

## การประเมินสภาพการจราจรจากข้อมูลสื่อสังคม Traffic Condition Estimation from Social-based Data

### อศมาภรณ์ สิทธิ<sup>1</sup>

#### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการใช้เทคโนโลยีการรับรู้ทางสังคมที่มีความเชื่อมโยงกับการจราจรบนท้องถนน โดยอาศัยเทคนิคการวิเคราะห์เหมืองข้อความที่เกี่ยวข้องกับการจราจรจากผู้ใช้เครือข่ายสังคม (social network) ในการพัฒนาเทคโนโลยีสารสนเทศดังกล่าวสามารถนำไปใช้สำหรับประเมินการจราจรเชิงพื้นที่ของข้อมูลสื่อสังคมออนไลน์ที่เชื่อมโยงกับระบบการคมนาคมขนส่งในปัจจุบัน แสดงให้เห็นถึงปัจจัยที่มีอิทธิพลต่อกิจกรรมต่างๆ ที่เกิดขึ้นบนท้องถนน ได้แก่ ความสามารถในการเข้าถึง โครงสร้างพื้นฐาน การให้บริการเครือข่ายอินเทอร์เน็ต ระบบคมนาคมขนส่ง และการบริการที่มีความสำคัญ เพื่อเป็นแนวทางพัฒนา แก้ไข และป้องกัน เพื่อลดจำนวนการเกิดอุบัติเหตุ ให้มีประสิทธิภาพยิ่งขึ้นซึ่งจะช่วยให้ผู้ใช้รถใช้ถนนสามารถขับขี่และสัญจรในพื้นที่ได้อย่างปลอดภัย

**คำสำคัญ:** การจราจร เหมืองข้อมูล สื่อสังคม

#### Abstract

This research aims to analyze the estimation of traffic condition from the geo-tagged information in Bangkok, Thailand using data mining technique. The significant of the research are implementing VGI information with for transportation application, and using natural language processing method to mine traffic events. The check-in information is utilized using database management system, text mining and pattern analysis in order to refer traffic events which can be related connectivity, accessibility, infrastructure, internet network services provider, transportation and other important facilities. The contribution of this research could be the baseline to shows the relationship with user behaviour regarding transportation tracking using geographical of social media technology.

**Keywords:** Traffic conditions, Social sensing, Text mining

#### บทนำ

โลกในยุคโลกาภิวัตน์มีการพัฒนาเทคโนโลยีสารสนเทศ ความก้าวหน้าทางเทคโนโลยีก่อให้เกิดสิ่งประดิษฐ์ใหม่ๆ การพัฒนาอย่างต่อเนื่องและการเปลี่ยนแปลงอย่างเป็นพลวัตทำให้ประชาชนได้รับข่าวสารผ่านสื่อที่รวดเร็วและทันเหตุการณ์ เครือข่ายอินเทอร์เน็ตที่ครอบคลุมทั่วโลกทำให้สามารถติดต่อกันได้อย่างไร้พรมแดน การพัฒนาเครื่องมือสื่อสารและการให้บริการเครือข่ายอินเทอร์เน็ตที่ครอบคลุมก่อให้เกิดความสะดวกในการดำเนินชีวิตประจำวันในยุคปัจจุบัน

เทคโนโลยีการรับรู้ทางสังคม (Social Network Service) เป็นบริการเครือข่ายสังคมออนไลน์บนเว็บไซต์ที่ให้ผู้คนทั่วโลกสามารถเข้ามาใช้บริการร่วมกันได้ กลุ่มคนเหล่านั้นสามารถทำความรู้จัก แลกเปลี่ยนความคิดเห็นหรือเชื่อมโยงข้อมูล

<sup>1</sup> อาจารย์ประจำภาควิชาภาษาไทย คณะสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

กิจกรรมความสนใจในทิศทางใดทิศทางหนึ่ง บริการสังคมออนไลน์มักจะประกอบไปด้วย การส่งข้อความ ภาพ เสียง วิดีโอ เป็นต้น Social Network ยังเป็นรูปแบบการใช้ชีวิตแบบใหม่ที่สามารถจัดจัดจำกัดทางด้านเวลา ระยะเวลา และงบประมาณต่างๆ จึงได้รับการตอบรับอย่างรวดเร็ว โดยมีจุดเด่น คือ ทำให้การสื่อสารมีประสิทธิภาพ สื่อสารได้หลากหลายรูปแบบ สามารถเข้าถึงกลุ่มเป้าหมายได้อย่างรวดเร็ว เช่น ใช้ในกรณีฉุกเฉินหรือการเกิดอุบัติเหตุ การให้บริการนำทาง การให้บริการข้อมูลต่างๆ หรือทางด้านการจราจรและระบบขนส่ง ได้นำมาใช้เพื่อแบ่งปันข้อมูลการจราจร สภาพความแออัด อุบัติเหตุ และสภาพอากาศที่ส่งผลต่อการจราจร (Napong, 2011) หรือพูดคุยตอบคำถาม ทำให้สามารถสื่อสารระหว่างบุคคลอย่างเป็นสาธารณะ

ในงานวิจัยนี้ เป็นการนำข้อมูลจากการแบ่งปันแลกเปลี่ยนข้อมูลตนเอง ข้อมูลภาพ เสียง วิดีโอ และสถานที่ในรูปแบบของการระบุตำแหน่ง จากโปรแกรมประยุกต์ที่ให้กลุ่มคนเข้ามามีส่วนร่วมจากหลากหลายแหล่งข้อมูล (Crowdsourcing) ผ่านการเรียกใช้เอพีไอ (Application programming interface : API) ในด้านของแพลตฟอร์มเว็บ โปรแกรมโปรแกรมมิ่ง (Web programming platform) เช่น เอชทีเอ็มแอลรุ่นที่ 5 (HTML5) เฟสบุ๊ค (Facebook) ฟลิคเกอร์ (Flickr) และทวิตเตอร์ (Twitter) ได้มีการสนับสนุนส่วนที่สนับสนุนการใช้งานสำหรับการบรรจุข้อมูลการระบุตำแหน่งทางภูมิศาสตร์ (Geo-location) ในการได้มาซึ่งข้อมูลระบุตำแหน่งทางภูมิศาสตร์ (Geotagged data) เข้าไปด้วย เทคโนโลยีเหล่านี้ช่วยอำนวยความสะดวกให้เกิดการสร้างข้อมูลทางภูมิศาสตร์ขึ้นและกระจายข้อมูลได้ ซึ่งมีการนำเสนอข้อมูลที่มีความสัมพันธ์ของรูปแบบการดำเนินชีวิตในเชิงพื้นที่และเชิงเวลา โดยมุ่งเน้นไปที่ข้อมูลที่สามารถเปลี่ยนแปลงได้ และไม่อยู่ในรูปของข้อมูลทางภูมิศาสตร์ เช่น ข้อมูล ดัชนีถ้อยคำ (Hashtag) โดยงานวิจัยได้นำเสนอวิธีการคือวิธีในการสร้างเครื่องมือเพื่อพัฒนาคลังข้อมูลเชิงพื้นที่โดยรวบรวมข้อมูล จัดเก็บข้อมูลและสร้างฐานข้อมูล จากแหล่งข้อมูลที่หลายหลาย มาใช้ในการวิเคราะห์ข้อมูลสภาพการจราจร โดยใช้แบบจำลองเบย์ในการจำแนกเพื่อให้มีความถูกต้องมากยิ่งขึ้นในการประเมินรูปแบบการจราจรประเภทต่างๆ ของผู้ใช้งานสื่อสังคมออนไลน์รวมถึงการวิเคราะห์ความสัมพันธ์ระหว่างรูปแบบการจราจรกับข้อมูลสื่อสังคมออนไลน์ในเชิงพื้นที่และเชิงเวลา

### วัตถุประสงค์ของการวิจัย

1. พัฒนาระบบฐานข้อมูลที่เกี่ยวข้องกับการจราจรจากข้อมูลสื่อสังคม
2. วิเคราะห์เหมืองข้อความและประเมินการจราจรจากข้อมูลสื่อสังคม

### สมมุติฐานของการวิจัย

1. ระบบฐานข้อมูลการจราจรจากข้อมูลระบุตำแหน่งสื่อสังคม
2. แบบจำลองอัตโนมัติสามารถสกัดและจำแนกสภาพการจราจรจากข้อมูลตัวอักษรและข้อมูลระบุตำแหน่งสื่อสังคม และเป็นพื้นฐานสำหรับการวิจัยซึ่งแสดงให้เห็นถึงความสัมพันธ์กับการติดตามพฤติกรรมหรือกิจกรรมของมนุษย์ในปัจจุบันกับการใช้เทคโนโลยีสื่อทางสังคมในทางภูมิศาสตร์

### ขอบเขตของการวิจัย

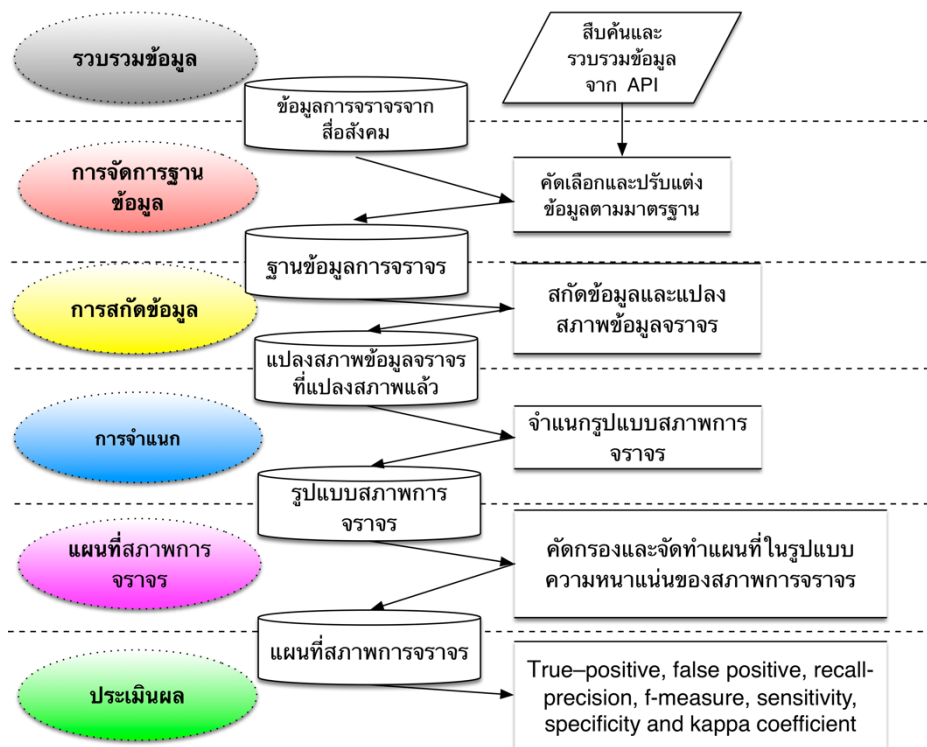
งานวิจัยนี้เป็นการศึกษารูปแบบการจราจรบนท้องถนนของผู้ใช้งานสื่อสังคมออนไลน์บริเวณกรุงเทพมหานครและปริมณฑลโดยมีการกำหนดขอบเขตการวิจัยใน 4 ลักษณะ คือ ขอบเขตเชิงพื้นที่ ขอบเขตเชิงเนื้อหา ขอบเขตเชิงระยะเวลา และขอบเขตด้านประชากร โดยมีรายละเอียดดังนี้

1. ขอบเขตเชิงพื้นที่ พื้นที่ศึกษารอบกลุ่มกรุงเทพมหานครบนถนนทางหลวง

2. ขอบเขตเชิงเนื้อหา ครอบคลุม 2 ประเด็นหลัก ได้แก่ พัฒนาระบบฐานข้อมูลระบุตำแหน่งสื่อสังคมออนไลน์และพัฒนาแบบจำลองอัตโนมัติในการประเมินการจราจรบนท้องถนนของผู้ใช้งานสื่อสังคมออนไลน์
3. ขอบเขตเชิงระยะเวลา ระยะเวลาของการวิจัย คือ การระบุตำแหน่งผ่านสื่อสังคมออนไลน์ พ.ศ. 2560
4. ขอบเขตด้านประชากร ประชากรที่เป็นเป้าหมายในการวิจัย ได้แก่ การระบุตำแหน่งของผ่านสื่อสังคมออนไลน์

**กรอบความคิดที่ใช้วิจัยและงานวิจัยที่เกี่ยวข้อง**

ในงานวิจัยนี้ เป็นการนำข้อมูลจากการแบ่งปันแลกเปลี่ยนข้อมูลตนเองจากข้อมูลทวิตเตอร์ (Twitter) ในการได้มาซึ่งข้อมูลระบุตำแหน่งทางภูมิศาสตร์ (Geotagged data) มาใช้ในการประเมินสภาพการจราจรบนท้องถนน โดยใช้แบบจำลองเบย์ในการจำแนกประเภทของเหตุการณ์ เพื่อให้มีความถูกต้องมากยิ่งขึ้นในการประเมินรูปแบบการจราจรประเภทต่างๆ ของผู้ใช้งานสื่อสังคมออนไลน์รวมถึงการวิเคราะห์ความสัมพันธ์ระหว่างรูปแบบการจราจรกับข้อมูลสื่อสังคมออนไลน์ในเชิงพื้นที่และเชิงเวลา ดังภาพที่ 1



ภาพที่ 1 วิธีดำเนินงานวิจัย

ทฤษฎีของเบย์ (bayes' theorem) เป็นการจำแนกแบบรู้จำประเภทหนึ่งที่น่าสนใจของความรู้ของความเป็นแบบมีเงื่อนไข เข้ามาช่วยในการหาคำตอบของการจำแนกและประเมินประเภท ซึ่งการเรียนรู้แบบเบย์ (bayesian learning) เป็นเทคนิคที่ใช้ทฤษฎีความน่าจะเป็น ตามกฎของเบย์ (bayes' Theorem) เพื่อหาสมมติฐานที่มีถูกต้องที่สุด ได้แก่ ความน่าจะเป็นก่อนหน้าสำหรับสมมติฐานหนึ่ง ๆ ที่มีลักษณะเหมือนหรือคล้ายคลึงกัน ร่วมกับข้อมูลจำนวน ในการนำเอาเทคโนโลยีทางด้านสื่อสังคมออนไลน์มาประยุกต์กับทฤษฎีของเบย์จะทำให้สามารถจำแนกประเภทของการดำรงชีวิตผ่านข้อมูลสื่อสังคมออนไลน์ที่มีความสัมพันธ์กันในเชิงพื้นที่และเชิงเวลา

ธนเดช และคณะ (2558) ได้ศึกษาการคำนวณสภาพการจราจรในเมืองใหญ่จากการใช้สื่อสังคมออนไลน์ร่วมกับข้อมูลระบบนำทางต่างๆ เช่น Google Map, Longdo Map เพื่อช่วยให้ข้อมูลการจราจรมีประสิทธิภาพมากขึ้น โดยวิเคราะห์

ต้นแบบข้อมูลในรูปแบบการสังเคราะห์ประโยชน์จากโปรแกรม TLexs และใช้ HMM Model ในการบันทึกและวิเคราะห์พฤติกรรมข้อมูลเพื่อแปลงสภาพข้อมูลเหล่านี้ให้เป็นข้อมูลการจราจร

รังสรรค์ สุขชัยรังสรรค์ (2553) ได้ศึกษาค้นคว้าการวิเคราะห์ชั้นข้อมูลพร้อมจัดทำแผนที่ข้อมูล GIS (สารสนเทศภูมิศาสตร์) และแสดงบริเวณอันตรายตามช่วงพิจารณา ของทางหลวงหมายเลข 1 ตอน(เชียงใหม่ – แม่จัน) โดยทำการรวบรวมสถิติอุบัติเหตุบนทางหลวงจากหน่วยงานที่เกี่ยวข้องและเก็บข้อมูลเพิ่มเติมประกอบด้วย ข้อมูลจุดเกิดอุบัติเหตุ ข้อมูลอุปกรณ์งานทาง ข้อมูลสายทาง ทำการเก็บข้อมูลเป็นระบบพิกัดควบคู่กับระบบกิโลเมตร ทำการถ่ายภาพด้วยกล้องดิจิทัลและบันทึกภาพเคลื่อนไหวจุดที่เกิดอุบัติเหตุรวมถึงภาพอุบัติเหตุใหญ่ที่เกิดขึ้นหลังจากนั้นนำข้อมูลทั้งหมดแปลงเป็นข้อมูล GIS แสดงผลการวิเคราะห์ข้อมูล GIS ในรูปแบบแผนที่ โดยการแบ่งลักษณะเป็นช่วงอันตราย( Black Section) และจุดอันตราย (Black Spot)

ภัทรสุดา วิชยพงศ์ (2554) ได้ศึกษาค้นคว้าการวิเคราะห์หาจุดเสี่ยงอันตรายของการเกิดอุบัติเหตุจราจรบนถนนสายหลักในเขตพื้นที่อำเภอเมือง จังหวัดนครราชสีมา ควบคู่กับการประยุกต์ใช้นาฬิกาอุบัติเหตุในการวางแผนด้านความปลอดภัย และแก้ไขปัญหาอุบัติเหตุให้กับหน่วยงานที่เกี่ยวข้องกับงานด้านอุบัติเหตุจราจร ผู้วิจัยได้ทบทวนวิธีการที่ใช้ในการกำหนดจุดอันตรายและได้เลือกใช้วิธี Rate Quality Control ในการกำหนดจุดอันตรายต่อการเกิดอุบัติเหตุเนื่องจากมีความเหมาะสมกับข้อมูลที่ใช้ในการวิเคราะห์ โดยงานวิจัยนี้ได้พัฒนาระบบที่ช่วยในการวิเคราะห์หาจุดเสี่ยงอันตรายบนถนนในรูปแบบของเว็บไซต์ ASP.NET โดยประยุกต์ใช้ Microsoft Visual Studio 2010 C# Express เป็นเครื่องมือในการพัฒนาระบบและมีการจัดเก็บข้อมูลในรูปแบบฐานข้อมูลสารสนเทศด้วย Microsoft SQL Server 2005 Express เพื่อความสะดวกรวดเร็ว ง่ายต่อการบันทึก วิเคราะห์ และจัดเก็บข้อมูล ซึ่งแสดงผลจุดเสี่ยงการเกิดอุบัติเหตุในรูปแบบแผนที่อิเล็กทรอนิกส์โดยใช้ Google Map

## วิธีดำเนินการวิจัย

### 1. พัฒนาระบบฐานข้อมูลที่เกี่ยวข้องกับการจราจรจากข้อมูลสื่อสังคม

งานวิจัยนี้รวบรวมข้อมูลโดยการสืบค้นข้อมูลจากเว็บเครือข่ายสังคม ทวิตเตอร์ ผ่านการเรียกใช้เอพีไอ (API) ของเว็บไซต์ ซึ่งคือ คำสั่ง (Code) ที่อนุญาตให้สามารถให้ผู้ให้บริการกับผู้ใช้บริการสื่อสารระหว่างกันได้ โดยรวบรวมข้อมูลผ่านโปรแกรมรหัสเปิดอาร์ (R programming) เพื่อดึงข้อมูลสาธารณะจากผู้ใช้งานทั่วไปที่มีการอัปโหลดในสื่อสังคมออนไลน์โดยมีการพัฒนาโปรแกรมเพื่อทำการดึงข้อมูล ในรูปแบบของ Representational state transfer หรือ REST ซึ่งเป็นการดึงข้อมูลผ่าน URL ในลักษณะของการร้องขอข้อมูลตามเงื่อนไขที่กำหนดเพื่อให้ได้รับข้อมูลส่งกลับมายังในรูปแบบของ Extensible Markup Language หรือ XML ซึ่งเป็นรูปแบบเอกสารที่เป็นตัวอักษรและรองรับการเข้ารหัส (unicode) ภาษาไทย ซึ่งในปัจจุบันยังคงเป็นข้อจำกัดในการวิเคราะห์และประมวลผลธรรมชาติของภาษา ข้อมูล twitter ที่สืบค้นได้จะเรียกว่าข้อมูลของข้อมูล (Metadata) ซึ่งประกอบไปด้วยคุณลักษณะ วันและเวลาที่อัปโหลดข้อมูล (date and time) ชื่อผู้ใช้งาน (username) รหัสผู้ใช้งาน (id) ภาษา (Language) แหล่งที่อัปโหลดข้อมูล (Source) และ ข้อความ (Text) ระบบบูรณาการฐานข้อมูลการจราจรนี้จะถูกพัฒนาให้อยู่ในรูปแบบฐานข้อมูล เพื่อให้สามารถนำมาใช้ในการวิเคราะห์และแปลงลักษณะข้อมูลในรูปแบบที่นำไปสู่การจำแนกประเภทสภาพการจราจรต่อไป

### 2. วิเคราะห์เหมืองข้อความและประเมินการจราจรจากข้อมูลสื่อสังคม

#### 2.1 วิเคราะห์เหมืองข้อความ

การประมวลผลข้อความขั้นต้นคือการคัดเลือก กรอง และจัดการข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปหาความสัมพันธ์ที่เกี่ยวข้องกับการจราจร โดยใช้เทคนิคดังต่อไปนี้

1) การลบคำที่ไม่สำคัญหรือไม่เกี่ยวข้องกับการจรรยา (stop words) ได้แก่ ตัวเลข ตัวอักษรภาษาอังกฤษ สัญลักษณ์ที่ไม่ใช่ตัวอักษร คำซ้ำต่างๆ

2) การตัดคำออกมาเป็นคำเดี่ยวๆ (tokenization) เพื่อแยกคำตามรูปคำหรือการอ่านออกเสียง สำหรับภาษาไทย เป็นภาษาที่ซับซ้อนเนื่องจากมีรูปแบบของคำที่ต่อเนื่อง ไม่มีการเว้นวรรคตามการอ่านหรือตามประโยค งานวิจัยนี้ได้ประยุกต์ใช้เทคนิคการตัดคำจากไลบรารี Thai Natural Language Processing (ThaiNLP) ซึ่งเป็นการสอนให้โปรแกรมคอมพิวเตอร์เข้าใจภาษามนุษย์นั่นเอง ซึ่งนับว่าเป็นประเด็นที่ค่อนข้างซับซ้อนและท้าทายอย่างมาก

3) การจัดลำดับความสำคัญของข้อมูล หลังจากตัดคำภาษาไทยออกเป็นกลุ่มคำหรือประโยคแล้ว ทำการวิเคราะห์ ความถี่ของคำที่พบและมีการกล่าวถึงข้อความนั้นบ่อยครั้ง เพื่อหาความสัมพันธ์ที่เกี่ยวข้องกับการสภาพการจรรยาให้อยู่ในรูปแบบถุงของข้อมูล (bag of words) จัดลำดับข้อมูลและนำมาแสดงผลในรูปแบบของดัชนีถ้อยคำก้อนเมฆ (tagcloud) และกราฟแสดงความถี่ของดัชนีถ้อยคำ (frequency of words)

## 2.2 การประเมินการจรรยาจากเหมืองข้อมูล

ทฤษฎีของเบย์ (bayes' theorem) เป็นการจำแนกแบบรู้จำประเภทหนึ่งที่สำคัญหลักการ ของความน่าจะเป็นแบบมีเงื่อนไข เข้ามาช่วยในการหาคำตอบของการจำแนกและประเมินประเภท ซึ่งการเรียนรู้ แบบเบย์ (bayesian learning) เป็นเทคนิคที่ใช้ทฤษฎีความน่าจะเป็น ตามกฎของเบย์ (bayes' Theorem) เพื่อหาสมมติฐานที่มีถูกต้องที่สุด ได้แก่ ความน่าจะเป็นก่อนหน้าสำหรับสมมติฐานหนึ่ง ๆ ที่มีลักษณะเหมือนหรือคล้ายคลึงกัน ร่วมกับข้อมูลจำนวน ในการนำเอาเทคโนโลยีทางด้านสื่อสังคมออนไลน์มาประยุกต์กับทฤษฎีของเบย์จะทำให้สามารถจำแนกประเภทของการดำรงชีวิตผ่านข้อมูลสื่อสังคมออนไลน์ที่มีความสัมพันธ์กันในเชิงพื้นที่และเชิงเวลา

ในการประเมินและคาดคะเนสภาพการจรรยาและอุบัติเหตุ ผ่านรูปแบบกิจกรรมกิจกรรมต่างๆ ที่เผยแพร่จากสื่อสังคมออนไลน์ ภายใต้การเปลี่ยนแปลงสภาพการจรรยาบนท้องถนน โดยประยุกต์กฎของ Bayes เป็นการคำนวณความน่าจะเป็นในการประเมินสภาพการจรรยาและอุบัติเหตุ  $P(c|d)$  ในสถานการณ์ต่างๆ ที่เกี่ยวข้องกับการจรรยาและอุบัติเหตุ  $P(c|d)$  จากข้อมูลสื่อสังคมออนไลน์จากฐานข้อมูลขนาดใหญ่ที่มีอยู่ (Asamaporn, 2016) ดังสมการ

$$P(c|d) = P(c) \frac{P(d|c)}{P(d)} .$$

ในการประเมินสภาพการจรรยาจากตัวแปรซึ่งเป็นดัชนีถ้อยคำที่รวบรวมมาได้ เป็นตัวแปรที่ไม่ต่อเนื่อง โดยใช้คุณลักษณะการปรับมาตรฐานความกว้างเท่ากันของข้อมูลให้เท่ากัน เราแบ่งช่วงของแต่ละตัวแปรเป็น เพื่อนำไปคำนวณและจัดกลุ่มประเภทสภาพการจรรยาให้เป็นหมวดหมู่ตามความน่าจะเป็นของข้อมูล อย่างไรก็ตาม ในการประมาณค่าพารามิเตอร์และจำเป็นต้องใช้ตัวจำแนก Naive Bayes ในสร้างแบบจำลองการรู้จำ เพื่อให้ความถูกต้องในการจำแนกมากที่สุด

## 2.3. การตรวจสอบความถูกต้อง

การตรวจสอบความถูกต้องจะเป็นการประเมินโดยใช้เทคนิคการสุ่มตัวอย่างในการกำหนดสภาพการจรรยาซึ่งจะสามารถตรวจสอบผลการวิเคราะห์จากแบบจำลองได้อย่างมีประสิทธิภาพในการนำข้อมูลไปใช้ประโยชน์ต่อไป

ผลการวิจัย

ผลการวิจัยของโครงการวิจัย ประกอบด้วย 2 ส่วนคือการรวบรวมข้อมูลและจัดทำคลังฐานข้อมูล และวิเคราะห์เหมืองข้อความ และประเมินการจรรยาจรจากข้อมูลสื่อสังคม

1. การรวบรวมข้อมูลและจัดทำคลังฐานข้อมูล

การพัฒนาโปรแกรมเพื่อทำการดึงและจัดทำคลังฐานข้อมูล twitter ปี 2560 โดยมีผลลัพธ์ตัวอย่างในการดึงข้อมูล และถูกทำการจัดเก็บเป็นฐานข้อมูล ซึ่งมีชุดข้อมูลที่ผู้ใช้งานอัปโหลดทั้งสิ้นจำนวน 20,000 ชุดข้อมูลจากการสืบค้นคำที่เกี่ยวข้องกับการจรรยาจร เช่น “JS100radio” “fm91bkkpro” “อุบัติเหตุ” “รถติด” ซึ่งในคลังฐานข้อมูลที่สืบค้นได้ประกอบไปด้วยหมายเลข ชื่อผู้ใช้งาน วันและเวลาที่อัปโหลดข้อมูล ดัชนีถ้อยคำที่อธิบายความคิดเห็น และพิกัดแสดงตำแหน่งของข้อมูลนั้นดังภาพที่ 2

CreatedAt	From.User	From.UserId	To.User	To.UserId	Language	Source	Text
2/20/19 10:37 PM	กาน-วิ-ลา @	4.834389e+09	NA	-1	th	<a href="http://twitter.com/download/iphone" rel="n...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...
2/20/19 10:32 PM	☺FN☺	4.223536e+09	NA	-1	th	<a href="http://twitter.com/download/android" rel="n...	RT @fm91traffipro: ปก. จ.ลำปาง สาขารักษาเพื่อ รักษาม...
2/20/19 10:30 PM	♥ซุซึนแะ♥	4.654320e+08	NA	-1	th	<a href="http://twitter.com/download/android" rel="n...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...
2/20/19 10:29 PM	FangZzz	1.309975e+09	NA	-1	th	<a href="http://twitter.com/download/android" rel="n...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...
2/20/19 10:29 PM	nothing left to hold	7.065149e+17	NA	-1	th	<a href="http://twitter.com/download/iphone" rel="n...	RT @fm91traffipro: วันที่ 20 ม.ค.62 เกิดเหตุชนกันบนทาง...
2/20/19 10:26 PM	Siwa_Tle	2.204207e+08	NA	-1	th	<a href="http://twitter.com/download/iphone" rel="n...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...
2/20/19 10:19 PM	invisible is nothing.	1.012002e+08	NA	-1	th	<a href="http://twitter.com/download/iphone" rel="n...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...
2/20/19 10:17 PM	สมเหนือ สหมลิณ	2.962359e+07	NA	-1	th	<a href="http://twitter.com" rel="nofollow">Twitter W...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...
2/20/19 10:17 PM	หนักใจจังมะนาว🍋	4.766332e+09	NA	-1	th	<a href="http://twitter.com/download/android" rel="n...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...
2/20/19 10:16 PM	KaDookKaDik	6.594794e+07	NA	-1	th	<a href="http://twitter.com/download/iphone" rel="n...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...
2/20/19 10:16 PM	amw13	1.080305e+08	NA	-1	th	<a href="https://mobile.twitter.com" rel="nofollow">T...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...
2/20/19 10:16 PM	Big wasabi	8.005280e+17	NA	-1	th	<a href="http://twitter.com/download/android" rel="n...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...
2/20/19 10:16 PM	Patcha	9.346317e+17	NA	-1	th	<a href="http://twitter.com/download/android" rel="n...	RT @fm91traffipro: เมื่อเวลา 21.51 น. ที่ผ่านมา เกิดเหตุ...

ภาพที่ 2 ตัวอย่างการรวบรวมข้อมูล

2. การวิเคราะห์เหมืองข้อความ

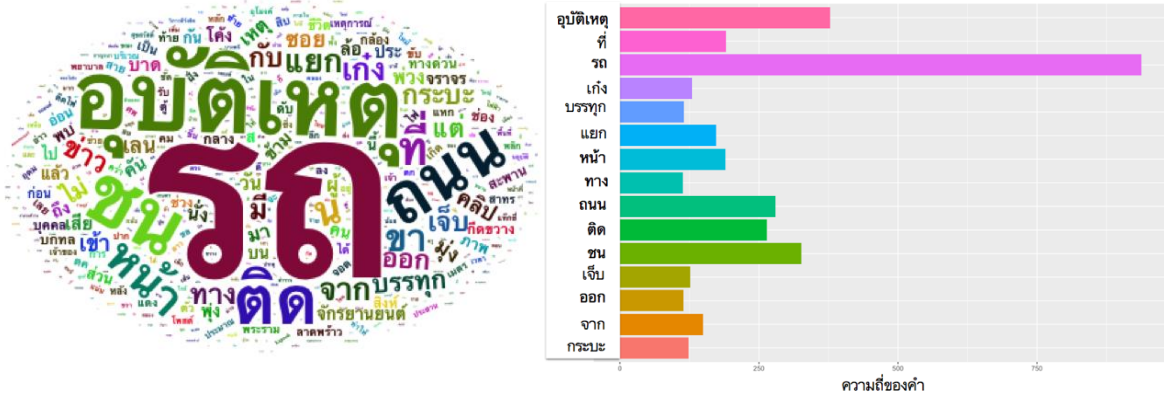
เทคนิคการประมวลผลข้อความจากดัชนีถ้อยคำ ประกอบไปด้วยขั้นตอน การเลือกข้อมูล การกรองข้อมูล และการจัดลำดับข้อมูล ให้อยู่ในรูปแบบคลังฐานข้อมูลเพื่อนำไปวิเคราะห์

- 1) การลบคำที่ไม่สำคัญหรือไม่เกี่ยวข้องกับการจรรยาจร (stop words) ได้แก่ ตัวเลข ตัวอักษรภาษาอังกฤษ สัญลักษณ์ที่ไม่ใช่ตัวอักษร คำซ้ำต่างๆ
- 2) การตัดคำออกมาเป็นคำเดี่ยวๆ (tokenization) เพื่อแยกคำตามรูปคำหรือการอ่านออกเสียง สำหรับภาษาไทย เป็นภาษาที่ซับซ้อนเนื่องจากมีรูปแบบของคำที่ต่อเนื่อง ดังภาพที่ 3

ID	Words
1	"ระยะไกล"   ทอยยเดินทางกลับ!!   ถ.มิตรภาพ   -     ถ.เอเชีย   ปริมาณรถเริ่มหนาแน่น
2	รพม.   เบียงช่องทางจราจร   2   จุด   บนถ.พชรพรม   9   และถ.รัชดาภิเษก   ตั้งแต่วันที่   110   ม.ค.   61   เป็นต้นไป
3	วังกันวัน!!   ไฟไหม้รถตู้ยูนไต   หน้า   โรงแรมรัตนโกสินทร์   ถ.ราชดำเนิน   เสียหายทั้งคัน
4	วังกันวัน!!   ไฟไหม้รถตู้ยูนไต   หน้า   โรงแรมรัตนโกสินทร์   ถ.ราชดำเนิน   เสียหายทั้งคัน
5	รถบัสชนรถบรรทุกตรงกลาง   ถ.มิตรภาพ   ช่วงกลางวัน   การจราจรติดขัดกว่า   110   กม.   ที่แยกแฉกฟาร์ม โชคชัย
6	"พบเพลิงไหม้รถบรรทุก   ไม่น่าใจมีใครรับผิดชอบอยู่ภายในหรือไม่?   จอดรถรีบลงไปดู   แล้วแจ้ง   #SOSI   ขอรถดับเพลิง!"
7	สาวออฟฟิศเป็นลม!   ขับรถต่อไม่ไหว   ต้องจอดรถริมถนนขอความช่วยเหลือ   #SOSI
8	ฝนตกหนักหน้าหนาวทั่วกทม.   น้ำท่วมซึ่งบนถนนหลายสาย   การจราจรติดขัดหนัก   คนกรุงไม่หันตั้งตัวรับมือไม่ทัน
9	รถบัสรับส่งนักเรียน   เสียหลักตกข้างทาง   เส้นทางน้ำตกเจ็ดสาวน้อย   จ.สระบุรี
10	18   ล้อชนกันรุนแรง!!   บนถ.เพชรเกษม   อ.ปากท่อ   จ.ราชบุรี
11	รถสองแถวตกเขา!!   ที่ภูทับเบิก   จ.เพชรบูรณ์   บาดเจ็บ   5   คน
12	รถบรรทุกชนท้ายรถทัวร์!!   เสียชีวิตติดภายใน   2   คน   บนถ.พหลโยธิน   ที่ตรงข้าม   ม.มหาจุฬาลงกรณราชวิทยาลัย
13	ไฟไหม้รถแก๊สหัวรถตู้ทั้งคัน   บน   ถ.ลาดพร้าว   ตรงข้ามที่ทำการ   ประชานิย์ฯ
14	รถกระบะถูกคนร้ายขับรถประเภตึงกลางถนนพหลโยธิน   ช่วงนคร   ชายคนขับเสียชีวิต
15	รถทัวร์พลิกตะแคง!!   บนทางหลวงหมายเลข226   จ.สุรินทร์   บาดเจ็บ   19   คน
16	ปิคอัพชนพระสงฆ์   มรณภาพ   1   รูป   บนถ.สายเอเชีย   ช่วงกม.   51   จ.อ่างทอง
17	รถทัวร์หนักท้องเที่ยวจีน   ตกจากเนินจอดรถเชิงทางเขีนพระธาตุโดยสุเทพ   จ.เชียงใหม่   มีผู้บาดเจ็บหลายคน

ภาพที่ 3 ตัวอย่างการตัดคำ

3) การนำข้อมูล twitter มาวิเคราะห์ความถี่ของคำที่เกิดขึ้นในรูปแบบถุงของข้อมูล (bag of words) และแสดงผลผ่านดัชนีถ้อยคำก้อนเมฆ (tagcloud) และการวิเคราะห์จำนวนความถี่ของข้อความ (frequency of words) ดังภาพที่ 4 พบว่า คำที่นิยมใช้ในการบรรยายคือ รถ อุบัติเหตุ ชน ถนน ติด กระบะ บรรทุก เจ็บ ขาเข้า ขาออก รถจักรยานยนต์ เป็นต้น จะเห็นได้ว่า คำเหล่านี้มาจากการที่ผู้ใช้งานเป็นผู้เผยแพร่ ซึ่งมีความเกี่ยวข้องกับสภาพการจราจรที่เกิดเป็นอย่างมาก



ภาพที่ 4 ดัชนีถ้อยคำก้อนเมฆ (tagcloud) และการวิเคราะห์จำนวนความถี่ของข้อความ

4) กฎของการหาความสัมพันธ์ (Association Rule Mining: ARM) ในงานวิจัยนี้ค้นหาความสัมพันธ์ระหว่างข้อความ twitter จากข้อมูลการระบุตำแหน่ง ซึ่งกำหนดให้เงื่อนไขของความสัมพันธ์คือ กลุ่มดัชนีถ้อยคำสำคัญ (K) ผลลัพธ์ที่เกิดขึ้น คือ กลุ่มดัชนีถ้อยคำต่างๆ (k) ที่เกี่ยวข้องกับดัชนีถ้อยคำสำคัญ การประเมินค่าของกฎจะใช้ค่าความเชื่อมั่น (confidence) ของข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องตามกฎของความสัมพันธ์ จากผลการประมวลผลข้อความในการจัดอันดับและวิเคราะห์ข้อความดัชนีถ้อยคำที่พบว่า กลุ่มดัชนีถ้อยคำสำคัญ (K) จะถูกแบ่งออกเป็น 2 ประเภทได้แก่ ประเภทที่เกี่ยวข้องกับสภาพจราจรและประเภทอื่นๆ ซึ่งสามารถเขียนตัวอย่างกฎความสัมพันธ์และแสดงค่าความเชื่อมั่นได้ ตารางที่ 1 แสดงตัวอย่างค่าความเชื่อมั่นจากกฎของการหาความสัมพันธ์

อุบัติเหตุ	ความเชื่อมั่น	บาดเจ็บ	ความเชื่อมั่น	รถนั่งส่วนบุคคล	ความเชื่อมั่น	ชน	ความเชื่อมั่น
คมชัดลึก	0.24	อุบัติเหตุ	0.64	ชนกับ	0.55	hotcilp	0.89
คลิปกล้องวงจรปิด	0.21	ปอดเด็กตั้ง	0.64	ถนนหทัยราษฎร์	0.48	รวมคลิป	0.8
ขอบคุณข่าว							
ภาพจากกล้องวงจรปิด	0.21	ชนกับรถแท็กซี่	0.59	รถไม่ป่วน	0.48	คลิปกล้องวงจรปิดขอบคุณข่าว	0.77
รวมคลิป	0.2	เลนซ้าย	0.59	รถชะลอตัว	0.48	ภาพจากกล้องวงจรปิด	0.77

อุบัติเหตุ	ความ เชื่อมั่น	บาดเจ็บ	ความ เชื่อมั่น	รถนั่งส่วนบุคคล	ความ เชื่อมั่น	ชน	ความ เชื่อมั่น
cctv	0.18	ประสาน	0.41	เยื้องท้ายราษฎร์ ชอยกีดขวางเลน กลาง	0.48	cctv	0.54
อย่างสดศพ	0.17	ก่อนถึงแยกทองหล่อ เหนือเมตร	0.41	กีดขวางเลนขวา	0.46	ซื้อทัวร์	0.45
hotcilp	0.17	ถนนเพชรบุรี	0.41	บก	0.39	พ่วงบรรทุกตู้ คอนเทนเนอร์	0.45
เสียชีวิตแล้ว	0.16	กีดขวางเลนขวาfm	0.41	ขาเข้า	0.34	cctv คลิป อุบัติเหตุ	0.45
คลิป	0.15	ปากซอยเพิ่มสิน อุบัติเหตุ	0.41	ถนนบรมราชชนนี	0.34	ขอบคุณ	0.45
ข่าวภูมิภาค	0.14	รถจักรยานยนต์คันชน กัน	0.41	ปากซอยบรมราช ชนนีรถกระบะ	0.34	อุบัติเหตุภาพ จากกล้องวงจร ปิด	0.45
ชน	0.14	รถจักรยานยนต์	0.33	รวมคัน	0.34	คลิป	0.4
ข่าวอาชญากรรม	0.14	ชนกับรถจักรยานยนต์	0.33	ตรงข้ามร้านเจ้เลี้ยง	0.28	ซึ่ง	0.25
หลังขับเก่งชนท้าย กระบะไฟไหม้	0.14	รถประจำทาง	0.33	ลงจากสะพานข้าม คลองบางกอกน้อย เมตร	0.28	คลิปกล้องวงจร ปิด	0.22
เศร้าอมหิตล	0.14	กีดขวางเลน	0.29	ก่อนถึงใต้ทางด่วน งามวงศ์วาน	0.28	อุบัติเหตุ	0.14

### 3. การประเมินสภาพการจราจรโดยแบบจำลองเบย์

ในการประเมินสภาพการจราจรจากตัวแปรซึ่งเป็นดัชนีถ้อยคำที่รวบรวมมาได้ เป็นตัวแปรที่ไม่ต่อเนื่อง โดยใช้คุณลักษณะการปรับมาตราส่วนความกว้างเท่ากันของข้อมูลให้เท่ากัน แบ่งช่วงของแต่ละตัวแปรเป็น เพื่อนำไปคำนวณและจัดกลุ่มประเภทสภาพการจราจรให้เป็นหมวดหมู่ตามความน่าจะเป็นของข้อมูล อย่างไรก็ตาม ในการประมาณค่าพารามิเตอร์และจำเป็นต้องใช้ตัวจำแนก Naive Bayes ในสร้างแบบจำลองการรู้จำ เพื่อให้ความถูกต้องในการจำแนกมากที่สุด โดยกำหนดประเภทการประเมินทั้งหมด 2 ประเภทคือ ประเภทที่เกี่ยวข้องกับการจราจรและประเภทอื่นๆ ผลการวิจัยพบว่า การประเมินสภาพการจราจรจากข้อมูลสื่อสังคมออนไลน์ทวีตเตอร์มีความถูกต้องสูงมากคิดเป็น ร้อยละ 97.50 เนื่องจากมีการเรียนรู้จากความน่าจะเป็นของคำที่พบบ่อยในข้อความซึ่งมีความบ่งชี้ชัดเจนเกี่ยวกับเหตุการณ์นั้นๆ ดังภาพที่ 5



	true Traffic	true Normal	class precision
pred. Traffic	31	1	96.88%
pred. Normal	0	10	100.00%
class recall	100.00%	90.91%	

ภาพที่ 5 ความถูกต้องจากการประเมินสภาพจราจรจากแบบจำลอง Bayes

**สรุปและอภิปรายผล**

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์เหมืองดัชนีข้อมูลเชิงตำแหน่งผ่านสื่อสังคมออนไลน์กรณีศึกษาแพลตฟอร์ม twitter โดยแบ่งการวิเคราะห์ออกเป็น 2 ขั้นตอนหลักได้แก่ การพัฒนาคลังฐานข้อมูลและการวิเคราะห์เหมืองดัชนีถ้อยคำเพื่อหาความสัมพันธ์ของข้อมูลจราจร เทคนิคการวิเคราะห์เหมืองดัชนีถ้อยคำประกอบไปด้วย 1) ประมวลผลข้อความจากคลังฐานข้อความ ได้แก่ การเลือกคุณลักษณะของข้อมูล การกรองข้อมูลและการจัดลำดับสำคัญของข้อมูล เพื่อนำมา 2) วิเคราะห์เหมืองดัชนีถ้อยคำ โดย กฎของการหาความสัมพันธ์ (Association Rule Mining) และการประเมินสภาพจราจรจากแบบจำลองความน่าจะเป็น Bayes ผลการวิจัยพบว่าในคลังฐานข้อมูลที่สืบค้นได้ประกอบไปด้วย หมายเลข ชื่อผู้ใช้งาน วัน และเวลาที่อัปโหลดข้อมูล ดัชนีถ้อยคำที่อธิบายความคิดเห็น และพิกัดแสดงตำแหน่งของข้อมูลซึ่งลักษณะข้อความที่มีการบรรยายร่วมกับการระบุตัวตนในแต่ละสถานที่ จากผลการประมวลผลข้อความในการจัดอันดับและวิเคราะห์ข้อความดัชนีถ้อยคำพบว่า กลุ่มดัชนีถ้อยคำสำคัญ (K) จะถูกแบ่งออกเป็น 2 ประเภทได้แก่ ประเภทที่เกี่ยวข้องกับสภาพจราจรและประเภทอื่นๆ ซึ่งคำที่นิยมในแสดงความคิดเห็นคือ รถ อุบัติเหตุ ชน ถนน ดิน กระจบะ บรรทุก เจ็บ ขาเข้า ขาออก รถจักรยานยนต์ เป็นต้น จะเห็นได้ว่าคำเหล่านี้มาจากการที่ผู้ใช้งานเป็นผู้เผยแพร่ ซึ่งมีความเกี่ยวข้องกับสภาพการจราจร ซึ่งเมื่อพิจารณาร่วมกับผลการจัดกลุ่มข้อมูล ค่าความเชื่อมั่นจากกฎของความสัมพันธ์จากดัชนีถ้อยคำสำคัญ และแผนที่ความหนาแน่นของข้อมูลระบุตำแหน่ง (heatmap) เชิงพื้นที่ มีความสอดคล้องกับสภาพการจราจร ในส่วนของการประเมินสภาพการจราจรจากข้อมูลสื่อสังคมออนไลน์ทวีตเตอร์มีความถูกต้องสูงมากคิดเป็น ร้อยละ 97.50 เนื่องจากมีการเรียนรู้จากความน่าจะเป็นของคำที่พบบ่อยในข้อความซึ่งมีความบ่งชี้ชัดเจนเกี่ยวกับสภาพการจราจรเช่นกัน จะเห็นได้ว่าพฤติกรรมหรือลักษณะการใช้งาน twitter มีความสอดคล้องกันอย่างเห็นได้ชัด ทั้งนี้ยังสะท้อนให้เห็นถึงข้อความที่สื่อถึงเหตุการณ์ที่เกิดขึ้น ณ สถานที่นั้นๆ อีกด้วย เช่น รถติด อุบัติเหตุ ไฟไหม้ สถานที่เกิดเหตุ และความคิดเห็นต่อเหตุการณ์

**ข้อเสนอแนะ**

ข้อเสนอแนะจากผลการวิจัยครั้งนี้ พบว่าการวิเคราะห์เหมืองข้อความจากข้อมูลสื่อสังคมออนไลน์ สามารถนำมาวิเคราะห์ร่วมกับข้อมูลตัวอักษร รูปภาพ ที่มีการแบ่งปันผ่านสื่อสังคมออนไลน์ ที่มีการระบุตำแหน่งอื่น ๆ ได้แก่ flickr instagram foursquare Panoramio หรือ openstreet view และการใช้เทคนิควิธีการวิเคราะห์อื่น ๆ เช่นการรู้จำและจำแนกประเภทของข้อมูลขั้นสูงจากปัญญาประดิษฐ์ (Artificial Intelligent) ทำให้สามารถสะท้อนพฤติกรรมของการใช้งานและช่วยความคิดเห็น ซึ่งจะช่วยส่งเสริมในการวางแผน พัฒนา ทั้งในด้านธุรกิจ การท่องเที่ยว การคมนาคม และโครงสร้างพื้นฐานต่างๆ รวมไปถึงการประยุกต์ในด้านการรายงานเหตุการณ์ การเตือนภัยและการเฝ้าระวัง เช่นสภาพการจราจร การเกิดอุบัติเหตุ หรือภัยพิบัติต่างๆ ในเชิงพื้นที่ให้ดียิ่งขึ้น การใช้เทคโนโลยีสื่อสังคมสำหรับภูมิสารสนเทศศาสตร์ เป็นเทคโนโลยีที่น่าสนใจอีกทางเลือกหนึ่ง เพื่อให้ทราบถึงลักษณะการรับรู้ทางสังคม จากการได้มาซึ่งข้อมูลสาธารณะซึ่งจะเป็นประโยชน์สำหรับการใช้งานและวางแผนเชิงพื้นที่ในอนาคตได้อย่างเหมาะสมและมีประสิทธิภาพต่อไป

## กิตติกรรมประกาศ

บทความวิจัยฉบับนี้ได้รับทุนสนับสนุนการวิจัยจากคณะสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประจำปีงบประมาณ 2561

## เอกสารอ้างอิง

- ชนเดช บุญวราพงศ์, สุภกิจ นุตยะกุล (2558). **การใช้สื่อสังคมออนไลน์เพื่อคำนวณสภาพการจราจรในเมืองใหญ่**. คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังฯ กรุงเทพฯ, Vol 4, No 2.
- พงศ์กร จันทราษ. (2550). **การจัดการฐานข้อมูล**. เอกสารประกอบการสอน รายวิชาการระบบฐานข้อมูล, ภาควิชาเทคโนโลยีสารสนเทศ : มหาวิทยาลัยฟาร์อีสเทอร์น.
- พิชิต วิจิตรบุญยรักษ์. (2553). **สื่อสังคมออนไลน์: สื่อแห่งอนาคต**. ค้นเมื่อ 23 กันยายน 2559, จาก <http://www.bu.ac.th>.
- สำนักงานราชบัณฑิตยสภา. (2556). **สื่อสังคม**. ค้นเมื่อ 23 กันยายน 2559, จาก <http://www.royin.go.th>.
- รังสรรค์ สุขชัยรังสรรค์ (2553) **การวิเคราะห์ชั้นข้อมูลพร้อมจัดทำแผนที่ข้อมูล GIS (สารสนเทศภูมิศาสตร์) และแสดงบริเวณอันตรายตามช่วงพิจารณา ของทางหลวงหมายเลข 1 ตอน (เชียงใหม่ – แม่จัน)**.
- ภัทรสุดา วิชยพงศ์ (2554) **การพัฒนานาฬิกาอุบัติเหตุโดยการประยุกต์ใช้ระบบสารสนเทศทางภูมิศาสตร์ในการระบุจุดเสี่ยงอันตรายบนถนน : กรณีศึกษาจังหวัดนครราชสีมา**. วิทยานิพนธ์, สาขาวิศวกรรมขนส่ง สำนักวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี.
- Wanichayapong, N., W. Pruthipunyaskul, W. Pattara-Atikom et al. (2011). **Social-based traffic information extraction and classification**. Paper presented at the ITS Telecommunications (ITST), 2011 11th International Conference On, pp.107-112.
- Sitthi, Asamaporn; Nagai, Masahiko; Dailey, Matthew; Ninsawat, Sarawut. (2016). **Exploring Land Use and Land Cover of Geotagged Social-Sensing Images Using Naive Bayes Classifier**. Sustainability 8, no. 9: 921.