
client_id :
'VYPDJ2Z5QROWDVGISDDRSS2MWQTARTSUU0LKLK3BONI03W76',
client_secret :
'TSJ5KW4EQYLGA5MKDKKgXFFJQJ5KGCUUCN3H3DWUJBTYJ0CQ',
 v :today()  }
export function searchVenues(geolocation){
    params.ll =geolocation
   return new Promise((resolve)=> {
request({url:'https://api.foursquare.com/v2/venues/search',
qs:params},(err, response, body)=> {
            if(err){ console.log(err); return; }
            resolve(body)
        })
})}
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรม 4.2 FoursquareService (ต่อ)

```
export function VenueDetail(venue_id){
    return new Promise((resolve)=> {

request({url:'https://api.foursquare.com/v2/venues/'+venue_id,
qs:params},(err, response, body)=> {
            if(err){ console.log(err); return; }
            resolve(body)
        })
})}
export function VenuePhoto(venue_id){
    return new Promise((resolve)=> {

request({url:'https://api.foursquare.com/v2/venues/'+venue_id+
'/photos', qs:params},(err, response, body)=> {
     if(err){ console.log(err); return; }
    resolve(body)})
})}
export function VenueHereNow(venue_id){
    return new Promise((resolve)=> {
request({url:'https://api.foursquare.com/v2/venues/'+venue_id+
'/herenow', qs:params},(err, response, body)=> {
        if(err){ console.log(err); return; }
      resolve(body)})
})}
export function nextVenue(venue_id){
   return new Promise((resolve)=> {

request({url:'https://api.foursquare.com/v2/venues/'+venue_id+
'/nextvenues', qs:params},(err, response, body)=> {
      if(err){ console.log(err); return; }
     resolve(body) })
    })
}
export function TrendingVenue(geolocation){
    params.ll =geolocation
    return new Promise((resolve)=> {

request({url:'https://api.foursquare.com/v2/venues/trending',
qs:params},(err, response, body)=> {
            if(err){ console.log(err); return; }
            resolve(body)})})}
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรม 4.2 FoursquareService (ต่อ)

```
export function venueTips(venue_id){
    params.sort ="recent"
    return new Promise((resolve)=> {

request({url:'https://api.foursquare.com/v2/venues/'+venue_id+
'/tips', qs:params},(err, response, body)=> {
       if(err){ console.log(err); return; }
            resolve(body) }) })}

export function VenueHours(venue_id){
    return new Promise((resolve)=> {

request({url:'https://api.foursquare.com/v2/venues/'+venue_id+
'/hours', qs:params},(err, response, body)=> {
            if(err){ console.log(err); return; }
          resolve(body) }) })}
```

4.2.3 ผลการทดลอง

จากการทดลองพบว่าโปรแกรม FoursquareService สามารถดึงข้อมูลจาก Foursquare API ได้โดยมีรูปแบบข้อมูลดังนี้

ตัวอย่าง 4.2.1 ผลลัพธ์ที่ได้จากการดึงข้อมูลรายละเอียดสถานที่จาก Foursquare


```
{
"venue":{
  "id":"4c034d0cf56c2d7fa6c71c66",
  "name":"King Mongkut's Institute of Technology
Ladkrabang (KMITL)(สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง)",
  "location":{
   "address":"Chalong Krung Rd.",
   "lat":13.727480824630678,
   "lng":100.77714220791155,
   "postalCode":"10520",
   "cc":"TH",0
```


เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่าง 4.2.1 ผลลัพธ์ที่ได้จากการดึงข้อมูลรายละเอียดสถานที่จาก Foursquare (ต่อ)


```
    "city":"Lat Krabang",
    "state":"Bangkok",
    "country":"Thailand",
    "formattedAddress":[
       "Chalong Krung Rd.",
       "Lat Krabang",
       "Bangkok 10520",
       "Thailand"
    ]
     },
    "categories":[
          {
    "id":"4bf58dd8d48988d1ae941735",
    "name":"University",
    "pluralName":"Universities",
    "shortName":"University",
    "icon":{
     "prefix":
"https://ss3.4sqi.net/img/categories_v2/education/default_",
     "suffix":".png"
          },
    "primary":true
          }
  ]
}
```


ตัวอย่าง 4.2.2 ผลลัพธ์ที่ได้จากการดึงข้อมูลจำนวนคนจาก Foursquare


```
{
 "hereNow":{
  "count":10,
  "items":[]
     }
   }
```


เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่าง 4.2.3 ผลลัพธ์ที่ได้จากการดึงข้อมูลรูปภาพจาก Foursquare


```
{  "photos": {
   "count": 30,
   "items": [
            {
    "id": "4fad5d94e4b0e981d8c9e871",
    "createdAt": 1336761748,
    "source": {
     "name": "Instagram",
     "url": "http://instagram.com"
             },
    "prefix": "https://irs1.4sqi.net/img/general/",
    "suffix": "/_0n9y_E5UQCy4S6wpMWIIiIj4w-XMDyfw5-
UhGqdX68.jpg",
    "width": 612,
    "height": 612,
    "user": {
     "id": "18932150",
     "firstName": "Atsadawut",
     "lastName": "Jungwat",
     "gender": "male",
             },
   }, ...
}
```


เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่าง 4.2.4 ผลลัพธ์ที่ได้จากการดึงข้อมูลช่วงเวลาที่ได้รับความนิยมจาก Foursquare


```
{ "popular": {
  "timeframes": [
          {
   "days": [ 5 ],
   "includesToday": true,
   "open": [
              {
    "start": "0700",
    "end": "2000"
              }
   ],
   "segments": []
         },
  ...
] } }
```


ตัวอย่าง 4.2.5 ผลลัพธ์ที่ได้จากการดึงข้อมูลคำแนะนำจาก Foursquare


```
{    "tips": {
  "count": 48,
  "items": [
          {
   "id": "52ad2268498eb6c1fe7c794d",
   "createdAt": 1387078248,
   "text": "i love kmitl :)",
   "type": "user",
   "user": {
    "id": "73727353",
    "firstName": "Perfume",
    "lastName": "C.Suebpasuk",
    "gender": "female",
    "photo": {
     "prefix": "https://irs1.4sqi.net/img/user/",
     "suffix": "/73727353-THTVMMBPDIJZGAZK.jpg"
              },
              }
  },
  ...
}
```


เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่าง 4.2.6 ผลลัพธ์ที่ได้จากการดึงข้อมูลสถานที่ที่ผู้คนนิยมไปต่อไปจาก Foursquare


```
{    "nextVenues": {
  "count": 4,
  "items":[
          {
   "id":"4f98ca42e4b028100302dbde",
   "name":"Megabangna (เมกาบางนา)",
   "contact":{ ... },
   "location":{ ... },
   "categories":[...],
   "verified":false,
   "stats":{
    "checkinsCount":258801,
    "usersCount":56145,
    "tipCount":243
          },
   "url":"http://www.mega-bangna.com",
          },
   ...
  ]
```


4.3 การทดสอบการดึงข้อมูลจาก Facebook

4.3.1 วัตถุประสงค์

เพื่อดึงข้อมูลรายระเอียด โพสต์ และ ความคิดเห็นในเพจที่เกี่ยวกับสถานที่ที่ต้องการค้นหาจาก Facebook Graph API และตรวจสอบรูปแบบของข้อมูลที่ได้รับ

4.3.2 วิธีการทดสอบ

สมัคร Facebook for developer และสร้าง App เพื่อทำการขอ API ID และ API Secret เพื่อใช้ในการดึงข้อมูล โดยการดึงข้อมูลจะใช้ fb-graph ซึ่งเป็นไลบรารี่ของ Node.js ที่ใช้เชื่อมต่อกับ Facebook Graph API และใช้เพจ “Faculty of Engineering,KMITL” มี PAGE_ID (USER_ID) คือ “157556534255462” เป็นคำค้นหา

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรม 4.3 FacebookService

```
import graph from 'fbgraph'
import request from 'request'
let access_token =
"139610759813141|f928e2e59299981a677116c967a20b1d"
graph.setAccessToken(access_token)
export function getFbDetail(userID){
    var params ={fields:"name,picture"}
    return new Promise((resolve)=>{
   graph.get(userID,params, (err, res)=> {
            resolve(res)
        })
    })
}
export function getFbFeed(userID,since,until){
    var params
   if(since){
   if(until) params ={fields:"message,created_time",since:
since,until:until,limit:100}
      else  params ={fields:
"message,created_time",since:since,limit:100}
    }
    else{
        params ={fields:"message,created_time",limit:100}
 }
    return new Promise((resolve,reject)=> {
        graph.get(userID+"/feed",params,(err ,res)=>{
            resolve(res)})
    })
}
export function getFbComment(postID){
    let params ={summary :1}
    return new Promise((resolve)=> {
     graph.get(postID+"/comments",params,(err,res)=> {
            resolve(res)})
    })
 }
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.3 ผลการทดลอง

จากการทดลองพบว่าโปรแกรม FacebookService สามารถดึงข้อมูลจาก Facebook Graph API ได้โดยมีรูปแบบข้อมูลดังนี้

ตัวอย่าง 4.3.1 ผลลัพธ์ที่ได้จากการดึงข้อมูลรายละเอียดเพจจาก Facebook

```
{
 "name":"Faculty of Engineering,KMITL",
 "id":"157556534255462"
}
```

ตัวอย่าง 4.3.2 ผลลัพธ์ที่ได้จากการดึงข้อมูลโพสต์จาก Facebook

```
{ "data":[
      {
  "message":"ในวันเสาร์ที่ ๑๙ พฤศจิกายน ๒๕๕๙ คณะวิศวกรรมศาสตร์ ร่วมกับ สมาคมศิษย์เก่าวิศวลาดกระบัง จัดงาน \"รำลึกพระมหากรุณาธิคุณ พระบาทสมเด็จพระปรมินทมหาภูมิพลอดุลยเดช องค์อัครนวัตกร\" ณ หอประชุม คณะวิศวกรรมศาสตร์ สจล. เริ่มตั้งแต่เวลา ๑๐.๐๐ น. เป็นต้นไป",
  "story":"Faculty of Engineering,KMITL added 2 new photos.",
  "created_time":"2016-11-10T05:32:12+0000",
  "id":"157556534255462_1388791947798575"
      },
 ....
 ]
}
```

ตัวอย่าง 4.3.2 ผลลัพธ์ที่ได้จากการดึงข้อมูลความคิดเห็นจาก Facebook

```
"data":[
      {
  "created_time":"2016-11-10T12:14:54+0000",
  "from":{
   "name":"Kitcha Suriyanakat",
   "id":"1161005057321079"
        },
  "message":"โหะๆ",
  "id":"1387937664550670_1389012324443204"
      },
...
]
}
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 การทดลองหาความสัมพันธ์ของข้อมูลที่ได้จากสื่อสังคมออนไลน์

4.4.1 วัตถุประสงค์

เพื่อนำความสัมพันธ์ของข้อมูลที่ได้จากสื่อสังคมออนไลน์แต่ละอันมาวิเคราะห์ว่าสามารถนำข้อมูลใดมาเข้ากระบวนการทำนายได้บ้าง โดยใช้ข้อมูลของ สยามพารากอน เนื่องจากเป็นสถานที่ที่ข้อมูลจำนวนมากที่สุดและเป็นสถานที่ยอดนิยม

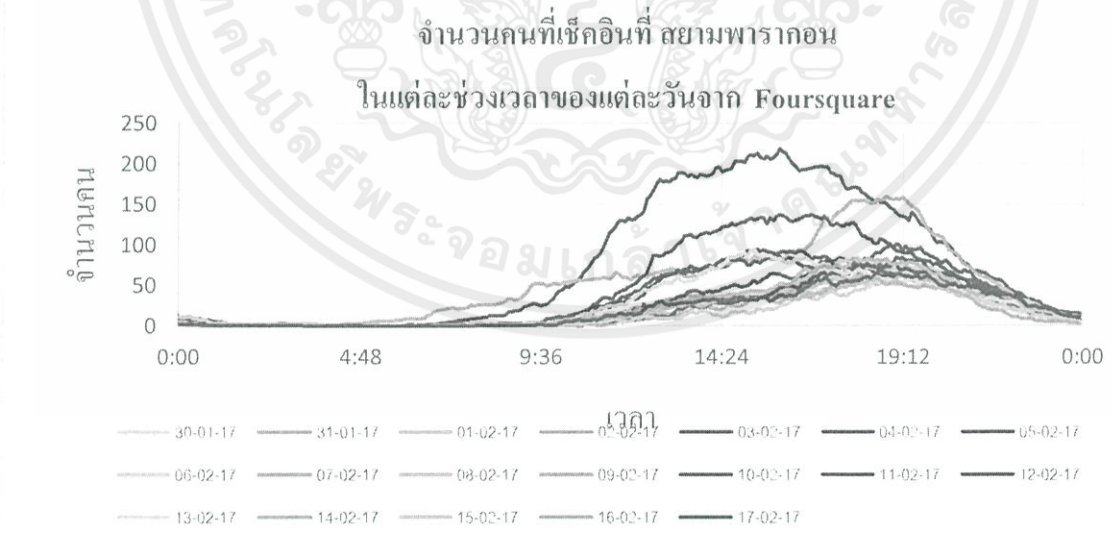
4.4.2 วิธีการทดสอบ

1) นำข้อมูลจำนวนคนที่เช็คอินที่ สยามพารากอน ที่เก็บไว้มาเข้ากระบวนการ Period Slicing และ Frequency Slicing เพื่อสร้างตาราง จากนั้นสร้างกราฟจากข้อมูลดังกล่าวเพื่อแสดงจำนวนคนที่เช็คอินในแต่ละช่วงเวลา
2) หาข้อมูลที่ได้มีความเป็นรูปแบบซ้ำวันหรือไม่
3) จากข้อมูลที่เหลือจากข้อมูลที่ได้ในข้อ 2) มีจำนวนข้อมูลทุกช่วงเวลา (ตามที่แบ่งใน Period Slicing) ใน 1 วันหรือไม่

4.4.3 ผลการทดลอง

4.4.3.1 ข้อมูลจำนวนคนที่เช็คอินในแต่ละสถานที่จาก Foursquare

จากการดึงข้อมูลจาก สยามพารากอน ตั้งแต่วันที่ 30/01/2017 จนถึง 17/02/2017 พบว่าข้อมูลมีความเป็นรูปแบบ มีข้อมูลทุกช่วงเวลา และมีการเพิ่มขึ้น-ลดลง สอดคล้องกับพฤติกรรมของผู้คนในสถานที่จริง ดังกราฟด้านล่าง



รูป 4.1 กราฟแสดงจำนวนคนที่เช็คอินที่ สยามพารากอน ในแต่ละช่วงเวลาของแต่ละวันจาก Foursquare

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.3.2 ข้อมูลจำนวนโพสต์ข้อความแนะนำสถานที่จาก Foursquare

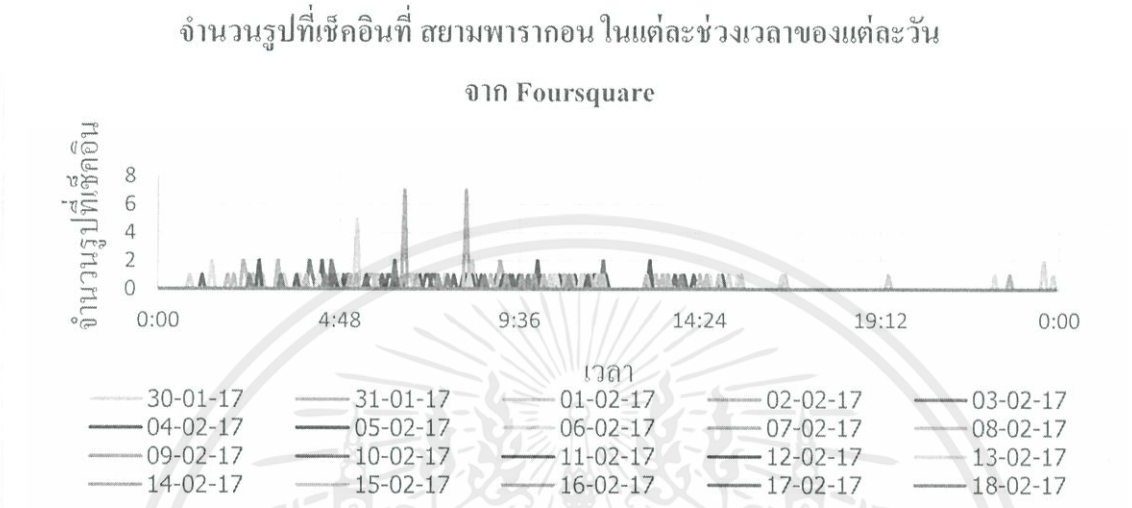
จากการดึงข้อมูลจาก สยามพารากอน ตั้งแต่ปี 2015 จนถึง 2017 พบว่ามีข้อมูลข้อความน้อยมาก และไม่มีรูปแบบ ดังกราฟด้านล่าง

รูป 4.2 กราฟแสดงจำนวนข้อความแนะนำสถานที่ สยามพารากอน ในแต่ละช่วงเวลาของแต่ละวันจาก Foursquare

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.3.3 ข้อมูลจำนวนรูปภาพที่ถ่ายและเช็คอินในแต่ละสถานที่จาก Foursquare

จากการดึงข้อมูลจาก สยามพารากอน ตั้งแต่วันที่ 30/01/2017 จนถึง 17/02/2017 พบว่ามีข้อมูลรูป 1-7 รูปในแต่ละวัน และไม่มีรูปแบบ ดังกราฟด้านล่าง

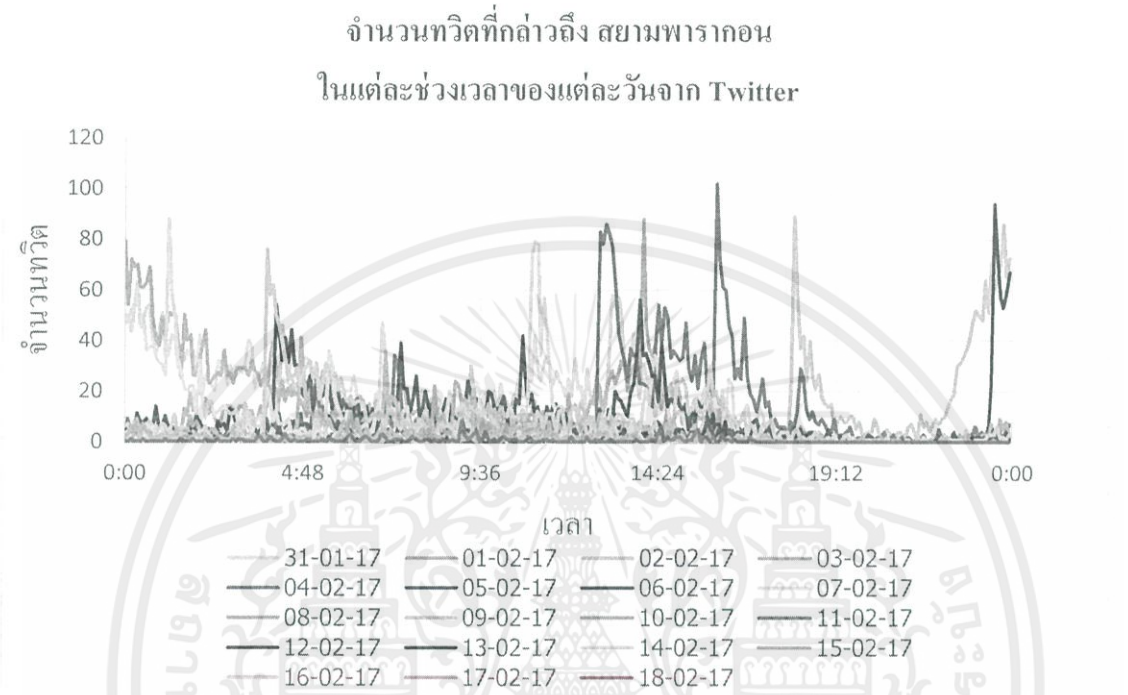


รูป 4.3 กราฟแสดงจำนวนรูปที่เช็คอินที่ สยามพารากอน ในแต่ละช่วงเวลาของแต่ละวันจาก Foursquare

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.3.4 ข้อมูลจำนวนทวิตที่กล่าวถึงแต่ละสถานที่จาก Twitter

จากการดึงข้อมูลทวิตจาก สยามพารากอน ตั้งแต่วันที่ 30/01/2017 จนถึง 18/02/2017 พบว่ามีข้อมูลที่ไม่มีรูปแบบ ดังกราฟด้านล่าง



รูป 4.4 กราฟแสดงจำนวนทวิตที่กล่าวถึง สยามพารากอน ในแต่ละช่วงเวลาของแต่ละวันจาก Foursquare

จากการหาความสัมพันธ์ของข้อมูลจากสื่อสังคมออนไลน์แต่ละอย่าง ในโครงงานนี้เลือกข้อมูลจำนวนคนที่เช็คอินในแต่ละสถานที่จาก Foursquare เนื่องจากเป็นข้อมูลที่มีรูปแบบและมีการเพิ่มขึ้น-ลดลง ใกล้เคียงกับจำนวนคนในสถานที่จริง

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5 การทดลองหาจำนวนวันที่ใช้เพื่อเป็นข้อมูลต้นในการทำนาย

4.5.1 วัตถุประสงค์

เพื่อทดสอบว่า ผลที่ได้จากการทำนายโดยใช้ข้อมูลจาก Foursquare ซึ่งจะใช้ข้อมูลต้นที่มีจำนวนวันแตกต่าง มีความแม่นยำมากเพียงใด เพื่อเลือกจำนวนวันที่เหมาะสมที่สุดในการทำนาย

4.5.2 วิธีการทดสอบ

1) ทำนายจำนวนคนของช่วงเวลา 5 นาทีถัดไป โดยใช้อัลกอริทึมในขั้นตอน Frequency Countingใส่ลงตาราง

ตาราง 4.1 จำนวนการเช็คอินในช่วงวันและเวลาต่างๆที่ สยามพารากอน ของ Foursquare

เวลา	วัน							
	1/3/2017	*2/3/2017*	*3/3/2017*	*4/3/2017*	*5/3/2017*	*6/3/2017*	*7/3/2017*	...
00:00	4	3	2	5	1	1	2	
...								
11.45	12	17	10	17	44	23	15	
11.50	12	15	13	18	47	22	16	
11.55	16	16	14	18	49	26	14	
12.00	12	14	14	17	54	28	18	
12.05	13	20	18	18	54	31	13	
...								
23:50	2	1	5	6	8	9	7	
23:55	5	6	10	12	12	9	3	

2) นำจำนวนคนที่มากที่สุดในข้อมูลจริงของ Foursquare ในช่วงเดียวกับที่ทำนาย มาใช้วิธีการคำนวณแบ่งออกเป็น 3 ช่วง ได้แก่ น้อย ปานกลาง มาก โดยใช้สูตร (3.1)
3) เปลี่ยนค่าที่ได้จากการทำนายซึ่งเป็นตัวเลขมาเป็นค่าของช่วง

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 4.2 ช่วงความหนาแน่นของการเช็คอินในช่วงวันและเวลาต่างๆ ที่ สยามพารากอน ของ Foursquare

เวลา	วัน							
	1/3/2017	*2/3/2017*	*3/3/2017*	*4/3/2017*	*5/3/2017*	*6/3/2017*	*7/3/2017*	...
00:00	น้อย	น้อย	น้อย	น้อย	น้อย	น้อย	น้อย	
...								
11.45	น้อย	น้อย	น้อย	น้อย	มาก	ปานกลาง	น้อย	
11.50	น้อย	น้อย	น้อย	น้อย	มาก	ปานกลาง	น้อย	
11.55	น้อย	น้อย	น้อย	น้อย	มาก	ปานกลาง	น้อย	
12.00	น้อย	น้อย	น้อย	น้อย	มาก	ปานกลาง	น้อย	
12.05	น้อย	ปานกลาง	น้อย	น้อย	มาก	ปานกลาง	น้อย	
...								
23:50	น้อย	น้อย	น้อย	น้อย	น้อย	น้อย	น้อย	
23:55	น้อย	น้อย	น้อย	น้อย	น้อย	น้อย	น้อย	

4) นำข้อมูลจริงของ Foursquare ในช่วงเดียวกับที่ทำนาย มาทำขั้นตอนที่ 3)

5) นำตารางที่ได้จาก ขั้นตอนที่ 3) และ 4) มาเปรียบเทียบกัน

ตาราง 4.3 ผลการทำนายในช่วงวันและเวลาต่างๆ ที่ สยามพารากอน ของ Foursquare

เวลา	วัน							
	1/3/2017	*2/3/2017*	*3/3/2017*	*4/3/2017*	*5/3/2017*	*6/3/2017*	*7/3/2017*	...
00:00	YES	YES	YES	YES	YES	YES	YES	
...								
11.45	YES	NO	YES	YES	YES	YES	YES	
11.50	YES	YES	YES	YES	YES	YES	YES	
11.55	YES	YES	YES	YES	NO	YES	YES	
12.00	YES	YES	YES	YES	YES	YES	NO	
12.05	YES	YES	YES	YES	YES	YES	YES	
...								
23:50	YES	YES	YES	YES	YES	YES	YES	
23:55	YES	YES	YES	YES	YES	YES	YES	

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6) จากนั้น นำมาคิดความแม่นยำ โดยใช้สูตร

$$\text{เปอร์เซ็นต์ความแม่นยำ} = \frac{\text{จำนวนช่วงเวลาที่ทำนายถูก}}{\text{จำนวนช่วงเวลาที่ทำนายทั้งหมด}} * 100 \qquad (4.1)$$

7) ทำนายโดยใช้ข้อมูลต้นเป็นจำนวน 1 ถึง 18 วัน จะได้ดังนี้

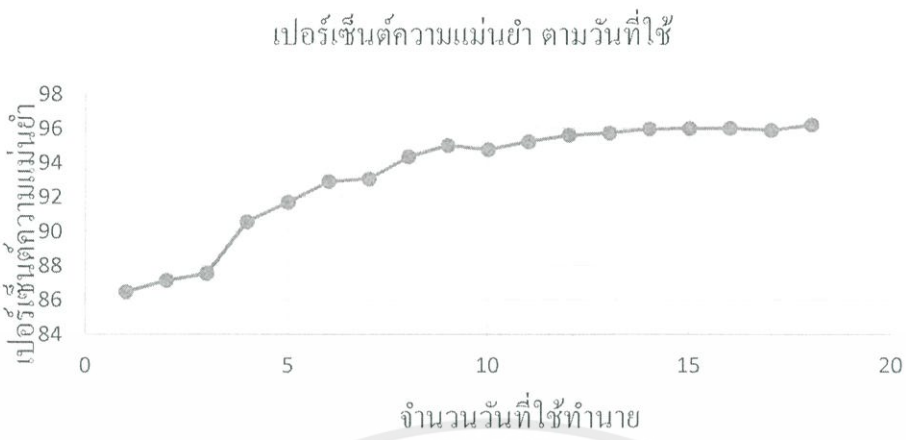
4.5.3 ผลการทดลอง

จากการทดลอง จะได้ค่าความแม่นยำของการใช้จำนวนวันเป็นข้อมูลต้นที่แตกต่างกัน ดังนี้

ตาราง 4.4 ค่าความแม่นยำเมื่อใช้จำนวนวันเป็นข้อมูลต้นแตกต่างกัน

จำนวนวันที่ใช้ทำนาย	เปอเซนต์ความแม่นยำ
1	86.48268093
2	87.13745271
3	87.54365541
4	90.56818182
5	91.70454545
6	92.90909091
7	93.06818182
8	94.36363636
9	95.02272727
10	94.77272727
11	95.25
12	95.61363636
13	95.72727273
14	95.97727273
15	96
16	96.02272727
17	95.90909091
18	96.22727273

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 4.5 กราฟแสดงเปอร์เซ็นต์ความแม่นยำตามจำนวนวันที่ใช้

โครงงานนี้จะเลือกใช้ 14 วันเนื่องจากในช่วง 8-18 วัน มีค่าความแม่นยำที่ใกล้เคียงกันมาก และ 14 วัน เป็นจำนวนวันที่ทำให้มีแต่ละวัน เป็นจำนวนเท่ากัน คือวันละ 2 ครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6 การทดลองหาค่าความแม่นยำของการทำนายความหนาแน่นในช่วง 5 นาทีถัดไปของแต่ละสถานที่

4.6.1 วัตถุประสงค์

เพื่อหาเปอร์เซ็นต์ความแม่นยำของแต่ละสถานที่ที่กำหนดไว้ ได้แก่ สยามพารากอน, สยามดิสคัฟเวอรี่, สยามเซ็นเตอร์, เอ็ม บี เค เซ็นเตอร์ และ เซ็นทรัลเวิลด์ จากนั้นดูว่าจำนวนของแต่ละช่วง ได้แก่ น้อย ปานกลาง มาก มีผลต่อเปอร์เซ็นต์ความแม่นยำหรือไม่

4.6.2 วิธีการทดสอบ

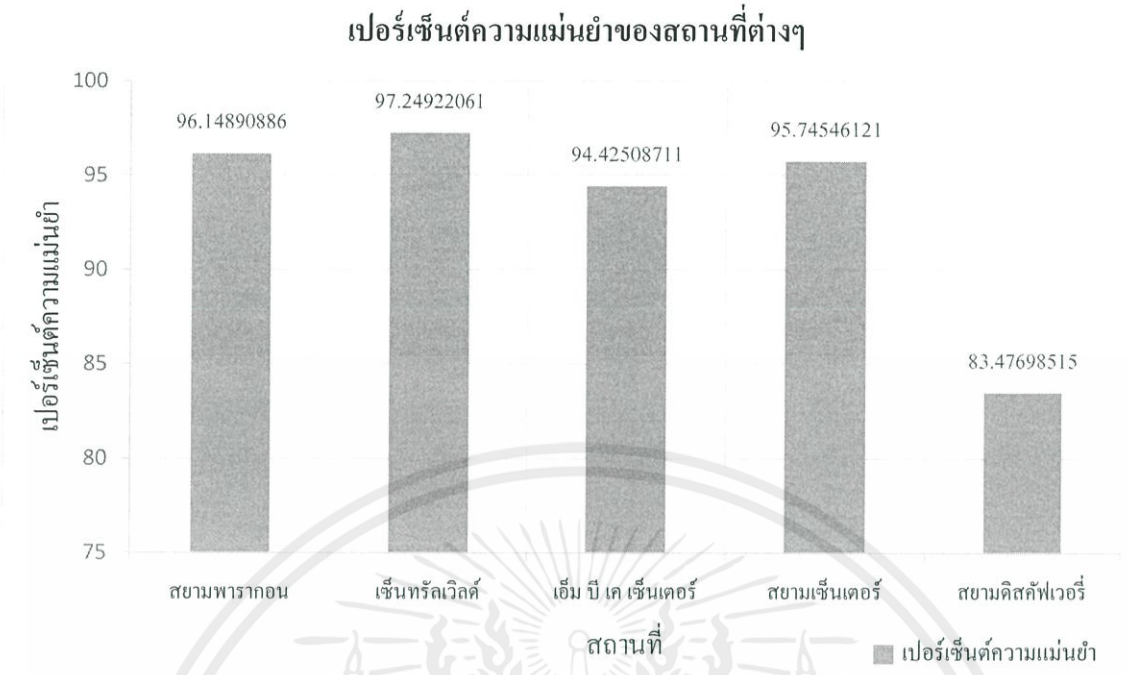
1) ทำการทดลองเช่นเดียวกับการทดลอง 4.4 ขั้นตอนที่ 1) – 6) ใช้ข้อมูลการทำนาย 5 นาที และ ใช้ข้อมูลย้อนหลัง 14 วัน
2) ทดลองซ้ำกับสถานที่ต่อไปนี้ คือ สยามพารากอน, สยามดิสคัฟเวอรี่, สยามเซ็นเตอร์, เอ็ม บี เค เซ็นเตอร์ และ เซ็นทรัลเวิลด์

4.6.3 ผลการทดลอง

ตาราง 4.6 เปอร์เซ็นต์ความแม่นยำของการทำนายความหนาแน่นในช่วง 5 นาทีถัดไป ในแต่ละสถานที่

สถานที่	เปอร์เซ็นต์ความแม่นยำ
สยามพารากอน	95.97826087
เซ็นทรัลเวิลด์	96.5
เอ็ม บี เค เซ็นเตอร์	94.4726087
สยามเซ็นเตอร์	95.43478261
สยามดิสคัฟเวอรี่	94.28509344
เฉลี่ย	95.33414912

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 4.6 กราฟแสดงเปอร์เซ็นต์ความแม่นยำของการทำนายความหนาแน่นในช่วง 5 นาทีถัดไป ในแต่ละสถานที่

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.7 การทดลองหาค่าความแม่นยำของการทำนายความหนาแน่นในช่วง 15 นาทีถัดไปของแต่ละสถานที่

4.7.1 วัตถุประสงค์

เพื่อหาเปอร์เซ็นต์ความแม่นยำของแต่ละสถานที่ที่กำหนดไว้ ได้แก่ สยามพารากอน, สยามดิสคัฟเวอรี่, สยามเซ็นเตอร์, เอ็ม บี เค เซ็นเตอร์ และ เซ็นทรัลเวิลด์ จากนั้นดูว่าจำนวนของแต่ละช่วง ได้แก่ น้อย ปานกลาง มาก มีผลต่อเปอร์เซ็นต์ความแม่นยำหรือไม่

4.7.2 วิธีการทดสอบ

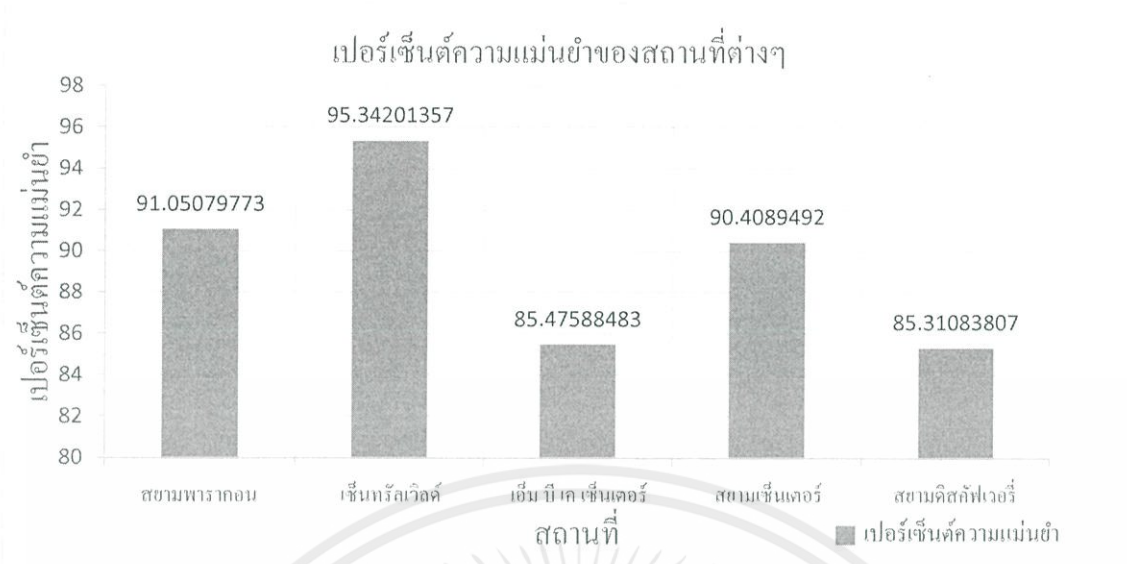
1) ทำการทดลองเช่นเดียวกับการทดลอง 4.4 ขั้นตอนที่ 1) – 6) ใช้ข้อมูลการทำนาย 15 นาที และ ใช้ข้อมูลย้อนหลัง 14 วัน
2) ทดลองซ้ำกับสถานที่ต่อไปนี้ คือ สยามพารากอน, สยามดิสคัฟเวอรี่, สยามเซ็นเตอร์, เอ็ม บี เค เซ็นเตอร์ และ เซ็นทรัลเวิลด์

4.7.3 ผลการทดลอง

ตาราง 4.7 เปอร์เซ็นต์ความแม่นยำของการทำนายความหนาแน่นในช่วง 15 นาทีถัดไปในแต่ละสถานที่

สถานที่	เปอร์เซ็นต์ความแม่นยำ
สยามพารากอน	91.05079773
เซ็นทรัลเวิลด์	95.34201357
เอ็ม บี เค เซ็นเตอร์	85.47588483
สยามเซ็นเตอร์	90.4089492
สยามดิสคัฟเวอรี่	85.31083807
เฉลี่ย	89.51769668

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 4.7 กราฟแสดงเปอร์เซ็นต์ความแม่นยำของการทำนายความหนาแน่นในช่วง 15 นาทีถัดไป ในแต่ละสถานที่

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.8 การทดลองการทำนายความเคลื่อนไหวของกลุ่มคน

4.8.1 วัตถุประสงค์

เพื่อทดสอบการทำงานของอัลกอริทึมการทำนายความเคลื่อนไหวของกลุ่มคน

4.8.2 วิธีการทดสอบ

1) ดึงข้อมูลการทำนายความเคลื่อนไหวของกลุ่มคนของสยามพารากอน เพราะเป็นสถานที่ยอดนิยม และมีการเช็คอินจำนวนมาก เป็นเวลา 1 วัน คือ 25/3/2560
2) นำข้อมูลที่ได้มาแสดงเป็นกราฟซึ่งแสดงจำนวนครั้งที่มีการทำนายว่ามีความเคลื่อนไหวของกลุ่มคน เป็นจำนวนกี่ครั้งในแต่ละช่วงเวลา

4.8.1 ผลการทดลอง

รูป 4.8 กราฟแสดงจำนวนครั้งที่ทำนายว่ามีการเคลื่อนไหวของกลุ่มคนในช่วงเวลาต่างๆ ของสยามพารากอน

จากกราฟ 4.7 จะเห็นว่าช่วงที่มีความเคลื่อนไหวของกลุ่มคนไปยังสถานที่อื่น ประกอบด้วย สยามดิสคัฟเวอรี่, สยามเซ็นเตอร์ และรถไฟฟ้าบีทีเอส สถานีสยาม ในช่วงเวลาประมาณ 9.00 น. ถึง 23.00 น. ซึ่งเป็นช่วงเวลาก่อนและหลังห้างสรรพสินค้า (สยาม พารากอน) จะเปิดปิดประมาณ 1 ชั่วโมง และ ไม่มีการเคลื่อนไหวของกลุ่มคนในช่วงเวลา เที่ยงคืน ถึง 8.00 น. ซึ่งเป็นช่วงเวลาที่ตรงกับการใช้ชีวิตประจำวัน

ดังนั้น จึงสรุปได้ว่า การทำนายความเคลื่อนไหวของกลุ่มคน มีความใกล้เคียงกับการใช้ชีวิตประจำวัน

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
บทสรุปและข้อเสนอแนะ

5.1 บทสรุป

ระบบทำนายความเคลื่อนไหวของกลุ่มคนจากสื่อสังคมออนไลน์ เป็นแอปพลิเคชั่นที่ใช้ทฤษฎีการวิเคราะห์ทางสถิติมาประยุกต์ใช้ในการทำนายหาความหนาแน่นและความเคลื่อนไหวของกลุ่มคน โดยแสดงแนวโน้มความหนาแน่นของแต่ละสถานที่ในอนาคต และแนวโน้มการเคลื่อนไหวของกลุ่มคนว่าจะไปสถานที่ใดต่อในเวลากำหนดไว้ เพื่อให้ผู้ใช้สามารถวางแผนการเดินทางได้

ใช้ข้อมูลจำนวนคนที่เช็คอินในสถานที่ต่างๆจาก Foursquare มาใช้ในการทำนาย เก็บข้อมูลแบบกระจายโดยใช้ Spark SQL และ Hadoop Distributed File System มาแบ่งตามช่วงเวลา ช่วงละ 5 นาทีของแต่ละวัน จากนั้นใช้ข้อมูลจำนวนคนที่เช็คอินย้อนหลัง 14 วัน เพื่อทำนายจำนวนคนในช่วงเวลาที่ต้องการ จากนั้นนำจำนวนคนมาหาช่วงความหนาแน่น ส่วนการเคลื่อนไหวของกลุ่มคนหาจากจำนวนคนที่เปลี่ยนแปลงไปในช่วงเวลาที่ทำนายของสถานที่ใกล้เคียง จากการทดสอบวัดค่าความถูกต้องของการทำนายความหนาแน่นของการทำนาย 5 นาทีถัดไป ได้ค่าความถูกต้องเฉลี่ย คือ ร้อยละ 95.3341491 ส่วนการทำนายความหนาแน่นของการทำนาย 15 นาทีถัดไป ได้ค่าความถูกต้องเฉลี่ย คือ ร้อยละ 89.51769668

5.2 ปัญหา และแนวทางแก้ไข

1) ข้อมูลการเช็คอินจาก Facebook ไม่เปิดให้ใช้ แก้ปัญหาโดยการดึงข้อมูลจากสื่อสังคมออนไลน์อื่นๆ คือ Twitter และ Foursquare
2) ข้อมูล ข้อความ Tweet จาก Twitter, ข้อความแนะนำสถานที่ และการเช็คอินด้วยรูปภาพจาก Foursquare ไม่เป็นรูปแบบที่แน่นอนทำให้ไม่สามารถนำมาทำนายได้ แก้ปัญหาโดยการใช้ข้อมูลจำนวนคนที่เช็คอิน จาก Foursquare ที่มีรูปแบบแน่นอน

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3 แนวทางการพัฒนาต่อ

1) เพิ่มตัวเลือกช่วงเวลาในการทำนายเพิ่มขึ้น เพื่อให้ตอบสนองต่อการใช้งานวางแผนการเดินทางได้มากขึ้น
2) พัฒนาอัลกอริทึมที่ใช้ทำนายเพื่อให้มีความแม่นยำมากขึ้น
3) ใช้ library ที่ช่วยหาเวลาในการเดินทางระหว่างสถานที่มาช่วยในการกำหนดขอบเขตสถานที่เพื่อทำนายความเคลื่อนไหวของกลุ่มคน
4) ใช้ข้อมูลจาก Twitter มาช่วยในการทำนายเหตุการณ์พิเศษที่มีผู้คนกล่าวถึงมาก เช่น งานลดราคาหรืองานอีเวนต์พิเศษ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] V. Kounev, "Where will I go next?: Predicting Future Categorical Check-ins in Location Based Social Networks", 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing, pp. 605-610, 2012.

[2] Tom White, "Hadoop The definitive guide," in O'Reilly Media,Inc., 2012.

[3] Apache, "Apache Hadoop," [Online]. Available: http://hadoop.apache.org/.

[4] Facebook Inc., "React," [Online]. Available: https://facebook.github.io/react/.

[5] Foursquare, "Foursquare API," [Online].Available: https://developer.foursquare.com.

[6] Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, Ion Stoica, "Spark: Cluster Computing with Working Sets", HotCloud 2010, June 2010.

[7] Apache Spark, "Apache Spark," [Online].Available: http://spark.apache.org/.

[8] Twitter, "REST APIs," [Online].Available: https://dev.twitter.com/rest/public.

[9] Wikipedia, "Computer cluster," [Online]. Available: https://en.wikipedia.org/wiki/Computer_cluster.

[10] Wikipedia, "Facebook," [Online]. Available: https://en.wikipedia.org/wiki/Facebook.

[11] Wordpress, "Location Based Service," [Online]. Available: https://goo.gl/axmxRj.

[13] Apache Software Foundation, "Apache Parquet," [Online]. Available: https://parquet.apache.org/.

[14] Daniel J. Abadi, Peter A. Boncz, Stavros Harizopo, "Column-oriented Database System", VLDB '09, 2009.

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[15] Foursquare, “About foursquare,” [Online]. Available: https://th.foursquare.com/about.

[16] Twitter, “Twitter company,” [Online]. Available: https://about.twitter.com/company.

[17] Wikipedia, "Twitter," [Online]. Available: https://goo.gl/XHR9ll.

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้