แนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data
กรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุ

ภวัญญา รักษาศิลป์

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการจัดการมหาบัณฑิต
วิทยาลัยการจัดการ มหาวิทยาลัยมหิดล
พ.ศ.2560

สารนิพนธ์

เรื่อง

แนวทางการวิเคราะห์ข้อมูลทางธุรกิจ

โดยใช้ Big Data : กรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุ

ได้รับการพิจารณาให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาการจัดการมหาบัณฑิต

วันที่ 23 พฤษภาคม พ.ศ. 2560

.....

นางสาวภวัญญา รักษาศิลป์

ผู้วิจัย

.....

ราชา มหากันธา

Ph.D.

อาจารย์ที่ปรึกษาสารนิพนธ์

.....

ผู้ช่วยศาสตราจารย์วินัย วงศ์สุรวัฒน์

Ph.D.

ประธานกรรมการสอบสารนิพนธ์

.....

ดวงพร อาภาศิลป์

Ph.D.

คณบดี

วิทยาลัยการจัดการ มหาวิทยาลัยมหิดล

.....

บุญยิ่ง คงอาชาภัทร

Ph.D.

กรรมการสอบสารนิพนธ์

กิตติกรรมประกาศ

สารนิพนธ์เรื่องแนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data : กรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุฉบับนี้ สามารถลุล่วงไปได้ด้วยดี ทั้งนี้ต้องขอขอบคุณ ดร.ราชา มหากันธา อาจารย์ที่ปรึกษางานสารนิพนธ์เป็นอย่างสูง ด้วยความกรุณาและความช่วยเหลืออย่างดีในการให้คำแนะนำต่างๆ ตลอดจนการตรวจสอบแก้ไขเนื้อหาในงานวิจัย

ผู้วิจัยขอขอบคุณผู้มีความรู้และประสบการณ์ทุกท่านไม่ว่าจะเป็นโปรแกรมเมอร์ที่ทำงานในประเทศไทยและโปรแกรมเมอร์ที่ทำงานในต่างประเทศ ที่เสียสละเวลาอันมีค่าในให้คำแนะนำ แก่ผู้วิจัยและขอขอบพระคุณอาจารย์ทุกท่านทั้งอาจารย์ใน CMMU และอาจารย์ที่สอนคอร์สสัมมนา Big Data สำหรับวิชาความรู้ต่างๆ ที่ให้แก่ผู้วิจัย ซึ่งเป็นประโยชน์อย่างมากในการทำงานวิจัย สุดท้ายผู้วิจัยหวังเป็นอย่างยิ่งว่าสารนิพนธ์ฉบับนี้จะเป็นประโยชน์และสามารถเป็นแนวทางธุรกิจหรือเป็นแนวทางแก่ผู้ที่มีความสนใจและทำการศึกษาเรื่องนี้เพิ่มเติมต่อไปในอนาคต ทั้งนี้หากมีข้อผิดพลาดประการใดผู้วิจัยขอรับไว้และขออภัยมา ณ ที่นี้

ภวัญญา รักษาศิลป์

สารบัญ

สารบัญ (ต่อ)

สารบัญตาราง

สารบัญภาพ

แนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data : กรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุ

USING BIG DATA FOR BUSINESS DATA ANALYSIS : THE CASE STUDY OF THE TWITTER AND ACCIDENT

ภวัญญา รักษาศิลป์ 5850389

กจ.ม.

คณะกรรมการที่ปรึกษาสารนิพนธ์ : ดร.ราชา มหากันธา, Ph.D., ดร.บุญยิ่ง คงอาชาภัทร, Ph.D., ผู้ช่วยศาสตราจารย์วินัย วงศ์สุรวัฒน์, Ph.D.

บทคัดย่อ

สารนิพนธ์ฉบับนี้จัดทำขึ้นเพื่อศึกษาประโยชน์และเครื่องมือของการใช้ Big Data เพื่อนำมาประยุกต์ใช้ ซึ่งผ่านการเก็บข้อมูลการทวีตอุบัติเหตุในเขตกรุงเทพมหานครจากทวิตเตอร์ จำนวน 16,350 ตัวอย่าง โดยนำมาประเมินโครงสร้างข้อมูล ตัวแปร การสกัดข้อมูลและแสดงผลการวิเคราะห์ในมิติต่างๆ ซึ่งจากผลการวิจัยการเลือกใช้โปรแกรมในการวิเคราะห์ Big Data นั้น จะใช้ Program R กับ API Twitter ในการดึงข้อมูล โดยสกัดข้อมูลร่วมกับ Sparklyr, SQL และ Package Rshiny ในการแสดงผลของข้อมูลการวิเคราะห์Big Data เช่น มิติของช่วงเวลาที่ทวีต, เขต, เพศของคนขับ, การทวีตสาเหตุ, ประเภทของรถที่เกิดอุบัติเหตุและความเสียหายที่เกิดขึ้น ทั้งนี้ข้อมูลที่ดึงได้มีความสอดคล้องกับรายงานสำนักงานตำรวจแห่งชาติ ในมิติของประเภทรถ,สาเหตุและระดับความเสียหายที่เกิดอุบัติเหตุ แต่ในส่วนของสาเหตุฯและระดับความเสียหายนั้น ข้อมูลของสำนักงานตำรวจฯมีรายละเอียดมากกว่า ส่วนข้อแตกต่างของข้อมูลจากการทำ Big Data นั้นมีความชัดเจนในเรื่องการทวีตรายงานอุบัติเหตุแยกย่อยตามเขตและเวลาที่เกิดเหตุและมิติความเสียหายที่เกิดขึ้นนั้นการวิเคราะห์ข้อมูล Big Data มีความละเอียดน้อยกว่าข้อมูลสำนักงานตำรวจฯ สำหรับข้อแนะนำหน่วยงานในการนำงานวิจัยไปใช้ประโยชน์ โดยสามารถนำงานวิจัยไปใช้เป็นพื้นฐานดูภาพรวมเครื่องมือหลักในการประมวลผลด้วย Big Data และเป็นแนวทางในการวิเคราะห์ข้อมูลที่ประหยัด รวดเร็วและวางแผนการเก็บข้อมูลในมิติต่างๆให้เพียงพอต่อการวิเคราะห์ด้วยการเปรียบจุดเด่นของข้อมูลในแต่ละ Platform, โครงสร้างของข้อมูล, โปรแกรมที่ใช้ในการดึงข้อมูล, การตรวจสอบความถูกต้องและการเปรียบเทียบมิติข้อมูลกับหน่วยงานรัฐ นอกจากนี้สามารถนำไปประยุกต์ใช้ประโยชน์ในรูปแบบอื่นๆ เช่น การวิเคราะห์ข้อมูลเพื่อสร้างแคมเปญทางการตลาดและการนำเสนอผลการวิเคราะห์ได้อย่างรวดเร็ว เป็นต้น

คำสำคัญ : แนวทางการวิเคราะห์Big Data/ ข้อมูลทวิตเตอร์/ อุบัติเหตุ

92 หน้า

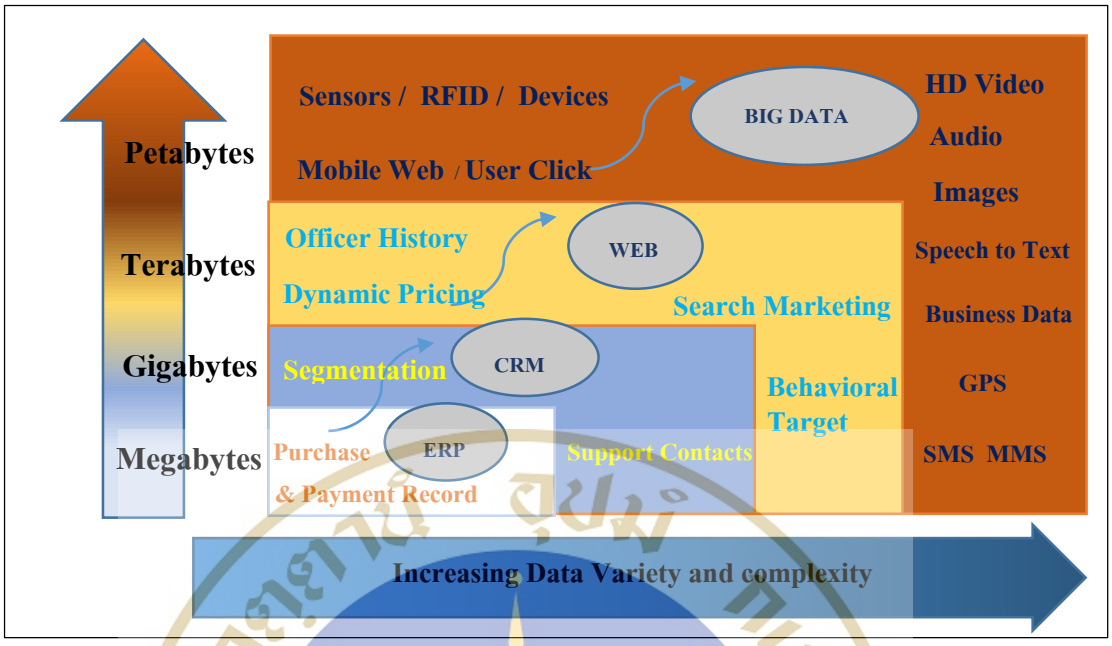
บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในยุคที่เทคโนโลยีและระบบต่างๆมีการพัฒนาก้าวหน้ามากอย่างรวดเร็ว ทำให้องค์กรมีการเก็บข้อมูลต่างๆอยู่มากมายมหาศาล โดยเฉพาะในต่างประเทศนั้นได้มีการนำข้อมูลมาใช้ประโยชน์ในระยะหนึ่งแล้ว สำหรับประเทศไทยเริ่มมีการนำข้อมูลเหล่านี้มาใช้ในการตัดสินใจเพื่อเกิดประโยชน์กับองค์กร (พสุ เดชะรินทร์, 2556) แต่ไม่มากนัก ซึ่งข้อมูลมากมายมหาศาล จะคุ้นเคยในชื่อของ Big Data

Big Data คือการที่มีข้อมูลปริมาณมาก ในระดับ เทราไบต์ (Tera Byte : TB) หรือระดับเพตะไบต์ (Peta Byte : PB) จากข้อมูลของ NECTEC Researcher Talk ครั้งที่ 2 ประจำปี 2558 กล่าวว่าคุณสมบัติของข้อมูลขนาดใหญ่ ควรต้องมีรายละเอียดดังนี้

1. ข้อมูลมีการเก็บไว้ในปริมาณ (Volumn) มากและต่อเนื่อง เช่น ข้อมูลมหาศาลจาก Facebook, Line, Twitter, Web Pages เป็นต้น
2. ข้อมูลที่มีความแตกต่าง หลากหลาย (Variety) เช่น ข้อมูลในฐานข้อมูล ข้อมูลภาพ ข้อมูลเสียง เป็นต้น
3. ข้อมูลมีการเกิดขึ้นอย่างรวดเร็ว (Velocity)
4. ข้อมูลสามารถสร้างประโยชน์ให้กับองค์กรและ มีคุณค่า(Value)กับส่วนที่เกี่ยวข้อง
5. ข้อมูลมีความน่าเชื่อถือ(Veracity)



ภาพที่ 1.1 เปรียบเทียบจำนวนข้อมูลที่มากขึ้นจนถึงระดับ Big Data

ที่มา : http://www.cyberthai.com/index.php/knowledge-center/97-what-big-data

จากรูปภาพที่ 1.1 คือภาพรวมในการเปรียบเทียบจำนวนข้อมูล แนวดิ่ง คือระดับของจำนวนข้อมูลที่เพิ่มขึ้น โดยเพิ่มจากระดับ เมกะไบต์ (Megabytes) จนถึง เพตะไบต์ (Petabytes) เรียงจากน้อยไปมาก และแนวราบคือ ความหลากหลาย และความซับซ้อนในรูปแบบของการเก็บข้อมูลลงระบบทั่วไป จนถึง ระดับของข้อมูลที่มีความหลากหลายและความซับซ้อนสูง เพิ่มขึ้นจากซ้ายไปขวา ซึ่งจะเห็นปริมาณการเพิ่มขึ้นของข้อมูลได้อย่างชัดเจน เช่นในระดับที่ข้อมูลมีเพียงเมกะไบต์ (Megabytes) และ จิกกะไบต์ (Gigabytes) ข้อมูลจะอยู่ในลักษณะของระบบ อีอาร์พี (ERP) คือ เป็นการวางแผนบริหารธุรกิจขององค์กร เช่น เก็บข้อมูลรายละเอียดการซื้อ การจ่ายเงิน วางแผนงานขาย งานผลิต งานทรัพยากรมนุษย์ งานบัญชี หรือ ระบบที่นำไว้สร้างความสัมพันธ์กับลูกค้า (CRM) เพื่อให้ลูกค้ามีความผูกพันกับสินค้า บริการ องค์กร เช่น ข้อมูลการติดต่อของลูกค้า วันเกิดลูกค้า เพื่อที่จะไปเสนอความพิเศษในวันพิเศษ เป็นต้น แต่หากข้อมูลนั้นเป็นระดับเทราไบต์ (Terabytes) หรือ เพตะไบต์ (Petabytes) นั้น ข้อมูลจะเป็นลักษณะของการทำเว็บไซต์ เพื่อค้นหาพฤติกรรมผู้บริโภค รวมถึงการทำโมบาย เว็บ (Mobile Web) ที่สามารถเข้าถึงผู้บริโภคและทำให้ข้อมูลนั้นเพิ่มขึ้นมากในระยะเวลาอันรวดเร็วจนถึงระดับ Big Data (Cyberthai, n.d.)

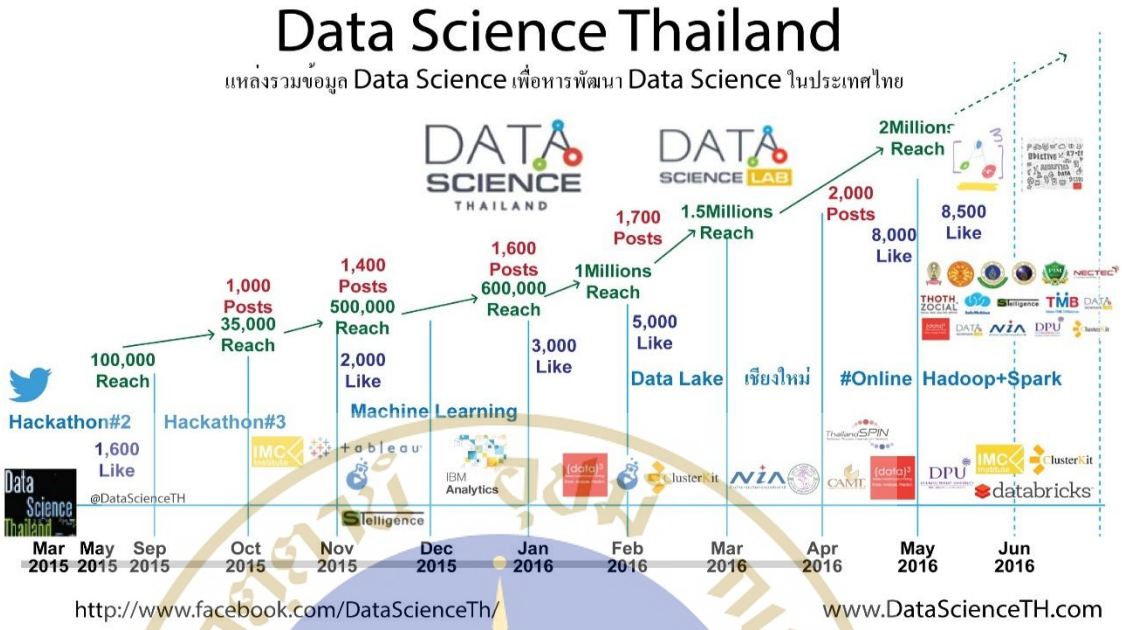
ข้อมูลต่างๆที่รวมเป็นข้อมูลขนาดใหญ่นั้น นอกจากที่องค์กรเกือบทุกแห่งเก็บเป็นปกติอยู่แล้วโดยปัจจุบันไม่ว่าจะเป็นตัวเลขทางด้านการเงิน ตัวเลขทางด้านการดำเนินงาน ข้อมูลเกี่ยวกับลูกค้า ข้อมูลเกี่ยวกับพนักงาน หรือ ข้อมูลในระบบ ERP ระบบฐานข้อมูล ระบบ Warehouse ฯลฯ

อีกทั้งไม่ว่าใครก็สามารถเป็นผู้ผลิตข้อมูลได้ เช่น การใช้โทรศัพท์มือถือ แท็บเล็ต คอมพิวเตอร์ตั้งโต๊ะ หรือโน๊ตบุ๊ค ในการสร้างข้อมูลโดยผ่าน Social Network เช่น เฟสบุค(Facebook) ทวิสเตอร์ (Twitter) อินสตาแกรม (Instragram) ยูทูป (Youtube) ดรอปบ๊อก (Dropbox) ฯลฯ ล้วนแล้วแต่เป็นการสร้างหรือเพิ่มปริมาณข้อมูลทั้งสิ้น เพียงแค่การเช็คอิน (check-in) ว่าอยู่ที่แห่งไหน ไลค์ (Like) โพสต์สเตตัส (Status-Posted) หรือทำกิจกรรมอะไรบางอย่างก็เป็นข้อมูลสำคัญที่ถ้าองค์กรรู้จักใช้ให้เป็นประโยชน์ก็จะช่วยองค์กรได้อย่างมาก

ไอที 24 ชั่วโมง (2559) ได้กล่าวว่า จากข้อมูลตลอดปี พ.ศ.2558 มีการโพสต์ข้อความสาธารณะทาง Social Media สูงถึง 2.6 ล้านข้อความ โดยคิดเฉลี่ยประมาณ 7 ล้านข้อความต่อวัน หรือ 82 ข้อความถูกแชร์และเกิดขึ้นใหม่ทุกๆวินาที โดยสถิติประชากรไทยที่ใช้ Social Media ใช้ Facebook 41 ล้านคน ซึ่งโตขึ้นจากปีก่อน 17% คิดเป็นประมาณ 60% ของประชากรไทย ส่วนการใช้ Facebook Page สูงถึง 7 แสนเพจ นอกจากนี้มีการใช้ทวิทเตอร์ (Twitter) 5.3 ล้านราย โดยที่มีการเติบโตขึ้นประมาณ 18% ส่วนคนไทยที่ใช้อินสตราแกรม (Instragram) มี 7.8 ล้านคน เติบโต 74% ใช้ไลน์ (Line) สูงถึง 33 ล้านคน ทั้งหมดนี้เป็นข้อมูลขนาดใหญ่อย่างเห็นได้ชัด

ประเทศไทยในยุคที่ทุกคนพกโทรศัพท์มือถือ ไอพอด (iPod) ไอแพด (iPad) มีคอมพิวเตอร์ที่สะดวกต่อการพกพา และมีการสื่อสารผ่านอุปกรณ์เหล่านี้มากขึ้นเช่นกัน ไม่ว่าจะเป็นอีเมล์ (Email) รวมถึงเครือข่ายทางสังคมอื่นๆ โดยเฉพาะFacebook และ Line เห็นได้เกือบทุกองค์กร ซึ่งเป็นโซเชียลแอพพลิเคชันยอดนิยมของไทย โดยการนำเครื่องมือเหล่านี้มาใช้เป็นส่วนหนึ่งของการทำการตลาดของบริษัทในยุคนี้ (TNT, 2558)

จากข้อมูลของ ไอที 24 ชั่วโมง และ Trends and technology สรุปได้ว่าข้อมูลของลูกค้าที่เพิ่มขึ้นอย่างรวดเร็วทำให้องค์กรต่างๆมีความจำเป็น และต้องการเทคโนโลยีสำหรับการจัดการข้อมูลมหาศาลเหล่านี้ เนื่องจากการประมวลผลข้อมูลที่รวดเร็วเพื่อการตัดสินใจสำหรับผู้บริหารในองค์กรเป็นสิ่งสำคัญที่สามารถนำมาซึ่งผลกำไรให้กับบริษัทได้ ทำให้หลายบริษัทเริ่มตื่นตัวกับการเตรียมรับมือกับ Big data โดยเฉพาะอย่างยิ่งบริษัทใหญ่ที่มีฐานข้อมูลลูกค้ามากเป็นทุนเดิมอยู่แล้ว แหล่งข้อมูลที่หลายๆบริษัทสนใจ อย่างเช่น Data Science Thailand ที่มีการรวบรวมข้อมูลแหล่งเรียนรู้เกี่ยวกับการวิเคราะห์ข้อมูลขนาดใหญ่พบว่า มีการเติบโตพุ่งขึ้นอย่างเห็นได้ชัด ดังภาพที่ 1.2



ภาพที่ 1.2 แหล่งรวมข้อมูล Data Science เพื่อการพัฒนา Data Science ในประเทศไทย

ที่มา : https://www.facebook.com/DataScienceTh

จากรูปภาพที่ 1.2 Data Scienceประเทศไทยได้มีการรวบรวมข้อมูลต่างๆ ตั้งแต่เดือนมีนาคม พ.ศ.2558 ถึงเดือนมิถุนายน พ.ศ.2559 นั้น พบว่าการเติบโตของคนที่กดถูกใจ Data Science Thailand จาก 1,600 Like เป็น 8,500 Like ผ่านช่องทาง Facebook คิดเป็นอัตราการเติบโต 431.25% ด้วยกิจกรรมการรวบรวม และเผยแพร่ความรู้ผ่านช่องทางต่างๆ ตั้งแต่พฤษภาคม ถึง มิถุนายน พ.ศ. 2559 มีการร่วมมือกับทางมหาวิทยาลัยชั้นนำของประเทศไทยมากขึ้น เช่น จุฬาลงกรณ์ มหาวิทยาลัย มหาวิทยาลัยธรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ มหาวิทยาลัยมหิดล และมหาวิทยาลัยอื่นๆ ทำให้จำนวนการเข้าถึงเพิ่มขึ้นอย่างรวดเร็ว ซึ่งอาจจะสรุปได้ว่ามีกลุ่มคนจำนวนมากที่ให้ความสนใจเกี่ยวกับข้อมูลBig Data (DataScienceTh, 2559)

การมีข้อมูลBig Data ที่ถูกเก็บโดย Social Media Online ทำให้ทราบถึงพฤติกรรมผู้บริโภคอย่างแท้จริงด้วยความรวดเร็ว และการนำข้อมูลมาวิเคราะห์ได้อย่างรวดเร็วนั้นเห็นความแตกต่างได้อย่างชัดเจนกับการเก็บข้อมูลในอดีตที่มีการสรุปผลข้อมูลจากประชากรทั้งหมด หรือทางสถิติเรียกว่าการสำมะโนประชากรซึ่งใช้เวลานานและค่าใช้จ่ายในการเก็บรวบรวมข้อมูลสูง (Pioneer, n.d.) ทำให้นักสถิติใช้เทคนิคการสุ่มตัวอย่างเพื่อเป็นตัวแทนที่ดีของประชากรทั้งหมด แต่อย่างไรก็ตามอาจจะมีหลายปัจจัยที่การสอบถามกลุ่มตัวอย่างไม่ทราบถึงพฤติกรรมผู้บริโภคโดยแท้จริง แต่ด้วยเทคโนโลยีปัจจุบันและข้อมูลที่มีอย่างมหาศาลนี้ทำให้การวิเคราะห์ข้อมูลทั้งหมดเป็นไปได้ และสามารถค้นหารายละเอียดที่สำคัญบางประการที่ไม่สามารถบอกได้จากกลุ่มตัวอย่าง

โดยเฉพาะอย่างยิ่งการค้นหาความสัมพันธ์ของข้อมูลที่จะทำให้องค์กรสามารถนำผลการวิเคราะห์ที่ได้ไปใช้ในการตลาด และการบริหาร เช่น การที่Google สามารถหาต้นตอการแพร่ระบาดของไข้หวัดสายพันธุ์ใหม่ที่เกิดการแพร่ระบาดอย่างรวดเร็วในอเมริกา โดย Google สามารถพยากรณ์การแพร่กระจายของไข้หวัดจากข้อมูลการค้นหาใน Google และสามารถระบุพื้นที่ที่เป็นต้นกำเนิดของการแพร่ระบาดได้(Big Data, 2559) โดยวรรณเพ็ญ บุญเพ็ญ กล่าวในเว็บไซต์TCDCว่า กูเกิลจะบันทึกที่อยู่ของผู้ค้นหาอาการและการรักษาโรคไข้หวัดใหญ่แบบเรียลไทม์จนได้เป็นข้อมูลที่สามารถคาดการณ์พื้นที่ที่อาจเกิดการแพร่ระบาดได้อย่างรวดเร็ว จะเห็นได้ว่ามีการนำข้อมูลเกี่ยวกับพฤติกรรมการใช้บริการ การค้นหาไปใช้เป็นประโยชน์ อย่างไรก็ตามยังมีอีกหลายผู้ใช้งาน และหลายผู้ประกอบการที่มักสนใจแต่การนำเทคโนโลยีไปใช้แต่ขาดศักยภาพในการใช้งานและวิเคราะห์ข้อมูล ซึ่งจากข้อมูล Big Data (2559) สามารถกล่าวสรุปได้ว่า Big Data สามารถนำมาให้ประโยชน์ได้อย่างหลากหลาย ที่เราอาจคาดคิดไม่ถึง จากกรณีดังกล่าวจะเห็นได้อย่างชัดเจนว่า แม้ Google จะไม่ได้อยู่ในธุรกิจที่มีความเกี่ยวข้องกับองค์การอนามัยโลก หรือเกี่ยวข้องกับหน่วยงานการยับยั้งการระบาดของไข้หวัดสายพันธุ์ใหม่ แต่เนื่องจากGoogle มีข้อมูลปริมาณมากที่ไหลเข้าสู่ฐานข้อมูลของบริษัทอยู่ตลอดเวลาอย่างต่อเนื่อง จึงสามารถช่วยหาต้นตอและยับยั้งการแพร่ระบาดของไข้หวัดสายพันธุ์ใหม่ได้

นอกจากนี้มีอีกธุรกิจหนึ่งที่น่าสนใจ คือธุรกิจประกันภัย ในส่วนของประกันอุบัติเหตุที่มีฐานข้อมูลของลูกค้าเป็นจำนวนมาก แต่ฐานข้อมูลนั้นไม่สามารถวิเคราะห์ได้ถึงพฤติกรรมผู้บริโภคโดยแท้จริง หลายครั้งที่บริษัทประกันภัยออกผลิตภัณฑ์ไม่ตอบโจทย์ผู้บริโภค เนื่องจากการออกผลิตภัณฑ์ในแต่ละครั้ง บริษัทฯ จะออกแบบตามความต้องการของตัวแทนที่มาบอกกล่าวอีกครั้งหนึ่ง หรือหากเป็นการสำรวจตลาดมักใช้เวลานานต่อการออก 1 ผลิตภัณฑ์ รวมถึงข้อมูลต่างๆที่มีการเก็บจากลูกค้าในระบบที่ถูกสร้างขึ้นมานานแล้ว ข้อมูลลูกค้าในระบบที่มีจำนวนมหาศาลหรือ Big Data นั้นไม่สอดคล้องกับพฤติกรรมของผู้บริโภคในปัจจุบัน ซึ่งการแก้ไขระบบแต่ละครั้งก็ไม่ใช่เรื่องง่าย ต้องใช้ต้นทุนทั้งเรื่องเงิน คน เวลา หรือทรัพยากรเป็นจำนวนมาก อีกทั้ง Softwareที่ใช้ส่วนใหญ่ล้วนมีราคาสูง แต่หากมีการเก็บและวิเคราะห์ข้อมูล Big Data อย่างมีประสิทธิภาพมากขึ้น บริษัทฯ จะสามารถเข้าใจผู้บริโภคมากขึ้น เพื่อสามารถออกผลิตภัณฑ์ที่ตอบโจทย์และช่วยเพิ่มรายได้ของบริษัทได้ยิ่งขึ้น ซึ่งมีบางบริษัทเริ่มมีการเก็บข้อมูลเพื่อต้องการรู้พฤติกรรมของลูกค้าในเชิงลึกมากขึ้น เช่น บริษัทประกันไทยวิวัฒน์ประกันภัยมีการเปิดให้ใช้บริการแอพพลิเคชัน ประกันรถเติมเงินไทยวิวัฒน์เปิดปิดได้ โดยที่ลูกค้าสามารถเปิดใช้และปิดการใช้เมื่อจอดรถตามความต้องการของตนเอง (DebuggingSoft, 2559) ในภาคธุรกิจอื่นก็สามารถหาประโยชน์ทางธุรกิจจาก Big Data ได้เช่นกันหากสามารถรวบรวมข้อมูลได้อย่างถูกต้องและวิเคราะห์ได้อย่างมีประสิทธิภาพ

(Infomobius, 2558) แต่การที่จะนำข้อมูลมหาศาลมาใช้วิเคราะห์เพื่อประโยชน์ทางธุรกิจได้นั้น ทางบริษัท หรือผู้ที่เกี่ยวข้องควรจะมีความรู้ความเข้าใจกับเกี่ยวกับข้อมูล Big Data แต่เนื่องจากความเข้าใจใน Big Data สำหรับคนไทยนับว่ายังไม่เป็นที่นิยม ในกรณีสำหรับผู้ที่สนใจใน Big Data เพื่อการศึกษาก็จะพบกับอุปสรรคกับข้อจำกัดของข้อมูล ไม่สามารถนำมาศึกษาได้เพราะประเทศไทยยังไม่มีข้อมูล Big Data แบบสาธารณะ (Open-Source) ที่เป็นข้อมูลเปิดให้ผู้ที่สนใจสามารถเข้าถึงได้มากนัก (อานนท์, 2559) อีกทั้งสำนักงานรัฐบาลอิเล็กทรอนิกส์(องค์การมหาชน) กล่าวว่าการจัดอันดับประเทศที่มีข้อมูลเปิดภาครัฐในปี พ.ศ. 2558 จาก The Global Open Data Index นั้น ประเทศไทยได้รับการจัดอันดับที่ 42 โดยมีการเปิดเผยข้อมูลเพียงแค่ 36% ดังนั้นปัญหาข้างต้นนี้ทำให้นักวิเคราะห์ไม่สามารถศึกษาเพื่อพัฒนาศักยภาพธุรกิจได้อย่างรวดเร็วและแม่นยำ

เทคโนโลยีที่นำมาใช้วิเคราะห์ข้อมูล Big Data มีจำนวนมาก ทำให้เกิดอุปสรรคกับผู้ที่ทำการศึกษา แต่อย่างไรก็ตามผู้ที่ต้องการนำมาใช้ควรจะเริ่มศึกษาเทคโนโลยีที่ถนัด ใกล้ตัว คุ้นเคย มีความยืดหยุ่นในเบื้องต้นให้เกิดความเชี่ยวชาญก่อนแล้วจึงนำไปต่อยอดพัฒนาทักษะอื่นๆจากเทคโนโลยีอื่นๆเพิ่มเติม และเพราะแต่ละเทคโนโลยีก็มีข้อจำกัด (Fusion idea, 2016)

จากความเป็นมาและปัญหาที่กล่าวมาข้างต้น จึงทำให้ผู้วิจัยมีความสนใจศึกษาการเริ่มต้นนำข้อมูล Big data มาประยุกต์ใช้กับธุรกิจเพื่อนำไปวิเคราะห์เพิ่มยอดขาย ขยายกิจกรรมทางการตลาดและการบริการที่ดียิ่งขึ้นให้กับธุรกิจ โดยเริ่มศึกษาข้อมูลเกี่ยวกับการเกิดอุบัติเหตุรถยนต์ในกรุงเทพมหานคร ซึ่งนำไปสู่การวางแผนและชี้นำแนวทางธุรกิจเพื่อให้มีการบริการลูกค้าในระดับดีเยี่ยมต่อไป

1.2 คำถามงานวิจัย

1. การใช้ Big Data มีรายละเอียดและประโยชน์อย่างไร
2. เครื่องมือหรือซอฟท์แวร์ที่นำมาใช้วิเคราะห์ Big Data อย่างง่าย และสะดวกมีอะไรบ้าง
3. การนำข้อมูล Big Data จาก Twitter อุบัติเหตุมาใช้มีข้อจำกัดอย่างไร

1.3 วัตถุประสงค์งานวิจัย

1. เพื่อศึกษารายละเอียดและประโยชน์ของการใช้ Big Data
2. เพื่อศึกษาหาเครื่องมือหรือซอฟท์แวร์ที่นำมาใช้กับการวิเคราะห์ Big Data
3. เพื่อศึกษาข้อจำกัดของการนำข้อมูล Big Data จาก Twitter มาใช้

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. บริษัทประกันภัย

1.1 สามารถนำวิธีการวิเคราะห์Big Data มาเป็นแนวทางในการวิเคราะห์ข้อมูลได้อย่างรวดเร็ว ด้วยงบประมาณที่ประหยัดกว่าเดิม

1.2 สามารถวางแผนการจัดเก็บฐานข้อมูลอย่างเป็นระบบและมีมิติของข้อมูลเพียงพอที่จะนำมาวิเคราะห์ได้ในอนาคต ลดต้นทุนในการพัฒนาระบบและการจัดเก็บข้อมูลได้มาก

2. ผู้บริโภค

ผู้บริโภคจะได้ผลิตภัณฑ์ที่ตอบสนองความต้องการอันเกิดจากการวิเคราะห์ข้อมูลจาก Big Dataได้มากยิ่งขึ้น

3. นักการตลาด

3.1 สามารถทำกิจกรรมทางการตลาดผ่านช่องทางต่างๆให้ตรงตามความต้องการของผู้ประกอบการและผู้บริโภคได้มากขึ้น เช่น สามารถทำการโฆษณา ประชาสัมพันธ์ได้ตรงประเด็น และตรงกลุ่มเป้าหมาย

3.2 สามารถปรับแผนการตลาดได้ตามสมัยนิยมไม่ต้องเสียเวลาในการทำแบบ สอบถาม หรือเสียเวลาหาข้อมูล เพราะสามารถเรียกใช้ข้อมูลได้อย่างทันที

4. กลุ่มคนที่ต้องการศึกษา Big Data ในประเทศไทย

สามารถอ่านและทำความเข้าใจพื้นฐานของ Big Dataได้ง่ายขึ้น เพื่อนำไปต่อยอดความรู้เพิ่มเติม หรืออาจจะเพื่อนำไปพัฒนาระบบของธุรกิจที่ตนเกี่ยวข้อง อีกทั้งประหยัดระยะเวลาในการค้นคว้าหาข้อมูลอีกด้วย

1.5 ขอบเขตงานวิจัย

ผู้วิจัยได้เลือกการวิจัยแบบกึ่งทดลอง ด้วยวิธีการใช้โปรแกรมสำเร็จรูป โดยกำหนดขอบเขตการวิจัย ดังนี้

1. ศึกษาแนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data เพื่อนำมาใช้ประโยชน์และเป็นแนวทางของการศึกษาข้อมูล Big Data จาก Twitter
2. กลุ่มประชากรที่ใช้ คือ ข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์ (Twitter)
3. ระยะเวลาการศึกษาตั้งแต่เดือนธันวาคม พ.ศ. 2559 ถึงมีนาคม พ.ศ.2560
4. พื้นที่ในการศึกษา คือ กรุงเทพมหานคร

1.6 นิยามศัพท์เฉพาะ

การวิเคราะห์บิ๊กดาตา (Big Data Analysis) หมายถึง เทคนิค หรือเทคโนโลยีในการกลั่น หรือวิเคราะห์ สกัด เอาคุณค่าออกมาจากข้อมูลขนาดใหญ่ ซึ่งเกินขอบเขตหรือขีดจำกัดของการจัดการข้อมูลแบบเดิม ๆ (Newplus, 2558)

เครื่องมือที่ใช้ในการวิเคราะห์ Big Data หมายถึง โปรแกรมต่างๆ เช่น อาร์ โปรแกรม (R Program) คือ โปรแกรมที่เป็นโปรแกรมฟรี (freeware) และเป็น Open Source ใช้ภาษา เรียกว่า R ในการเขียน สำหรับการคำนวณทั้งทางสถิติ คณิตศาสตร์ การวิเคราะห์ธุรกิจ การเงิน เศรษฐศาสตร์ เภสัชศาสตร์ การแพทย์ ภูมิศาสตร์ และแทบทุกสาขาวิชาที่ใช้การคำนวณหรือการวิเคราะห์ข้อมูลเป็นฐาน(อานนท์, 2559) ฮาดูป (Hadoop) คือ แพล็ตฟอร์ม(Platform) สำหรับการจัดเก็บ และประมวลผลข้อมูลขนาดใหญ่ (Big Data) ซึ่งสามารถรองรับการขยายตัวของข้อมูล และ มีความน่าเชื่อถือสูง โดยสามารถทำการประมวลผลแบบกระจายโดยผ่านเครื่องคอมพิวเตอร์มากมายที่อยู่ในคัสเตอร์ (Cluster) อีกทั้งสามารถขยายระบบด้วยการเพิ่มเครื่อง คอมพิวเตอร์เป็นหลักร้อยหลักพันมากกว่าการขยายเครื่องเดียวให้มีประสิทธิภาพที่สูงขึ้น (Somkiat, 2015) แมพรีดิวซ์ (MapReduce) คือ กรอบการทำงาน (Framework) ในการเขียนโปรแกรมแบบหนึ่งที่ช่วยในงานประมวลผลที่มีชุดของข้อมูล จำนวนมาก เป็นการทำงานแบบขนาน ซึ่งจะอาศัยเครื่องคอมพิวเตอร์หลายๆเครื่องช่วยกันทำงาน (Band1gun, 2559) และไฮฟ์ (Hive) คือ เครื่องมือใช้เตรียมข้อมูลที่เป็นลักษณะคลังข้อมูล (Data warehouse) บน Hadoop โดยมีการกำหนดรายการเตรียมไว้ทำให้สามารถทำการสืบค้น (Query) โดยใช้ภาษาที่เรียก Hive QL ซึ่งมีลักษณะคล้ายภาษา SQL (Prezi, 2014)

ข้อมูล หมายถึง ข้อเท็จจริงหรือเหตุการณ์ต่างๆ ที่เกิดขึ้น อาจจะเป็นตัวเลข ตัวอักษร สัญลักษณ์หรือรายละเอียดซึ่งอาจอยู่ในรูปแบบของภาพ เสียง (iM2Market, 2558) ซึ่งในที่นี้หมายถึง ข้อมูลการเกิดอุบัติเหตุผ่านช่องทางทวิตเตอร์ ที่เกิดขึ้นในเขตกรุงเทพมหานคร

ธุรกิจประกันภัย หมายถึง ธุรกิจประกันวินาศภัยและธุรกิจประกันชีวิตบริษัทใดบริษัทหนึ่ง ซึ่งรับประกันต่อความสูญเสียหรือความเสียหายต่อบุคคลหรือกลุ่มบุคคล โดยสัญญาว่าจะจ่ายชดเชยให้แก่ผู้เอาประกันภัย หรือผู้รับผลประโยชน์กรณีผู้เอาประกันภัยมีการเสียชีวิต หรือมีความเสียหายต่างๆ

บทที่ 2
แนวคิด ทฤษฎี และวรรณกรรมที่เกี่ยวข้อง

ในการศึกษาความสำคัญของการวิเคราะห์ข้อมูล Big Data เพื่อเป็นแนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data กรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุ ในเขตกรุงเทพมหานครในครั้งนี้ ผู้วิจัยได้ศึกษาแนวความคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง เพื่อเป็นพื้นฐานในการวิจัย โดยแบ่งออกเป็น 3 ส่วนดังนี้

2.1 ความหมายของคำสำคัญ

2.1.1 Big Data

2.1.2 การนำข้อมูลBig Dataไปใช้ประโยชน์ทางธุรกิจ

2.2 แนวคิดและทฤษฎีที่เกี่ยวข้อง

2.2.1 แนวคิดและทฤษฎีของข้อมูลขนาดใหญ่ (Big Data)

2.2.2 แนวคิดการใช้โปรแกรมและกระบวนการวิเคราะห์ Big Data

2.2.3 แนวคิดการนำไปใช้และพัฒนาแนวทางธุรกิจ

2.3 งานวิจัยที่เกี่ยวข้อง

2.31 งานวิจัยที่เกี่ยวข้องในประเทศ

2.31 งานวิจัยที่เกี่ยวข้องต่างประเทศ

2.1 คำสำคัญ

2.1.1 Big Data

พนิดา ตันศิริ. (2556) ได้ให้ความหมายว่า Big Data คือปริมาณข้อมูลและสารสนเทศที่เกิดจากการใช้งานของบริษัท รวมทั้งการเติบโตของการใช้บริการสื่อสังคมออนไลน์ (Social Media) ผ่านเครือข่ายสังคม (Social Network) ทำให้มีข้อมูลมหาศาลเกิดขึ้นตลอดเวลาโดยไม่มีรูปแบบหรือไม่มีโครงสร้างแน่นอน

สำรวย กมลายุตต์. (2557). ได้ให้ความหมายว่า Big Data ไม่เพียงแต่มีข้อมูลปริมาณมหาศาลเท่านั้น แต่ยังเป็นคำที่ใช้อธิบายข้อมูลที่อยู่ในรูปแบบที่หลากหลาย และมีที่มาจากแหล่งต่างๆ ทั้งระบบคอมพิวเตอร์ที่ใช้งานในองค์กร เครื่องจักร อุปกรณ์ที่ควบคุมด้วยคอมพิวเตอร์ เซ็นเซอร์ต่างๆที่สร้างข้อมูลและจัดเก็บข้อมูลตลอดเวลาอย่างต่อเนื่อง รวมถึงอุปกรณ์เคลื่อนที่และสื่อสังคมออนไลน์ อีกทั้ง Big Data เป็นพัฒนาการที่ต่อยอดมาจากคลังข้อมูลและการทำธุรกิจอย่างชาญฉลาด

Niels Mouthaan. (2012). ได้ให้ความหมายว่า Big Data คือข้อมูลที่มีความซับซ้อน มีจำนวนมาก มีความหลากหลาย และมีความรวดเร็วของข้อมูล หรือข้อมูลนั้นมีความเกี่ยวข้องสัมพันธ์กับข้อมูลอื่นซึ่งทำให้มีความยากต่อการจัดการฐานข้อมูลหากใช้การจัดการหรือใช้เครื่องมือแบบเดิมๆ

ดังนั้นจึงกล่าวสรุปได้ว่า Big Data หมายความว่า ข้อมูลมหาศาลที่มีการเกิดขึ้นอย่างรวดเร็วตลอดเวลา มีความซับซ้อน และมีโครงสร้างที่ไม่แน่นอน โดยมีรูปแบบที่หลากหลาย

2.1.2 การนำข้อมูลBig Dataไปใช้ประโยชน์ทางธุรกิจ

สุพล พรหมมาพันธุ์. (2557) ได้ให้ความหมายว่า การนำข้อมูล Big Dataไปใช้ประโยชน์ทางธุรกิจ คือ การหาความสัมพันธ์ของข้อมูลที่สำคัญออกมาจากฐานข้อมูลขนาดใหญ่ที่มีอยู่ตามแหล่งข้อมูลต่างๆบนอินเตอร์เน็ต ทำให้ได้เปรียบคู่แข่งขัน

อานนท์ ศักดิ์วรวิชญ์. (2559) ได้ให้ความหมายว่า การนำข้อมูล Big Dataไปใช้ประโยชน์ทางธุรกิจ คือ การค้นหารูปแบบ (Pattern) ต่างๆของข้อมูล เพื่อเอาไปประยุกต์ใช้งาน ซึ่งการวิเคราะห์ธุรกิจ หมายถึงการวิเคราะห์ตะกร้า (Market Basket Analysis) และพยายามหากฎความเชื่อมโยง (Association rule) ระหว่างสินค้าในตะกร้าของคนที่เข้ามาซื้อของในร้านค้า ว่า Stock Keeping Unit ใด สัมพันธ์กับ Stock Keeping Unit ใด ทำให้เราสามารถวางสินค้าใกล้กัน มีการบริหารพื้นที่ในร้าน (Space Management) ได้ดีขึ้น ขายสินค้าที่ควรต้องขายด้วยกันได้มากขึ้น ได้ยอดขายมากขึ้น ลูกค้าพอใจหาของในร้านได้ง่าย

วิกเตอร์ เมเยอร์ ชอนเบอร์เกอร์ และ เคนเน็ต ซูเคีย. (2556) ได้ให้ความหมายว่า การนำข้อมูลBig Dataไปใช้ประโยชน์ทางธุรกิจ คือ การจัดการกับข้อมูลมหาศาลเพื่อสรรหาแนวคิดใหม่ๆที่แตกต่างไปจากแนวคิดเดิมๆและยังเพิ่มคุณค่าทางการตลาด และองค์กร

ดังนั้นจึงสรุปได้ว่า การนำข้อมูลBig Data ไปใช้ประโยชน์ทางธุรกิจ หมายถึง การหารูปแบบหรือความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ เพื่อนำไปประยุกต์ใช้งาน นำพาซึ่งแนวคิดใหม่ๆที่แตกต่างไปจากแนวคิดเดิมและสามารถเพิ่มคุณค่าให้กับธุรกิจ

2.2 แนวคิดและทฤษฎีที่เกี่ยวข้อง

2.2.1 แนวคิดและทฤษฎีของข้อมูลขนาดใหญ่ (Big Data)

“Data is big when data size becomes part of the problem” จากข้อความดังกล่าวให้ข้อสังเกตเกี่ยวกับ Big Data ว่าหากข้อมูลที่มีอยู่ในองค์กรไม่สามารถจัดการได้ด้วยเทคโนโลยีที่มีอยู่แล้วหรือเริ่มมีปัญหาเกี่ยวกับการจัดการข้อมูล กล่าวคือ ข้อมูล ขนาดใหญ่ หลากหลาย เปลี่ยนแปลงรวดเร็ว ยากต่อการประมวลผลและวิเคราะห์ หมายถึง การเริ่มเผชิญกับ Big Data ซึ่งปัญหาที่เกิดขึ้นอาจจะเป็นระบบจัดเก็บข้อมูล (Storage) ที่มีอยู่เริ่มไม่พอ และความเร็ว (Speed) ที่ใช้ในการประมวลผลมีประสิทธิภาพแย่ลง (ปิยะภัสร์ โรจน์รัตนวาณิชย์, 2556)

เมื่อกล่าวถึง Big Data ตามนิยามแรกเริ่มเดิมของ Doug Laney (Vice President ของบริษัท Gartner) จะมี 3 มุม ได้แก่ Volume, Velocity และ Variety หรือ 3Vs แต่ IBM7 เสนอมุมมองอีกเรื่องคือความไม่แน่นอนของข้อมูลเป็นอีกมิติที่ควรพิจารณาเมื่อต้องทำงานกับข้อมูล Big Data ดังนี้

1) Volume (Scale of Data) หมายถึง ขนาดของข้อมูล เช่น Terabytes Zettabytes

2) Velocity (Analysis of Streaming Data) หมายถึง การเปลี่ยนแปลง การเคลื่อนไหวของข้อมูล

3) Variety (Different form of Data) หมายถึง ความหลากหลายของข้อมูล มีการผสมผสานของข้อมูลหลาย ๆ ประเภท เช่น ข้อความ Video Streaming ภาพ เสียง Sensor เป็นต้น

4) Veracity (Uncertainty of Data) หมายถึง ความไม่แน่นอนของข้อมูลที่เข้ามา

นอกจากนี้ NECTEC Researcher Talk ครั้งที่ 2 ประจำปี 2558 กล่าวถึงคุณสมบัติของข้อมูลขนาดใหญ่ Big Data คือการที่มีข้อมูลปริมาณมาก ในระดับ เทราไบต์ (Tera Byte : TB) หรือระดับเพตะไบต์ (Peta Byte : PB) ควรต้องมีรายละเอียดดังนี้

1. ข้อมูลมีการเก็บไว้ในปริมาณ (Volumn) มากและต่อเนื่อง เช่น ข้อมูลมหาศาลจาก Facebook, Line, Twitter และWeb Pages เป็นต้น
2. ข้อมูลที่มีความแตกต่าง หลากหลาย (Variety) เช่น ข้อมูลในฐานข้อมูล ข้อมูลภาพ ข้อมูลเสียง เป็นต้น
3. ข้อมูลมีการเกิดขึ้นอย่างรวดเร็ว (Velocity)
4. ข้อมูลสามารถสร้างประโยชน์ให้กับองค์กรและ มีคุณค่า(Value)กับส่วนที่เกี่ยวข้อง
5. ข้อมูลมีความน่าเชื่อถือ(Veracity)

ประเภทของข้อมูล Big Data

ประเภทของข้อมูล Big Data หมายถึง การจำแนกลักษณะของข้อมูลที่จะนำมาใช้วิเคราะห์ตามรูปแบบของข้อมูล เนื่องจากข้อมูลที่เราใช้นั้นอาจมาจากหลายแหล่งข้อมูล ซึ่งแต่ละแหล่งข้อมูลนั้นจะมีการนิยามลักษณะข้อมูลที่ต่างกันหรือข้อมูลมีลักษณะไม่เหมือนกันแต่มีความหมายเดียวกัน เป็นต้น ซึ่งลักษณะสิ่งรบกวน (Noise) ที่เกิดขึ้นเหล่านี้มีโอกาสพบมากในแหล่งข้อมูลที่สามารถให้ผู้ใช้งานระบุข้อมูลเองได้ เช่น ข้อมูลจาก Twitter ข้อมูล Comment หรือ Review จากเว็บไซต์ เป็นต้น (ปิยะภัสร์ โรจน์รัตนวาณิชย์, 2556) จึงอาจแบ่งประเภทข้อมูล Big Data ออกเป็น 2 แบบใหญ่ๆ คือ

1) ข้อมูลที่มีโครงสร้าง (Structured Data)

ส่วนใหญ่เป็นข้อมูล Transaction เช่น การซื้อขาย ที่บันทึก ได้จาก Point of Sales (POS) ข้อมูลจากระบบการติดต่อสื่อสารของระบบคอมพิวเตอร์ (Log) เช่น การใช้อินเทอร์เน็ต Banking การทำธุรกรรมทางการเงิน เป็นต้น

2) ข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data)

ส่วนใหญ่เป็นข้อมูล Text ที่มาจาก Twitter หรือ Facebook, ข้อมูล Email, ข้อมูลสอบถามหรือร้องเรียนจาก Call Center, ข้อมูลรูปภาพ ข้อมูลข่าว หรือแม้แต่พวกไฟล์เสียง วีดีโอ จาก YouTube เป็นต้น ซึ่งเทคโนโลยีที่ใช้ในการวิเคราะห์ข้อมูลเหล่านี้ก็คือ Text Mining

2.2.2 แนวคิดการใช้โปรแกรมและกระบวนการวิเคราะห์ Big Data

โปรแกรมที่ใช้สำหรับ Big Data หมายถึง เทคโนโลยีหรือเครื่องมือในการจัดการกับข้อมูลที่ซับซ้อนผ่านการเชื่อมต่อกับระบบเพื่อเชื่อมโยงสู่ระบบประมวลผลของข้อมูลปริมาณมาก ทั้งนี้โปรแกรมที่นำมาใช้จะขึ้นอยู่กับลักษณะหรือประเภทของข้อมูลที่นำมาเพื่อใช้ส่งต่อสู่กระบวนการวิเคราะห์ (iNnovationLab, 2557)

Big Data เทคโนโลยี หมายถึง กระบวนการและการวิเคราะห์ Big Data ผ่านเทคโนโลยีและการใช้เทคนิคความสามารถในการจัดเก็บข้อมูลที่ไม่มีมีโครงสร้างและกึ่งโครงสร้าง รวมถึงการประยุกต์ใช้ในการวิเคราะห์ขั้นสูงและเทคโนโลยีการแสดงผลข้อมูลของพฤติกรรมผู้บริโภคในเชิงลึก โดยมี 3 เทคโนโลยีหลัก ที่โดดเด่น คือ MapReduce, Hadoop และ NoSQL เพื่อส่งเสริมการวิเคราะห์ธุรกิจและจัดการข้อมูลทางการตลาด (Ramesh Sharda, Dursun Delen and Dfraim Turban, 2014) ดังนี้

1. MapReduce คือ เทคโนโลยีจากผู้ผลิต Google ที่ใช้สำหรับจัดการข้อมูลขนาดใหญ่ โดยมีหลักการทำงานหลักๆ 2 อย่าง คือการแมพ (Map) ที่เป็นการจัดการข้อมูลด้วยการแบ่งประเภทของข้อมูลออกมาย่อยๆเป็นหมวดหมู่ และการ Reduce คือ การวิเคราะห์ผล

2. Hadoop คือ Open source framework สำหรับกระบวนการจัดเก็บและวิเคราะห์ข้อมูลขนาดใหญ่ จัดทำขึ้นโดย Yahoo โดยมีแรงบันดาลใจมาจาก MapReduce โดย Hadoopสามารถจัดการแบ่งข้อมูลขนาดใหญ่ให้ออกเป็นหลายๆ Node เพื่อวิเคราะห์ข้อมูลแบบคู่ขนานและเป็นการแบ่งงานกันทำระหว่างข้อมูลในเวลาเดียวกัน ซึ่ง Hadoop นั้นสามารถทำงานกับหลายๆโปรแกรม เช่น Hive ที่หลักการทำงานคล้ายSQL และถูกพัฒนาโดย Facebook เป็นต้น

3. NoSQL คือ เทคโนโลยีที่มีการทำงานคล้ายกับ Hadoop หรืออาจมีการทำงานร่วมกับ Hadoop (Ramesh Sharda, 2014) โดยเป็นอีกแนวทางหนึ่งในการจัดการและออกแบบฐานข้อมูล สำหรับข้อมูลขนาดใหญ่ที่อยู่อย่างกระจัดกระจายหลากหลายรูปแบบ (Somkiat, 2014)

2.2.3 แนวคิดนำไปใช้และพัฒนาแนวทางธุรกิจ

แนวคิดนำไปใช้และพัฒนาแนวทางธุรกิจ หมายถึง ความต้องการใช้ประโยชน์จากข้อมูลเหล่านั้น การจะนำเอาข้อมูล Big Data มาใช้งานให้เกิดประโยชน์ต้องมีความพร้อมหลายด้าน ไม่ว่าจะเป็นด้านเทคโนโลยี ด้านเทคนิค และด้านบุคลากร หากไม่มีความพร้อมหรือจัดการไม่เหมาะสม ข้อมูลเหล่านั้นอาจจะกลายเป็นขยะในการจัดการข้อมูลไปในทันที ซึ่งBig Data เป็นประโยชน์ต่อการใช้งานหลายประการ เช่น การใช้งานข้อมูลเกี่ยวกับการค้นคว้า วิจัย เอกสาร เครือข่ายทางสังคม หรือข้อมูลเฉพาะต่างๆ เช่น โรงพยาบาล คลังข้อมูลต่างๆ เป็นต้น ซึ่ง Big Data นี้เหมาะสำหรับการนำมาวิเคราะห์ข้อมูลดิบ หรือข้อมูลกึ่งโครงสร้างต่างๆ นำไปใช้ในการวิเคราะห์พฤติกรรมลูกค้าหรือธุรกิจที่เกี่ยวข้อง เพื่อหาการแก้ไขหรือหาวิธีการจัดการให้ธุรกิจเป็นไปตามที่คาดหวัง ไม่ว่าจะเป็นด้านธุรกิจที่จะเพิ่มโอกาสทำให้เกิดนวัตกรรมด้านเทคนิคที่สามารถรวบรวมและจัดเก็บข้อมูลได้ง่ายยิ่งขึ้น ซึ่งปัจจุบันนี้มีเครื่องมือที่ได้รับความนิยมเข้ามามีส่วนช่วยในการจัดการก็คือ Hadoop ที่ถูกพัฒนามาจาก Open Source Technology สามารถเก็บข้อมูลขนาดใหญ่และนำไปประมวลผลได้ แต่การวิเคราะห์ Big Data นั้นเป็นเพียงแค่การวิเคราะห์ข้อมูลดิบแบบย่อยเท่านั้น หากต้องการข้อมูลที่เจาะลึกมากขึ้นไปอีกก็ต้องเพิ่มขั้นตอนการวิเคราะห์แบบ Analytics ที่จะทำให้ได้ข้อมูลในเชิงลึกมากขึ้นไปอีกด้วย (สุภัค ลายเลิศ, 2558)

นอกจากนี้ สุภัค (2558) ยังกล่าวเพิ่มเติมอีกว่า แนวโน้มในการนำ Big Data Analytics มาใช้งานนั้นมีเป้าหมายรองรับการขยายตัวธุรกิจอย่างชัดเจนดังนี้ ต้องมีความเร็วในการนำข้อมูลมาวิเคราะห์และใช้งานได้รวดเร็วและง่ายพอที่จะรับมือกับผู้ใช้งานแผนกต่างๆ ได้ และด้วยปริมาณ

ข้อมูลที่เติบโตอย่างรวดเร็วเช่นกัน หากสามารถนำมารวบรวมและวิเคราะห์ได้ก็จะเป็นการใช้ประโยชน์จากสินทรัพย์ขององค์กรเพื่อสร้างก้าวต่อไปที่แข็งแกร่งได้อีก และยังหากสามารถทำให้บริษัทสามารถเข้าใจลูกค้าของตนได้ลึกซึ้ง มีภาพรวมธุรกิจ และวิเคราะห์ตอบโจทย์เพื่อสร้าง Customer Experience ที่ให้ความแตกต่างทางธุรกิจได้ ก็จะยิ่งเป็นการต่อยอดธุรกิจให้เหนือคู่แข่ง นอกจากนี้สิ่งสำคัญการนำข้อมูลเหล่านี้มาใช้งาน ต้องสามารถตอบสนองทิศทางของแต่ละแผนกได้ดี และไม่ควรใช้งานยากเกินไป ทางที่ดี คือ ต้องง่ายจนแทบไม่ต้องเทรนด์หรือสอนเลย เพื่อให้เข้าถึงจิตใจลูกค้า ผู้บริโภคและเป้าหมายของบริษัทได้แม้รายละเอียดเล็กน้อย ทุกวันนี้การทำการค้าที่เข้าถึงใจผู้บริโภคและมีความคล่องตัวในการให้การบริการแบบหลากหลายช่องทางนั้นคือกุญแจแห่งความสำเร็จ

กวี รุจิรัตน์. (2556) กล่าวว่า การนำไปใช้และวิธีพัฒนาแนวทางธุรกิจวิธีหนึ่งที่จะช่วยนำข้อมูลกลับมาใช้ใหม่ได้คือ การออกแบบระบบการจัดเก็บข้อมูลที่สามารถนำข้อมูลไปขยายผลได้ในอีกหลากหลายช่องทาง เช่น ผู้ค้าปลีกบางรายติดตั้งกล้องวงจรปิดภายในร้าน ไม่ใช่เพื่อจับขโมยอย่างเดียวเท่านั้น แต่เพื่อสำรวจจำนวนลูกค้าที่เข้ามาใช้บริการในร้านและทราบถึงรายละเอียดพฤติกรรมของลูกค้าที่หยุดดูสินค้าเพื่อออกแบบผังที่ดีที่สุดของร้านค้าและนำข้อมูลไปต่อยอดทางการตลาด

อีกทั้งอภิมหาข้อมูลยังเอื้อประโยชน์สำหรับบริษัทขนาดเล็กที่ไม่จำเป็นต้องมีทรัพยากรทางกายภาพมากนักและไม่ต้องอาศัยเงินทุนจำนวนมาก แต่สามารถซื้อสิทธิการใช้ข้อมูลโดยที่ไม่จำเป็นต้องเป็นเจ้าของเอง และดำเนินการวิเคราะห์บน Platform คอมพิวเตอร์ในราคาต่ำและจ่ายค่าลิขสิทธิ์เป็นเปอร์เซ็นต์จากรายได้

2.3 บทความและงานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้องกับความสำคัญของการวิเคราะห์ข้อมูล Big Data เพื่อการวางแผนและชี้นำแนวทางของธุรกิจฯ สามารถแบ่งออกได้เป็น ผลงานวิจัยที่เกี่ยวข้องในประเทศและผลงานวิจัยที่เกี่ยวข้องในต่างประเทศ ดังนี้

2.3.1 งานวิจัยที่เกี่ยวข้องในประเทศ

ณิชจิรัชย์ ตั้งคำ. (2556). ได้ศึกษาเรื่อง แผนธุรกิจเพื่อระบบวิเคราะห์ Big Data ของบริษัทฮิวเลตต์-แพคการ์ดประเทศไทย สำหรับธุรกิจธนาคาร โดยมีวัตถุประสงค์เพื่อเพิ่มประสิทธิภาพในการทำตลาดของระบบวิเคราะห์ Big Data ให้กับกลุ่มลูกค้าธนาคาร และเพิ่มรายได้จากระบบวิเคราะห์ Big Dataให้กับบริษัท อีกทั้งเพื่อเป็นต้นแบบในการนำไปใช้ประยุกต์กับกลุ่ม

อุตสาหกรรมอื่น ซึ่งผู้วิจัยได้ทำการศึกษาและวิจัยเกี่ยวกับแผนธุรกิจให้กับบริษัท ฮิวเลตต์-แพ็คการ์ด ในการนำเสนอระบบวิเคราะห์ Big Dataให้กับธนาคารในประเทศไทย โดยใช้เทคโนโลยีในการวิเคราะห์ Big Data ในตลาดโลก ข้อมูลล่าสุดในปี พ.ศ. 2557 และใช้ข้อมูลทางการเงินในระบบวิเคราะห์ Big Data ด้วยวิธีการประมาณการจากกำไรที่เกิดขึ้นในอดีตเป็นอัตราส่วน เนื่องจากไม่สามารถแตกตัวเลขทางการเงินเป็นต้นทุนย่อยๆได้ ในส่วนของกระบวนการในการวิจัยนั้นได้ศึกษาและประเมินทรัพยากรรวมถึงความสามารถที่ประเทศไทยไทยทำได้ ซึ่งทรัพยากรที่ประเทศไทยมีคือ ระบบในการวิเคราะห์ Big Dataของบริษัท ที่ประกอบด้วยซอฟต์แวร์หลายชนิด ทรัพยากรด้านบุคคลโดยเน้นไปที่ความสามารถของบุคคลที่จะทำตลาดด้านนี้ได้ อีกทั้งศึกษาความต้องการของกลุ่มลูกค้าธนาคาร ใช้วิธีสัมภาษณ์ผู้บริการด้านกลยุทธ์ขององค์กร และผู้บริการด้านเทคโนโลยีสารสนเทศ เพื่อทราบถึงความต้องการนำระบบวิเคราะห์ Big Data เข้ามาใช้อย่างจริงจังในองค์กร โดยแบ่งการสัมภาษณ์ออกเป็น 2 กลุ่มใหญ่ คือ ธนาคารพาณิชย์และธนาคารรัฐบาล เพื่อศึกษาถึงความต้องการว่าเหมือนหรือแตกต่างกันอย่างไร โดยมีธนาคารเพื่อการเกษตรและสหกรณ์การเกษตรเป็นตัวแทนของธนาคารรัฐบาล นอกจากนี้หลังจากที่ได้ข้อมูลตลาดและข้อมูลภภายในองค์กรครบถ้วนแล้วจึงวางแผนธุรกิจเพื่อปรับปรุงการดำเนิการของบริษัทฯ และเสนอแผนกลยุทธ์การตลาดที่เหมาะกับสถานการณ์ต่างๆที่อาจเกิดขึ้นจากสภาพแวดล้อมการแข่งขันในปัจจุบัน จากการศึกษาพบว่า การช่วยให้ธุรกิจธนาคารเล็งเห็นประโยชน์มหาศาลของการนำการวิเคราะห์ Big Data มาใช้ในธุรกิจ เป็นหัวใจในการทำการตลาด อีกทั้งแนวโน้มของเทคโนโลยีไม่เพียงแต่จะมีแค่ Big Data เท่านั้น ยังมีเรื่องของ Cloud Computing ที่จะมาลดต้นทุนให้กับบริษัท ดังนั้นแนวทางการทำตลาดแบบ Big Data on cloud น่าจะเป็นหนทางทำการตลาดที่มีประสิทธิภาพสูงสุด อีกทั้งยังพบว่าการที่บริษัทมีบุคคลากรที่มีความรู้ ความเชี่ยวชาญในผลิตภัณฑ์ของตนเองเป็นจุดแข็งของบริษัททางด้านเทคโนโลยีแต่การสร้างความพึงพอใจให้ลูกค้าจำเป็นต้องรู้จักธุรกิจของลูกค้าอย่างลึกซึ้ง ดังนั้นสิ่งที่บริษัทเอชพีและบริษัททางด้านเทคโนโลยีขาดไปคือ ผู้เชี่ยวชาญหรือผู้ประสานงานให้เกิดความเข้าใจดังกล่าวและถ่ายทอดออกมาเป็นประโยชน์โดยภาพรวมที่บริษัทเป็นมากกว่าคู่ค้าขาย แต่เป็นผู้ร่วมทำธุรกิจและจะเติบโตไปด้วยกัน อีกทั้ง งานระบบวิเคราะห์ Big Data จำเป็นกับธุรกิจทั่วไปในทุกๆอุตสาหกรรม เช่น ธุรกิจการสื่อสาร หากสามารถวิเคราะห์พฤติกรรมการใช้งาน รวมถึงแนวทางการใช้ชีวิตของลูกค้าแต่ละรายได้ ก็สามารถเสนอการใช้บริการที่เหมาะสมที่สุดได้ อีกทั้งสามารถส่งเสริมการขายสินค้าหรือบริการที่มีความจำเป็นต่อลูกค้าแตะละรายเพิ่มเติมได้ โดย Big Data นั้นเป็นเทรนด์ใหม่ที่ทุกๆอุตสาหกรรมนำมาใช้ให้เกิดความได้เปรียบทางการแข่งขัน

ทวีวัฒน์ ขนาน. (2558). ได้ศึกษาเรื่อง ระบบวิเคราะห์ข้อมูลขนาดใหญ่เพื่อสนับสนุนการตัดสินใจในการบริหารช่องทางการให้บริการของธนาคารผ่านเครื่องรับจ่ายเงินอัตโนมัติ โดยมีวัตถุประสงค์เพื่อ สร้างฐานข้อมูลสำหรับจัดเก็บข้อมูลขนาดใหญ่ที่จะนำมาใช้เพื่อการสนับสนุนการตัดสินใจในการบริหารช่องทางการให้บริการของธนาคารผ่านเครื่องรับจ่ายอัตโนมัติ และสร้างรายงานการวิเคราะห์ข้อมูลขนาดใหญ่ของการให้บริการของธนาคารผ่านเครื่องรับจ่ายเงินอัตโนมัติที่จัดเก็บในฐานข้อมูล อีกทั้งยังเพื่อให้ผู้บริหารมีรายงานการวิเคราะห์ที่สามารถนำมาช่วยในการตัดสินใจในการบริหารช่องทางการให้บริการของธนาคารผ่านเครื่องรับจ่ายเงินอัตโนมัติ โดยผู้วิจัยได้นำเครื่องจ่ายเงินอัตโนมัติที่อยู่ในเขตกรุงเทพมหานครเท่านั้นมาเป็นตัวอย่าง จำนวนทั้งสิ้น 10 เครื่อง ซึ่งเป็นเครื่องที่มีการใช้งานอย่างต่อเนื่อง ไม่มีการหยุดให้บริการ และเครื่องจ่ายเงินนั้นฯต้องมีการกำหนดการเติบเงินที่ชัดเจน เช่น 3 วัน 5 วัน หรือ 10 วัน เป็นต้น โดยมีขั้นตอนการดำเนินงานทั้งหมด 5 ขั้นคือ วิเคราะห์ระบบ ออกแบบระบบ พัฒนาระบบ ทดสอบระบบและจัดทำเอกสาร มีรายละเอียดต่างๆดังนี้ ขั้นตอนที่ 1 คือ การวิเคราะห์ระบบ(System Analysis) โดยการศึกษาและทำความเข้าใจในกระบวนการทำงานของระบบงานปัจจุบันอย่างละเอียด จากนั้นวิเคราะห์ระบบและความต้องการของผู้ใช้งาน รวมถึงกำหนดวัตถุประสงค์และขอบเขตโครงการ เทคโนโลยีที่ใช้ วิธีแก้ไขปัญหาพร้อมรายละเอียด ขั้นตอนการดำเนินงาน ระยะเวลาที่ใช้และประโยชน์ที่คาดว่าจะได้รับ ขั้นตอนที่ 2 คือ ออกแบบระบบ (System Design) คือ การออกแบบโมเดลโดยโครงสร้างข้อมูลจะอยู่ในรูปแบบของ HDFS เพื่อให้สามารถใช้ Hadoop Cluster ทำ Map Reduce กับข้อมูลได้ รวมถึงออกแบบรายงานให้ตรงตามกับผู้ใช้งาน ขั้นตอนที่ 2 คือ การพัฒนาระบบ (System Development) โดยการพัฒนาโปรแกรมและนำข้อมูลเข้าสู่ระบบ Hadoop Cluster อีกทั้งพัฒนาการเรียกดูข้อมูลขนาดใหญ่และพัฒนารูปแบบรายงานเพื่อช่วยในการวิเคราะห์และตัดสินใจของผู้บริหาร (Report Preparation) ขั้นตอนที่ 4 คือ การทำสอบระบบ (System Testing) โดยการทดสอบระบบงานใหม่ที่ได้พัฒนาเพื่อตรวจสอบให้เป็นไปตามที่ได้ออกแบบไว้และค้นหาข้อผิดพลาดต่างๆที่อาจจะเกิดขึ้นได้ รวมถึงการปรับปรุงแก้ไขข้อผิดพลาดด้วย ขั้นตอนที่ 5 คือ การจัดทำเอกสาร (Documentation) โดยจัดทำคู่มือการใช้งานระบบใหม่ รวมทั้งเอกสารประกอบการดำเนินงานโครงการ เป็นต้น นอกจากนี้ผู้วิจัยได้ใช้เทคโนโลยีพัฒนาระบบ โดยด้าน Software ใช้ระบบปฏิบัติการ Ubuntu12.04.3 LTS, ระบบปฏิบัติการBig Data Hadoop Ecosystem, ระบบจัดการฐานข้อมูล Oracle Express Edition 11g Release 2, เครื่องมือในการแสดงผลระบบวิเคราะห์ข้อมูลขนาดใหญ่ด้วย Tableau Destop 9.1, เครื่องมือที่ใช้ในการทำ Data Mining Rapid Miner Studio 6.0 และ RStudio ส่วนด้วน Hardware ใช้หน่วยประมวลผลกลาง (CPU) Intel Core i5 2.40GHz, หน่วยความจำ 8.0 GB และ Hard Disk 256 GB โดยจากผลการวิจัยพบว่า ระบบวิเคราะห์ข้อมูลขนาด

ใหญ่เพื่อสนับสนุนการตัดสินใจในการบริหารช่องทางการให้บริการธนาคารผ่านเครื่องรับจ่ายเงินอัตโนมัตินั้น เพื่อให้เกิดประโยชน์สูงสุดต่อบริษัทแบ่งออกเป็น 3 ระบบหลัก คือ 1. ระบบวิเคราะห์เครื่องเติมเงินในเครื่องรับจ่ายเงินอัตโนมัติ 2. ระบบวิเคราะห์ความพร้อมในการให้บริการของเครื่องรับจ่ายเงินอัตโนมัติ และ 3. ระบบวิเคราะห์พฤติกรรมของลูกค้าในการใช้งานเครื่องรับจ่ายเงินอัตโนมัติ เทคโนโลยีที่ใช้ในโครงการนี้เป็นชุดโปรแกรมของ Hadoop Ecosystem สำหรับจัดเก็บข้อมูลขนาดใหญ่ เพื่อนำมาสนับสนุนการตัดสินใจในการบริหารช่องทางการให้บริการของธนาคารผ่านเครื่องรับจ่ายเงินอัตโนมัติได้จริง และสามารถสร้างรายงานการวิเคราะห์ข้อมูลขนาดใหญ่ได้ รวมถึงสามารถพยากรณ์การถอนเงินสดในแต่ละวันจากเครื่องรับจ่ายเงินอัตโนมัติ พร้อมทั้งสามารถหากฎความสัมพันธ์ของการเสียของเครื่องรับจ่ายเงินอัตโนมัติได้ นอกจากนี้ยังพบปัญหาในการพัฒนาโครงการ แบ่งเป็น 2 ส่วนหลัก คือ 1. ปัญหาด้านการวิเคราะห์ระบบ ได้แก่ การจัดเก็บข้อมูลขององค์กรตัวอย่างนั้นถูกเก็บในลักษณะที่กระจัดกระจาย มีหลากหลายรูปแบบ ไม่มีความสัมพันธ์กันในแบบที่เหมาะสม อีกทั้งข้อมูลนั้นถูกบันทึกด้วยพนักงานหลายคนซึ่งบันทึกกันคนละรูปแบบไม่ครบถ้วน และ 2. ปัญหาด้านเทคนิค คือ ข้อจำกัดของเครื่องมือที่ใช้ในการพัฒนาของงานวิจัยนี้ เช่น Hadoop Ecosystem มีการทำงานค่อนข้างซับซ้อนตั้งแต่ติดตั้งโปรแกรม อีกทั้งโปรแกรมอื่นๆ ต้องมีความเหมาะสมของแต่ละเทคนิค รวมถึงผู้พัฒนาระบบไม่มีประสบการณ์ในการใช้เครื่องมือด้าน Business Intelligence มาก่อนทำให้ต้องใช้เวลาในการศึกษา แต่ทั้งนี้ ระบบวิเคราะห์พฤติกรรมลูกค้าในการใช้งานเครื่องรับจ่ายเงินอัตโนมัติหรือ ATM มีการพัฒนาขึ้นตามแนวคิดของ Big Data สามารถเพิ่มโอกาสในการแข่งขันให้กับธนาคาร นอกจากนี้กสามารถวิเคราะห์ Big Data เพื่อติดตามและวิเคราะห์การตอบสนองของผู้ใช้บริการ ผ่านการติดตาม Social Media Twitter, Facebook, Youtube ซึ่งช่วยประเมินในส่วนของโฆษณาการตลาดใหม่ๆ ผลิตภัณฑ์และบริการ เพื่อตอบสนองลูกค้าในเชิงบวกและลดการตอบสนองในเชิงลบ ผ่านการวิเคราะห์ Big Data

ข้าวทิพย์ ตันติวรวงศ์. (2558). ได้ทำการศึกษาเรื่องการนำเสนอข้อมูลขนาดใหญ่ด้วยแท็บโบล (Visualizing Big Data using Tableau) โดยมีวัตถุประสงค์เพื่อ ศึกษาเทคโนโลยีที่เกี่ยวข้องกับวิธีการนำเสนอข้อมูลขนาดใหญ่แบบต่างๆจากแท็บโบล และเพื่อประยุกต์ใช้โดยมีกระบวนการรับข้อมูลขนาดใหญ่เข้ามาผ่าน Hadoop Framwork หรือ Data Warehouse ที่เหมาะสมและจำลองสถานการณ์การทำงานของทั้งกระบวนการด้วยชุดข้อมูลตัวอย่าง โดยมีวิธีการดำเนินงาน คือ ศึกษาเอกสาร คู่มือการใช้งานที่มีมาก่อน, ติดตั้งโปรแกรมแท็บโบลลงบนคอมพิวเตอร์, ทดลองการเชื่อมต่อในกระบวนการรับและนำเสนอข้อมูลขนาดใหญ่ ทั้งนี้งานวิจัยนี้ได้ใช้ข้อมูลของ NYC Taxi Data เฉพาะส่วน Yellow Taxi ประจำเดือนมิถุนายน พ.ศ. 2558 จำนวน 12,324,935 แถว ซึ่งเป็นข้อมูลที่ได้มาจาก NYC Taxi & Limousine Commission มีขนาดประมาณ 1.69 GB และจัดเก็บข้อมูล

บน Amazon Web Serviceซึ่งเป็นบริการบน Cloud นอกจากนี้ยังมีการใช้ Google Maps APIs ร่วมกับโปรแกรมแท็บโบลเพื่อแสดงสถานที่ตามพิกัดต่างๆที่รถแท็กซี่รับ-ส่งผู้โดยสารและเพิ่มความสวยงามของแผนที่ด้วยโปรแกรม Mapbox ที่มีอยู่ในแท็บโบล ผู้วิจัยได้สุ่มตัวอย่างจำนวน 5,000 แถวจากคลังข้อมูลซึ่งเป็นข้อมูลจำนวนน้อยที่สามารถนำเสนอบนคอมพิวเตอร์ได้ โดยแบ่งข้อมูลออกเป็น 3 กลุ่ม คือ พิกัดละติจูดลองติจูด, พื้นที่รอบๆบริเวณขนาดเล็กพิกัดรัศมีไม่เกิน 0.2 ไมล์ และพื้นที่ขนาดใหญ่รัศมีเกิน 2 ไมล์ ผลงานวิจัยพบว่า มีเทคโนโลยีต่างๆที่โปรแกรมแท็บโบลสามารถทำได้ เช่น การใช้งานร่วมกับการบริการ Google Maps API, การใช้บริการแผนที่ของ Mapbox, การรวบรวมข้อมูล(Blend Data), การแบ่งกลุ่มด้วยการสร้างGroup, การสร้าง Set การ Calculated Field, การแบ่งความแตกต่างด้วยสีและขนาด, ตลอดจนการสร้าง Dashboard พร้อมกับการสร้างปฏิสัมพันธ์กับผู้ใช้ เป็นต้น นอกจากนี้ยังได้ข้อสรุปว่า กล่าวว่า การทำงานกับข้อมูลขนาดใหญ่ไม่จำเป็นต้องทำงานกับข้อมูลทั้งหมด แต่ควรสุ่มตัวอย่างข้อมูลด้วยปริมาณที่เหมาะสม เพื่อใช้เป็นตัวแทนประชากร อย่างไรก็ตามการแสดงผลด้วยข้อมูลขนาดใหญ่ต้องใช้เครื่องคอมพิวเตอร์ที่มีประสิทธิภาพและสามารถรองรับการทำงานกับข้อมูลจำนวนมาก ดังนั้นโปรแกรมแท็บโบลจึงมีเมนูสุ่มตัวอย่างที่รับประกันการสุ่มตัวอย่างที่ดีที่สุดเพื่อให้ได้กลุ่มตัวอย่างที่น้อยและนำไปใช้งานได้ ซึ่งการทำงานกับข้อมูลขนาดเล็กนั้นยังมีความเหมาะสมกับเครื่องคอมพิวเตอร์ทั่วไป ทำให้ทำงานได้รวดเร็วขึ้นและไม่สิ้นเปลืองพื้นที่ในการจัดเก็บ

สุวิมล ประทุม. (2555). ได้ศึกษาการปรับปรุงประสิทธิภาพของระบบที่มีฐานข้อมูลขนาดใหญ่ โดยมีวัตถุประสงค์เพื่อให้การทำงานของระบบมีประสิทธิภาพ โดยศึกษาถึงความสัมพันธ์ของข้อมูล โครงสร้างของข้อมูล การเข้าถึงข้อมูล และกระบวนการที่โปรแกรมใช้ประยุกต์เรียกฐานข้อมูล อีกทั้งงานวิจัยนี้ได้นำเสนอวิธีการปรับปรุงประสิทธิภาพของระบบที่มีฐานข้อมูลขนาดใหญ่ และเสนอแนะแนวทางการปรับแต่งและออกแบบฐานข้อมูลที่เหมาะสมของระบบที่ใช้ฐานข้อมูลของกรณีศึกษา ซึ่งได้ทดสอบกับระบบรับบุคคลเข้าศึกษาในสถาบันอุดมศึกษา ประกอบด้วยระบบย่อย 3 ระบบ คือ ระบบการรับสมัครบุคคลเข้าศึกษาในสถาบันอุดมศึกษา ระบบการรวมคะแนนและจัดลำดับการคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา และระบบประกาศผลการคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา โดยใช้ฐานข้อมูลออราเคิลและยึดหลักการปรับปรุงประสิทธิภาพของฐานข้อมูลตามแนวทางที่นำเสนอ 2 แนวทาง คือ การสร้างดรรชนีที่เหมาะสม และการปรับปรุงคำสั่งเอสคิวแอล (SQL)โดยใช้พาราเลลฮินต์ พบว่าแนวทางที่นำเสนอนั้นสามารถช่วยให้ประสิทธิภาพของระบบดีขึ้น ได้แก่ ระบบการรับสมัครบุคคลเข้าศึกษาในสถาบันอุดมศึกษาดีขึ้นร้อยละ 44.42 ระบบการรวมคะแนนและจัดลำดับการคัดเลือกดีขึ้นร้อยละ 51.12 และระบบประกาศผลการคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาดีขึ้นร้อยละ 33.89

ดังนั้นระบบที่ฐานข้อมูลขนาดใหญ่จำเป็นต้องมีการออกแบบฐานข้อมูลที่เหมาะสม ทั้งนี้ต้องศึกษาความสัมพันธ์ของข้อมูล โครงสร้างข้อมูล การเข้าถึงข้อมูลและกระบวนการที่โปรแกรมประยุกต์จะเรียกใช้ฐานข้อมูล เป็นต้น

ปิยะภัสร์ โรจน์รัตนวาณิชย์ (2556) ได้ทำการศึกษาเรื่อง แนวทางการคุ้มครองข้อมูลใน Big Data : ศึกษาประเด็นความเป็นส่วนตัวและความมั่นคงปลอดภัยของข้อมูล (Guideline for Data Protection in Big Data: Privacy and Data Security) โดยมีวัตถุประสงค์เพื่อศึกษาวิเคราะห์ปัญหาและผลกระทบของการใช้ข้อมูลใน Big Data ที่เกี่ยวข้องกับความเป็นส่วนตัว และความมั่นคงปลอดภัยของข้อมูล และวิเคราะห์กฎหมายไทยที่เกี่ยวข้องกับ Big Data เรื่องความเป็นส่วนตัว และความมั่นคงปลอดภัยของข้อมูล นอกจากนี้ยังเพื่อเปรียบเทียบกฎหมายประเทศสหรัฐอเมริกา และประเทศอังกฤษตามแนวทางของสหภาพยุโรปที่เกี่ยวข้องกับ Big Data เรื่องความเป็นส่วนตัว และความมั่นคงปลอดภัยของข้อมูล อันจะนำมาสู่การวิเคราะห์เปรียบเทียบกับกฎหมายไทยที่เป็นอยู่ในปัจจุบัน เป็นต้น โดยดำเนินการวิจัยศึกษากฎหมายที่เกี่ยวข้องกับ Big Data ทั้งในประเทศและต่างประเทศ และรวบรวมผลการวิเคราะห์จากเอกสารทั้งหมด พบว่า การจัดเก็บรวบรวมและวิเคราะห์ข้อมูลอย่างแม่นยำและรวดเร็ว ทั้งภาพ เสียง ตัวอักษร ตัวเลข และอื่น ๆ มีความหลากหลายและมากมายจนระบบฐานข้อมูลเดิมไม่ สามารถจัดการได้ รวบรวมไว้เป็นข้อมูลมหาศาล เรียกว่า Big Data แม้ว่าการใช้งาน Big Data จะ เต็มไปด้วยประโยชน์มากมายแต่ก็อาจเปรียบดังเหรียญที่มีสองด้าน เนื่องจากคุณสมบัติเฉพาะตัวบาง ประการที่สามารถสร้างผลกระทบในเชิงลบต่อบุคคลได้ และสิ่งหนึ่งที่ต้องระลึกถึงอยู่เสมอคือ ความ มั่นคงปลอดภัยของข้อมูลที่ต้องมีการป้องกันไม่ให้เกิดความเสียหายหรือรั่วไหล ตลอดจนตระหนักถึง การใช้งานข้อมูลทั้งของตนเองและขององค์กรอย่างเหมาะสม หลังจากได้ทำการศึกษากฎหมายต่าง ๆ ทั้งที่มีผลใช้บังคับอยู่ ในปัจจุบันและที่กำลังอยู่ในระหว่างการพิจารณาของรัฐสภาพบว่ากฎหมายที่ให้ความคุ้มครองความเป็นส่วนตัวในข้อมูลส่วนบุคคล (Personal Information) และการรักษาความมั่นคงปลอดภัยของข้อมูล (Data Security) ก็มีการให้ความคุ้มครองในลักษณะเป็นกฎหมายเฉพาะเรื่อง แต่ยังขาด กฎหมายคุ้มครองข้อมูลส่วนบุคคลเพื่อบังคับใช้กับภาคเอกชนเป็นการทั่วไป เป็นต้น

ธนพร สิทธิชัยวิเศษ (2557) ได้ทำการศึกษาเรื่องการวิเคราะห์ข้อมูลขนาดใหญ่เพื่อการดำเนินการโดยใช้เอสเอพี ฮานา (Actionable Big Data Analytics by Using SAP HANA) โดยมีวัตถุประสงค์เพื่อศึกษาซอฟต์แวร์ เอสเอพีฮานาและนำซอฟต์แวร์เอสเอพี ฮานา ไปใช้ในการวิเคราะห์ข้อมูลและแก้ปัญหาธุรกิจ โดยใช้ข้อมูลตัวอย่าง (Sample Data) จากข้อมูลจริงในธุรกิจ โดยผู้วิจัยได้ศึกษาหลักการเบื้องต้นของซอฟต์แวร์ เอสเอพี ฮานา จากเอกสารและข้อมูลทางอินเตอร์เน็ต ทั้งฟังก์ชันและการทำงานของซอฟต์แวร์ และหาข้อมูลการใช้งานเพิ่มเติมจากอินเตอร์เน็ต รวมถึง

ทดลองติดตั้งและใช้งานฟังก์ชันสำคัญของซอฟต์แวร์ฯ พบว่า เอสเอพี ฮานา สามารถประมวลผลแบบใหม่ที่เรียกว่า In-memory Computing ได้พร้อมทั้งมีเทคโนโลยีที่มาช่วยเพิ่มประสิทธิภาพในการทำงานในด้านการเก็บข้อมูลแบบ Column-based และสามารถนำมาใช้วิเคราะห์แก้ไขปัญหาทางธุรกิจได้จริง แต่อย่างไรก็ตามทางผู้วิจัยมีความคิดเห็นว่า เอสเอพี ฮานา ยังไม่เหมาะสมกับผู้ใช้ทั่วไป เนื่องจากยังเป็นซอฟต์แวร์ใหม่ และมีการอัพเดทบ่อย เช่น กรณีคู่มือบนหน้าเว็บต์ไซต์มีการเปลี่ยนแปลงบ่อยครั้ง หากไม่มีผู้เชี่ยวชาญช่วยเหลืออาจทำให้เกิดปัญหาในการนำมาใช้ได้

สำรวย กมลายุตต์ (2557) ได้รายงานผลโครงการศึกษาเพิ่มเติมด้าน Big Data Governance and Big Analytic โดยมีวัตถุประสงค์เพื่อวางแผนกลยุทธ์ในการแข่งขันทางธุรกิจ โดยผู้วิจัยได้วิเคราะห์ประเภทของ Big Data ออกมา 5 ประเภทหลัก ได้แก่ ข้อมูลบนเว็บและสื่อสังคม (web and social media data), ข้อมูลระหว่างเครื่องจักรกับเครื่องจักร (machine-to-machine data), ข้อมูลรายการที่เกิดจากการทำธุรกรรม (big transaction data), ข้อมูลไบโอเมทริกซ์ (biometrics data) ข้อมูลชีวภาพหรือกายภาพของผู้นั้น, ข้อมูลที่เกิดจากมนุษย์ (human-generated data) เช่นการติดต่อสื่อสารอุปกรณ์ไร้สาย อีกทั้งผู้รายงานได้กล่าวถึงความรู้เบื้องต้นเกี่ยวกับฐานข้อมูลที่ทำงานในหน่วยความจำสำหรับระบบงานประยุกต์เชิงวิเคราะห์ อีกทั้งภาพรวมการจำลองข้อมูลของ SAP HANA และมีตัวอย่างการใช้งาน เช่น การนำ Big Data ไปใช้ประโยชน์ด้านการแพทย์และสาธารณสุข โดยจากกรณีศึกษาพบว่า ข้อมูลทางการแพทย์และสาธารณสุขเป็นข้อมูลอีกแหล่งที่มีปริมาณมากหรือเป็น Big Data ที่มีความสำคัญต่อการช่วยชีวิตผู้ป่วย โดย Big Data ดังกล่าวต้องได้รับการจัดการให้สามารถนำไปใช้ประโยชน์ได้อย่างเต็มที่ ซึ่งมีตัวอย่างที่เห็นได้ชัด เช่นการรักษาผู้ป่วยโรคหัวใจเรื้อรังนั้น แพทย์ต้องหมั่นตรวจสอบการเต้นของหัวใจโดยอ่านจากข้อมูลการตรวจ EKG (Electrocardiogram) ที่พิมพ์ออกมาและหาข้อมูลที่มีความผิดปกติในกราฟที่พิมพ์ออกมา ซึ่งข้อมูลกราฟที่พิมพ์ออกมาหากมีการพิมพ์อย่างต่อเนื่อง 10 ชั่วโมงความยาวของกระดาษจะยาวประมาณ 2 ไมล์ แต่ในความจริงแล้วต้องการแค่เฉพาะข้อมูลที่ผิดปกติเท่านั้น ทำให้มีนักวิจัยจากสถาบัน MIT ได้พัฒนาแบบจำลองคอมพิวเตอร์เพื่อเป็นการวิเคราะห์สิ่งเหล่านี้ โดยใช้เทคนิ์การทำเหมืองข้อมูล และการเรียนรู้ของเครื่องจักร (Machine learning) ในการวิเคราะห์กลั่นกรองข้อมูลที่ไม่ได้แสดงความผิดปกติของหัวใจออกไป และผลการวิจัยพบว่า ข้อมูลที่ผิดปกติมีด้วยกัน 3 กรณีที่จะนำไปสู่ความเสี่ยงสูงของการเกิดโรคหัวใจเฉียบพลันของผู้ป่วยในช่วงหนึ่งปีข้างหน้า อีกทั้งสามารถช่วยเพิ่มความสามารถให้แก่แพทย์ในการวินิจฉัยข้อมูลของผู้ป่วยโรคหัวใจมากขึ้นและลดความผิดพลาดในการอ่านข้อมูลลดลงได้ถึงร้อยละ 70 นอกจากนี้แล้วแบบจำลองดังกล่าวยังสามารถคัดกรองผู้ป่วยที่มีประมาณร้อยละ5 ที่จำเป็นต้องใช้เครื่องกระตุ้นหัวใจฝังในร่างกายได้ด้วย

พนิดา ตันศิริ (2556) ได้ทำการศึกษาเรื่อง ข้อมูลขนาดใหญ่กับความท้าทาย โดยมีวัตถุประสงค์เพื่อแก้ปัญหาและการจัดการข้อมูลขนาดใหญ่ให้เกิดประโยชน์ต่อการนำไปใช้สร้างโอกาสของธุรกิจ รวมทั้งการลดความเสี่ยงในการทำงานได้อย่างมีประสิทธิภาพ โดยผู้วิจัยได้ทำการศึกษาการได้มาของข้อมูล Big Data จากงานวิจัยและบริษัทต่างๆ รวมถึงศึกษาบทบาทของผู้บริหารในการจัดการกับ Big Data จากสิ่งพิมพ์และบทความที่ได้มาจากการสำรวจความคิดเห็นของ IDC Digital Universe รวมถึงประโยชน์จากการใช้ Big Data จากการประชุมเพื่อสร้างความร่วมมือในการวิจัย Big Data กระทรวงการศึกษา วัฒนธรรม การกีฬา วิทยาศาสตร์และ

เทคโนโลยี (MEXT) จากประเทศญี่ปุ่น และ National Science Foundation (NSF) ประเทศสหรัฐอเมริกา และมีการศึกษาตัวอย่างผลิตภัณฑ์และเทคโนโลยีที่ใช้จัดการ Big Data ของบริษัท Oracle, Dell, IBM, SAP และ SAS เป็นต้น จากการศึกษาพบว่า ด้วยปริมาณข้อมูลที่เกิดขึ้นอย่างมหาศาลทั้งจากข้อมูลประจำวันและข้อมูลจำเครือข่ายสังคมผ่านอุปกรณ์ต่างๆทำให้หลายองค์กรกำลังเผชิญกับความท้าทายในการจัดการข้อมูลมหาศาลที่มีแนวโน้มว่าจะมากขึ้นอย่างต่อเนื่องและเพื่อให้เกิดประโยชน์ต่อการนำข้อมูลขนาดใหญ่มาใช้วิเคราะห์การดำเนินธุรกิจ จึงเป็นแรงผลักดันที่ทำให้หลายบริษัทหาแนวทางในการพัฒนาผลิตภัณฑ์โดยการนำเทคโนโลยีและนวัตกรรมใหม่ๆ มาใช้วิเคราะห์ข้อมูลขนาดใหญ่ เพื่อนำผลจากการวิเคราะห์ที่ได้มาปรับปรุงและวางแผนการทำงานของธุรกิจ ตลอดจนเพื่อลดความเสี่ยงในการทำงานและลดค่าใช้จ่ายที่สูงในการจัดซื้อและการปรับปรุงฮาร์ดแวร์ที่จำเป็นต่อการรองรับข้อมูลจำนวนมาก การจัดการและการวิเคราะห์ข้อมูลขนาดใหญ่หรือ Big Data จะทำให้เกิดโอกาสในการดำเนินธุรกิจที่สอดคล้องกับความเปลี่ยนแปลงของข้อมูลที่มีอย่างทันทีและตลอดเวลาได้อย่างมีประสิทธิภาพ

2.3.2 งานวิจัยที่เกี่ยวข้องในต่างประเทศ

Bohdan Stryk (2015) ได้ทำการศึกษาเรื่องการเตรียมความพร้อมในการจัดการองค์กรและข้อมูลขนาดใหญ่เพื่อให้องค์กรบรรลุเป้าหมาย กรณีศึกษา Delphi (How do organizations prepare and clean big data to achieve better data better data governance, a Delphi study) โดยมีวัตถุประสงค์เพื่อสร้างประสบการณ์และความพร้อมให้กับส่วนงาน IT และเพื่อให้องค์กรใช้ข้อมูลจาก Big Data อย่างมีประสิทธิภาพ ทั้งข้อมูลที่มีโครงสร้างและไม่มีโครงสร้างด้วยการจัดการกับข้อมูลที่มีอยู่ให้พร้อมสำหรับทำการวิเคราะห์ โดยการเก็บข้อมูลจากการสัมภาษณ์แบบกลุ่ม ด้วยลำดับคำถาม 3 ชุดคำถาม โดยรอบที่ 1 มีคำถามทั้งหมด 25 ข้อ เรื่องทั่วไปเกี่ยวกับ Big Data และรอบ 2, 3 เป็นคำถามที่เกี่ยวข้องกับประสบการณ์และการทำความสะอาดข้อมูล Big Data รวมถึงการจัดการกับข้อมูลตั้งแต่การ Backup, การเข้าถึงข้อมูล, ความปลอดภัยของข้อมูล รวมถึงการเชื่อถือได้

ในข้อมูล จากกลุ่มตัวอย่าง 15 คน ที่มีประสบการณ์ทางด้าน IT และมีประสบการณ์การทำงานอย่างน้อย 5 ปีทางด้าน IT และอย่างน้อย 2 ปีทางด้าน Big Data ผลการศึกษาพบว่า รอบที่ 1 กลุ่มตัวอย่างส่วนใหญ่ได้ให้ความสำคัญเกี่ยวกับเรื่องความปลอดภัยของการจัดเก็บข้อมูล Big Data, กระบวนการและการเข้าถึงข้อมูล นอกจากนี้ประมาณ 50% ของกลุ่มตัวอย่าง ได้ให้ความเห็นว่า Cloud มีความจำเป็นในการเก็บข้อมูลขนาดใหญ่ แต่ไม่มีความปลอดภัยมากนัก อีกทั้งยังให้ความเห็นที่ตรงกันว่าเป้าหมายสูงสุดของการทำความสะอาดข้อมูลขนาดใหญ่เพื่อสามารถนำมาวิเคราะห์ข้อมูลและสร้างรายได้ให้กับองค์กรได้ ส่วนผลการศึกษาในคำถามรอบที่ 2 และ 3 ที่มีการถามเกี่ยวกับประสบการณ์ในการทำความสะอาดข้อมูลและจัดการข้อมูล สรุปว่า จากการสอบถามส่วนใหญ่แล้วให้ความสำคัญกับการจัดการและความสะอาดของข้อมูลเนื่องจากสามารถสร้างกำไรในอนาคตให้กับธุรกิจ แต่ที่อย่างไรก็ตามการเก็บข้อมูลขนาดใหญ่บน Cloud Storage ไม่มีความปลอดภัย ความลับของข้อมูล และการดูแลรักษาข้อมูลเป็นอย่างมาก จึงมีความกังวลกับเรื่องการหางบประมาณในการจัดเก็บข้อมูลมากกว่าการประมวลผลของข้อมูล นอกจากนี้ยังมีความกังวลเกี่ยวกับการขาดความรู้ในการทำงานกับข้อมูลขนาดใหญ่ และความหลากหลายของSoftware และ Hardware ที่จะนำมาใช้ช่วยในการทำความสะอาดข้อมูล เป็นต้น

Stephanie F. Hood-Clark (2016) ได้ศึกษาอิทธิพลต่อการใช้และพฤติกรรมในการใช้ข้อมูล Big Data โดยมีวัตถุประสงค์เพื่อตรวจสอบปัจจัยที่เกี่ยวข้องกับการใช้ข้อมูล Big Data งานวิจัยนี้เป็นงานวิจัยเชิงปริมาณ โดยการเก็บข้อมูลผ่านแบบสอบถามจากกลุ่มตัวอย่าง 120 คน และเป็นแบบสอบถามจำนวน 111 ตัวอย่าง ซึ่งประกอบด้วย 2 vice presidents IT, 11 IT director, 29 IT managers, 12 programmer, 11 IT supervisors , 43 manager และด้านอื่นๆ ผลการศึกษาพบว่ากลุ่มตัวอย่างมีความสัมพันธ์กับการใช้งานข้อมูลขนาดใหญ่ รวมถึงกลุ่มตัวอย่างมองว่าข้อมูล Big Data มีความความสำคัญกับผู้บริหารและการตัดสินใจในองค์กร นอกจากนี้ยังสรุปได้ว่าทัศนคติในของผู้ใช้งานBig Data มีความสำคัญในการทำงาน อีกทั้งประโยชน์ในการใช้งาน ทัศนคติและผู้มีอำนาจในการตัดสินใจในองค์กรมีความสัมพันธ์กันในการใช้ข้อมูล Big Data รวมถึงการที่รับรู้ถึงประโยชน์ในการใช้ Big Data มีความสัมพันธ์กับความตั้งใจในการนำBig Data เข้ามาใช้งาน แต่อย่างไรก็ตามการใช้Big Dataนั้นมีความซับซ้อนและไม่ง่ายที่จะนำมาใช้ในองค์กร

Pouria Pirzadeh (2015) ได้ศึกษางานวิจัยเรื่อง การประเมินผลการปฏิบัติงานฐานข้อมูลขนาดใหญ่ (On the Performance Evaluation of Big Data Systems) โดยมีวัตถุประสงค์เพื่อให้เข้าใจความแตกต่างทางการปฏิบัติงานและเพื่อการประมวลผลในการวิเคราะห์จากการจัดเก็บข้อมูล นอกจากนี้ยังเพื่อมีส่วนร่วมในพื้นที่การปฏิบัติงานข้อมูล Big Data ในมุมมองที่แตกต่างโดยแบ่งลักษณะของการทำงานวิจัยออกเป็น 2 ส่วนหลักคือ ข้อมูลการเปรียบเทียบการให้บริการของระบบ

การทำงานบนCloud และ การประมวลผลคำสั่งและการดึงข้อมูลตามลักษณะของข้อมูลที่มีความแตกต่างกัน จากผลงานวิจัยพบว่า สถาปัตยกรรมและการออกแบบการตัดสินใจของระบบ Big Data ที่มีการเติบโตอย่างรวดเร็วได้สร้างความท้าทายที่สำคัญในการสร้างเกณฑ์มาตรฐานสำหรับการประเมินและเปรียบเทียบแพลตฟอร์มของระบบข้อมูลที่มีขนาดใหญ่ โดยการประเมินนั้นต้องมีความครอบคลุม และเพียงพอในแง่มุมต่างๆ ทั้งในแง่ของลักษณะของข้อมูลและปริมาณข้อมูลที่จะใช้ทำงาน เพื่อทำให้ทราบถึงตัวชี้วัดประสิทธิภาพในการทำงานและสิ่งที่ต้องแก้ไขเพื่อบรรลุเป้าหมายที่กำหนดไว้

Ylli Sadikaj (2016) ได้ทำการศึกษาเกี่ยวกับบริการประกันสุขภาพส่วนบุคคลโดยใช้ Big Data (PERSONALIZED HEALTH INSURANCE SERVICES USING BIG DATA) โดยมีวัตุประสงค์เพื่อลดจำนวนขั้นตอนการเปรียบเทียบแผนประกันและช่วยจัดแผนประกันให้เหมาะสมกับบุคคล ซึ่งในการศึกษาครั้งนี้ใช้ซอฟต์แวร์ในการรวบรวมข้อมูลผ่านหน้าเว็บไซต์ของบริษัทประกันโดยดึงข้อมูลผู้ให้บริการ แผนประกันสุขภาพต่างๆ ซึ่งเป็นข้อมูลที่มีจำนวนมากและไม่มีโครงสร้าง จากผลงานวิจัยพบว่า cloud framework ที่ช่วยในการจัดเก็บข้อมูลขนาดใหญ่และช่วยไปสู่การประมวลผล อีกทั้งยังช่วยให้ผู้ใช้งานระบุแผนประกันสุขภาพตามเกณฑ์ที่ต้องการได้ทั้งในด้านของความคุ้มครองในแผนประกันและค่าใช้จ่าย นอกจากนี่ยังมีเทคนิคในการทำ Cluster ที่ช่วยลดระยะเวลาในการเปรียบเทียบแผนประกันด้วยการใช้เทคโนโลยีเข้าช่วยลดจำนวนกลุ่มในการเปรียบเทียบ โดยที่โปรแกรมสามารถจับคู่ความต้องการกับความคุ้มครองที่มีในแผนประกันที่มีอยู่ด้วยการจัดกลุ่มแบบ DBSCANที่มีความแม่นยำสูง ซึ่งเป็นประโยชน์ต่อนักวิจัยอื่นในการเสนอและนำแผนประกันสุขภาพได้อย่างมาก

Mechael O. Ojo (2016) ผลสรุปจากการวิเคราะห์ Big Data กรณีศึกษาความท้าทายใน Big Data Analytics (BDA) เกี่ยวกับการดำเนินงานเกี่ยวกับธุรกิจการบริการทางการเงิน โดยมีวัตถุประสงค์เพื่อจัดอันดับความท้าทายของผลกระทบที่เป็นอุปสรรคทางด้านการเงิน และเพื่อเป็นแนวทางในการสร้างกรอบการวัดผลในการดำเนินงานที่เป็นประโยชน์กับองค์กร โดยในการเลือกกลุ่มตัวอย่างของเอกสารที่จะนำมาใช้ในงานวิจัย แบ่งออกเป็น 5 กลุ่ม คือ 1.) Industry leading BDA solutions providers (e.g., IBM, CSC, HP, Dell, SAP, Teradata, Oracle, SAS, Accenture, PWC, Deloitte, McKinsey, etc.) 2.) Independent technology research and advisory companies. (e. g. Gartner,Forrester, Nielsen, IDC, etc.) 3.) Financial Services Companies and BDA solutions users in the financial services industry. 4.) Financial Services experts (and their publications. e.g., The American Banker Journal). 5.) Academia – Academic research on Big Data and Big Data Analytics ทั้งนี้รวมทั้งหมด 75 เอกสาร เพื่อหาความท้าทาย 10 ข้อ ของการดำเนินงานใน BDAของ

อุตสาหกรรมบริการทางการเงิน และหาสิ่งที่เป็นความรุนแรงหรือผลกระทบด้วยการวัดผลเชิงปริมาณในบริษัทจากแบบสอบถามสำรวจความคิดเห็น โดยผลการศึกษาคือ 77.8%ของผู้ตอบแบบสอบถามในภาคการบริการทางการเงินซึ่งประกอบด้วย 30.6%จากธนาคาร, 27.8%จากการรับประกันภัย และ 19.4% จากการให้บริการทางการเงินอื่นๆ โดย 11.1% ได้ให้ความเห็นว่าสิ่งที่เป็นปัจจัยหลักแห่งความสำเร็จหรือความล้มเหลวของ Big Data Analytics แบ่งออกป็น 4 กลุ่มหลัก คือ กลยุทธ์, กระบวนการ, คน และเทคโนโลยี จากการวิเคราะห์จะเห็นได้ว่า กลยุทธ์และคนเป็นปัจจัยสูงสุดที่เป็นความท้าทายในการดำเนินงานตามลำดับ ตามด้วยกระบวนการ และเทคโนโลยีเป็นอันดับสุดท้ายซึ่งมีความขัดแย้งกับความคิดเห็นของเจ้าของงานวิจัยอย่างน่าประหลาดใจที่ผลสรุปของงานวิจัยนั้นไม่ได้ให้ความสำคัญกับเทคโนโลยีมากนั้นทั้งๆที่เป็นยุคของ Big Data จึงอาจสรุปได้ว่าปัจจัยสำคัญก่อนที่จะมีการใช้เทคโนโลยี Big Data มาใช้วิเคราะห์ในองค์กรนั้น ควรจะคำนึงกระบวนการในการทำงานและทรัพยากรบุคคลเป็นอันดับแรกก่อน

ZHENNING (JIMMY) XU (2016) ได้ทำการศึกษาเรื่อง 3 เรื่องเกี่ยวกับ Analytics ข้อมูลขนาดใหญ่, การตลาดแบบดั้งเดิม การวิเคราะห์การค้นพบความรู้ใหม่และประสิทธิภาพของผลิตภัณฑ์ โดยมีวัตถุประสงค์เพื่อให้นักวิจัยและผู้ปฏิบัติงานวางแผนสำหรับ Big Data Analytics และTraditional Marketing Analytics ในอนาคต เพื่อค้นหาความรู้ใหม่ๆในการพัฒนาผลิตภัณฑ์ โดยมีกระบวนการในการศึกษาจากกลุ่มตัวอย่าง จากผู้จัดการบริษัทของอเมริกาและญี่ปุ่น โดยเน้นผู้ตอบแบบสอบถามที่เป็นหน่วยงานการตลาด, ผลิตภัณฑ์, แบรนด์ และวิจัยทางการตลาด จากการสำหรับด้วยแบบสอบถาม 270 ตัวอย่าง จากผลการศึกษาพบว่า แม้ว่าบริษัทจะมีการลงทุนกับการวิเคราะห์ข้อมูลขนาดใหญ่มากขึ้น แต่เครื่องมือใหม่ๆที่ใช้ในการวิเคราะห์นั้นยังคงไม่ได้มีความเข้าใจกันอย่างท่องแท้และแพร่หลาย ดังนั้นผู้ปฏิบัติการเกี่ยวกับการตลาดยังคงตระหนักถึงความสามารถของข้อมูลขนาดใหญ่และความรู้ความสามารถของผู้ใช้งานและวิเคราะห์ อย่างไรก็ตามทางบริษัทให้ความสำคัญและทราบถึงการนำ Big Data มาวิเคราะห์เพื่อค้นพบความรู้และผลิตภัณฑ์ใหม่ๆแต่บริษัทอาจจะยังไม่มีเป้าหมายชัดเจนว่าจะนำ Big Data มาใช้ในส่วนใด นอกจากนี้วัฒนธรรมองค์กรมีผลอย่างมากในการทำการตลาดโดยการนำ Big Data เข้ามาใช้ในงานวิเคราะห์

จากการศึกษางานวิจัยที่เกี่ยวข้องทั้งในและต่างประเทศพบว่า การวิเคราะห์และใช้ Big data สามารถนำมาวิเคราะห์ธุรกิจ และวิเคราะห์พฤติกรรมลูกค้าเพื่อตอบสนองความต้องการของลูกค้าได้ แต่ทั้งนี้ข้อมูลที่ถูกจัดเก็บไว้เป็น Big Dataนั้น เป็นข้อมูลที่ไม่มีโครงสร้างที่ชัดเจน จึงต้องใช้เวลาทำความสะอาดข้อมูลและจัดข้อมูลให้เป็นโครงสร้างเพื่อพร้อมที่จะนำไปใช้ในการประมวลผลและวิเคราะห์ข้อมูล

โดยเทคโนโลยีที่นำมาใช้ในการวิเคราะห์ Big Data เช่น Hadoop มีความซับซ้อน และต้องใช้เทคโนโลยีหรือโปรมแกรมอื่นๆร่วมในการวิเคราะห์ด้วย อีกทั้ง กระบวนการของแต่ละบริษัท และทรัพยากรบุคคล รวมถึงความรู้ความสามารถของผู้ใช้งาน Big Dataนั้นมีน้อย เป็นสิ่งที่น่ากังวลใจ จึงเป็นปัจจัยหลักร่วมกับการนำเทคโนโล Big Dataมาใช้ในองค์กร แต่อย่างไรก็ตาม บริษัทต่างๆให้ความสนใจใน Big Dataมาก และมีการลงทุนกับเทคโนโลยีเหล่านี้เพื่อนำมาใช้ในการวิเคราะห์ แต่บางบริษัทอาจจะไม่มีความเข้าใจและไม่มีเป้าหมายว่าจะนำ Big Data ไปวิเคราะห์ในส่วนใดก่อน ซึ่งผู้วิจัยจะนำประเด็นที่ได้เหล่านี้ไปใช้ในการออกแบบเครื่องมือที่ใช้ในงานวิจัยต่อไป

บทที่ 3
วิธีการดำเนินการวิจัย

การศึกษางานวิจัยแนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data ซึ่งเป็นกรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุ เป็นงานวิจัยแบบกึ่งทดลอง (Quasi - experimental research design) โดยผู้วิจัยได้มีการนำเสนองานวิจัยเป็นส่วนๆ ดังนี้

3.1 กรอบขั้นตอนการวิจัย

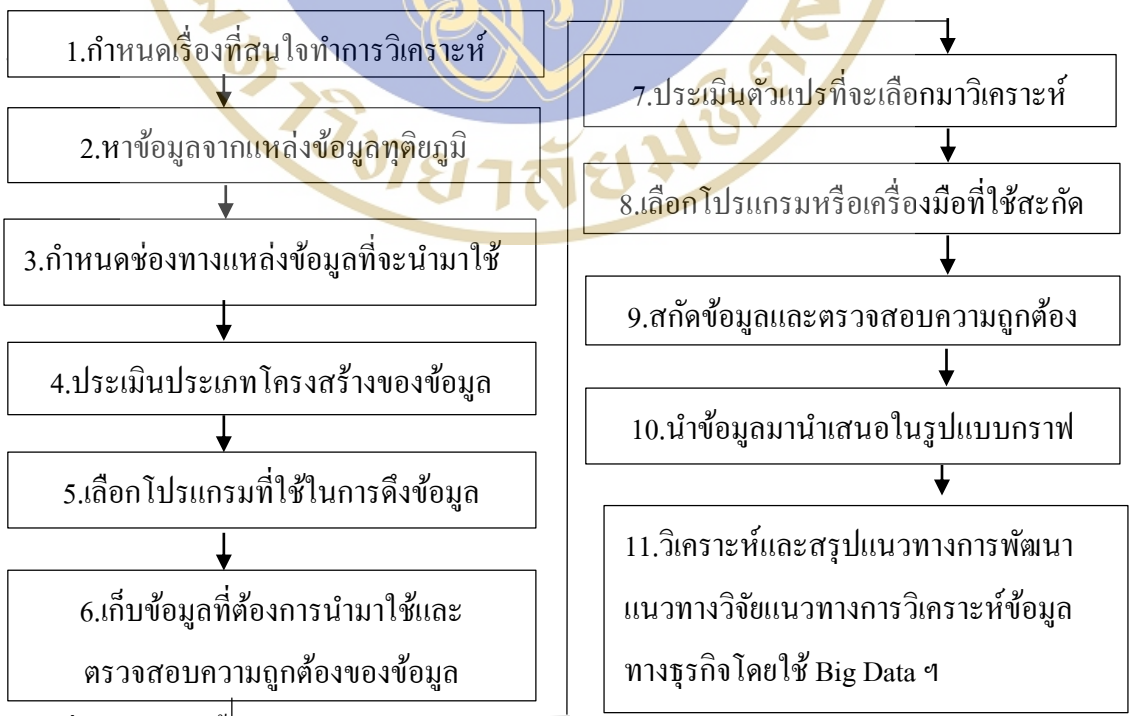
3.2 กลุ่มเป้าหมาย

3.3 เครื่องมือที่ใช้ในการทำงานวิจัย

3.4 การเก็บรวบรวมข้อมูล

3.5 การวิเคราะห์ข้อมูล

3.6 การแสดงผลข้อมูลและสถิติที่ใช้ในการวิเคราะห์ข้อมูล



ภาพที่ 3.1 : กรอบขั้นตอนการวิจัย

ที่มา : ผู้วิจัย

ผู้วิจัยได้กำหนดกรอบขั้นตอนในการวิจัยพัฒนาไว้ 11 ขั้นตอน ดังนี้

1. กำหนดเรื่องที่สนใจในการวิเคราะห์เป็นจุดเริ่มต้นในการหาข้อมูลให้ตรงตามวัตถุประสงค์ โดยงานวิจัยนี้ได้กำหนดเกี่ยวกับข้อมูลอุบัติเหตุที่อยู่ในทวิตเตอร์เพื่อนำมาจำลองการวิเคราะห์ข้อมูลของเทคโนโลยี Big Data

2. ศึกษาข้อมูลที่เกี่ยวข้องกับงานวิจัย Big Dataจากแหล่งข้อมูลทุติยภูมิ เช่น วารสาร งานวิจัยต่างๆ หนังสือ บทความ สัมมนาต่างๆ บทสัมภาษณ์ผู้บริหาร สำหรับประเทศไทย แหล่งข้อมูลที่เกี่ยวกับ Big Data นั้นสามารถหาอ่านได้จากสื่อออนไลน์ เช่น Facebook Big Data Thailand, Facebook Big Data Experience Center, Facebook Big Data Analytics by True, Facebook Data Science Thailand เป็นต้น ที่มีการเขียนบทความทั้งภาษาไทย และแบ่งปันข้อมูลที่น่าสนใจจากเว็บไซต์ต่างประเทศมาไว้ที่ Facebook เหล่านี้ นอกจากนี้หากสนใจศึกษาข้อมูลในเชิงลึกที่เป็นภาษาอังกฤษ Big Data Analytics by True (2559) กล่าวว่าสามารถเรียนออนไลน์ หรือดาวน์โหลดข้อมูลที่เป็นหนังสือต่างๆ ซึ่งได้รวบรวมข้อมูลจากมหาวิทยาลัยชั้นนำทั่วโลก ได้จากเว็บไซต์ Datascience.community รวมถึงศึกษาการนำข้อมูลออกจากทวิตเตอร์โดยใช้โปรแกรม R จาก r-bloggers (Leonardo Toglia, 2016) ที่นำมาใช้ในการศึกษาข้อมูลงานวิจัยครั้งนี้

3. กำหนดช่องทางของแหล่งข้อมูลที่จะนำมาใช้ในการวิเคราะห์ข้อมูลผ่านระบบของ Big Data สำหรับกรณีศึกษาในงานวิจัยนี้ใช้ข้อมูลการเกิดอุบัติเหตุที่ถูกทวีตผ่านช่องทางทวิตเตอร์

4. ประเมินข้อมูลที่สนใจศึกษาจากแหล่งข้อมูลของทวิตเตอร์ โดยข้อมูลของทวิตเตอร์นั้นเป็นข้อมูลที่ไม่มีโครงสร้างซึ่งอยู่ในรูปของข้อความทวีตเรื่องการเกิดอุบัติเหตุในเขตกรุงเทพมหานคร ทั้งนี้จะใช้เวลามากประมาณ 80%ของเวลาทั้งหมดในการจัดรูปแบบแยกข้อมูลเพื่อเตรียมไปใช้ในการวิเคราะห์ (สัมมนาLearning Big Data : Hadoop & Spark, 2559)

5. โปรแกรมที่จะนำมาใช้ดึงข้อมูลจากทวิตเตอร์ สามารถดึงข้อมูลผ่าน Application Programming Interface (API) หรือคือช่องทางการเชื่อมต่อของระบบ (MeeWebFree, 2010) ที่ทางทวิตเตอร์มีให้ใช้ฟรี โดยที่ทางผู้ใช้งานสามารถไปสมัครที่ dew.twitter.com เพื่อได้ชื่อและรหัสในการเข้าถึงข้อมูล ทั้งนี้ในงานวิจัยนี้ได้ใช้ R Program ในการดึงข้อมูลร่วมกับ API ของ Twitter

6. เก็บรวบรวมข้อมูลที่ต้องการนำมาใช้ผ่านการดึงข้อมูลจากโปรแกรมข้างต้น ตามช่องทางที่ได้กำหนดไว้ โดยทำการดึงมา 100 ชุดข้อมูล เพื่อนำมาตรวจสอบเบื้องต้นว่าข้อมูลที่เก็บมานั้นตรงกับความต้องการหรือไม่ เมื่อตรวจสอบความถูกต้องแล้วจึงดึงข้อมูลทั้งหมดออกมาตามข้อจำกัดของทวิตเตอร์ที่ให้ดึงข้อมูลได้ไม่เกิน 10 วัน และ 15ครั้ง ทุกๆ 15 นาที (API Rate Limit, n.d.)

7. จากข้อมูลที่ดึงออกมาทั้งหมดนั้น จะประกอบด้วยข้อมูลมากมายทั้งที่จำเป็นเป็นต้องใช้ ดังนั้นจึงทำการประเมินข้อมูลทั้งหมดว่าควรนำข้อมูลหรือตัวแปรใดในข้อมูลมาวิเคราะห์บ้าง สำหรับข้อมูลที่ใช้ในงานวิจัยในครั้งนี้ ข้อมูลที่จำเป็นในการวิเคราะห์ ได้แก่ วันที่ สถานที่และเวลาที่เกิดอุบัติเหตุ เวลาที่ใช้ในขณะทวีตในทวิตเตอร์ ซึ่งจำกัดพื้นที่ในเขตกรุงเทพมหานครเท่านั้น โดยมีการสกัดข้อมูลที่เป็นข้อมูลที่เกิดจากการทวีต ออกมาแยกโครงสร้างข้อมูลให้ชัดเจน เพื่อง่ายต่อการวิเคราะห์ อีกทั้งทำการหาข้อมูลจากแหล่งอื่นๆเพิ่มเติมนอกจากในทวิตเตอร์ ได้แก่ ชื่อเขต แขวง และพิกัดต่างๆ ของเขต กรุงเทพฯ จาก เว็บไซต์ Data.go.th เพื่อนำมาจับคู่พิกัดสถานที่ ในการแสดงผลข้อมูล เป็นต้น

8. โปรแกรมหรือเครื่องมือที่ใช้ในการสกัดข้อมูลนั้น จะใช้ Package sparklyr ซึ่งอยู่ภายใต้ RStudio เขียนโดยภาษา R โดยใช้คอมพิวเตอร์โน๊คบุ๊คที่มี RAM 16 GB, ระบบ 64 bit, หน่วยความจำ 100 GB และ Processor core i7

9. สกัดข้อมูลโดยคัดกรองข้อมูลที่จำเป็นหรือที่ต้องการออกมาเป็นโครงสร้างเพื่อง่ายต่อการเตรียมการวิเคราะห์ข้อมูล รวมถึงตรวจสอบข้อมูลที่กรองหรือสกัดออกมาว่าถูกต้องตามข้อมูลเดิม และตรงตามความต้องการของผู้วิจัยหรือไม่ เนื่องจากการสกัดข้อมูลนั้นต้องเขียนโปรแกรมในการสกัดข้อมูลออกมาโดยทำงานร่วมกับโปรแกรมเมอร์ ซึ่งจากประสบการณ์การทำงานของผู้วิจัยนั้น โปรแกรมเมอร์ส่วนใหญ่อาจจะไม่มีความเข้าใจข้อมูลหรือธุรกิจที่เกี่ยวข้องมาก นักอาจจะเข้าใจความต้องการของ ผู้ใช้งานผิดพลาดกันได้

10. นำข้อมูลที่ได้มาวิเคราะห์และเสนอเป็นกราฟ เพื่อง่ายต่อการเข้าใจ ซึ่งสามารถใช้โปรแกรม R ในการวิเคราะห์และนำเสนอข้อมูลด้วยกราฟต่างๆ

11. วิเคราะห์และสรุปแนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data กรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุ เพื่อประโยชน์ในด้านการนำไปใช้กับธุรกิจและการพัฒนาในด้านการวิเคราะห์ข้อมูลด้วยเทคโนโลยีใหม่ๆได้อย่างมีประสิทธิภาพ

3.2 กลุ่มเป้าหมายที่ใช้ในงานวิจัย

สำหรับกลุ่มเป้าหมายที่ใช้ในงานวิจัยครั้งนี้ ได้แก่ ผู้ที่ใช้ทวิตเตอร์และมีการโพสต์ข้อความการเกิดอุบัติเหตุในทวิตเตอร์ อันได้แก่ จราจรและอุบัติเหตุ (@wichansuriyo), สวพ. FM91(@fm91trafficpro), ศูนย์อุบัติเหตุ (@MOT_1356), Traffy. in. th (@traffy), JS100 (@js100radio), เพื่อนเดินทาง (@travel_friendss) และทวีตอยู่ในเขตกรุงเทพมหานคร

3.3 เครื่องมือที่ใช้ในงานวิจัย

ในงานวิจัยครั้งนี้ผู้วิจัยได้ออกแบบเครื่องมือที่ใช้ในงานวิจัยออกเป็น 3 ส่วน ได้แก่

ส่วนที่ 1 เครื่องมือในการทำ Big Data ประกอบด้วย

1.1 Notebook

1.2 ระบบ Ubuntu

1.3 ระบบ Hadoop

1.4 เครื่องมือวิเคราะห์Apache Spark

1.5 โปรแกรมRStudio และภาษา R

ส่วนที่ 2 แบบตรวจสอบรายการข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์ประกอบไปด้วย

ตารางที่ 3.1 แบบตรวจสอบรายการ (Check List)

แบบตรวจสอบรายการ (Check List)	ใช่	ไม่ใช่	หมายเหตุ
กำหนดช่องทางของแหล่งข้อมูลที่จะนำมาใช้			
ช่องทางของแหล่งข้อมูลมาจากทวิตเตอร์			
ช่องทางของแหล่งข้อมูลมาจากเฟซบุ๊ค			
ช่องทางของแหล่งข้อมูลมาจากเว็บไซต์			
ช่องทางของแหล่งข้อมูลมาจากแหล่งอื่นๆ (โปรดระบุที่หมายเหตุ)			
ประเภทโครงสร้างของข้อมูล			
ข้อมูลที่นำมาวิเคราะห์ประเภทมีโครงสร้าง เช่น ตารางแบ่งแยกข้อมูลออกเป็นColumn ต่างๆอย่างชัดเจน			
ข้อมูลที่นำมาวิเคราะห์ประเภทไม่มีโครงสร้าง เช่น ไฟล์ข้อความ, ภาพ, เสียง เป็นต้น			
รายละเอียดข้อมูลที่นำมาวิเคราะห์			
ข้อมูลมีรายละเอียดสถานที่เกิดเหตุ			
ข้อมูลมีรายละเอียดเวลาที่ทำการทวีตการเกิดเหตุ			
ข้อมูลมีรายละเอียดประเภทของรถที่เกิดเหตุ			

ที่มา : ผู้วิจัย

ส่วนที่ 3 เปรียบเทียบระหว่างรูปแบบการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data กับ ข้อมูลที่ไม่ได้ใช้ Big Data

3.4 การเก็บรวบรวมข้อมูล

งานวิจัยในครั้งนี้ได้มีวิธีการในการเก็บรวบรวมข้อมูล ดังนี้

ดำเนินการรวบรวมและศึกษาข้อมูลจากแหล่งข้อมูลทุติยภูมิ (Secondary Data) ได้แก่

- เอกสารบทความ สิ่งตีพิมพ์ งานสัมมนา วิทยานิพนธ์ สารนิพนธ์ รวมไปถึงข้อมูลจากอินเตอร์เน็ตที่เกี่ยวข้องเพื่อเป็นข้อมูลพื้นฐานสำหรับงานวิจัย

- ข้อมูลการเกิดอุบัติเหตุที่อยู่ในทวิตเตอร์ ในเขตกรุงเทพมหานครและดำเนินการลงโปรแกรม สกัดข้อมูล และวิเคราะห์ข้อมูลที่เกี่ยวข้อง โดยใช้เวลาในเดือนธันวาคม พ.ศ. 2559 ถึงมีนาคม พ.ศ. 2560

3.5 การวิเคราะห์ข้อมูล

หลังจากที่ผู้วิจัยได้ทำการเก็บรวบรวมข้อมูลจากแหล่งต่างๆข้างต้น ทั้งจากทวิตเตอร์และงานสัมมนาดังกล่าวจะถูกนำมาตรวจสอบและวิเคราะห์ผล ดังนี้

1. การตรวจสอบข้อมูล (Editing) ผู้วิจัยทำการตรวจสอบถึงความสมบูรณ์ของข้อมูลที่ได้รับและทำการคัดแยกข้อมูลที่ไม่สมบูรณ์ออกไป

2. การวิเคราะห์ข้อมูลที่ได้จากการเก็บข้อมูลทวีตการเกิดอุบัติเหตุ ตามทวิตเตอร์ที่กำหนด ด้วยการใช้ เทคโนโลยี Spark และ โปรแกรม R

3.6 การแสดงผลข้อมูลและสถิติที่ใช้ในการวิเคราะห์ข้อมูล

งานวิจัยครั้งนี้ได้ใช้การแสดงผลข้อมูล โดยใช้กราฟต่างๆจากวิธีการทางสถิติเชิงพรรณนา (Descriptive Statistic) ที่ทำขึ้นด้วย โปรแกรม R ได้แก่ ฐานนิยม (Mode) ซึ่งนำมาแสดงข้อมูล เวลา, สถานที่ และประเภทของรถที่มีการเกิดอุบัติเหตุมากที่สุด

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การศึกษาวิจัยเรื่อง “แนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data กรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุ” เป็นการวิจัยโดยเก็บข้อมูลจากแหล่งข้อมูลทุติยภูมิ (Secondary Data) จากข้อความที่ทำการทวีตในทวิตเตอร์ โดยการวิเคราะห์ข้อมูลได้ผลการวิจัยแบ่งออกเป็น 3 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 ผลการวิเคราะห์เครื่องมือในการทำ Big Data

ขั้นตอนที่ 2 ผลการวิเคราะห์แบบตรวจสอบรายการข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์

ขั้นตอนที่ 3 ผลการวิเคราะห์เปรียบเทียบระหว่าง รูปแบบข้อมูลที่ได้จากการทำ Big Data กับผลที่ได้จากข้อมูลของสำนักงานตำรวจแห่งชาติ

ขั้นตอนที่ 1 ผลการวิเคราะห์เครื่องมือในการทำ Big Data

ข้อมูล Big Data สามารถใช้เครื่องมือต่างๆได้หลายเครื่องมือขึ้นอยู่กับข้อมูลและเป้าหมายงานที่ต้องการวิเคราะห์ อีกทั้งขึ้นอยู่กับความถนัดของการเลือกใช้โปรแกรมที่เหมาะสมกับผู้วิจัยและเหมาะสมกับข้อมูลที่นำมาใช้ในงานวิจัย ซึ่งงานวิจัยนี้ได้ศึกษาการใช้เครื่องมือต่างๆที่ใช้ในการทำ Big Data โดยมีทั้งหมด 4 เครื่องมือหลัก ดังนี้

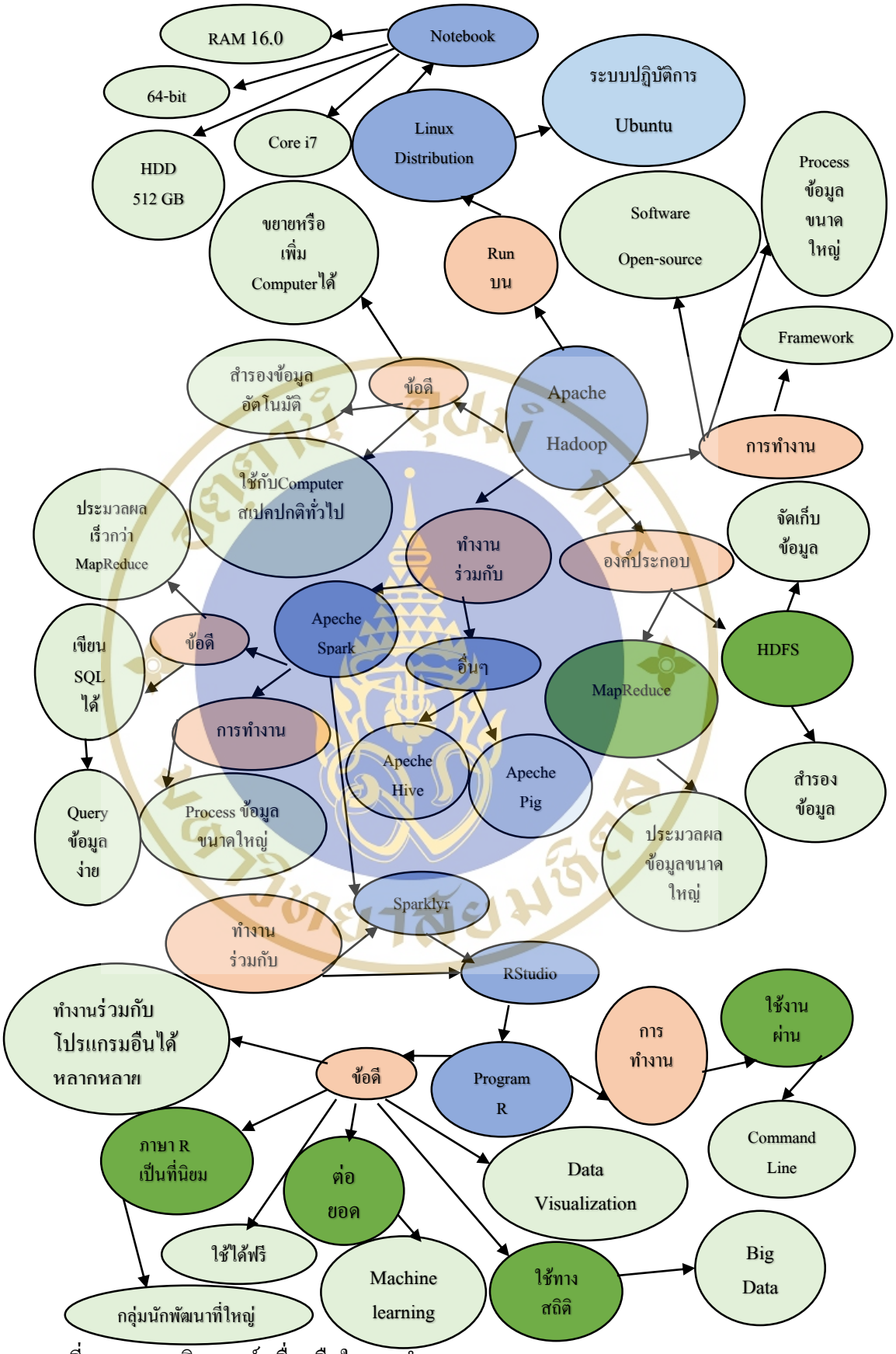
1.1 Notebook

1.2 ระบบ Ubuntu

1.3 Software Apache Hadoop

1.4 โปรแกรม Rstudio และภาษา R

จาก 4 เครื่องมือหลักที่กล่าวข้างต้น ผู้วิจัยได้สรุปแผนภาพ กระบวนการ (Process) การเชื่อมต่อการใช้งาน เช่น มีลักษณะการทำงานอย่างไร ต้องทำงานร่วมกับโปรแกรมใดบ้าง รวมถึงจุดเด่นหรือข้อดีของโปรแกรมดังกล่าว สามารถสรุปได้เป็นแผนภาพดังนี้



ภาพที่ 4.1 ผลการวิเคราะห์เครื่องมือในการทำ Big Data

ที่มา : ผู้วิจัย, 2560

จากภาพที่ 4.1 ที่ได้จากการศึกษางานวิจัยในครั้งนี้ผ่านการอ่านบทความ คอร์สสัมมนา รวมถึงการทดลองทำโปรแกรม ผู้วิจัยสรุปเป็นแผนภาพได้ว่า การใช้ Apeche Hadoop Run บน Linux Distribution ด้วยระบบปฏิบัติการ Ubuntu เนื่องจากเป็นระบบปฏิบัติการที่ฟรี ใช้ร่วมกับ Hadoop ได้ และใช้ Notebook ที่มี RAM 16.0 GB, 64 bit, HDD 512 GB และเป็น Core i7 โดยที่การทำงานเป็นลักษณะของ Software Open-source บน Framework ที่สามารถ process ข้อมูลขนาดใหญ่ได้ ส่วนข้อดีของ Hadoop นั้นจุดเด่นคือสามารถใช้กับเครื่องคอมพิวเตอร์ที่มีคุณสมบัติทั่วไป อีกทั้งสามารถเพิ่มขยายจำนวนคอมพิวเตอร์เชื่อมต่อกับระบบในอนาคต รวมถึงสามารถสำรองข้อมูลได้แบบอัตโนมัติ ทั้งนี้องค์ประกอบของ Hadoop จะแบ่งออกเป็น 2 ส่วนหลักๆคือส่วนที่ไว้จัดเก็บข้อมูลที่เรียกว่า HDFS (Hadoop Distributed File System) และส่วนที่ไว้สำหรับประมวลผล ที่เรียกว่า MapReduce อีกทั้ง Hadoop สามารถทำงานร่วมกับโปรแกรมได้อย่างหลากหลายแล้วแต่ความเหมาะสมของการวิเคราะห์ข้อมูล โดยผู้วิจัยได้ทำการติดตั้งทั้ง Apeche Hive และ Apeche Spark แต่เลือกการทำการวิจัยหลักในงานวิจัยนี้เป็นในส่วนของ Apeche Spark และSparklyr โดยเน้นการทำงานในส่วนของ Sparklyr เป็นหลัก

ในส่วนของ Apeche Spark ทางผู้วิจัยได้ติดตั้งแบบ Spark Standalone คือ สามารถทดสอบการประมวลผลจากการติดตั้งในคอมพิวเตอร์เครื่องเดียว ทั้งนี้ Spark สามารถทำงานร่วมกับ Rstudio ได้ โดยมีจุดเด่นคือ สามารถประมวลผลได้เร็วกว่า MapReduce ของ Hadoop และสามารถเขียน SQL เพื่อ Query ข้อมูลได้สะดวกและง่ายขึ้นแทนที่จะใช้แค่คำสั่งเป็น Command Line ใน RStudio ซึ่งการทำงานของ Spark นั้น สามารถประมวลผล ข้อมูลที่มีขนาดใหญ่ได้เช่นกัน โดยจะใช้ร่วมกันกับ Hadoop หรือไม่ก็ได้

ทั้งนี้สำหรับ RStudio นั้นสามารถใช้ Package ที่ชื่อว่า Sparklyr ในการเชื่อมต่อ RStudio กับ Spark ให้ทำงานร่วมกันได้ โดยที่RStudio มีโปรแกรมที่เรียกว่า ProgramR และภาษาที่ใช้คือ ภาษา R ที่ใช้งานผ่าน Command Line และมีจุดเด่นหรือข้อดี คือ สามารถทำงานร่วมกับโปรแกรมอื่นได้อย่างหลากหลาย และสามารถทำงานได้ในหลายระบบปฏิบัติการ เช่น Linux, Windows หรือ MacOS เป็นต้น อีกทั้งภาษา R เป็นที่นิยมมากในกลุ่มนักพัฒนา (Developer) ขนาดใหญ่ ที่สำคัญคือ เป็นโปรแกรมที่ใช้ได้ฟรี และสามารถต่อยอดในการทำ Machine Learning ได้ รวมถึงเป็นที่นิยมในการใช้คำนวณข้อมูลทางสถิติและนิยมใช้ร่วมในการประมวลผล Big Data นอกจากนี้ ProgramR ยังมีฟังก์ชันการใช้งานในส่วนการแสดงผล (Visualization) ที่หลากหลายและสวยงาม เช่นการทำกราฟต่างๆ หรือการทำข้อมูลลงบนแผนที่ผ่าน Package ที่หลากหลายที่ Program R มีให้ Download ใช้งานได้ฟรี

อย่างไรก็ตามในการทำข้อมูลBig Data ไม่จำเป็นต้องใช้งานครบทุกเครื่องมือ ทั้งนี้ขึ้นอยู่กับลักษณะข้อมูล ขอบเขตที่จะใช้ในการทำงานวิจัยหรือการประมวลผลข้อมูล เป็นต้น

ขั้นตอนที่ 2 ผลการวิเคราะห์แบบตรวจสอบรายการข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์

ผลการวิเคราะห์แบบตรวจสอบรายการข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์ของผู้วิจัย

ส่วนที่ 1 การกำหนดช่องทางของแหล่งข้อมูลที่จะนำมาใช้

จากข้อมูลแหล่งต่างๆที่ค้นหามาตามแหล่งที่มาด้านล่างของตาราง สรุปได้ว่าทั้งทั่วโลกและในประเทศไทยมีการใช้ Social Media Platforms ที่แตกต่างกันออกไป แต่อย่างไรก็ตาม 4 อันดับแรกที่นิยมใช้กันมาก ได้แก่ Twitter, Facebook, Instagram (Gary Vayerchuk, 2013) และรายงาน 3 อันดับแรก Facebook, Twitter และYoutube สำหรับผู้ใช้ในประเทศไทย (socialbakers, 2017) ผู้วิจัยจึงได้ทำการเปรียบเทียบเงื่อนไขการใช้งานในแต่ละ Platform ซึ่งแบ่งออกเป็น 4 จำพวกหลัก ได้แก่ การใช้เพื่อการโฆษณา, ใช้เพื่อSocial Interaction, ความรวดเร็วในการเข้าถึง อ่าน กระจาย รวมถึงการ Load ข้อมูล และ Style ของผู้ใช้งานที่แตกต่างกันไป โดยมีรายละเอียดดังนี้

ตารางที่ 4.1 แสดงช่องทางของแหล่งข้อมูล

เงื่อนไขที่มี	Twitter	Facebook	Youtube	Instagram
แชร์ข้อความ ภาพ วีดีโอ Location	แชร์ได้ทั้งหมด	แชร์ได้ทั้งหมด	VDO เท่านั้น	แชร์ได้ทั้งหมด
นิยมใช้สำหรับการตลาดและการโฆษณา	ไม่นิยมใช้	นิยมใช้	นิยมใช้	นิยมใช้
นิยมใช้เล่นเกม ดูความเคลื่อนไหวต่างๆไปเรื่อยๆ	ไม่นิยมใช้เล่นเกม	นิยมใช้	ไม่นิยมใช้เล่นเกม	ไม่นิยมใช้เล่นเกม
สำหรับผู้ที่นิยมโพสต์ภาพจำนวนมาก หรือดูภาพรวม	ไม่นิยมใช้ดูภาพรวม	ไม่นิยมใช้ดูภาพรวม	ไม่นิยมใช้ดูภาพรวม	นิยมใช้ดูภาพ/ภาพรวม
Social Interaction	ใช้ Retweet	ไม่ใช้ Retweet	ไม่ใช้ Retweet	ไม่ใช้ Retweet
นิยมใช้Hashtags (#) ในการเข้าถึงข้อมูลหรือบทสนทนา	นิยมใช้ # มานาน	เริ่มนิยมใช้#	ไม่นิยมใช้#	เริ่มนิยมใช้#

ตารางที่ 4.1 แสดงช่องทางของแหล่งข้อมูล (ต่อ)

เงื่อนไขที่มี	Twitter	Facebook	Youtube	Instagram
ใช้สำหรับบทสนทนายาวๆ	จำกัดคำ	เน้นคำยาวๆ	เน้นVDO	เน้นภาพ
ความรวดเร็วในเรื่องการแชร์ข้อมูลต่อ	เน้นแชร์รวดเร็ว	แชร์บางส่วน	แชร์บางส่วน	แชร์บางส่วน
ความเร็วในการLoad ข้อมูล	รวดเร็ว	ช้า	ช้า	ช้า
โพสต์เฉพาะข้อความเท่านั้น	เน้นข้อความสั้น	เน้นข้อความยาว	เน้นVDO	เน้นภาพ
โพสต์ข้อความสั้นกระชับได้ใจความ (เนื่องจากพิมพ์ได้เพียง 140 ตัวอักษร)	สั้นกระชับ	ยาว	เน้นVDO	เน้นภาพ
ใช้สื่อสารกันเป็นหลัก เช่น การพูดคุย (Read and broadcast)	Read, broadcast	Read, Like,Share	เน้นดู	เน้นดู
ง่ายต่อการอ่านเข้าใจในข้อความทันที	เข้าใจง่าย	อ่านนาน	เน้นVDO	เน้นภาพ
ค้นหาข้อความของใครก็ได้ที่เราไม่จำเป็นต้องติดตามคนๆนั้น	นิยม Public	นิยม Privacy	เน้นVDO	เน้นภาพ
ใช้ติดต่อกับเพื่อนและครอบครัวในชีวิตจริง	นิยม Public	นิยม Privacy	นิยม Public	นิยม Public

ที่มา : https://blog.bufferapp.com/5-points-where-you-shouldnt-confuse-twitter-with-facebook
http://www.danmarkel.com/
https://www.inc.com/magazine/201311/gary-vaynerchuk/how-to-master-the-four-major-social-media-platforms.html
http://www.puwadon.com/post.php?id=18
https://www.quora.com/What-is-the-difference-between-Twitter-and-Facebook
https://www.statista.com/statistics/284483/thailand-social-network-penetration/
https://www.thanop.com/hashtag/

จากตารางที่ 4.1 พบว่า การกำหนดช่องทางของแหล่งข้อมูลที่จะนำมาใช้นั้น แต่ละแหล่งข้อมูลจะมีรูปแบบการใช้งานที่แตกต่างกันออกไป ซึ่งกรณีนี้หากต้องการดึงข้อมูลเกี่ยวกับอุบัติเหตุ แหล่งข้อมูลจากทวิตเตอร์เป็นแหล่งข้อมูลที่มีความเหมาะสมที่สุด เนื่องจากเป็น Platform ที่สามารถโพสต์ได้ทั้งข้อความ รูปภาพ VDO และ Location และมีจุดเด่นในด้าน Social Interaction ข้อมูลที่ใช้ทวิตนั้นสั้นกระชับได้ใจความสำคัญ เนื่องจากถูกจำกัดไว้ที่ 140 ตัวอักษร นอกจากนี้ยังมีในส่วนของความเร็วในการโหลด อ่าน และกระจายข้อมูลอย่างชัดเจน เหมาะสมกับเหตุการณ์

อุบัติเหตุที่ควรเป็นข้อความที่สั้นกระชับได้จำความและให้ประเด็นสำคัญกับผู้อ่านอย่างรวดเร็ว ส่วนข้อมูลจาก Facebook เหมาะสมกับข้อความยาวๆ ใช้สื่อสารกับครอบครัวและเพื่อนๆสนิทอย่างแท้จริง เน้นการดูข้อมูลผ่านๆไปเรื่อยๆและนิยมใช้ในการโฆษณาทางการตลาด ส่วนข้อมูล Youtube จะเป็น VDO ที่เป็นตัวแทนในการอธิบายข้อความหรือเนื้อหา และ Instagram เหมาะสมกับการโพสต์ ภาพหลายๆภาพ ซึ่งทั้ง Youtube และ Instagram นั้นต่างนิยมใช้ในการโฆษณาทางการตลาดด้วยกันทั้งสิ้น

ทั้งนี้สำหรับงานวิจัยนี้ ผู้วิจัยจึงเลือกใช้ข้อมูลจากทวิตเตอร์และได้ใช้ข้อมูลของศูนย์ข่าวที่แจ้งเกี่ยวกับข่าวอุบัติเหตุใน กทม. ผ่านช่องทางทวิตเตอร์ ตั้งแต่เดือน 3 ธันวาคม พ.ศ. 2559 ถึง 3 มีนาคม พ.ศ. 2560 มาเป็นตัวอย่างในการนำ Big Data เข้ามาใช้ในธุรกิจ ซึ่งเป็นข้อมูลที่ได้มาจากทวิตเตอร์ทั้งหมด 6 แหล่ง ได้แก่ fm91trafficpro, js100radio, MOT_1356, traffy, travel_friendss, wichansuriyo จากทวิตเตอร์ทั้งหมดที่มีการทวิตเกี่ยวกับการจราจรและอุบัติเหตุบนถนนที่มีการอัพเดทข้อมูลอยู่ตลอด ข้อมูลเหล่านี้มีขนาดประมาณ 2 GB โดยข้อมูลทั้งส่วนใหญ่มีรูปแบบเป็นลักษณะของข้อมูลไม่มีโครงสร้าง (Unstructured) และมีลักษะเป็นข้อความ (Text)ที่อยู่บนทวิตเตอร์ดังนี้

ภาพที่ 4.2 แสดงภาพตัวอย่างทวิตเตอร์ของข่าวจราจร สวพ.FM91 หรือชื่อทวิตเตอร์fm91 trafficpro
ที่มา : Twitter fm91trafficpro

ส่วนที่ 2 ประเภทโครงสร้างของข้อมูล

เนื่องจากข้อมูลนั้นมีการทวีตจากหลายแหล่ง ซึ่งแต่ละแหล่งมีลักษณะข้อมูลหรือการนิยามข้อมูลที่แตกต่างกัน จึงมีความจำเป็นต้องจำแนกลักษณะของข้อมูลที่จะนำมาใช้วิเคราะห์ ซึ่งอาจแบ่งประเภทข้อมูลออกเป็น 2 ประเภท คือ ข้อมูลที่มีโครงสร้างชัดเจน (Structured Data) เช่น ข้อมูลที่ออกจากระบบของทวิตเตอร์โดยมีแบ่งแยกลักษณะต่างๆออกมาในแต่ละ Field อย่างชัดเจน และข้อมูลไม่มีโครงสร้าง (Unstructured Data) โดยข้อมูลส่วนใหญ่คือข้อความ (Text) ที่มาจาก ทวิตเตอร์

ทั้งนี้ข้อมูลที่ต้องการเก็บเพื่อนำมาใช้เป็นตัวอย่างนั้นเป็นข้อมูลที่มาจากทวิตเตอร์ ซึ่งสามารถดึงข้อมูลผ่าน Application Programming Interface (API) ที่ทวิตเตอร์มีให้ใช้ฟรี โดยสามารถ

สมัครใช้งานได้จาก https://apps.twitter.com/app/new และดูตัวอย่างเพิ่มเติมได้จาก http://socialmedia-class.org/twittertutorial.html ซึ่งจากการสมัครใช้ API Twitter จะได้รายละเอียดในหน้า Keys and Access Tokens ที่เป็นสิ่งสำคัญที่สุดในการนำไปใช้เชื่อมต่อกับ Program R เพื่อดึงข้อมูล โดยมีรายละเอียดตามภาพที่ 4.3

ภาพที่ 4.3 แสดงข้อมูล Keys และ Access Token ใน Application Management

ที่มา : https://apps.twitter.com/app/new

เมื่อทำการเชื่อมต่อและดึงข้อมูลผ่าน Program R แล้วจะได้ผลเป็นตารางดังภาพที่ 4.4 และทำการวิเคราะห์โครงสร้างของข้อมูล เพื่อทำการกรองและคัดแยกข้อมูลให้เป็นระเบียบ

	text	created	id	screenName	retweet	longitude	latitude
1	09.06น. อุบัติเหตุ ถนนวิภาวดีรังสิต ขาออก ช่องท	2017-02-10 09:07	8298742925904	fm91trafficpro	1	NA	NA
2	กรณี อุบัติเหตุ ถนนมอเตอร์เวย์(ทล.7) มุ่งหน้ากรุงเ	2017-02-10 07:37	8298516813325	fm91trafficpro	3	NA	NA
3	07.08 น.กรณีอุบัติเหตุ ถ.มอเตอร์เวย์ กม.1+300 ล่า	2017-02-10 07:11	8298451071468	fm91trafficpro	9	NA	NA
4	06.38 น.อุบัติเหตุ ถ.มอเตอร์เวย์ ขาเข้า กม.1+300	2017-02-10 06:39	8298372624896	fm91trafficpro	3	NA	NA
5	05.43 น. คืบหน้ากรณีอุบัติเหตุรถชนกัน ด่วนบูรพาวิถี	2017-02-10 05:46	8298237725804	fm91trafficpro	2	NA	NA
6	00.49น. ถนนนครอินทร์ ขาเข้า บนสะพานข้ามแยกบ	2017-02-10 00:54	8297503449093	fm91trafficpro	0	NA	NA
7	18.55น. ก่อนถึงวงเวียนรัชโยธิน เล็กน้อยถ.พหลโยธิ	2017-02-09 18:55	8296599046672	fm91trafficpro	3	NA	NA
8	16.54 น. ปากซอยประชาชื่น-นนทบุรี3 มุ่งหน้าแจ้งวัฒน	2017-02-09 16:55	8296297764952	fm91trafficpro	15	NA	NA
9	13.57น. ถนนติวานนท์ ขาเข้า หน้า รพ.ทรวงอก มี	2017-02-09 13:58	8295852087687	fm91trafficpro	1	NA	NA
10	11.32น. ถนนประชาธิปไตย บริเวณ แยกคุรุสภา มีอ	2017-02-09 11:32	8295484495614	fm91trafficpro	3	NA	NA

ภาพที่ 4.4 แสดงตัวอย่างตารางข้อมูลที่ดึงออกมาจากทวิตเตอร์ สวพ. FM91
ที่มา : ผู้วิจัย, 2560

จากตารางที่ 4.2 พบว่า ข้อมูลหลักที่นำมาใช้วิเคราะห์ 16,530 ตัวอย่าง ที่เกิดจากการทวีตข้อความในทวีตเตอร์ที่กล่าวมาข้างต้น สามารถสรุปประเภทข้อมูลในแต่ละ Field ได้ดังนี้

ตารางที่ 4.2 แสดงผลการวิเคราะห์การจำแนกประเภทโครงสร้างของข้อมูล

Field	ข้อมูลมีโครงสร้าง	ข้อมูลไม่มีโครงสร้าง	หมายเหตุ
-ข้อความที่เกิดจากการทวีต (Text)		-เป็นประโยคยาว	-ต้องทำการกรองข้อมูลเพื่อนำมาใช้วิเคราะห์
-วันที่และเวลาที่ทำการทวีต (Create)	-วันที่กับเวลาอยู่ในข้อมูล Columnเดียว		-มีความชัดเจน แต่ต้องแยก Format วันที่และเวลาออกจากกัน
-รหัสประจำตัวผู้ใช้ Twitter (Id)	-Id ชัดเจนใน Columnเดียว		-ไม่นำมาใช้ในการวิเคราะห์
-จำนวนครั้งที่ทำการรีทวีต (Retweet)	-นับจำนวน Retweet		-ไม่นำมาใช้ในการวิเคราะห์
-Location (Longitude)	-แยกพิกัดชัด		-ส่วนมากMissing Data
-Location (Latitude)	-แยกพิกัดชัด		ส่วนมากMissing Data

ที่มา : ผู้วิจัย, 2560

จากตารางที่ 4.2 สรุปได้ว่า วันที่และเวลาที่ทำการทวีต ใน Field ของ Column “Create” มีความชัดเจนของข้อมูลและเป็นข้อมูลประเภทมีโครงสร้าง และ Location ทั้ง Longitude และ Latitude เป็นประเภทข้อมูลที่มีโครงสร้างเช่นกันแต่ไม่มีข้อมูลปรากฏ (Missing Data) จึงไม่สามารถใช้ข้อมูลจาก Field นี้ได้ ส่วนรหัสประจำตัวผู้ใช้ Twitter และจำนวนครั้งที่ทำการทวีตเป็นข้อมูลที่มีโครงสร้าง แต่ผู้วิจัยไม่ได้นำข้อมูลส่วนนี้มาใช้ ซึ่งจะกล่าวในลำดับถัดไปในเรื่องของข้อมูลที่มีความจำเป็นต้องเก็บมาใช้วิเคราะห์ต้องมีอะไรบ้าง อย่างไรก็ตามมีเพียงข้อมูล Field เดียวที่เป็นข้อมูลประเภทไม่มีโครงสร้าง ได้แก่ ข้อความที่เกิดจากการทวีต ซึ่งข้อมูลนี้มีความสำคัญเนื่องจากข้อความเหล่านี้ประกอบด้วยข้อมูลที่เป็นประโยชน์ในการนำมาใช้วิเคราะห์แต่อยู่ในลักษณะของข้อความทวีตยาวๆ ไม่แบ่งแยกข้อมูลเป็นส่วนๆ จึงต้องทำการสกัดข้อมูล เพื่อนำมาใช้ประโยชน์ต่อไป

ส่วนที่ 3 วิธีการคัดเลือกข้อมูลที่ต้องนำมาใช้ในการวิเคราะห์

เนื่องจากข้อมูลที่ดึงมาจากทวิตเตอร์นั้นไม่เพียงพอต่อการวิเคราะห์ข้อมูลเพราะข้อมูลที่สามารถนำมาใช้ได้โดยไม่ต้องสกัดข้อมูลเพิ่มเติมมีเพียงไม่กี่ Field ทางผู้วิจัยจึงได้นำข้อมูลของสถิติคดีอุบัติเหตุการจราจรทางบก จำแนกตามสาเหตุการเกิดอุบัติเหตุจากบุคคล สาเหตุจากสิ่งแวดล้อม และ สาเหตุจากอุปกรณ์ที่ใช้ขับขี่ กรุงเทพมหานคร (บช.น.) พ.ศ. 2549 – พ.ศ.2558 ของสำนักงานตำรวจแห่งชาติ มาเป็นต้นแบบในการเก็บข้อมูลสาเหตุการเกิดอุบัติเหตุ และเนื่องจากสาเหตุการเกิดอุบัติเหตุจากการเก็บสถิติข้อมูลมีแนวโน้มของสาเหตุไปในทิศทางเดียวกันตลอด 10 ปี และข้อมูลสถิติการรับแจ้งคดีอุบัติเหตุการจราจรทางบก จำแนกตามประเภทรถ ความเสียหาย และผู้ต้องหา กรุงเทพมหานคร กองบัญชาการตำรวจนครบาล(บช.น.) ปีงบประมาณ พ.ศ.2550 – พ.ศ. 2557 สามารถดูข้อมูลเพิ่มเติมได้ที่ http://service.nso.go.th/nso/web/statseries/statseries21.html ทางผู้วิจัยจึงขอยกตัวอย่างข้อมูล 5 ปีย้อนหลัง 5 อันดับแรกในแต่ละสาเหตุหลักที่มีจำนวนการเกิดการอุบัติเหตุ (Case) มากที่สุดในเขตกรุงเทพมหานคร ซึ่งอย่างไรก็ตามทางสำนักงานตำรวจแห่งชาติได้แบ่งการเกิดอุบัติเหตุออกเป็น 5 มิติ คือ สาเหตุจากบุคคล สาเหตุจากสิ่งแวดล้อม สาเหตุจากอุปกรณ์ขับขี่ ประเภทของรถที่เกิดอุบัติเหตุต่างๆ และความเสียหายที่เกิดกับบุคคล ซึ่งสรุปได้ดังนี้

3.1 ข้อมูลด้านสาเหตุจากบุคคล

ตารางที่ 4.3 แสดงผลการวิเคราะห์ Case การเกิดอุบัติเหตุแบ่งตามสาเหตุจากบุคคลที่เกิดมากที่สุดตั้งแต่ปี พ.ศ. 2554 – พ.ศ. 2558 ในเขตกรุงเทพมหานคร

สาเหตุการเกิดอุบัติเหตุ	ปีที่เกิดอุบัติเหตุ					% of 2558
	2554	2555	2556	2557	2558	
ไม่ยอมรถที่มีสิทธิไปก่อน	2,712	3,590	3,028	2,877	3,249	22.06%
ขับรถเร็วเกินกว่ากฎหมายกำหนด	3,593	3,397	2,640	2,660	2,921	19.83%
ขับรถตัดหน้ากระชั้นชิด	2,474	1,854	1,412	1,202	1,525	10.36%
ขับรถตามกระชั้นชิด	1,057	881	788	880	792	5.38%
อื่นๆ	11,307	10,993	9,143	7,930	6,240	42.37%
รวมสาเหตุจากบุคคล	**21,143**	**20,715**	**17,011**	**15,549**	**14,727**	**100.00%**

ที่มา : สำนักงานตำรวจแห่งชาติ

จากตารางที่ 4.3 สรุปได้ว่าสาเหตุที่มีจำนวนสูงสุดจากบุคคลในเขตกรุงเทพมหานคร ได้แก่ การไม่อมรถที่มีสิทธิไปก่อน, การขับรถเร็วกว่ากฎหมายกำหนด, การขับรถตัดหน้ากระชั้นชิด และ การขับรถตามกระชั้นชิด ซึ่งมีเปอร์เซ็นในการเกิดอุบัติเหตุเทียบจากปี พ.ศ. 2558 คือ 22.06%, 19.83%, 10.36%, 5.38% ตามลำดับ และสาเหตุอื่นๆ 42.37%

3.2 ข้อมูลด้านสาเหตุจากสิ่งแวดล้อม

ตารางที่ 4.4 แสดงผลการวิเคราะห์ Case การเกิดอุบัติเหตุแบ่งตามสาเหตุจากสิ่งแวดล้อมที่เกิดมากที่สุด ตั้งแต่ปี พ.ศ. 2554 – พ.ศ. 2558 ในเขตกรุงเทพมหานคร

สาเหตุการเกิดอุบัติเหตุ	ปีที่เกิดอุบัติเหตุ					% of 2558
	2554	2555	2556	2557	2558	
ถนนแคบ	489	679	57	298	278	17.82%
ถนนลื่น	1,357	1,353	2,647	231	102	6.54%
มีฝนตก	180	104	35	57	52	3.33%
ถนนชำรุด	57	51	769	71	23	1.47%
อื่นๆ	3,657	3,147	668	1,439	1,105	70.83%
รวมสาเหตุจากสิ่งแวดล้อม	**5,740**	**5,334**	**4,176**	**2,096**	**1,560**	**100.00%**

ที่มา : สำนักงานตำรวจแห่งชาติ

จากตารางที่ 4.4 สรุปได้ว่าสาเหตุจากสิ่งแวดล้อมที่มีจำนวนสูงสุดในเขตกรุงเทพมหานคร ได้แก่ ถนนแคบ, ถนนลื่น, มีฝนตกและ ถนนชำรุด ซึ่งมีเปอร์เซ็นในการเกิดอุบัติเหตุเทียบจากปี พ.ศ. 2558 คือ 17.82%, 6.54%, 3.33%, 1.47% ตามลำดับ และสาเหตุอื่นๆ 70.83%

3.3 ข้อมูลด้านสาเหตุจากอุปกรณ์ที่ใช้ขับขี่

ตารางที่ 4.5 แสดงผลการวิเคราะห์ Case การเกิดอุบัติเหตุแบ่งตามสาเหตุจากอุปกรณ์ที่ใช้ขับขี่ที่เกิดมากที่สุด ตั้งแต่ปี พ.ศ. 2554 – พ.ศ. 2558 ในเขตกรุงเทพมหานคร

สาเหตุการเกิดอุบัติเหตุ	ปีที่เกิดอุบัติเหตุ					% of 2558
	2554	2555	2556	2557	2558	
ระบบห้ามล้อขัดข้อง	1,257	1,352	741	307	124	8.95%
ระบบบังคับเลี้ยวขัดข้อง	28	44	37	13	12	0.87%
ยางแตก	25	30	14	13	10	0.72%
ยางเสื่อมสภาพ	134	342	51	29	9	0.65%
อื่นๆ	4,266	3,148	951	1,291	1,230	88.81%
รวมสาเหตุจากอุปกรณ์ที่ใช้ขับขี่	**5,710**	**4,916**	**1,794**	**1,653**	**1,385**	**100.00%**

ที่มา : สำนักงานตำรวจแห่งชาติ

จากตารางที่ 4.5 สรุปได้ว่าสาเหตุจากสิ่งแวดล้อมที่มีจำนวนสูงสุดในเขตกรุงเทพมหานคร ได้แก่ ระบบห้ามล้อขัดข้อง, ระบบบังคับเลี้ยวขัดข้อง, ยางแตกและ ยางเสื่อมสภาพ ซึ่งมีเปอร์เซ็นในการเกิดอุบัติเหตุเทียบจากปี พ.ศ. 2558 คือ 8.95%, 0.87%, 0.72%, 0.65% ตามลำดับ และสาเหตุอื่นๆ 88.81%

นอกจากนี้ยังจำเป็นต้องจำแนกตามประเภทรถและความเสียหาย ดูข้อมูลเพิ่มเติมได้จาก http://service.nso.go.th/nso/web/statseries/statseries21.html ซึ่งขอสรุปได้ดังนี้

3.4 ข้อมูลประเภทรถ

ตารางที่ 4.6 แสดงผลการวิเคราะห์ ประเภทรถที่เกิดอุบัติเหตุมากที่สุด ตั้งแต่ปี พ.ศ. 2553 – พ.ศ. 2557 ในเขตกรุงเทพมหานคร

ประเภท	ปีที่เกิดอุบัติเหตุ					% of 2557
	2553	2554	2555	2556	2557	
รถยนต์นั่ง	12,628	12,902	12,070	11,041	10,338	41.49%
รถจักรยานยนต์	10,196	10,084	9,725	9,091	8,534	34.25%
รถบรรทุกขนาดเล็ก (ปิคอัพ)	3,513	3,420	2,773	2,512	2,303	9.24%
แท็กซี่	2,589	2,838	2,762	2,290	1,851	7.43%
อื่นๆ	2,598	2,602	2,681	2,266	1,890	7.59%
รวมประเภทรถ	**31,524**	**31,846**	**30,011**	**27,200**	**24,916**	**100.00%**

ที่มา : สำนักงานตำรวจแห่งชาติ

จากตารางที่ 4.6 สรุปได้ว่าจำนวนรถที่เกิดอุบัติเหตุมากที่สุดในเขตกรุงเทพมหานคร คือ รถยนต์นั่ง, รถจักรยานยนตร์, รถบรรทุกขนาดเล็ก (ปิคอัพ) และรถแท็กซี่ ซึ่งมีเปอร์เซ็นการเกิดอุบัติเหตุ 41.49%, 34.25%, 9.24% และ 7.43% ตามลำดับ ส่วนประเภทรถอื่นๆจำนวน 7.59%

3.5 ข้อมูลประเภทความเสียหาย

ตารางที่ 4.7 แสดงผลการวิเคราะห์ ประเภทความเสียหายของการเกิดอุบัติเหตุ ตั้งแต่ปี พ.ศ. 2553 – พ.ศ. 2557 ในเขตกรุงเทพมหานคร

ความเสียหายที่เกิดขึ้นกับ บุคคลประเภท	ปีที่เกิดอุบัติเหตุ					% of 2557
	2553	2554	2555	2556	2557	
ตาย	**252**	**400**	**294**	**276**	**224**	
ชาย	203	302	224	204	181	80.80%
หญิง	49	98	70	72	43	19.20%
บาดเจ็บสาหัส	**310**	**288**	**258**	**201**	**166**	
ชาย	209	186	172	123	107	64.46%
หญิง	101	102	86	78	59	35.54%
บาดเจ็บเล็กน้อย	**6,391**	**7,424**	**7,206**	**6,589**	**6,267**	
ชาย	4,143	4,695	4,484	4,228	4,075	65.02%
หญิง	2,248	2,729	2,722	2,361	2,192	34.98%

ที่มา : สำนักงานตำรวจแห่งชาติ

จากตารางที่ 4.7 สรุปได้ว่าประเภทความเสียหายในเขตกรุงเทพมหานคร แบ่งเป็น 3 ความเสียหายหลัก ได้แก่ ตาย, บาดเจ็บสาหัส และบาดเจ็บเล็กน้อย ซึ่งทั้ง 3 ประเภทนั้น ผู้ชายมีจำนวนความเสียหายมากกว่าผู้หญิงอย่างมาก ยกตัวอย่างในปี พ.ศ. 2557 กรณีตาย ชายมีจำนวน 80.80% และหญิงมีจำนวน 19.20% ส่วนกรณีบาดเจ็บสาหัส ชายมีจำนวน 64.46% และหญิงมีจำนวน 35.54% กรณีบาดเจ็บเล็กน้อย ชายมีจำนวน 65.02% และหญิงมีจำนวน 34.98%

ดังนั้นผู้วิจัยจึงวิเคราะห์ข้อมูล 5 มิติหลักด้วยข้อมูลที่ดึงมาจากทวิตเตอร์ที่ดีควรจะมีข้อมูลการเกิดอุบัติเหตุที่เกิดจาก ดังนี้

1. สาเหตุจากบุคคล ได้แก่ การไม่ยอมรถที่มีสิทธิไปก่อน, การขับรถตัดหน้ากระชั้นชิดและการขับรถตามกระชั้นชิด

2. สาเหตุจากสิ่งแวดล้อม ได้แก่ ถนนแคบ, ถนนลื่น, มีฝนตกและถนนชำรุด

3. สาเหตุจากการอุปกรณ์ที่ใช้ขับขี่ ได้แก่ ระบบห้ามล้อขัดข้อง, ระบบบังคับเลี้ยวขัดข้อง, ยางแตกและยางเสื่อมสภาพ

4. ประเภทของรถที่เกิดเหตุ ได้แก่ รถยนต์, รถจักรยานต์ และรถบรรทุกขนาดเล็ก

5. ความเสียหายที่เกิดกับบุคคล ได้แก่ การตาย, บาดเจ็บสาหัส และบาดเจ็บเล็กน้อย

ส่วนที่4 โปรแกรมที่ใช้ในการดึงข้อมูล

ผู้วิจัยได้เลือกโปรแกรมที่มีทักษะในการใช้งานและสามารถเชื่อมต่อกับโปรแกรมอื่นๆ ได้อย่างหลากหลาย และเป็นโปรแกรมที่ใช้งานฟรี สามารถทำงานได้ในหลายระบบปฏิบัติการ รวมถึงมี Community ที่ใหญ่ เป็นที่ยอมรับในหมู่นักวิจัย (Somkiat, 2557) โดยเลือกโปรแกรม R ในการดึงข้อมูลจาก API ของ Twitter โดยใช้โปรแกรมทั้งสองส่วนประกอบกัน ดังที่ได้กล่าวไว้ข้างต้น โดยสรุปโปรแกรมที่ใช้ตามลำดับดังนี้

เนื่องจากโปรแกรม R เป็นโปรแกรมที่ฟรี มีความนิยมสูง จึงมีฟังก์ชันการใช้งานที่หลากหลายมาก หากต้องการใช้งานในส่วนใดก็ติดตั้ง เพิ่มฟังก์ชันการใช้งานในส่วนนั้นๆ โดยการติดตั้งนั้นเรียกว่าการติดตั้ง R Packages ซึ่งการประมวลผลและการแสดงผลต่างๆจากโปรแกรม

RStudio มีความจำเป็นต้อง Install Packages ตามที่มีความจำเป็นต้องใช้ในแต่ละขั้นตอน เช่น หากต้องการแสดงผลเกี่ยวกับการพล็อต (Plot) กราฟ ต้องทำการติดตั้งPackage ด้วยการใช้คำสั่ง install.packages(“ggplot2”) เป็นต้น ซึ่งในเครื่องหมาย “……. ” นั้นคือชื่อของ Packagesที่ต้องการทำการติดตั้ง ทั้งนี้ในการใช้งานในงานวิจัยนี้ ผู้วิจัยจึงได้สรุปผลการติดตั้ง Package ทั้งหมด ดังนี้

ตารางที่ 4.8 แสดงการสรุปผลการติดตั้ง Package ใน RStudio

ชื่อ Packages	**คำอธิบาย**
dplyr	ใช้ในการจัดการข้อมูล ก่อนนำไปวิเคราะห์ต่อ เช่น การกรองข้อมูล, การลบข้อมูลที่ซ้ำซ้อนออก, การดึงเฉพาะข้อมูลที่ต้องการมาใช้งาน (Somkiat, 2016) (A grammar of data manipulation)
ggplot2	ใช้ในการแสดงผลข้อมูลในรูปแบบ graphic (Somkiat, 2016) (Create elegant data visualizations using the grammar of graphics)
jsonlite	ใช้ในการเพิ่มประสิทธิภาพการปฏิสัมพันธ์ระหว่างโปรแกรมและ web API (Cran.r-project, 2016) (A robust, high performance JSON parser and generator for R)
leaflet	ใช้ในการสร้างแผนที่ และย่อ ขยาย แผนที่ (rstudio.github, 2016) (Create interactive web maps with the Javascript ‘Leaflet’ library)
lubridate	ใช้ในการจัดการข้อมูลเรื่องวันที่ (date) และ เวลา (time) เช่น การแปลงรูปแบบ (format) ของข้อมูลให้เป็นข้อมูลวันที่ที่มีโครงสร้างเหมือนกันเพื่อนำมาใช้ในการประมวลผล สรุปจำนวนข้อมูล เป็นต้น (Cran.r-project, 2016) (Make dealing with dates a little easier)
shiny	เป็น application ที่ใช้ในการสร้างผลลัพธ์ (output) ให้แสดงได้อย่างสวยงามจากการใส่ข้อมูล (input) เข้าไป(Cran.r-project, n.d.) (Web application framework for R)

ตารางที่ 4.8 แสดงการสรุปผลการติดตั้ง Package ใน RStudio (ต่อ)

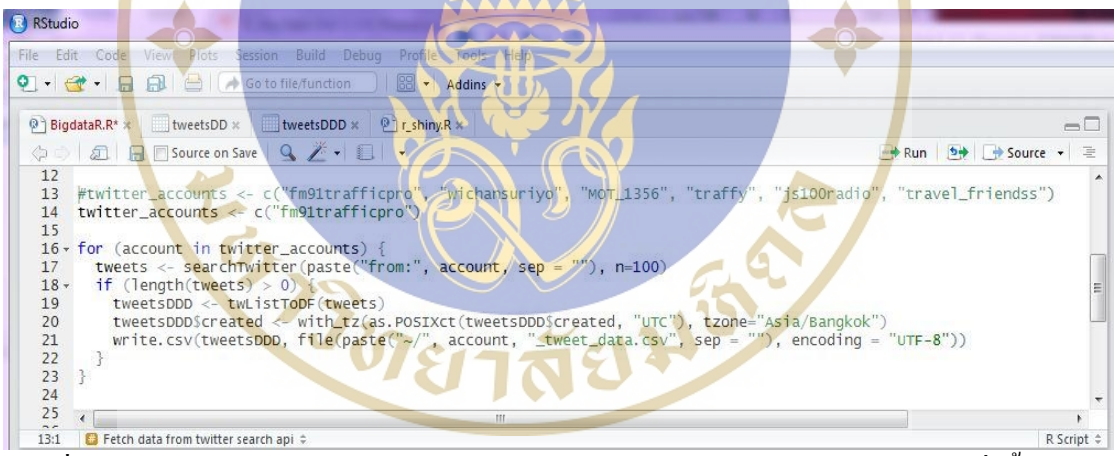
ชื่อ Packages	**คำอธิบาย**

sparklyr	ใช้ในการทำงานร่วมกันกับ Apeche Spark เพื่อการจัดการข้อมูลที่ดีและรวดเร็วมากยิ่งขึ้น (r-bloggers, 2016) (R interface to apache spark)
twitterR	ใช้ในการเข้าถึง Twitter API เพื่อนำข้อมูลมาวิเคราะห์ (Cran.r-project,n.d) (R based twitter client)

ที่มา : ผู้วิจัย, 2560

ส่วนที่ 5 การตรวจสอบความถูกต้องของข้อมูลที่ดึงมาใช้เบื้องต้น

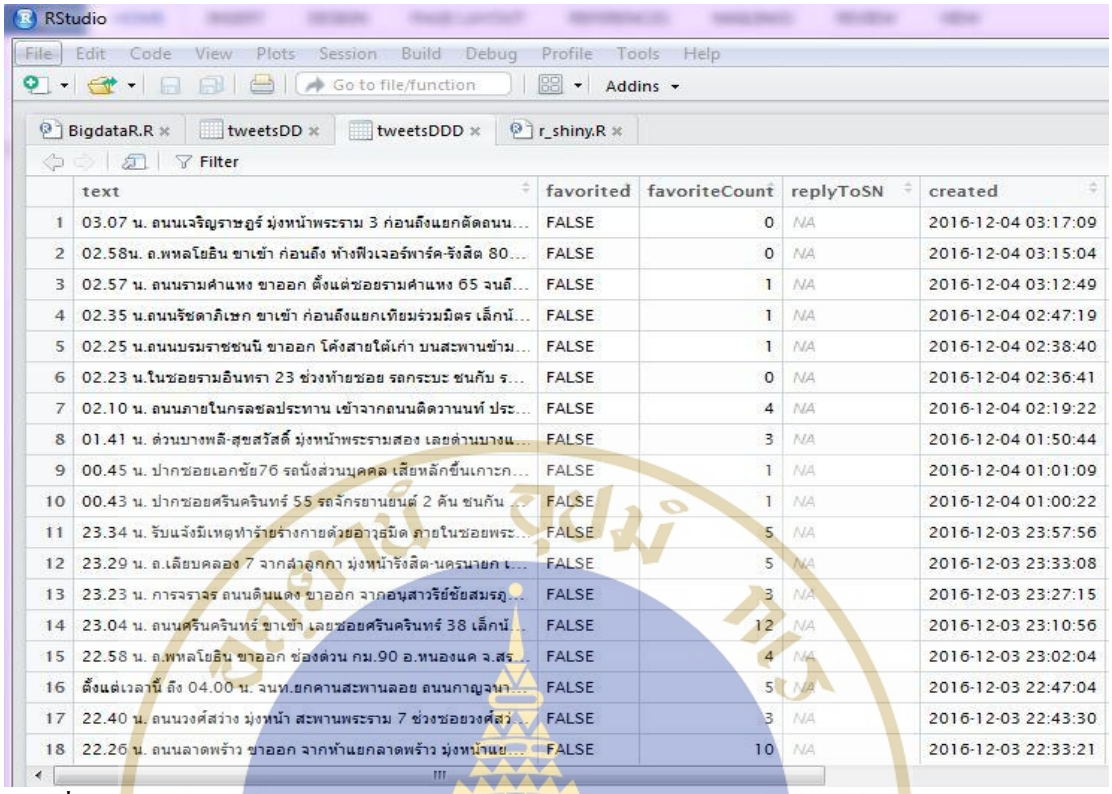
หลังจากที่มีการเชื่อมต่อ R Program กับ Twitter แล้วนั้น ได้ทำการเก็บข้อมูลการจราจรและอุบัติเหตุจาก fm91trafficpro, js100radio, MOT_1356, traffy, travel_friendss, wichansuriyo จากทวิตเตอร์ทั้งหมดที่มีการทวีตเกี่ยวกับการจราจรและอุบัติเหตุบนถนน โดยเริ่มจากการเลือก fm91trafficpro มาเป็นตัวอย่างกลุ่มแรกในการเก็บข้อมูล 100 ตัวอย่าง ทั้งนี้มีการแปลงค่า encoding = “UTF-8” เพื่อให้สามารถอ่านภาษาไทยได้ดียิ่งขึ้น จากนั้นทำการบันทึกเป็นfile CSV จากคำสั่ง write.csv



ภาพที่ 4.5 แสดงเก็บข้อมูลจาก Fm91 มา 100 ตัวอย่าง และปรับค่าให้อ่านภาษาไทยได้ดียิ่งขึ้น

ที่มา : ผู้วิจัย, 2560

เมื่อได้ข้อมูลมาแล้ว สามารถเรียกไฟล์ข้างต้นเปิดดูข้อมูลดังกล่าวด้วยคำสั่ง Read.csv จะได้ข้อมูลตามภาพ



	text	favorited	favoriteCount	replyToSN	created
1	03.07 น. ถนนเจริญราษฎร์ มุ่งหน้าพระราม 3 ก่อนถึงแยกตัดถนน...	FALSE	0	NA	2016-12-04 03:17:09
2	02.58น. ถ.พหลโยธิน ขาเข้า ก่อนถึง ห้างฟิวเจอร์พาร์ค-รังสิต 80...	FALSE	0	NA	2016-12-04 03:15:04
3	02.57 น. ถนนรามคำแหง ขาออก ตั้งแต่ซอยรามคำแหง 65 จนถึ...	FALSE	1	NA	2016-12-04 03:12:49
4	02.35 น.ถนนรัชดาภิเษก ขาเข้า ก่อนถึงแยกเทียมร่วมมิตร เล็กน้...	FALSE	1	NA	2016-12-04 02:47:19
5	02.25 น.ถนนบรมราชชนนี ขาออก โค้งสายใต้เก่า บนสะพานข้าม...	FALSE	1	NA	2016-12-04 02:38:40
6	02.23 น.ในซอยรามอินทรา 23 ช่วงท้ายซอย รถกระบะ ชนกับ ร...	FALSE	0	NA	2016-12-04 02:36:41
7	02.10 น. ถนนภายในกรลชลประทาน เข้าจากถนนติดวานนท์ ประ...	FALSE	4	NA	2016-12-04 02:19:22
8	01.41 น. ด่วนบางพลี-สุขสวัสดิ์ มุ่งหน้าพระรามสอง เลยด่านบางแ...	FALSE	3	NA	2016-12-04 01:50:44
9	00.45 น. ปากซอยเอกชัย76 รถนั่งส่วนบุคคล เสียหลักขึ้นเกาะก...	FALSE	1	NA	2016-12-04 01:01:09
10	00.43 น. ปากซอยศรีนครินทร์ 55 รถจักรยานยนต์ 2 คัน ชนกัน ...	FALSE	1	NA	2016-12-04 01:00:22
11	23.34 น. รับแจ้งมีเหตุทำร้ายร่างกายด้วยอาวุธมีด ภายในซอยพระ...	FALSE	5	NA	2016-12-03 23:57:56
12	23.29 น. ถ.เลียบคลอง 7 จากสำโรงกา มุ่งหน้ารังสิต-นครนายก ร...	FALSE	5	NA	2016-12-03 23:33:08
13	23.23 น. การจราจร ถนนดินแดง ขาออก จากอนุสาวรีย์ชัยสมรภู...	FALSE	3	NA	2016-12-03 23:27:15
14	23.04 น. ถนนศรีนครินทร์ ขาเข้า เลยซอยศรีนครินทร์ 38 เล็กน้...	FALSE	12	NA	2016-12-03 23:10:56
15	22.58 น. ถ.พหลโยธิน ขาออก ช่องด่วน กม.90 อ.หนองแค จ.สร...	FALSE	4	NA	2016-12-03 23:02:04
16	ตั้งแต่เวลานี้ ถึง 04.00 น. จนท.ยกคานสะพานลอย ถนนกาญจนา...	FALSE	5	NA	2016-12-03 22:47:04
17	22.40 น. ถนนวงศ์สว่าง มุ่งหน้า สะพานพระราม 7 ช่วงซอยวงศ์สว่...	FALSE	3	NA	2016-12-03 22:43:30
18	22.26 น. ถนนลาดพร้าว ขาออก จากห้าแยกลาดพร้าว มุ่งหน้าแย...	FALSE	10	NA	2016-12-03 22:33:21

ภาพที่ 4.6 แสดงตัวอย่างการเรียกดูข้อมูลจาก Fm91 จากfile CSV
ที่มา : ผู้วิจัย, 2560

ตรวจสอบข้อมูลที่ได้ดึงมากับ application twitter ว่าข้อมูลที่ได้มานั้นตรงกันหรือไม่ โดยผู้วิจัยได้สุ่มการตรวจสอบข้อมูลลำดับแรก ข้อมูลลำดับที่ 50 และลำดับที่ 100 โดยนำเทคนิคที่ได้จากการทำงานด้านข้อมูลกับบริษัทประกันภัยมาใช้ เพราะหากข้อมูลที่สุ่มมีความผิดพลาดมักจะเห็นได้ชัดจากข้อมูลแรกและข้อมูลท้ายสุด เป็นต้น ดังนั้นเมื่อตรวจสอบความถูกต้องของข้อมูลเรียบร้อยแล้ว สามารถดึงข้อมูลตามจำนวนที่ต้องการได้ โดยผู้วิจัยได้เปลี่ยนการดึงจาก 100 เป็น 16,530 ข้อมูลโดยประมาณ และเก็บข้อมูลกับองค์กรคลื่นวิทยุที่มีการทวีตเรื่องการเกิดอุบัติเหตุบนถนนทั้งหมด 6 แหล่ง ตามที่ได้กล่าวไปข้างต้น เป็นระยะเวลา 3 เดือน

ส่วนที่ 6 รายละเอียดข้อมูลที่นำมาวิเคราะห์

ตารางที่ 4.9 แสดงรายละเอียดข้อมูลที่นำมาวิเคราะห์

No.	วันที่ Tweet	เวลาที่Tweet	เขตที่เกิดเหตุ	เพศของคนขับ	สาเหตุจาก						ประเภทรถ			ความเสียหาย		
					บุคคล			สิ่งแวดล้อม			อุปกรณ์ที่ใช้ขับขี่					
				หญิง ชาย	ไม่ยอมรถที่มีสิทธิ์ไปก่อน	ขับรถตัดหน้ากระชั้นชิด	การขับรถตามกระชั้นชิด	ถนนแคบ	ถนนลื่น	ถนนที่ไฟไม่ตกและถนนชำรุด	ระบบห้ามล้อขัดข้อง	ระบบบังคับเลี้ยวขัดข้อง	ยางแตก	รถยนต์	รถจักรยานยนต์	รถบรรทุก	ตาย	บาดเจ็บสาหัส	บาดเจ็บเล็กน้อย
1																			
2																			
3																			
4																			
5																			

ที่มา : ผู้วิจัย, 2560

จากตารางที่ 4.9 พบว่าตัวแปรที่มีความเกี่ยวข้องและจำเป็นต้องใช้ในการวิเคราะห์ข้อมูล ได้แก่ วันที่ Tweet, เวลาที่ Tweet, เขตที่เกิดเหตุ, เพศของคนขับ, สาเหตุการเกิดอุบัติเหตุจากบุคคล สิ่งแวดล้อม หรืออุปกรณ์ที่ใช้ขับขี่, ประเภทรถที่เกิดอุบัติเหตุ และประเภทความเสียหาย ซึ่งรายละเอียดตามตาราง ผู้วิจัยจึงได้ทำการจัดรูปแบบข้อความเหล่านั้นให้อยู่ใน Column เป็นโครงสร้างที่ชัดเจนและง่ายต่อการนำไปวิเคราะห์มากยิ่งขึ้น ตัวอย่างข้อความ “9.35น. อุบัติเหตุ ถ.รัชดาภิเษก ขาออก ก่อนถึงแยกรัชโยธินเล็กน้อย รถจักรยานยนต์ชนกับรถนั่งส่วนบุคคลขวางเลนขวา...” โดยข้อมูลดังกล่าว เช่น เวลาที่เกิดเหตุ คือ 9.35น. จึงต้องทำการดึงข้อมูลนี้ออกมาจากข้อความ เพื่อนำมาใช้ เป็นต้น ส่วนตัวแปรอื่นๆที่จำเป็นต้องนำมาใช้ก็ต้องแยกออกจากข้อความเหล่านี้เช่นกัน

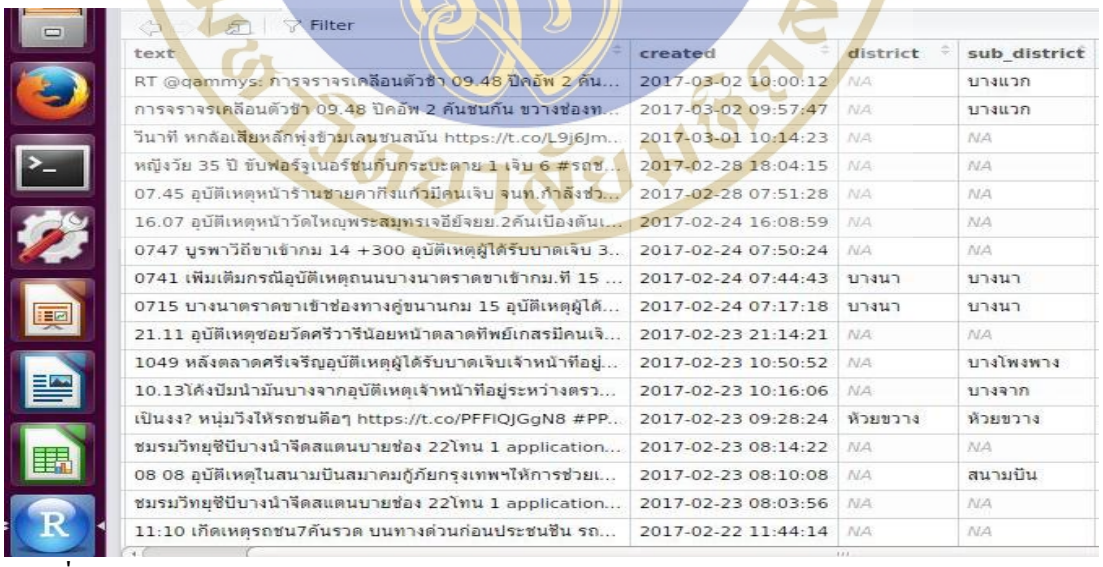
ในส่วนของสถานที่ เขต และพิกัด ทางผู้วิจัยได้นำข้อมูลต้นแบบจากเว็บไซต์ Data.go.th เพื่อนำมาจับคู่กับสถานที่ที่มีอยู่ในข้อความทวีตตามภาพ เนื่องจากข้อมูลเดิมที่ได้จากการดึงออกมาจากทวิตเตอร์นั้นไม่แสดงข้อมูลละติจูดและลองจิจูด

ตารางที่ 4.10 แสดงตัวอย่างการนำข้อมูลพิกัดของเขตมาเป็นต้นแบบในการจับคู่แสดงผลข้อมูล

ลำดับ	รหัสไปรษณีย์	จังหวัด	เขต	ละติจูด	ลองจิจูด
1	10200	กรุงเทพมหานคร	พระนคร	13.7560243	100.4986793
2	10300	กรุงเทพมหานคร	ดุสิต	13.7726943	100.5099262
3	10530	กรุงเทพมหานคร	หนองจอก	13.8559883	100.8619689
4	10500	กรุงเทพมหานคร	บางรัก	13.7262395	100.5267991
5	10220	กรุงเทพมหานคร	บางเขน	13.864387	100.6146434
6	10240	กรุงเทพมหานคร	บางกะปิ	13.7650282	100.6473885
7	10330	กรุงเทพมหานคร	ปทุมวัน	13.7401666	100.5352367
8	10100	กรุงเทพมหานคร	ป้อมปราบศัตรูพ่าย	13.7514954	100.5108479
9	10260	กรุงเทพมหานคร	พระโขนง	13.6910654	100.614025
10	10510	กรุงเทพมหานคร	มีนบุรี	13.8130488	100.731339
11	10520	กรุงเทพมหานคร	ลาดกระบัง	13.7277339	100.7486314
12	10120	กรุงเทพมหานคร	ยานนาวา	13.7171256	100.5151327

ที่มา : Data.go.th

หลังจากได้ข้อมูลเขตต้นแบบทั้งหมด 50 เขต ของกรุงเทพมหานครแล้วนั้น จะนำข้อมูลเหล่านั้นมาจับกับข้อมูลประโยคการทวีตที่ได้จากทวิตเตอร์ เพื่อนำไป Plot สถานที่การเกิดอุบัติเหตุ



ภาพที่ 4.7 แสดงตัวอย่างการแมพ (Map) ข้อมูลและแสดงผลละติจูดและลองจิจูด

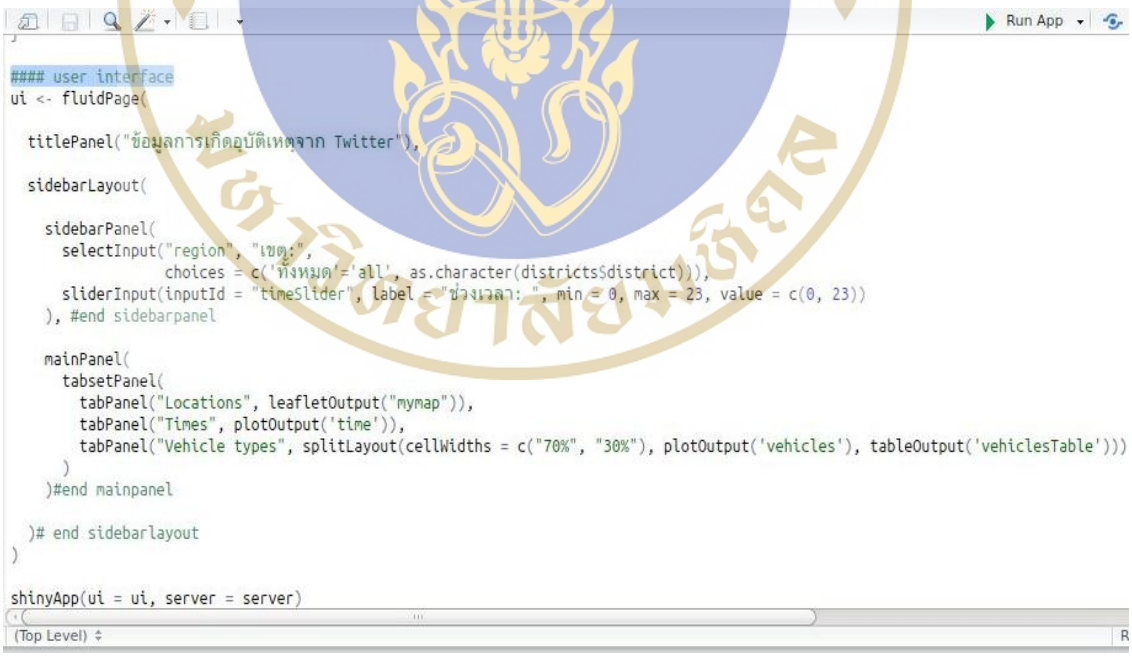
ที่มา : ผู้วิจัย, 2560

ส่วนที่ 7 การวิเคราะห์และนำเสนอเป็นกราฟหรือตาราง

หลังจากที่ได้รวมข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์ที่ได้เลือกไว้ และทำการปรับข้อมูลจากตัวหนังสือ (Text) เป็นข้อมูลที่แบ่งเป็นโครงสร้างอย่างชัดเจน ทั้งหมด 6 ส่วน (ตัดส่วนวันที่ Tweet ออกเนื่องจากการเก็บข้อมูลมีเพียงระยะเวลาสั้นๆไม่สามารถดู Seasonal ของการเกิดอุบัติเหตุได้) โดยเริ่มจากโปรแกรมที่นำเสนอ ผ่านPackage shiny ของโปรแกรม R โดยแบ่งการนำเสนอออกเป็น 3 ส่วนหลัก ได้แก่ สถานที่ (Location), เวลาต่างๆ(Times) และประเภทของรถ (Vehicle types) และอีก 3 ส่วนนำเสนอในรูปของตาราง ผ่าน SQLได้แก่ เพศของคนขับ, สาเหตุของอุบัติเหตุ, และความเสียหายในส่วนของการเสียชีวิต ดังนี้

7.1 การนำเสนอในส่วนของสถานที่ที่เกิดอุบัติเหตุ (Location)

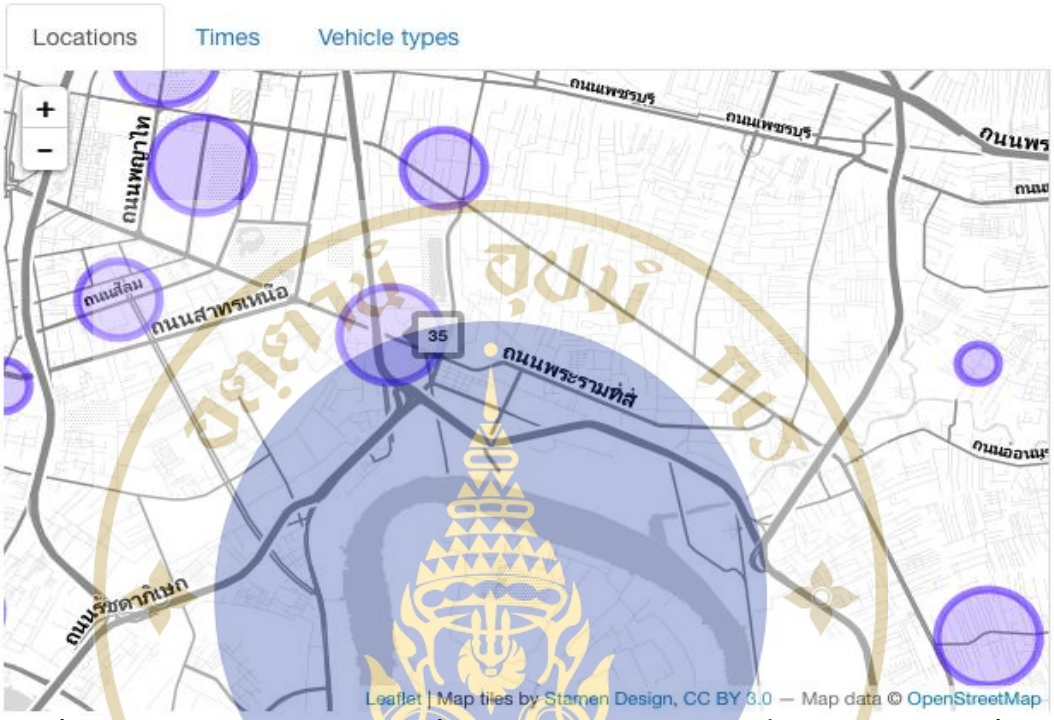
ในส่วนนี้จะเป็นการนำเสนอการเกิดอุบัติเหตุในสถานที่ต่างๆ ทั้งหมด 50 เขต ในกรุงเทพมหานคร โดยการ plot ลงในแผนที่ตามโปรแกรม R Package ของShiny ซึ่งแผนที่นี้จะสามารถ Zoom in หรือ Zoom out เขตที่ต้องการดูการแสดงผล โดยผู้วิจัยได้ใส่เป็นจุดวงกลมสีม่วงตามรัศมีเขตต่างๆที่มีการทวีตการเกิดอุบัติเหตุ ขนาดจุดเล็กแสดงถึงการเกิดอุบัติเหตุน้อย และขนาดจุดใหญ่แสดงถึงจำนวนทวีตการเกิดอุบัติเหตุมาก นอกจากนั้นเมื่อนำเม้าท์ไปชี้ที่จุดวงกลมจะมีการแสดงค่าจำนวนการเกิดเหตุในภาพด้วยเพื่อความชัดเจนในการแสดงผลมากยิ่งขึ้น



ภาพที่ 4.8 แสดงคำสั่งโปรแกรมเพื่อให้แสดงผลค่าการเกิดอุบัติเหตุในแผนที่

ที่มา : ผู้วิจัย, 2560

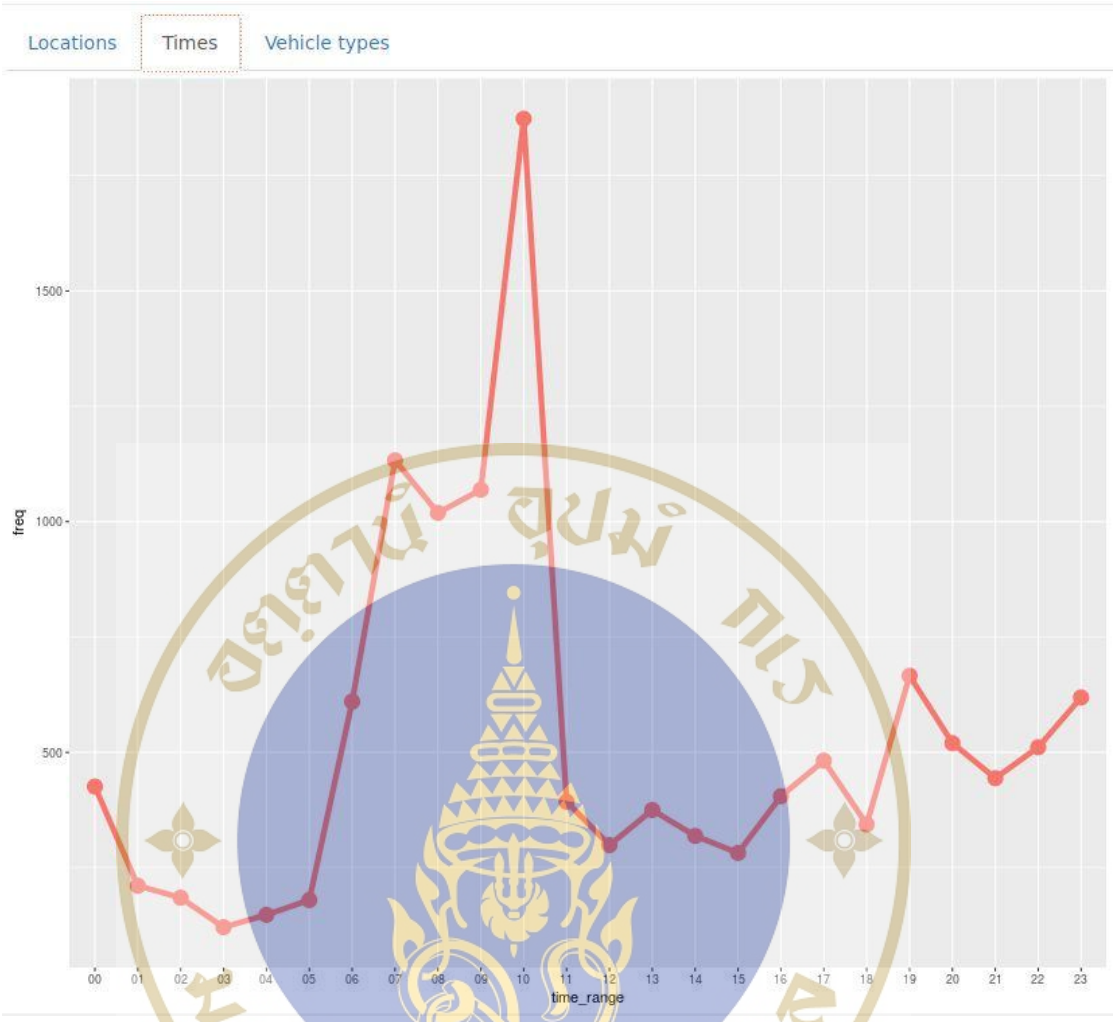
จากภาพแสดงการเกิดอุบัติเหตุจำนวน 50 เขต ในกรุงเทพมหานคร ที่ถูกPlot ลงในแผนที่เพื่อความสวยงามและเป็นการนำเสนอที่เห็นภาพได้อย่างชัดเจน และสามารถใช้เม้าท์วางที่จุดวงกลมของเขตที่ต้องการทราบข้อมูล ตัวอย่าง เขตพระรามที่สี่ มีจำนวนการเกิดอุบัติเหตุ 35 ครั้ง ในรอบ 3 เดือน คือตั้งแต่เดือน ธันวาคม พ.ศ. 2559 จนถึง กุมภาพันธ์ พ.ศ. 2560 ตามภาพที่ 4.9



ภาพที่ 4.9 แสดงตัวอย่างการขยายแผนที่ (Zoom in) การเกิดอุบัติเหตุที่ถูก Plot ลงในแผนที่
ที่มา : ผู้วิจัย, 2560

7.2 การนำเสนอในส่วนของเวลาที่ทำการทวีต (Times)

นอกจากนี้ภายในโปรแกรม Rstudio ยังสามารถนำข้อมูลมา Plot สร้างกราฟดูความสัมพันธ์ระหว่างเวลากับจำนวนที่เกิดเหตุในรอบระยะเวลา 3 เดือน คือตั้งแต่ธันวาคม พ.ศ. 2559 ถึง ต้นเดือนมีนาคม พ.ศ. 2560 ได้อีกด้วย ซึ่งจากกราฟพบว่าเวลาที่มีการทวีตการเกิดอุบัติเหตุมากที่สุดคือ ช่วงเวลาประมาณ 10 นาฬิกา รองลงมาคือ 7.00 น. 9.00 น. และ 8.00 น. ตามลำดับดังภาพที่ 4.10



ภาพที่ 4.10 แสดงจำนวนและเวลาที่ทวีตการเกิดอุบัติเหตุที่ถูก Plot ลงในกราฟเส้น

ที่มา : ผู้วิจัย, 2560

7.3 การนำเสนอในส่วนของประเภทและจำนวนรถที่เกิดอุบัติเหตุ ตามช่วงเวลาและเขตต่างๆ (Vehicle types)

อีกทั้งผู้วิจัยได้เพิ่มส่วนการนำเสนอประเภทและจำนวนรถที่เกิดอุบัติเหตุ ตามช่วงเวลาและเขตต่างๆ จะเห็นได้ว่าบางเขต เช่น เขตดินแดง มีการทวีตการเกิดอุบัติเหตุที่สูงถึง 47 ครั้ง ในรอบ 3 เดือน ตั้งแต่ 3 ธันวาคม พ.ศ. 2559 ถึง 3 มีนาคม พ.ศ. 2560 ซึ่งแตกต่างจากบางเขต เช่น เขตทวีวัฒนาที่ไม่มีการทวีตการเกิดอุบัติเหตุเลยในรอบ 3 เดือนดังกล่าว

อย่างไรก็ตามRProgram ยังไม่สามารถอ่านภาษาไทยได้ดีนักและในบางครั้งเมื่อมีการใช้ภาษาไทยโปรแกรมจะผิดพลาด (Error) และไม่สามารถประมวลผลได้ ดังนั้นผู้วิจัยจึงได้ใช้ภาษาอังกฤษในการใช้เพื่อแสดงผล (Output) เสนอบนกราฟดังกล่าว และเพื่อง่ายต่อความเข้าใจผู้วิจัยได้แบ่งประเภทของรถ (Car type) ในการนำเสนอไว้ดังนี้

ตารางที่ 4.11 แสดงรายละเอียดของตัวแปรที่ใช้ในการแสดงผลในกราฟ

ชื่อ	รายละเอียด
Car_type	ประเภทของรถ
Freq	จำนวนครั้งที่ทำการทวีตอุบัติเหตุ
Car	รถเก๋ง, เก๋ง
Pickup	รถกระบะ, กระบะ
Motorcycle	รถจักรยานยนต์, จักรยานยนต์, รถมอเตอร์ไซต์, มอเตอร์ไซต์, รถมอไซต์, มอไซต์
Truck	รถบรรทุก, บรรทุก, รถพ่วง, รถสิบล้อ, สิบล้อ
Other	รถ, รถยนต์

ที่มา : ผู้วิจัย, 2560

ภาพที่ 4.11 แสดงตัวอย่างการทวีตการเกิดอุบัติเหตุเขตดินแดงตั้งแต่ช่วงเวลา 0.00 – 24.00 น.

ที่มา : ผู้วิจัย, 2560

จากภาพที่ 4.12 พบว่า การเกิดอุบัติเหตุในเขตดินแดงการทวีตอุบัติเหตุรถยนต์มีจำนวนมากถึง 33 ครั้ง คิดเป็นร้อยละ 70 แต่ส่วนใหญ่ไม่ระบุประเภทของรถ จะระบุแค่เพียงว่ารถ หรือรถยนต์เป็นจำนวนมาก รองลงมาเป็นรถเก๋ง 7 ครั้ง คิดเป็นร้อยละ 15, รถกระบะ 5 ครั้ง คิดเป็นร้อยละ 11, รถจักรยานยนต์ 2 ครั้ง คิดเป็นร้อยละ 4 ตามลำดับ นอกจากนี้ยังไม่พบการเกิดอุบัติเหตุรถบรรทุกจากการทวีตในรอบ 3 เดือน ตั้งแต่ 3 ธันวาคม พ.ศ. 2559 ถึง 3 มีนาคม พ.ศ. 2560

นอกจากนี้เพื่อการเห็นภาพรวมได้มากยิ่งขึ้นทางผู้วิจัยจึงได้ทำตารางการสรุปภาพรวมของการทวีตการเกิดอุบัติเหตุแบ่งตามเขต และช่วงเวลาต่างๆจากข้อมูลที่เก็บมาจากทวิตเตอร์ ตั้งแต่ 3 ธันวาคม พ.ศ. 2559 ถึง 3 มีนาคม พ.ศ. 2560

ตารางที่ 4.12 แสดงผลการสรุปภาพรวมของการทวีตการเกิดอุบัติเหตุแบ่งตามเขต และช่วงเวลาต่างๆจากข้อมูลที่เก็บมาจากทวิตเตอร์ ตั้งแต่ 3 ธันวาคม พ.ศ. 2559 ถึง 3 มีนาคม พ.ศ. 2560

รายชื่อเขต	เวลา									รวม
	0.01 - 3.00	3.01 - 6.00	6.01 - 9.00	9.01 - 12.00	12.01 - 15.00	15.01 - 18.00	18.01 - 21.00	21.00 - 24.00	N/A	
ลาดพร้าว	8	2	33	26	12	20	4	13	42	160
พญาไท	5	6	46	72	8	6	3	2		148
บางนา	2	2	37	22	13	13	13	4	11	117
ดินแดง	8	5	12	4	11	8	2	4	8	62
สาทร	1		27	4	4	16	2	4		58
บางเขน	4	2	10	10	9	3	1	1		40
คลองเตย		1	21	3		1	3	3	3	35
หลักสี่		2	2	22		5	1			32
ธนบุรี	2		6	9	7					24
ห้วยขวาง	1		2	3	3	2	6	1	5	23
พระโขนง	7	2		8	1	1	3			22
ตลิ่งชัน			9	3	1			4	4	21
บางพลัด	1	2	4	6				2	4	19
พระนคร		2	3		4	3				12
ลาดกระบัง		1	3	2			2	2	2	12
บางแค	2		2	1	4		1		1	11
บางบอน			1	5		1		1	1	9
สายไหม				2			6		1	9
บางกะปิ			6					1	1	8
ปทุมวัน		3	1		1	1			2	8
ราชเทวี		1	3		4					8
บางขุนเทียน	1			1	1	1	2		1	7
บางกอกน้อย	1					1		4		6
บางรัก	2		4							6
วัฒนา	1				4					5
หนองแขม	1						1		3	5

ตารางที่ 4.11 แสดงผลการสรุปภาพรวมของการทวีตการเกิดอุบัติเหตุแบ่งตามเขต และช่วงเวลาต่างๆจากข้อมูลที่เก็บมาจากทวิตเตอร์ ตั้งแต่ 3 ธันวาคม พ.ศ. 2559 ถึง 3 มีนาคม พ.ศ. 2560 (ต่อ)

รายชื่อเขต	เวลา									รวม
	0.01 - 3.00	3.01 - 6.00	6.01 - 9.00	9.01 - 12.00	12.01 - 15.00	15.01 - 18.00	18.01 - 21.00	21.00 - 24.00	N/A	
คลองสาน								4		4
ประเวศ		2					1	1		4
มีนบุรี			2	2						4
ยานนาวา		1	1			2				4
หนองจอก							4			4
ดุสิต			1		2					3
ทุ่งครุ		1		2						3
นวมินทร์									3	3
ราษฎร์บูรณะ		1		1					1	3
สวนหลวง			2						1	3
ภาษีเจริญ		1						1		2
คันนายาว									1	1
จอมทอง						1				1
ทวีวัฒนา									1	1
บางซื่อ								1		1
บึงกุ่ม							1			1
วังทองหลาง							1			1
อื่นๆ	621	336	2,253	2,828	738	760	1,088	1,272	5,544	15,440
รวม	**668**	**373**	**2,491**	**3,036**	**827**	**845**	**1,145**	**1,325**	**5,640**	**16,350**

ที่มา : ผู้วิจัย, 2560

จากตารางที่ 4.11 สรุปได้ว่าเขตที่มีการทวีตการเกิดอุบัติเหตุสูงที่สุดอย่างเห็นได้ชัด คือ เขตลาดพร้าว เขตพญาไท และเขตบางนา โดยมีการทวีตทั้งหมด 160, 148 และ 117 ครั้ง ตามลำดับ โดยเขตลาดพร้าวและเขตบางนา มีเวลาที่ทำการทวีตจำนวนสูงสุด คือช่วงเวลา 6.01 – 9.00 น. จำนวน 33 และ 37 ครั้งตามลำดับ และเขตพญาไทมีเวลาที่ทำการทวีตจำนวนสูงสุด คือ ช่วงเวลา 9.01 – 12.00 น. จำนวน 72 ครั้ง

7.4 การนำเสนอในส่วนเพศของคนขับในกรณีอุบัติเหตุดังกล่าว

จากข้อมูลที่เก็บจากทวิตเตอร์นั้นมีจำนวนน้อยมากที่มีการทวีตถึงเพศของผู้ขับขี่ โดยมีการทวีตเป็นชาย 6 ครั้ง และหญิง 6 ครั้ง จึงไม่สามารถนำข้อมูลมาวิเคราะห์ได้เนื่องจากข้อมูลที่มีอยู่ในข้อความที่เก็บจากทวีตเตอร์ดังกล่าวนั้นไม่เพียงพอและไม่ครบถ้วน

7.5 การนำเสนอในส่วนของสาเหตุของอุบัติเหตุ

จากข้อมูลที่เก็บจากทวิตเตอร์นั้นมีจำนวนน้อยมากที่มีการทวีตถึงสาเหตุในการเสียชีวิตหรือเกิดอุบัติเหตุ เช่น มีการทวีตคำเกี่ยวกับการตัดหน้ากระชั้นชิด จำนวน 4 ครั้ง มีการทวีตกรณีอุบัติเหตุจากถนนลื่น 14 ครั้ง เป็นต้น จึงไม่สามารถนำข้อมูลมาวิเคราะห์ได้เนื่องจากข้อมูลที่มีอยู่ในข้อความที่เก็บจากทวีตเตอร์ดังกล่าวนั้นไม่เพียงพอและไม่ครบถ้วน

7.6 การนำเสนอในส่วนของความเสียหาย

จากข้อมูลที่เก็บจากทวิตเตอร์นั้น ผู้วิจัยขอยกตัวอย่างกรณี การนำเสนอกรณีความเสียหายในส่วนของการเสียชีวิต เนื่องจากเป็นความเสียหายที่สร้างความรุนแรงมากที่สุดเมื่อเทียบกับความเสียหายอื่นๆ ที่ต้องคำนึงถึงเป็นอันดับแรก และความเสียหายกรณีอื่นๆ เช่น การบาดเจ็บสาหัส หรือบาดเจ็บเล็กน้อย จะเป็นประเภทของความเสียหายที่มีระดับความรุนแรงรองลงมา

ตารางที่ 4.13 แสดงตัวอย่างข้อมูลในส่วนของความเสียหายในกรณีเสียชีวิต แบ่งตามเขต ในกรุงเทพมหานคร

เขต	จำนวนครั้งที่ Tweetการเสียชีวิต	เขต	จำนวนครั้งที่ Tweet การเสียชีวิต
ลาดพร้าว	5	ดินแดง	1
ดอนเมือง	4	ดุสิต	1
สาทร	3	บางขุนเทียน	1
คลองเตย	2	บางเขน	1
บางนา	2	บางพลัด	1
ปทุมวัน	2	พระนคร	1
พระโขนง	2	หลักสี่	1
คลองสาน	1	ห้วยขวาง	1
		รวม	**29**

ที่มา : ผู้วิจัย, 2560

จากตารางที่ 4.12 สรุปได้ว่า กรณีความเสียหายที่รุนแรงที่สุดคือ ในส่วนของการเสียชีวิตนั้น มีการทวีตถึงกรณีเสียชีวิตตามเขตต่างๆ 16 เขต รวม 29 ครั้ง เป็นเขตลาดพร้าว สูงที่สุด รองลงมาเป็นการทวีตเขตดอนเมือง และสาทร 5 , 4 และ 3 ครั้งตามลำดับ ส่วนเขต สาทร, คลองเตย, บางนา, ปทุมวันและ เขตพระโขนง มีการทวีตกรณีเสียชีวิตเขตละ 2 ครั้งและเขตคลองสาน, ดินแดง, ดุสิต, บางขุนเทียน, บางเขน, บางพลัด, พระนคร, หลักสี่และห้วยขวาง มีการทวีตกรณีเสียชีวิตเขตละ 1 ครั้ง

ขั้นตอนที่ 3 ผลการวิเคราะห์เปรียบเทียบระหว่าง รูปแบบข้อมูลที่ได้การทำ Big data จากทวิตเตอร์กับผลที่ได้จากข้อมูลของสำนักงานตำรวจแห่งชาติ

จากผลการวิเคราะห์เครื่องมือในการทำ Big Data และ ผลการวิเคราะห์แบบตรวจสอบรายการข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์ที่ได้กล่าวมาข้างต้น สามารถนำมาเปรียบเทียบรูปแบบข้อมูลที่ได้การทำ Big data จากทวิตเตอร์กับผลที่ได้จากข้อมูลของสำนักงานตำรวจแห่งชาติ ดังนี้

ตารางที่ 4.14 แสดงผลการวิเคราะห์เปรียบเทียบระหว่าง รูปแบบข้อมูลที่ได้การทำ Big data จากทวิตเตอร์กับผลที่ได้จากข้อมูลของสำนักงานตำรวจแห่งชาติ

รายละเอียด	รูปแบบแหล่งข้อมูล จากการทำ Big Data	รูปแบบแหล่งข้อมูล จากสำนักงานตำรวจแห่งชาติ	หมายเหตุ / ข้อจำกัด
- การนำเสนอรายงานอุบัติเหตุแยกย่อยตามเขต -การนำเสนอรายงานอุบัติเหตุแยกย่อยตามวัน/เวลา -การนำเสนอรายงานอุบัติเหตุแยกย่อยตามประเภทของรถ	-แยกตามเขต -แยกวัน เวลา -แยกประเภทรถ	-แยกตามจังหวัด -ไม่มีการแยกวัน เวลา -แยกประเภทรถ	-สำนักงานตำรวจแห่งชาติจะรายงานข้อมูลตามจังหวัด -ข้อมูลจากทวิตเตอร์มีรายละเอียดของวันและเวลา -ข้อมูลของสำนักงานตำรวจฯรจัดประเภทของรถในรายงานแล้วทำให้ไม่ทราบประเภทรถที่แท้จริง ซึ่งอาจเกิดจากการให้คำนิยามประเภทรถที่ไม่เหมือนกัน
-สาเหตุของการเกิดอุบัติเหตุ	-มีข้อมูลสาเหตุบางส่วน	-มีข้อมูลสาเหตุชัดเจน	-ข้อมูลจากทวิตเตอร์ในการทำ Big Data มักไม่แจ้งสาเหตุของการเกิดอุบัติเหตุและมีการนิยามการเกิดอุบัติเหตุที่แตกต่างกันมาก ทำให้จัดประเภทสาเหตุการเกิดได้ยาก

ตารางที่ 4.13 แสดงผลการวิเคราะห์เปรียบเทียบระหว่าง รูปแบบข้อมูลที่ได้การทำ Big data จากทวิตเตอร์กับผลที่ได้จากข้อมูลของสำนักงานตำรวจแห่งชาติ (ต่อ)

รายละเอียด	รูปแบบแหล่งข้อมูล		หมายเหตุ / ข้อจำกัด
	จากการทำ Big Data	จากสำนักงานตำรวจแห่งชาติ	
-ระดับความเสียหาย	-ระบุไว้บางข้อความในทวีต	-มีรายงานชัดเจน	-ข้อมูลสำนักงานตำรวจแห่งชาติมีการระบุระดับความเสียหายอย่างระเอียดชัดเจน เนื่องจากเป็นข้อมูลจากการสำรวจจริง ณ จุดเกิดเหตุ
-รายละเอียดเพิ่มเติมของความเสียหาย	-มีรายละเอียดเพิ่มในข้อความ Tweet	-ไม่มีรายละเอียดเพิ่ม มีแค่ส่วนข้อมูลที่รายงาน	-ข้อมูลสำนักงานตำรวจแห่งชาติเป็นข้อมูลสรุปไม่สามารถดูรายละเอียดเพิ่มเติมในแต่ละกรณีได้ เนื่องจากข้อมูลเป็น Confidential
-ความรวดเร็วของการเกิดข้อมูล (Real time)	-มีการทวีตการเกิดอุบัติเหตุบน Socialตลอด	-รอรายงานประจำปีจากสำนักงานตำรวจ	-ข้อมูลทวิตเตอร์รวดเร็ว ทันที ณ เวลาเกิดเหตุ แต่ข้อมูลของสำนักงานตำรวจฯเนื่องจากเป็นหน่วยงานรัฐกระบวนการที่ได้มาซึ่งข้อมูลและรายงานผลข้อมูลมีความล่าช้ามาก
-รายละเอียดเพิ่มเติมของความเสียหาย	-มีรายละเอียดเพิ่มในข้อความ Tweet	-ไม่มีรายละเอียดเพิ่ม มีแค่ส่วนข้อมูลที่รายงาน	-ข้อมูลสำนักงานตำรวจแห่งชาติเป็นข้อมูลสรุปไม่สามารถดูรายละเอียดเพิ่มเติมในแต่ละกรณีได้ เนื่องจากข้อมูลเป็น Confidential

ตารางที่ 4.13 แสดงผลการวิเคราะห์เปรียบเทียบระหว่าง รูปแบบข้อมูลที่ได้การทำ Big data จากทวิตเตอร์กับผลที่ได้จากข้อมูลของสำนักงานตำรวจแห่งชาติ (ต่อ)

รายละเอียด	รูปแบบแหล่งข้อมูล จากการทำ Big Data	รูปแบบแหล่งข้อมูล จากสำนักงานตำรวจแห่งชาติ	หมายเหตุ / ข้อจำกัด
-ความรวดเร็วของการเกิดข้อมูล (Real time)	-มีการทวีตการเกิดอุบัติเหตุบน Socialตลอด	-รอรายงานประจำปีจากสำนักงานตำรวจ	-ข้อมูลทวิตเตอร์รวดเร็ว ทันที ณ เวลาเกิดเหตุ แต่ข้อมูลของสำนักงานตำรวจฯเนื่องจากเป็นหน่วยงานรัฐกระบวนการที่ได้มาซึ่งข้อมูลและรายงานผลข้อมูลมีความล่าช้ามาก
-ความน่าเชื่อถือของข้อมูล	-น้อย	- มาก	-ความน่าเชื่อถือของข้อมูลทวิตเตอร์ ไม่มากเท่ากับความน่าเชื่อถือของหน่วยงานรัฐ ข้อมูลจากทวิตเตอร์ไม่สามารถบอกจำนวนอุบัติเหตุที่แท้จริงได้ บอกได้เพียงแค่ว่ามีการทวีตเกี่ยวกับอุบัติเหตุกี่ครั้ง
-การเปิดเผยข้อมูล	-จาก Social มาก	-น้อย	-ข้อมูลจากทวิตเตอร์มีการเปิดเผยข้อมูลสู่สาธารณะมากกว่า แต่มีข้อจำกัดระดับหนึ่งของปริมาณที่จะดึงข้อมูลหากเป็นข้อมูลฟรี
-ความรวดเร็วของการนำมาแสดงผล (Visualization)	-ถ้าข้อมูลเปลี่ยนแปลงจุดที่แสดงในแผนที่เปลี่ยนอัตโนมัติ	-ใช้เวลานำข้อมูลที่ได้มา Plot กราฟแสดงผล	-ข้อมูลจาก Big data พร้อมต่อการประมวลผลที่รวดเร็วมากกว่า ถ้าเป็นข้อมูลของสำนักตำรวจต้องผ่านกระบวนการกรอกข้อมูลและเข้าสู่โปรแกรมเพื่อแสดงผล ซึ่งหลายขั้นตอน

ตารางที่ 4.13 แสดงผลการวิเคราะห์เปรียบเทียบระหว่าง รูปแบบข้อมูลที่ได้การทำ Big data จากทวิตเตอร์กับผลที่ได้จากข้อมูลของสำนักงานตำรวจแห่งชาติ (ต่อ)

รายละเอียด	รูปแบบแหล่งข้อมูล จากการทำ Big Data	รูปแบบแหล่งข้อมูล จากสำนักงานตำรวจแห่งชาติ	หมายเหตุ / ข้อจำกัด
-รูปแบบการแสดงผล	-นำเสนอในรูปแบบหลากหลาย	-นำเสนอเฉพาะตาราง	-โปรแกรมที่ใช้ในการประมวลผล Big Data มีความสวยงาม และมีเครื่องมือการนำเสนอในแบบใหม่ๆอยู่เสมอ เนื่องจากมี Community ในการพัฒนาโปรแกรม อยากใช้โปรแกรมไหนก็แค่เรียก Packageนั้นมาใช้ร่วมกับ Rstudio
-การรองรับข้อมูลจำนวนมากในอนาคต	-โปรแกรมที่ใช้รองรับข้อมูลจำนวนมากได้	-Excelไม่สามารถรองรับข้อมูลจำนวนมากได้ ต้องใช้โปรแกรมอื่นร่วม	-โปรแกรมที่ใช้งานในการทำ Big Data ถูกออกแบบมาให้สามารถรองรับข้อมูลจำนวนมาก และประมวลผลได้อย่างรวดเร็วมากกว่า
-การเชื่อมต่อและการใช้งานร่วมกับโปรแกรมอื่นๆอย่างหลากหลายและมีประสิทธิภาพ	-R Program สามารถเชื่อมต่อโปรแกรมอื่นๆได้อย่างหลากหลาย	-Excel เชื่อมต่อโปรแกรมอื่นได้ แต่อาจจะไม่มีประสิทธิภาพกับข้อมูลขนาดใหญ่	-โปรแกรมที่ใช้งานในการทำ Big Data ถูกออกแบบมาให้สามารถใช้งานกับโปรแกรมอื่นได้อย่างหลากหลาย รองรับข้อมูลปริมาณมาก และประมวลผลได้อย่างมีประสิทธิภาพ

ที่มา : ผู้วิจัย, 2560

จากตารางที่ 4.13 สรุปผลการวิเคราะห์เปรียบเทียบระหว่างรูปแบบข้อมูลที่ได้จากการทำ Big Data กับผลที่ได้จากข้อมูลของสำนักงานตำรวจแห่งชาติ ในด้านของรูปแบบรายละเอียดของรายงานอุบัติเหตุแยกย่อยตามเขต และแยกย่อยตามวัน เวลา นั้น แหล่งข้อมูลที่ใช้มาทำ Big Data มีความละเอียดมากกว่า สำนักงานตำรวจแห่งชาติรายงานสถานที่แบ่งตามจังหวัด และไม่มีการเปิดเผยข้อมูลวันที่และเวลา และไม่มีการแสดงรายละเอียดของมูลความเสียหายรายย่อย อย่างไรก็ตามข้อมูลจากการทำ Big Data และข้อมูลสำนักงานตำรวจแห่งชาติมีการนำเสนอรายงานอุบัติเหตุแยกย่อยตามประเภทของรถ เช่นเดียวกัน แต่ข้อมูลการนำเสนอจากการจัดประเภทรถด้วย Big Data มีความละเอียดมากกว่าสำนักงานตำรวจแห่งชาติ อีกทั้งข้อมูลของสำนักงานตำรวจแห่งชาติการจัดประเภทของรถมาในรายงานแล้วทำให้ไม่ทราบประเภทรถในรายละเอียด ซึ่งอาจเกิดจากการให้คำนิยามประเภทรถที่ไม่เหมือนกันกับข้อมูลที่ใช้กับ Big Data ส่วนสาเหตุของการเกิดอุบัติเหตุนั้น ข้อมูลจากทั้งสองแหล่งมีการนำเสนอทั้งคู่ แต่ข้อมูลที่นำมาประมวลผลใน Big Data มีข้อจำกัดเนื่องจากข้อมูลจากทวิตเตอร์ในการทำ Big Data มักไม่แจ้งสาเหตุของการเกิดอุบัติเหตุและมีการนิยามการเกิดอุบัติเหตุที่แตกต่างกันมาก ทำให้จัดประเภทสาเหตุได้ยาก และมีข้อจำกัดในด้านของระดับความเสียหายของการเกิดอุบัติเหตุไม่ละเอียดชัดเจนเท่าข้อมูลของสำนักงานตำรวจแห่งชาติที่มีการระบุระดับความเสียหายอย่างระเอียดชัดเจน เนื่องจากเป็นข้อมูลจากการสำรวจจริง ณ จุดเกิดเหตุ

ในส่วนของการใช้โปรแกรมร่วมกับข้อมูล การเก็บข้อมูลที่ใช้เทคโนโลยี Big Data มีความรวดเร็วของการเกิดข้อมูล (Real time) มากกว่าข้อมูลที่มาจากสำนักงานตำรวจแห่งชาติ แต่อย่างไรก็ตาม Big data มีข้อจำกัดในเรื่องของความซับซ้อนของข้อมูล ความยากในการสกัดข้อมูล กล่าวคือ ข้อมูลจาก Big Data มีความซับซ้อนมากกว่า เนื่องจากการทวีตของแต่ละบุคคลมีมาตรฐานแตกต่างกัน ข้อมูลมีปริมาณมาก มีการไหลเข้าระบบตลอดเวลาและต่อเนื่องเป็นรายวินาทีในทวิตเตอร์

อย่างไรก็ตามข้อมูลที่นำมาใช้ในการทำ Big Data นั้น เป็นข้อมูลจากทวิตเตอร์ที่มีการเปิดเผยข้อมูลสู่สาธารณะมากกว่า มีความรวดเร็วของการนำมาแสดงผล (Visualization) และมีความสวยงาม อีกทั้งโปรแกรมที่นำมาใช้สามารถรองรับข้อมูลจำนวนมากที่จะเกิดขึ้นได้ในอนาคต รวมถึงการเชื่อมต่อและการใช้งานร่วมกับโปรแกรมอื่นๆอย่างหลากหลายและมีประสิทธิภาพ แต่มีข้อจำกัดที่มีความสำคัญอย่างหนึ่งคือ เรื่องของความน่าเชื่อถือของข้อมูล โดยความน่าเชื่อถือของข้อมูลทวิตเตอร์ ไม่มากเท่ากับความน่าเชื่อถือของหน่วยงานรัฐ ข้อมูลจากทวิตเตอร์ไม่สามารถบอกจำนวนอุบัติเหตุที่แท้จริงได้ บอกได้เพียงแค่ว่ามีการทวีตเกี่ยวกับอุบัติเหตุกี่ครั้งเป็นต้น

ในส่วนของข้อมูล Twitter นั้น มีข้อจำกัดในด้านของความไม่น่าเชื่อถือของข้อมูล และการให้คำจำกัดความของประโยคที่ใช้ในการทวีตนั้นไม่มีมาตรฐานในการทวีต ทำให้การดึงข้อมูล

ทำได้ไม่ครบถ้วน รวมถึงกรณีข้อจำกัดของการใช้ทรัพยากรบุคคลและตัวโปรแกรมที่ซับซ้อน ยากต่อการแบ่งกลุ่มคำหรือแยกข้อความออกจากประโยคตามความหมายต่างๆที่ต้องการ

บทที่ 5
สรุปผลการวิจัย อภิปรายผล

การศึกษางานวิจัยเรื่อง “แนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data ซึ่งเป็นกรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุ” เป็นงานวิจัยแบบกึ่งทดลอง (Quasi - experimental research design) ไม่มีการควบคุมปัจจัยต่างๆแต่เป็นการทดลองใช้โปรแกรมกับข้อมูลส่วนที่มีอยู่แล้ว โดยมีวัตถุประสงค์เพื่อศึกษารายละเอียดและประโยชน์ของการใช้ Big Data รวมถึงศึกษาเครื่องมือหรือซอฟท์แวร์ที่นำมาใช้กับการวิเคราะห์ข้อมูล Big Data และ การนำ Big Data มาประยุกต์ใช้กับธุรกิจ โดยกลุ่มเป้าหมายที่ใช้ในงานวิจัยครั้งนี้คือ ผู้ที่ใช้ทวิตเตอร์และมีการโพสต์ข้อความการเกิดอุบัติเหตุในทวิตเตอร์ อันได้แก่ จราจรและอุบัติเหตุ (@wichansuriyo), สวพ.FM91(@fm91trafficpro), ศูนย์อุบัติเหตุ (@MOT_1356), Traffy.in.th (@traffy), JS100 (@js100radio), เพื่อนเดินทาง (@travel_friendss) และทวีตอยู่ในเขตกรุงเทพมหานคร จำนวน 16,530 ตัวอย่าง โดยดำเนินการเก็บข้อมูลจากทวิตเตอร์ในช่วง 3 ธันวาคม พ.ศ. 2559 ถึง 3 มีนาคม พ.ศ. 2560 เป็นตัวอย่างในการนำ Big Data ที่เป็นแนวทางในการใช้ในธุรกิจ โดยผู้วิจัยสรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะแบ่งออกเป็นดังนี้

1. สรุปผลการวิจัย

ส่วนที่ 1 การวิเคราะห์เครื่องมือในการทำ Big Data

ส่วนที่ 2 การวิเคราะห์ข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์

ส่วนที่ 3 การวิเคราะห์เปรียบเทียบระหว่างรูปแบบข้อมูลที่ได้จากการทำ Big data กับผลที่ได้จากข้อมูลของสำนักงานตำรวจแห่งชาติ

2. การอภิปรายผลการศึกษา
3. ข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ส่วนที่ 1 การวิเคราะห์เครื่องมือในการทำ Big Data

การทำข้อมูล Big Data นั้นสามารถใช้ได้หลายเครื่องมือ เช่น การใช้ Hadoop ที่สามารถ process ข้อมูลขนาดใหญ่ได้ สามารถใช้กับเครื่องคอมพิวเตอร์ที่มีคุณสมบัติระดับปานกลางทั่วไป สามารถเพิ่มขยายจำนวนคอมพิวเตอร์เชื่อมต่อกับระบบในอนาคต และสามารถสำรองข้อมูลได้แบบอัตโนมัติ โดยแบ่งการทำงานออกเป็น 2 ส่วนหลัก คือส่วนเก็บข้อมูล กับส่วนประมวลผล หรือสามารถเชื่อมต่อ Spark หรือใช้แค่เฉพาะ Spark ที่สามารถทำงานร่วมกับ R Studio ได้ โดยมีจุดเด่น คือ สามารถประมวลผลได้เร็วกว่า MapReduce ของ Hadoop และสามารถเขียน SQL เพื่อ Query ข้อมูลได้สะดวกและง่ายขึ้น เหมือนในงานวิจัย สามารถประมวลผล ข้อมูลที่มีขนาดใหญ่ได้เช่นกัน โดยจะใช้ร่วมกันกับ Hadoop หรือไม่ก็ได้

สำหรับ RStudio นั้นสามารถใช้ Package ที่ชื่อว่า Sparklyr ในการเชื่อมต่อ RStudio กับ Spark ให้ทำงานร่วมกันได้ โดยที่RStudio มีโปรแกรมที่เรียกว่า ProgramR ข้อดี คือ สามารถทำงานร่วมกับโปรแกรมอื่นได้อย่างหลากหลาย และสามารถทำงานได้ในหลายระบบปฏิบัติการ อีกทั้งเป็นโปรแกรมที่ใช้ได้ฟรี และสามารถต่อยอดในการทำ Machine Learning หรือ ข้อมูล Big Dataในแบบอื่นๆได้ โดยมีฟังก์ชันการใช้งานในด้านการแสดงผลที่หลากหลายและสวยงาม อย่างไรก็ตามการทำข้อมูลBig Data ไม่จำเป็นต้องใช้งานครบทุกเครื่องมือ ทั้งนี้ขึ้นอยู่กับลักษณะข้อมูล ขอบเขตที่จะใช้ในการทำงานวิจัยหรือการประมวลผลข้อมูล

ส่วนที่ 2 การวิเคราะห์ข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์

จากผลการวิจัยพบว่า การกำหนดช่องทางของแหล่งข้อมูลเกี่ยวกับอุบัติเหตุนั้น แหล่งข้อมูลที่มีความเหมาะสมที่สุดคือ ทวิตเตอร์ ซึ่งสามารถโพสต์ข้อความ รูปภาพ VDO และสถานที่ อีกทั้งด้วยข้อจำกัดของทวิตเตอร์ที่สามารถโพสต์ข้อความได้ 140 ตัวอักษร ทำให้เนื้อหาข้อมูลที่โพสต์มีความกระชับ ได้ใจความสำคัญ และรวดเร็วต่อการนำมาใช้งาน โดยทำการเก็บข้อมูลข่าวที่แจ้งเกี่ยวกับอุบัติใน กทม. ผ่านช่องทางทวิตเตอร์ทั้งหมด 6 แหล่ง ได้แก่ fm91trafficpro, js100radio, MOT_1356, traffy, travel_friendss และ wichansuriyo ผ่าน API ที่ใช้งานฟรีของทวิตเตอร์ ด้วยการดึงข้อมูลจาก Program Rstudio ทั้งหมด 16,530 ตัวอย่าง ทั้งนี้ข้อมูลที่ได้ทำการเก็บมาที่สามารถนำมาใช้ประโยชน์ในการวิเคราะห์ได้นั้น สามารถแบ่งออกเป็น 2 ประเภทหลัก คือ ข้อมูลที่มีโครงสร้างชัดเจน เช่น วันที่และเวลาที่ทำการทวีต และข้อมูลที่ไม่มีโครงสร้างแต่สามารถนำมาสะกัดใช้ประโยชน์ได้ คือ ข้อความที่เกิดจากการทวีต ส่วนข้อมูลมูลอื่นๆที่ดึงมานั้น เช่น พิกัดสถานที่ ส่วนใหญ่มีข้อมูลไม่ครบถ้วนจึงไม่สามารถนำมาใช้ในการวิเคราะห์ได้

อย่างไรก็ตามได้นำข้อมูลสถิติการเกิดอุบัติเหตุของสำนักงานตำรวจแห่งชาติ มาเป็นต้นแบบในการเก็บหรือดึงข้อมูลจากข้อความในทวิตเตอร์ ทั้งหมด 5 ด้านหลัก ได้แก่ สาเหตุจาก

บุคคล สาเหตุจากสิ่งแวดล้อม สาเหตุจากอุปกรณ์ขับขี่ ประเภทของรถที่เกิดอุบัติเหตุต่างๆ และความเสียหายที่เกิดกับบุคคล โดยข้อมูลสำหรับปี พ.ศ. 2558 ด้านสาเหตุจากบุคคล การเกิดอุบัติเหตุในกรุงเทพมหานครส่วนใหญ่เกิดจากการที่ไม่ยอมรถที่มีสิทธิไปก่อน คิดเป็นร้อยละ 22.06 และข้อมูลด้านสิ่งแวดล้อมส่วนมากเกิดจาก ถนนแคบ คิดเป็นร้อยละ17.82 ด้านสาเหตุจากอุปกรณ์ขับขี่ เกิดอุบัติเหตุมากที่สุด คิดเป็นร้อยละ 8.95 ส่วนประเภทรถที่เกิดอุบัติเหตุมากที่สุด คือประเภทรถยนต์นั่ง คิดเป็นร้อยละ 41.49 โดยความเสียหายเหล่านี้ ไม่ว่าจะเป็นการตาย, บาดเจ็บสาหัส หรือ บาดเจ็บเล็กน้อย จะเป็นเพศชายสูงกว่าเพศหญิงจำนวนมาก จากนั้นจึงทำการดึงข้อมูล 5 มิติหลัก ที่มีสถิติการเกิดอุบัติเหตสูงสุดจากข้อความทวิตเตอร์ คือ สาเหตุจากบุคคล, สาเหตุจากสิ่งแวดล้อม, สาเหตุจากการอุปกรณ์ที่ใช้ขับขี่, ประเภทของรถที่เกิดเหตุ และความเสียหายที่เกิดกับบุคคล

การเลือกโปรแกรมที่ใช้ในการดึงข้อมูลนั้นได้เลือกจากโปรแกรมที่ผู้วิจัยมีทักษะการใช้งานและตัวโปรแกรมสามารถเชื่อมต่อกับโปรแกรมอื่นๆที่ใช้กับข้อมูล Big Data ได้อย่างหลากหลาย อีกทั้งสามารถใช้งานได้ฟรีไม่เสียค่าใช้จ่าย โดยมีลำดับการใช้โปรแกรม ดังนี้คือ 1. ใช้โปรแกรม R ในการดึงข้อมูลผ่าน API Twitter 2. ข้อมูลจาก API Twitter จะให้ข้อมูลที่เป็นโครงสร้างระดับหนึ่ง 3. ใช้Program Rในการสกัดข้อมูลที่ไม่มีโครงสร้าง เช่น ข้อความ Twitter ร่วมกับโปรแกรมอื่นที่สามารถดึงข้อมูลเป็น SQL เช่น Sparklyr, mysql เป็นต้น อย่างไรก็ตามการดึงข้อมูลข้างต้นจากทวิตเตอร์ที่ได้มาเป็น File.CSV นั้นต้องตรวจสอบความถูกต้องของข้อมูลด้วยว่าถูกต้องครบถ้วนตามแหล่งที่มาของทวิตเตอร์และต้องผ่าน Process การสกัด เตรียมข้อมูลให้พร้อมกับการวิเคราะห์

เมื่อได้ข้อมูลที่พร้อมแล้วจึงนำมาวิเคราะห์และนำเสนอเป็นกราฟหรือตาราง โดยเริ่มจากโปรแกรมที่นำเสนอ ผ่านPackage shiny ของโปรแกรม R โดยแบ่งการนำเสนอออกเป็น 3ส่วนหลัก ได้แก่ สถานที่ (Location), เวลาต่างๆ(Times) และประเภทของรถ (Vehicle types) และอีก 3 ส่วนที่นำเสนอในรูปของตาราง ผ่าน SQLได้แก่ เพศของคนขับ, สาเหตุของอุบัติเหตุ, และความเสียหายในส่วนของการเสียชีวิต ซึ่งภาพรวมการทวีตการเกิดอุบัติเหตุของทุกเขตนั้น การทวีตการเกิดอุบัติเหตุมากที่สุดคือ ช่วงเวลาประมาณ 10 นาฬิกา รองลงมาคือ 7.00 น. 9.00 น. และ 8.00 น. และเขตลาดพร้าว มีการทวีตการเกิดอุบัติเหตุสูงที่สุด รวม 160 ครั้งในรอบ 3 เดือน ตั้งแต่ 3 ธันวาคม พ.ศ. 2559 ถึง 3 มีนาคม พ.ศ. 2560 โดยทวีตเวลา 6.01 – 9.00 น. เป็นจำนวนมากที่สุดถึง 33 ครั้ง แต่อย่างไรก็ตามมีการทวีตถึงการเกิดอุบัติเหตุที่เขตลาดพร้าว แต่ไม่บอกเวลาการเกิดเหตุถึง 42 ครั้ง นอกจากนี้ในส่วนของการแสดงผลสามารถนำเสนอจำนวนและเขตที่ทำการทวีตอุบัติเหตุผ่านแผนที่ (Google map) ด้วยโปรแกรม R ที่ชื่อ Package Rshinyได้อย่างสวยงามและดูได้ง่าย รวมถึงการ

นำเสนอข้อมูลการทวีตการเกิดอุบัติเหตุจากทวิตเตอร์เป็นกราฟแท่งโดยสามารถเลือกเขต เวลา ที่ต้องการสำหรับการนำเสนอจำนวนของประเภทของรถที่ถูกทวีตการเกิดอุบัติเหตุในทวิตเตอร์

อย่างไรก็ตามการนำเสนอในส่วนของเพศคนขับและสาเหตุของความเสียหายต่อบุคคลนั้นไม่สามารถทำได้เนื่องจากมีข้อมูลที่สามารถสกัดออกจากข้อความในทวิตเตอร์ได้จำนวนน้อยมาก ผู้วิจัยจึงสามารถนำเสนอได้เฉพาะความเสียหายในส่วนของการเสียชีวิตที่เป็นความเสียหายที่สร้างความรุนแรงมากที่สุดเมื่อเทียบกับความเสียหายอื่นๆที่บริษัทประกันต้องคำนึงถึงเป็นอันดับแรก โดยเขตลาดพร้าวมีการทวีตถึงความเสียหายที่มีการเสียชีวิตมากที่สุด จำนวนทั้งหมด 5 ครั้ง

ส่วนที่ 3 การวิเคราะห์เปรียบเทียบระหว่าง รูปแบบข้อมูลที่ได้จากการทำ Big data กับผลที่ได้จากข้อมูลของสำนักงานตำรวจแห่งชาติ

จากผลการวิเคราะห์เปรียบเทียบระหว่างข้อมูลที่ได้จากการทำ Big Data มีข้อดีที่มีมากกว่าในเรื่องของรายละเอียดย่อยในการนำเสนอรายงาน เช่น อุบัติเหตุการทวีตแยกย่อยตามเขตต่างๆ เวลาต่างๆ และรายละเอียดเพิ่มเติมที่สามารถดูข้อมูลได้เป็นแต่ละกรณี รวมถึงจุดเด่นในเรื่องของความรวดเร็วของข้อมูล สามารถรองรับข้อมูลได้ปริมาณมากทั้งในปัจจุบันรวมถึงการรองรับข้อมูลจำนวนมากในอนาคต และมีความรวดเร็ว อีกทั้งมีความสวยงามของการนำมาแสดงผล นอกจากการทำ Big Data นี้ยังสามารถเชื่อมต่อและการใช้งานร่วมกับโปรแกรมอื่นๆอย่างหลากหลายและมีประสิทธิภาพ แต่อย่างไรก็ตามข้อจำกัดหลักที่สำคัญของการใช้ข้อมูลทวิตเตอร์ในการทำ Big Data คือ เรื่องของความน่าเชื่อถือของข้อมูลไม่มากเท่ากับความน่าเชื่อถือของหน่วยงานรัฐ ข้อมูลจากทวิตเตอร์ไม่สามารถบอกจำนวนอุบัติเหตุที่แท้จริงได้และส่วนใหญ่ไม่แจ้งสาเหตุการเกิดอุบัติเหตุ บอกได้เพียงแค่ว่ามีการทวีตเกี่ยวกับอุบัติเหตุกี่ครั้ง นอกจากนี้ข้อจำกัดในข้อมูล Big Data ยังมีในเรื่องของความซับซ้อนของเครื่องมือการวิเคราะห์และข้อมูลที่ไม่มีโครงสร้างทำให้ยากต่อการประมวลผล เป็นต้น

5.2 การอภิปรายผลการศึกษา

จากผลการศึกษางานวิจัยแนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data ซึ่งเป็นกรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุ สามารถอภิปรายผลโดยใช้แนวคิดและทฤษฎีที่เกี่ยวข้องดังนี้

จากการสรุปข้อมูลจากการวิเคราะห์ข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์

จากวัตถุประสงค์ข้อที่ 1 เพื่อศึกษารายละเอียดและประโยชน์ของการใช้ Big Data

ด้านการตรวจสอบข้อมูลที่ได้ดึงมากับ application twitter ว่าข้อมูลที่ได้มานั้นตรงกันหรือไม่ โดยผู้วิจัยได้สุ่มการตรวจสอบข้อมูลลำดับแรก ข้อมูลลำดับที่ 50 และลำดับที่ 100 โดยดึงข้อมูล 100 ตัวเอย่าง จาก 16,530 ตัวอย่าง โดยนำเทคนิคที่ได้จากการทำงานด้านข้อมูลกับบริษัทประกันภัยมาใช้ เพราะหากข้อมูลที่สุ่มมีความผิดพลาดมักจะเห็นได้ชัดจากข้อมูลแรกและข้อมูลท้ายสุดซึ่งใกล้เคียงกับงานวิจัยของ ข้าวทิพย์ ตันติวรวงศ์ (2558) ที่ระบุว่า การทำงานกับข้อมูลขนาดใหญ่ไม่จำเป็นต้องทำงานกับข้อมูลทั้งหมด แต่ควรสุ่มตัวอย่างข้อมูลด้วยปริมาณที่เหมาะสม

ด้านประโยชน์ของโปรแกรม Program R มีข้อดี คือ สามารถทำงานร่วมกับโปรแกรมอื่นได้อย่างหลากหลาย และสามารถทำงานได้ในหลายระบบปฏิบัติการ อีกทั้งเป็นโปรแกรมที่ใช้ได้ฟรี และสามารถต่อยอดในการทำ ซึ่งสอดคล้องกับงานวิจัยของ พนิดา ตันศิริ (2556) ที่ระบุว่า แนวทางในการพัฒนาผลิตภัณฑ์โดยการนำเทคโนโลยีและนวัตกรรมใหม่ๆ มาใช้วิเคราะห์ข้อมูลขนาดใหญ่ เพื่อนำผลจากการวิเคราะห์ที่ได้มาปรับปรุงและวางแผนการทำงานของธุรกิจ ตลอดจนลดค่าใช้จ่ายที่สูงในการจัดซื้อและการปรับปรุงฮาร์ดแวร์ที่จำเป็นต่อการรองรับข้อมูลจำนวนมาก

จากวัตถุประสงค์ข้อที่ 2 เพื่อศึกษาหาเครื่องมือหรือซอฟท์แวร์ที่นำมาใช้กับการวิเคราะห์ Big Data

จากวิเคราะห์เครื่องมือในการทำ Big Data

ด้านข้อมูลจาก Social Media หากต้องการดึงข้อมูลเกี่ยวกับอุบัติเหตุ แหล่งข้อมูลจากทวิตเตอร์เป็นแหล่งข้อมูลที่มีความเหมาะสม ซึ่งเป็น Platform ที่สามารถโพสต์ได้ทั้งข้อความ รูปภาพ VDO และ Location และมีจุดเด่นในด้าน Social Interaction โดยข้อมูลที่ใช้ทวีตนั้นสั้นกระชับได้ใจความสำคัญ เนื่องจากถูกจำกัดไว้ที่ 140 ตัวอักษร นอกจากนี้ยังมีในส่วนของความเร็วในการโหลด อ่าน และกระจายข้อมูลอย่างชัดเจน เหมาะสมกับเหตุการณ์อุบัติเหตุที่ควรเป็นข้อความที่สั้นกระชับได้จำความและประเด็นสำคัญกับผู้อ่านอย่างรวดเร็ว ส่วนข้อมูลจาก Facebook เหมาะสมกับข้อความยาวๆ ใช้สื่อสารกับครอบครัวและเพื่อนๆสนิทอย่างแท้จริง เน้นการดูข้อมูลผ่านๆไปเรื่อยๆและนิยมใช้ในการโฆษณาทางการตลาด ส่วนข้อมูล Youtube จะเป็น VDO ที่เป็นตัวแทนในการอธิบายข้อความหรือเนื้อหา และInstagram เหมาะสมกับการโพสต์ภาพหลายๆภาพ ซึ่งทั้ง Youtube และ Instagram นั้นต่างนิยมใช้ในการโฆษณาทางการตลาด ซึ่งใกล้เคียงกับงานวิจัยของ ทวีวัฒน์ ขนาน (2558) ที่ระบุว่า สามารถวิเคราะห์ Big Data เพื่อติดตามและวิเคราะห์การตอบสนองของผู้ใช้บริการ ผ่านการติดตาม Social Media Twitter, Facebook, Youtube ซึ่งช่วยประเมินในส่วนของโฆษณาการตลาดใหม่ๆ

ด้านของฐานข้อมูลและตัวแปรที่มีความเกี่ยวข้องและจำเป็นต้องใช้ในการวิเคราะห์ข้อมูลในงานวิจัยนี้ ได้แก่ วันที่ Tweet, เวลาที่ Tweet, เขตที่เกิดเหตุ, เพศของคนขับ, สาเหตุการเกิดอุบัติเหตุจากบุคคล สิ่งแวดล้อม หรืออุปกรณ์ที่ใช้ขับขี่, ประเภทรถที่เกิดอุบัติเหตุ และประเภทความเสียหาย ซึ่งผู้วิจัยจึงได้ทำการจัดรูปแบบข้อความเหล่านั้นให้อยู่ใน Column เป็นโครงสร้างที่ชัดเจนและง่ายต่อการนำไปวิเคราะห์มากยิ่งขึ้นซึ่งสอดคล้องกับงานวิจัยของ สุวิมล ประทุม (2555) ที่ระบุว่าระบบที่ฐานข้อมูลขนาดใหญ่จำเป็นต้องมีการออกแบบฐานข้อมูลที่เหมาะสมทั้งนี้ต้องศึกษาความสัมพันธ์ของข้อมูล โครงสร้างข้อมูล การเข้าถึงข้อมูลและกระบวนการที่โปรแกรมประยุกต์จะเรียกใช้ฐานข้อมูล

จากวัตถุประสงค์ข้อที่ 3 เพื่อศึกษาข้อจำกัดของการนำข้อมูล Big Data จาก Twitter มาใช้

ด้าน Platform ที่เป็นแหล่งข้อมูล สำหรับประเทศไทยและทั่วโลก มีการใช้ Social Media Platforms ที่แตกต่างกันออกไป แต่อย่างไรก็ตาม 4 อันดับแรกที่นิยมใช้กันมาก ได้แก่ Twitter, Facebook, Instagram (Gary Vayerchuk, 2013) และรายงาน 3 อันดับแรก Facebook, Twitter และYoutube สำหรับผู้ใช้ในประเทศไทย (socialbakers, 2017) ผู้วิจัยจึงได้ทำการเปรียบเทียบเงื่อนไขการใช้งานในแต่ละ Platform ซึ่งแบ่งออกเป็น 4 จำพวกหลัก ได้แก่ การใช้เพื่อการโฆษณา, ใช้เพื่อ Social Interaction, ความรวดเร็วในการเข้าถึง การอ่าน กระจายข้อมูล รวมถึงการ Load ข้อมูล และ Style ของผู้ใช้งานที่แตกต่างกันไป สอดคล้องกับงานวิจัยของ Pouria Pirzadeh (2015) ที่ระบุว่าสถาปัตยกรรมและการออกแบบการตัดสินใจของระบบ Big Data ที่มีการเติบโตอย่างรวดเร็วได้สร้างความท้าทายที่สำคัญในการสร้างเกณฑ์มาตรฐานสำหรับการประเมินและเปรียบเทียบแพลตฟอร์มของระบบข้อมูลที่มีขนาดใหญ่ โดยการประเมินนั้นต้องมีความครอบคลุม และเพียงพอในแง่มุมต่างๆ ทั้งในแง่ของลักษณะของข้อมูลและปริมาณข้อมูลที่จะใช้ทำงาน เพื่อทำให้ทราบถึงตัวชี้วัดประสิทธิภาพในการทำงานและสิ่งที่ต้องแก้ไขเพื่อบรรลุเป้าหมายที่กำหนดไว้

ด้านประเภทโครงสร้างของข้อมูล โดยข้อมูลประเภทไม่มีโครงสร้างในงานวิจัย ได้แก่ ข้อความที่เกิดจากการทวีต ซึ่งข้อมูลนี้มีความสำคัญเนื่องจากข้อความเหล่านี้ประกอบด้วยข้อมูลที่เป็นประโยชน์ในการนำมาใช้วิเคราะห์แต่อยู่ในลักษณะของข้อความทวีตยาวๆ ไม่แบ่งแยกข้อมูลเป็นส่วนๆ จึงต้องทำการสกัดข้อมูล เพื่อนำมาใช้ประโยชน์ต่อไป ซึ่งสอดคล้องกับงานวิจัยของ Ylli Sadikaj (2016) ที่ระบุว่า การศึกษาเกี่ยวกับบริการประกันสุขภาพส่วนบุคคลโดยใช้ Big Data การศึกษาครั้งนี้ใช้ซอฟต์แวร์ในการรวบรวมข้อมูลผ่านหน้าเว็บไซต์ของบริษัทประกัน โดยดึงข้อมูลผู้ให้บริการ แผนประกันสุขภาพต่างๆ ซึ่งเป็นข้อมูลที่มีจำนวนมากและไม่มีโครงสร้าง

ด้านการสกัดข้อมูล การทวีตในส่วนของความเสียหายเป็นข้อมูลที่ต้องสกัดออกมาอย่างซับซ้อน จึงสกัดหรือกรองได้ข้อมูลจำนวนน้อย และอาจไม่ทราบความเสียหายที่แท้จริง เนื่องจากคนทวีตอาจจะประเมินจากสถานการณ์ที่พบเห็นหรือการบอกต่อ อีกทั้งเมื่อได้สอบถามจากผู้เชี่ยวชาญด้านข้อมูลและนักคณิตศาสตร์ประกันภัย ให้ความเห็นว่า ข้อความที่ใช้ในการทวีตมีการจับความหมายเพื่อนำเอาข้อมูลออกมาใช้ได้ยาก เนื่องจากข้อมูลมีความซับซ้อน ต้องใช้เทคนิคในการดึงข้อมูลอื่นเพิ่มเติม เช่น การทำ Deep Learning หรือ Machine learning ซึ่งต้อง Model ในการออกแบบให้คลอบคลุมจำนวนมาก ซึ่งสอดคล้องกับงานวิจัยของ สำรวย กมลายุตต์ (2557) ที่ระบุว่า นักวิจัยสถาบัน MIT ได้จำลองคอมพิวเตอร์เพื่อทำการวิเคราะห์โดยใช้เทคนิคการทำเหมืองข้อมูลและการเรียนรู้ของเครื่องจักร (Machine learning) ในการวิเคราะห์กลั่นกรองข้อมูล ช่วยเพิ่มความสามารถให้แก่แพทย์ในการวินิจฉัยข้อมูลของผู้ป่วยโรคหัวใจมากขึ้นและลดความผิดพลาดในการอ่านข้อมูลลดลงได้ถึงร้อยละ 70

5.3 ข้อเสนอแนะ

จากการศึกษาแนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data ซึ่งเป็นกรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุ ผู้วิจัยได้ทำการวิเคราะห์ผลงานวิจัยและได้นำเอาผลการวิจัยมาสรุปเป็นข้อเสนอแนะ 2 ส่วน คือ ข้อเสนอแนะในการนำผลการวิจัยไปใช้และข้อเสนอแนะในการทำวิจัยครั้งต่อไป ดังนี้

5.3.1ข้อเสนอแนะในการนำผลการวิจัยไปใช้

5.3.1.1 บริษัทประกันหรือบริษัทที่มีความสนใจในการเรียนรู้เครื่องมือในการทำ Big Data สามารถนำผลการวิจัยไปใช้เป็นพื้นฐานความรู้ในภาพรวมของเครื่องมือหลักๆ ในการสร้างระบบประมวลผลด้วย Big Data เช่น สเปคโดยประมาณของ Notebook และระบบปฏิบัติการ Ubuntu ที่นิยมใช้เป็นเครื่องมือหลักของโปรแกรมที่จะใช้กับ Big Data รวมถึงความสามารถและการทำงานหลักของ Software Apache Hadoopว่านำไปใช้งานในรูปแบบใด และมี Software ใดที่สามารถทดแทนกันได้ เช่น การใช้ Sparklyr ผ่านโปรแกรมของ Rstudio ทดแทนในส่วนการประมวลผลแทน Hadoop นอกจากนี้ยังเป็นแนวทางโปรแกรม R ในด้านของการนำเสนอผลการวิเคราะห์ (Visualization) อีกด้วย

5.3.1.2 บริษัทประกัน สามารถนำผลวิจัยไปใช้เป็นแนวทางในการนำ Big Data มาใช้วิเคราะห์ข้อมูลได้อย่างรวดเร็ว ด้วยงบประมาณที่ประหยัดด้วยการใช้โปรแกรมที่ไม่เสีย

ค่าใช้จ่าย อย่างเช่น โปรแกรมR และสามารถนำไปใช้วางแผนการเก็บข้อมูลในมิติต่างๆให้เพียงพอต่อการนำมาวิเคราะห์ให้ครบถ้วนและง่ายต่อการนำมาใช้งาน รวมถึงเป็นแนวทางในการใช้งานในด้านต่างๆ เช่น

ด้านการกำหนดช่องทางของแหล่งข้อมูลที่จะนำมาใช้ ควรมีแนวทางในการกำหนดอย่างไร แต่ละแหล่งข้อมูลมีจุดเด่นและความเหมาะสมกับวัตถุประสงค์ของการนำไปวิเคราะห์อย่างไร โดยเฉพาะข้อมูลจากแหล่ง ทวิตเตอร์ มีจุดเด่นของ Platform และข้อมูลที่เกิดจากทวิตเตอร์อย่างไร

ด้านประเภทโครงสร้างของข้อมูล ควรมีแนวทางในการแบ่งประเภทข้อมูลอย่างไร รวมถึงรายละเอียดของข้อมูลในแต่ละ Field มีความแตกต่างของข้อมูลอย่างไรและเพราะเหตุใด

ด้านโปรแกรมที่ใช้ในการดึงข้อมูล ควรมีลำดับขั้นในการใช้โปรแกรมอย่างไร โปรแกรมใดบ้างที่ควรใช้พร้อมกันและใช้เพื่ออะไร เช่น ใช้เพื่อสกัดข้อมูล หรือ การใช้ Rshiny ในการแสดงผลของข้อมูล เป็นต้น

ด้านการตรวจสอบความถูกต้องของข้อมูลในเบื้องต้น ควรมีการแบ่งการตรวจสอบเบื้องต้นอย่างไรก่อนนำข้อมูลทั้งหมดไปใช้

ด้านรายละเอียดข้อมูลที่นำมาใช้วิเคราะห์ ควรมีมิติของข้อมูลที่เกี่ยวข้องในด้านต่างๆอย่างไร และใช้เกณฑ์ใดเป็นมาตรฐานเพิ่มเติมในการดึงข้อมูลในแต่ละมิติ

5.3.1.3 บริษัทประกันหรือบริษัทที่มีความสนใจ สามารถนำแนวทางในการนำเสนอ เช่น รูปแบบการแสดง Visualization ต่างๆ ไปปรับใช้ตามข้อมูลและความเหมาะสมที่ต้องการนำเสนอให้กับผู้บริหารหรือผู้ที่ใช้งานเป็นประจำได้ เช่น การนำเสนอในรูปแบบของแผนที่ (Map) ที่มีข้อมูลแบ่งตามเขต โดยให้ขนาดของวงแทนจำนวนการทวีตอุบัติเหตุ และสามารถเรียกดูผลในแต่ละเขตได้ตามต้องการ รวมถึงการทำกราฟแท่งให้แสดงผลเปลี่ยนไปตามข้อมูลที่ต้องการดูตามเขตและเวลาที่ต้องการอย่างรวดเร็ว เป็นต้น

5.3.1.4 ผู้บริหารหรือบริษัทที่มีความสนใจใน Big Data สามารถทราบถึงข้อจำกัดได้เบื้องต้น ก่อนตัดสินใจศึกษาข้อมูลเชิงลึกของการนำ Big Data เข้ามาใช้กับองค์กรได้ รวมถึงบริษัทประกันสามารถทราบถึงการดึงรายละเอียดข้อมูลอุบัติเหตุว่ามีข้อจำกัดอย่างไรในส่วนของสาเหตุการเกิดอุบัติเหตุ

5.3.1.5 นักการตลาดสามารถนำผลการวิจัยในส่วนของประโยชน์ที่สามารถวิเคราะห์ข้อมูลจาก Big Data ที่อยู่บนทวิตเตอร์ไปสร้างแคมเปญทางการตลาดได้อย่างรวดเร็ว ลดอัตราการเกิดอุบัติเหตุตามเขตต่างๆ ลดอัตราการตาย ที่เป็นค่าใช้จ่ายหลักของบริษัท

ประกันชีวิตได้ อีกทั้งนักการตลาดสามารถใช้เป็นแนวทางในการให้ผู้บริโภคป้อนข้อมูลเกี่ยวกับสาเหตุการเกิดอุบัติเหตุผ่านทวิตเตอร์ได้อย่างไร เพื่อเป็นข้อมูลในการวิเคราะห์พฤติกรรมที่นำไปสู่การเกิดอุบัติเหตุ นอกจากนี้ในส่วนของธุรกิจอื่น เช่น ธุรกิจห้างสรรพสินค้า หรือธุรกิจธนาคารที่มีสาขาจำนวนมาก นักการตลาดสามารถเก็บข้อมูลจากสาขามาใช้ในการวิเคราะห์อย่างรวดเร็วแทนที่จะรอสาขาต่างๆส่งข้อมูลมา ลดปัญหาเรื่องระยะเวลาในการรอ Report และลดปัญหาในเรื่องของรูปแบบของการทำรายงานที่แตกต่างกันในแต่ละสาขา รวมถึงสามารถดูผลการวิเคราะห์และให้ข้อแนะนำที่สาขานั้นๆควรปรับปรุงพัฒนาได้ทันทีที่ข้อมูลนั้นมีการ Update และสามารถปรับการดูการแสดงผลกราฟต่างๆได้อย่างยืดหยุ่น ไม่ว่าจะเป็นข้อมูลการซื้อขายผลิตภัณฑ์ เช่น เวลาที่ลูกค้าเข้าสาขา การใช้จ่ายซื้อผลิตภัณฑ์ของลูกค้าทั้งเวลาที่สูงสุดและต่ำสุด รวมถึงเรื่องที่ลูกค้ามีปัญหาและเขียน Comment สาขา มากที่สุดที่จำเป็นต้องจัดการอย่างเร่งด่วน เป็นต้น

5.3.2ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

5.3.2.1การทำวิจัยในครั้งนี้ เป็นการวิจัยเฉพาะกลุ่มเป้าหมายที่มีการทวีตในเขตกรุงเทพมหานคร ในช่วงระยะเวลาหนึ่ง จึงทำให้ผลการวิจัยไม่สามารถนำมาใช้แทนสถิติการเกิดอุบัติเหตุทั้งหมดในเขตกรุงเทพมหานคร เพราะอาจจะเกิดอุบัติเหตุแต่ไม่มีการทวีต จึงควรเก็บข้อมูลจากแหล่งอื่นๆเพิ่มเติม โดยอาจจะเป็นข้อมูลที่เกิดจากหน่วยงานภาครัฐที่น่าเชื่อถือ

5.3.2.2 การกรองข้อมูลเพื่อนำมาใช้ในการวิเคราะห์นั้น เป็นการกรองข้อมูลอย่างง่าย ไม่ละเอียดมากนักและมีข้อจำกัดในเรื่องการกรองข้อมูลภาษาไทยที่เกิดจากการทวีตโดยการสะกดคำผิดของคนที่ทวีต ทำให้ไม่สามารถเก็บข้อมูลเหล่านี้มาวิเคราะห์ได้ครบ ทั้งนี้ควรจะใช้เครื่องมืออื่นๆเข้ามาช่วย เช่นการใช้ Deep Learning หรือ Machine Learning ที่เป็นการคาดเดาการสะกดคำ คล้ายลักษณะการทำงาน Search หาคำใน Google ได้อย่างถูกต้อง แม้ผู้พิมพ์จะพิมพ์ผิดหรือไม่ครบถ้วนก็สามารถคาดเดาความหมายได้ ซึ่งสามารถลดปัญหาข้อมูลที่เป็นขยะได้มาก ซึ่งความสามารถเหล่านี้โปรแกรม R สามารถทำต่อยอดได้

5.3.2.3 สามารถนำโปรแกรม R ไปวิเคราะห์ข้อมูล โดยใช้งานร่วมกับ Sparklyr เพื่อทำให้ข้อมูลประมวลผลเร็วขึ้น และเก็บข้อมูลปริมาณมากด้วยการเชื่อมต่อเทคโนโลยี Hadoop โดยการแบ่งออกเป็นหลายๆ Cluster หรือสามารถนำขั้นตอนการวิจัยในครั้งนี้ไปใช้กับเครื่องมืออื่นๆเพิ่มเติมได้ เช่น Hive, Apeche Spark, Cloudera และ Amezon Web Services เป็นต้น รวมถึงการใช้วิเคราะห์ข้อมูลแบบ Real-time โดยการดึงข้อมูลผ่านระบบ Cloud

5.3.2.4 สามารถนำงานวิจัยไปต่อยอดการทำ Visualization แสดงกราฟต่างๆในแบบอื่นๆ ผ่านPackage ของโปรแกรม R ได้อย่างสวยงาม แปลกใหม่ และเข้าใจง่าย

5.3.2.5 ควรมีบุคลากรและทรัพยากรที่มากพอในการวิเคราะห์เพิ่มเติม รวมถึงเทคนิคในการประมวลผลข้อมูล เพื่อประสิทธิภาพของการวิเคราะห์ข้อมูล

บรรณานุกรม

กวี รุจิรัตน์. (2556). *Big Data อภิมหาข้อมูล* (หน้า 196). สำนักพิมพ์ทรูไลฟ์.

ข่าวจราจร สวพ.FM91. (2016). *ตัวอย่างทวิตเตอร์ของข่าวจราจร สวพ.FM91 หรือชื่อทวิตเตอร์fm91.* ค้นเมื่อ 5 มกราคม 2560, จาก https://twitter.com/fm91trafficpro?lang=en

ข้าวทิพย์ ตันติวรวงศ์. (2558). *การนำเสนอข้อมูลขนาดใหญ่ด้วยแท็บโบล*. ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี, จุฬาลงกรณ์มหาวิทยาลัย.

ณิชจิรัชย์ ตั้งคำ. (2556). *แผนธุรกิจเพื่อระบบวิเคราะห์ Big Data ของบริษัทฮิวเลตต์-แพคการ์ดประเทศไทย สำหรับธุรกิจธนาคาร*. ปริญญาบริหารธุรกิจมหาบัณฑิต, คณะพาณิชยศาสตร์และการบัญชี, จุฬาลงกรณ์มหาวิทยาลัย.

ทวีวัฒน์ ขนาน. (2558). *ระบบวิเคราะห์ข้อมูลขนาดใหญ่เพื่อสนับสนุนการตัดสินใจในการบริหารช่องทางการให้บริการของธนาคารผ่านเครื่องรับจ่ายอัตโนมัติ*. หลักสูตรวิทยาศาสตร์มหาบัณฑิต, สาขาวิชาระบบสารสนเทศทางการจัดการ,คณะพาณิชยศาสตร์และการบัญชี, จุฬาลงกรณ์มหาวิทยาลัย.

ทีเอ็นที มีเดียแอนเน็ตเวิร์คจำกัด. (2558). 5 เทรนด์ใหม่ของลูกค้าที่คุณจำเป็นต้องรู้. ค้นเมื่อ 3 ตุลาคม 2559, จาก http://www.tnt.co.th/en/news/126-5-new-trend-of-customers-want-you-know

ธนพร สิทธิชัยวิเศษ. (2557). *การวิเคราะห์ข้อมูลขนาดใหญ่เพื่อการดำเนินการโดยใช้เอสเอพี ฮานา (Actionable Big Data Analytics by Using SAP HANA)*. ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี, จุฬาลงกรณ์มหาวิทยาลัย.

ปิยะภัสร์ โรจน์รัตนวาณิชย์. (2556). *แนวทางการคุ้มครองข้อมูลใน Big Data: ศึกษาประเด็นความเป็นส่วนตัวและความมั่นคงปลอดภัยของข้อมูล* (หน้า18). กรุงเทพมหานคร : นิติศาสตร์มหาบัณฑิต มหาวิทยาลัยกรุงเทพ.

พนิดา ตันศิริ. (2556). *ข้อมูลขนาดใหญ่กับความท้าทาย*. ค้นเมื่อ 1 มกราคม 2560, จาก http://www.bu.ac.th/knowledgecenter/executive_journal/jan_mar_13/pdf/aw03.pdf

บรรณานุกรม (ต่อ)

พสุ เดชะรินทร์. (2556). Big Data หรือ อภิมหาข้อมูล. ค้นเมื่อ 3 ตุลาคม 2559, จาก http://library.acc.chula.ac.th/PageController.php?page=FindInformation/ArticleACC/2556/Pasu/BangokBiznews/B2901131

ภูวดล เหมวิวรรณ. (2011). ความแตกต่างระหว่างFacebook และTwitter. ค้นเมื่อ 17 เมษายน 2560, จาก http://www.puwadon.com/post.php?id=18

วรรณเพ็ญ บุญเพ็ญ. (2558). 2009 Influenza (H1N1). ค้นเมื่อ 14 ตุลาคม 2559, จาก https://www.tcdc.or.th/upload/iblock/d99/d99204aef476ee28093842b5c6954065.pdf

วิกเตอร์ เมเยอร์ ชอนเบอร์เกอร์ และ เคนเน็ต ซูเคีย. (2556). *Big Data อภิมหาข้อมูล* (หน้า 24). กรุงเทพมหานคร : สำนักพิมพ์ทรู ไลฟ์ .

สำนักงานสถิติแห่งชาติ. (2553). *อุบัติเหตุการจราจรทางบกจากสำนักงานตำรวจแห่งชาติ*. ค้นเมื่อ 15 เมษายน 2560, จาก http://service.nso.go.th/nso/web/statseries/statseries21.html http://service.nso.go.th/nso/web/statseries/statseries21.html

สำรวย กมลายุตต์. (2557), *แบบรายงานผลโครงการศึกษาเพิ่มเติมด้าน Big Data Governance and Big Analytic ณ Ludwigshafen University of Applied Sciences (Hochschule Ludwigshafen am Rhein) เมือง Ludwigshafen ประเทศสหพันธ์สาธารณรัฐเยอรมัน*น. ค้นเมื่อ 5 มกราคม 2560, จาก http://libarts.stou.ac.th/UploadedFile/รายงานBig%20Data%20Analyticฉบับสมบูรณ์.pdf

สุภัค ลายเลิศ. (2558). *Big Data ความสำคัญที่ใหญ่กว่าชื่อ*. ค้นเมื่อ 7 มกราคม 2560, จาก http://www.telecomjournalthailand.com/big-data-%E0%B9%83%E0%B8%AB%E0%B8%8D%E0%B9%88%E0%B8%81%E0%B8%A7%E0%B9%88%E0%B8%B2%E0%B8%8A%E0%B8%B7%E0%B9%88%E0%B8%AD/

สุวิมล ประทุม. (2555). *การปรับปรุงประสิทธิภาพของระบบที่มีพื้นฐานข้อมูลขนาดใหญ่*. วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย.

บรรณานุกรม (ต่อ)

อานนท์ ศักดิ์วรวิชญ์. (2560). 3 R ของ Big Data ภาครัฐ. ค้นเมื่อ 19 กุมภาพันธ์ 2560, จาก http://www.manager.co.th/daily/ViewNews.aspx?NewsID=9600000017378

อานนท์ ศักดิ์วรวิชญ์. (2560). *BIG DATA, BUSINESS ANALYTICS, DATA SCIENCES กับการวิจัยเชิงคุณภาพ*. ค้นเมื่อ 7 มกราคม 2560, จาก http://as.nida.ac.th/th/index.php/researchs/intelligance-and-information/item/563-big-data-qualitative-research

ไอที 24 ชั่วโมง. (2559). สรุปสถิติผู้ใช้ Social Media ในไทยจากงาน Thailand Zocial Awards 2016. ค้นเมื่อ 3 ตุลาคม 2559, จาก https://www.it24hrs.com/2016/stat-social-media-thai-populations/

Chatri Ngambenchawong. (2559). Digital Thailand Big Data in Business. ค้นเมื่อ 3 ตุลาคม 2559, จาก http://naiwaen.debuggingsoft.com/2016/06/digital-thailand-big-data-in-business/

Business Analytics and Intelligence, GSAS, NIDA. (2016). Retrieved November 10, 2016, from https://www.facebook.com/BusinessAnalyticsNIDA/videos/1541641359470302/

Dandmarkel. (2016). เฟสบุ๊คโซเชียลมิเดียสุดฮิต. ค้นเมื่อ 17 เมษายน 2560, จาก http://www.danmarkel.com/

Data Science Thailand. (2016). *Big Data Analytics by True*. ค้นเมื่อ 3 ธันวาคม 2559, จาก https://web.facebook.com/DataScienceTh/posts/659351514216487?_rdc=1&_rdr

Data Scienceในประเทศไทย. (2559). แหล่งรวมข้อมูล Data Science เพื่อการพัฒนา Data Science ในประเทศไทย. ค้นเมื่อ 14 ตุลาคม 2559, จาก https://www.facebook.com/DataScienceTh

Data.go.th. (2016). *ข้อมูลพิกัด LAT/LONG ที่ตั้งตำบล*. ค้นเมื่อ 15 กุมภาพันธ์ 2560, จาก https://data.go.th/

Fusion idea. (2559). หลงทางสู่ Big Data. ค้นเมื่อ 3 ตุลาคม 2559, จาก http://www.fusionidea.biz/lose-big-data/

InfoMobius. (2558). BIG DATA ช่วยเพิ่มคุณค่าให้ธุรกิจได้อย่างไร. ค้นเมื่อ 3 ตุลาคม 2559, จาก http://www.infomobius.com/2015/03/how-big-data-delivers-business-values/

บรรณานุกรม (ต่อ)

iNnovationLab. (2014). *ทำความรู้จักกับ Big Data*. ค้นเมื่อ 1 มกราคม 2560, จาก http://nawattakam.blogspot.com/2014/07/big-data.html

Meewebfree. (n.d.). *API คืออะไร เกี่ยวข้องกับการทำเว็บอย่างไร*. ค้นเมื่อ 3 ธันวาคม 2559, จาก http://meewebfree.com/site/basic-website/274-what-is-api

Pioneer. (n.d.). การเก็บรวบรวมข้อมูล และการวิเคราะห์ข้อมูล. ค้นเมื่อ 14 ตุลาคม 2559, จาก http://pioneer.netserv.chula.ac.th/~jaimorn/re8.htm

Prasertcbs. (2016). *การใช้โปรแกรม R: การติดตั้ง R และ RStudio บน Ubuntu 16.04 (Install R and RStudio on Ubuntu)*. Retrieved April 10, 2016, from https://www.youtube.com/watch?v=4zeRI_3OwrY&feature=youtu.be

Somkiat. (2016). สรุปการเรียนขั้นพื้นฐานของภาษา R. Retrieved April 9, 2017, from http://www.somkiat.cc/learn-basic-of-r-programming/

Somkiat. (2557). *สวัสดีกับภาษา R กันหน่อยสิ*. ค้นเมื่อ 15 มีนาคม 2560, จาก http://www.somkiat.cc/hello-world-with-r/

Thanachart. (2016). การประมวลผล Big Data ใช้เทคโนโลยีไหนดี?. ค้นเมื่อ 2 พฤษจิกายน 2559, จาก https://thanachart.org/2016/02/11/%E0%B8%81%E0%B8%B2%E0%B8%A3%E0%B8%9B%E0%B8%A3%E0%B8%B0%E0%B8%A1%E0%B8%A7%E0%B8%A5%E0%B8%9C%E0%B8%A5-big-data-%E0%B8%84%E0%B8%A7%E0%B8%A3%E0%B9%83%E0%B8%8A%E0%B9%89%E0%B9%80%E0%B8%97%E0%B8%84%E0%B9%82/

Thanop. (2014). Hashtag คืออะไร และ วิธีการใช้ #Hashtag ที่เหมาะ. ค้นเมื่อ 27 เมษายน 2560, จาก https://www.thanop.com/hashtag/

Bohdan Stryk. (2015). HOW DO ORGANIZATIONS PREPARE AND CLEAN BIG DATA TO ACHIEVE BETTER DATA GOVERNANCE? A DELPHI STUDY. Capella University : Degree Doctor of Philosophy.Niels Mouthaan. (2012), *Business Information Systems*. Retrieved January 5, 2016, from http://nielsmouthaan.nl/big-data-thesis.pdf

บรรณานุกรม (ต่อ)

David Loshin. (2013). *Big Data Analytics*. Elsevier Science, Morgan Kaufmann.

Edgarauiz. (2016). *Setup a Spark 2.0 Cluster + R in AW*. Retrieved November 10, 2016, from https://edgarsdatalab.com/2016/08/25/setup-a-spark-2-0-cluster-r-on-aws/

Garrett Grolemund, Hadley Wickham. (2012). *Do more with dates and times in R with lubridate 1.3.0*. Retrieved April 9, 2017, from https://cran.r-project.org/web/packages/lubridate/vignettes/lubridate.html

Gary Vayerchuk. (2013). How to Master the 4 Big Social-Media Platforms. Retrieved April 5, 2017, from https://www.inc.com/magazine/201311/gary-vaynerchuk/how-to-master-the-four-major-social-media-platforms.html

Hadley Wickham. (2016). *Package 'gtable'*. Retrieved April 9, 2017, from https://cran.r-project.org/web/packages/gtable/gtable.pdf

Heuristicandrew. (2011). *Basic line chart with ggplot2*. Retrieved November 2, 2016, from https://www.r-bloggers.com/basic-line-chart-with-ggplot2/

Jaroen Ooms, Duncan Temple Lang, Lloyd Hilaiel. (2017). *Package 'jsonlite'*. Retrieved April 9, 2017, from https://cran.r-project.org/web/packages/jsonlite/jsonlite.pdf

Jjallaire. (2016). *Sparklr – R interface for Apache Spark*. Retrieved April 9, 2017, from https://www.r-bloggers.com/sparklyr-r-interface-for-apache-spark/

John Taveras. (2014). *How to Make a PivotTable in R*. Retrieved November 2, 2016, from https://www.rforexcelusers.com/make-pivottable-in-r/

Leo Widrich . (2016). 5 Important Differences Between Twitter And Facebook. Retrieved April 17, 2017, from https://blog.bufferapp.com/5-points-where-you-shouldnt-confuse-twitter-with-facebook

บรรณานุกรม (ต่อ)

Leonardo Toglia. (2016). *Extract Twitter Data Automatically using Scheduler R package*. Retrieved December10, 2016. from https://www.r-bloggers.com/extract-twitter-data-automatically-using-scheduler-r-package/

Mechael O. Ojo. (2016). *BIG DATA ANALYTICS SOLUTIONS: THE IMPLEMENTATION CHALLENGES IN THE FINANCIAL SERVICES INDUSTRY*. University of Delaware : MBA, Master of Business Administration.

Pouria Pirzadeh. (2015). *On the Performance Evaluation of Big Data Systems*. University of California, Irvine : Doctor of Philosophy in Computer Science.

Quora. (2017). What is the difference between Twitter and Facebook? Retrieved April 17, 2017, from https://www.quora.com/What-is-the-difference-between-Twitter-and-Facebook

RStudio. (2016). *Leaflet for R*. Retrievd April 9, 2017, from https://rstudio.github.io/leaflet/

Socialbakers. (2017). *March 2017 Social Marketing Report Thailand*. Retrieved April 15, 2017, from https://www.socialbakers.com/resources/reports/thailand/2017/march/

Statista. (2016). Penetration of leading social networks in Thailand as of 4th quarter 2016. Retrieved April 17, 2017, from https://www.statista.com/statistics/284483/thailand-social-network-penetration/

Stephanie F. Hood-Clark. (2016). *INFLUENCES ON THE USE AND BEHAVIORAL INTENTION TO USE BIG DATA.* Capella University : the Degree Doctor of Philosophy.

Twitter. (2016). Create an application. Retrieved January 9, 2017, from https://apps.twitter.com/app/new

https://apps.twitter.com/app/new

Wei Xu. (2016). Twitter API tutorial. Retrieved January 9, 2017, from http://socialmedia-class.org/twittertutorial.html

บรรณานุกรม (ต่อ)

Ylli Sadikaj. (2016). *PERSONALIZED HEALTH INSURANCE SERVICES USING BIG DATA*. A Thesis Submitted to the Graduate Faculty of the North Dakota State University of Agriculture and Applied Science, Degree of MASTER OF SCIENCE, Major Department : Electrical and Computer Engineering.

ZHENNING (JIMMY) XU. (2016). *THREE ESSAYS ON BIG DATA ANALYTICS, TRADITIONAL MARKETING ANALYTICS, KNOWLEDGE DISCOVERY, AND NEW PRODUCT PERFORMANCE*. The University of Texas :Master of Business Administration.

ภาคผนวก

ภาคผนวก ก
เครื่องมือที่ใช้ในการทำวิจัย
แบบตรวจสอบรายการ
เรื่อง “แนวทางการวิเคราะห์ข้อมูลทางธุรกิจโดยใช้ Big Data : กรณีศึกษาข้อมูลทวิตเตอร์อุบัติเหตุกับบริษัทประกันภัย

งานวิจัยนี้เป็นงานวิจัยแบบกึ่งทดลอง (Quasi - experimental research design) โดยมีแบบตรวจสอบรายการข้อมูลการเกิดอุบัติเหตุจากทวิตเตอร์เพื่อใช้เป็นเครื่องมือในการตรวจสอบข้อมูลในการทำวิจัยของ นักศึกษาปริญญาโท สาขาวิชาการตลาด วิทยาลัยการจัดการ มหาวิทยาลัยมหิดล ดังนี้

แบบตรวจสอบรายการ (Check List)

แบบตรวจสอบรายการ (Check List)	ใช่	ไม่ใช่	หมายเหตุ
กำหนดช่องทางของแหล่งข้อมูลที่จะนำมาใช้			
ช่องทางของแหล่งข้อมูลมาจากทวิตเตอร์			
ช่องทางของแหล่งข้อมูลมาจากเฟซบุ๊ค			
ช่องทางของแหล่งข้อมูลมาจากเว็บไซต์			
ช่องทางของแหล่งข้อมูลมาจากแหล่งอื่นๆ (โปรดระบุที่หมายเหตุ)			
ประเภทโครงสร้างของข้อมูล			
ข้อมูลที่นำมาวิเคราะห์ประเภทมีโครงสร้าง เช่น ตารางแบ่งแยกข้อมูลออกเป็นColumn ต่างๆอย่างชัดเจน			
ข้อมูลที่นำมาวิเคราะห์ประเภทไม่มีโครงสร้าง เช่น ไฟล์ข้อความ, ภาพ, เสียง เป็นต้น			
รายละเอียดข้อมูลที่นำมาวิเคราะห์			
ข้อมูลมีรายละเอียดสถานที่เกิดเหตุ			
ข้อมูลมีรายละเอียดเวลาที่ทำการทวีตการเกิดเหตุ			
ข้อมูลมีรายละเอียดประเภทของรถที่เกิดเหตุ			

ภาคผนวก ข

โปรแกรมที่เกี่ยวข้องในการใช้งาน

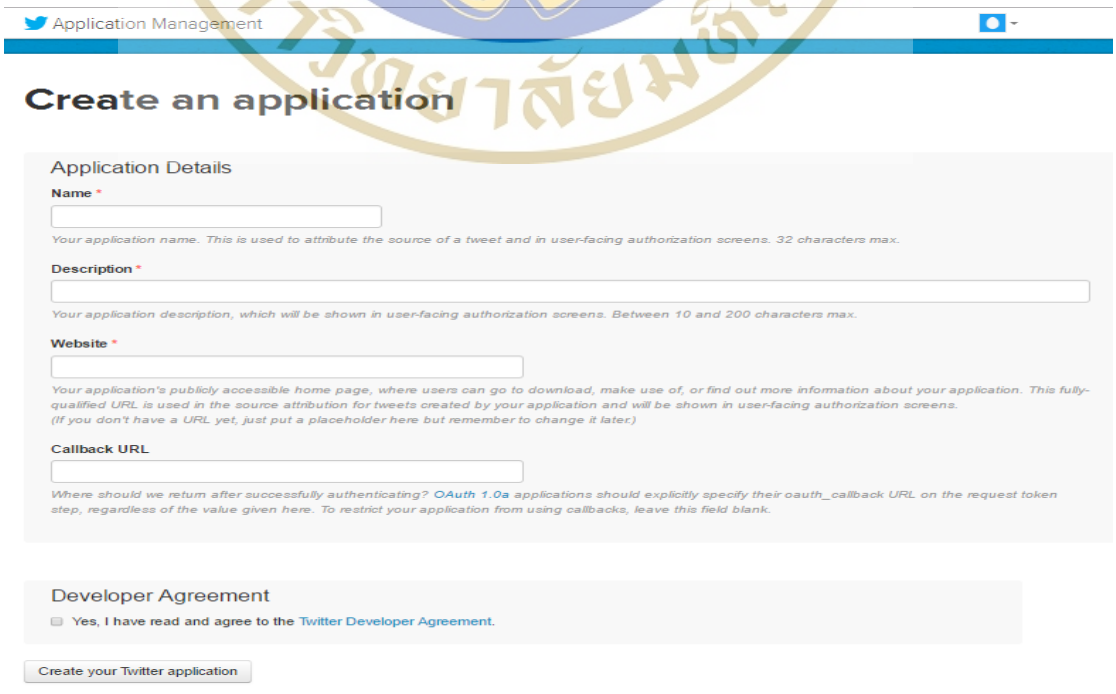
โปรแกรมหลักที่ต้องติดตั้งและขั้นตอนที่มีความเกี่ยวข้องในงานวิจัยนี้ ขอยกตัวอย่างให้เห็นภาพชัดเจนมากขึ้น ดังนี้

1. โปรแกรมAPI Twitter
2. โปรแกรม R Studio บนระบบปฏิบัติการ Ubuntu
3. การดึงข้อมูลจาก API Twitter ด้วย RStudio
4. การนำเสนอกราฟบนแผนที่ในส่วนของสถานที่ที่เกิดอุบัติเหตุ (Location)

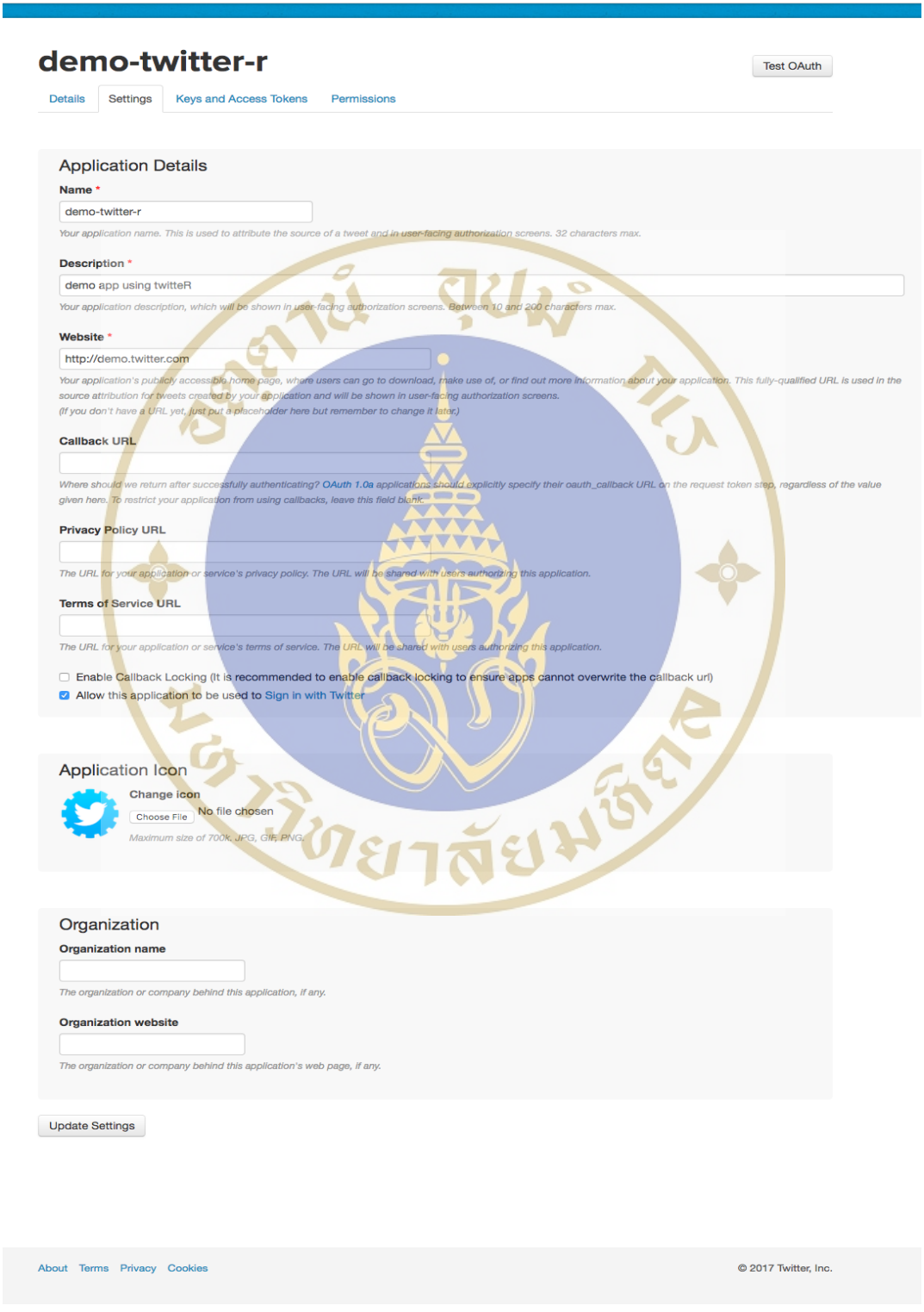
1. โปรแกรมAPI Twitter

เนื่องจากข้อมูลที่ต้องการเก็บเพื่อนำมาใช้เป็นตัวอย่างนั้นเป็นข้อมูลที่มาจากทวิตเตอร์ซึ่งสามารถดึงข้อมูลผ่าน Application Programming Interface (API)ที่ทวิตเตอร์มีให้ใช้ฟรี โดยสามารถสมัครใช้งานตามขั้นตอนดังนี้

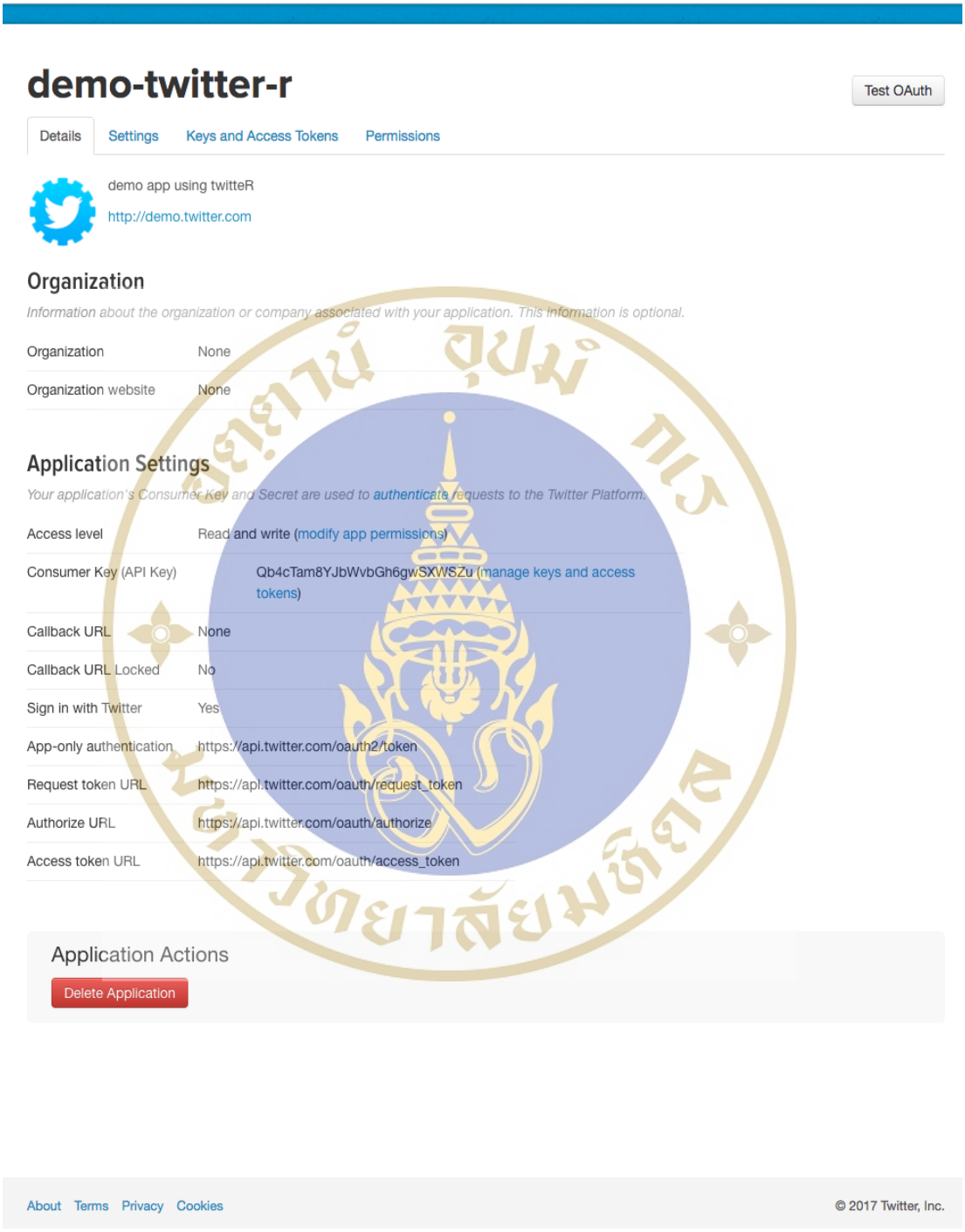
1. ภาพรวมการกรอกรายละเอียดที่ใบสมัครบนระบบออนไลน์หน้าเว็บไซต์ https://apps.twitter.com/app/new หลังจากที่กรอกรายละเอียดตามที่Twitter ได้กำหนดแล้ว เลือกถูกใน Check-box ตามข้อตกลง (Yes, I have read and agree to the Twitter Develop Agreement) จากนั้นกด Create your Twitter application



จากนั้นในส่วนของ Application Management จะแสดงรายละเอียดการตั้งค่าตามรูปที่ X ที่ผู้วิจัยได้สร้างตัวอย่างใบสมัครไว้ จากนั้นกดเลือกอนุญาตให้ Application นี้เข้าถึง Twitter เมื่อทำการ Sign in (Allow application to be use to sign in with Twitter)



ในหน้าของ Application Management ในส่วนของแถบรายละเอียด (Details) จะแสดงข้อมูลรายละเอียดที่ทางผู้วิจัยได้สมัครไว้เป็นตัวอย่างตามรูป



จากการสมัครใช้ API Twitter จะได้รายละเอียดในหน้า Keys and Access Tokens ที่เป็นสิ่งสำคัญที่สุดในการนำไปใช้เชื่อมต่อกับโปรแกรมอื่นเพื่อดึงข้อมูล โดยมีรายละเอียดตามภาพ

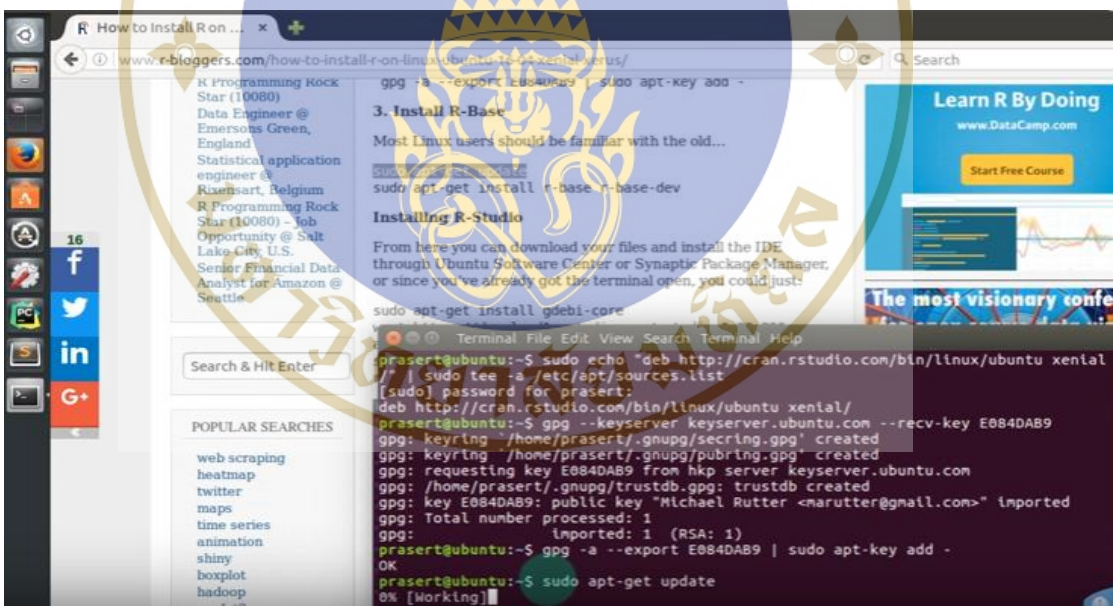
2. โปรแกรม R Studio บนระบบปฏิบัติการ Ubuntu

เนื่องจาก R Studio สามารถเป็นได้ทั้งโปรแกรม ส่วนประมวลผลทางสถิติและส่วนแสดงผลในรูปแบบกราฟ อีกทั้งยังเป็น Open-source ที่ฟรี สามารถทำงานได้ในหลายระบบปฏิบัติการ รวมถึงมี Community ที่ใหญ่ เป็นที่ยอมรับในหมู่นักวิจัย (Somkiat, 2557) ผู้วิจัยจึงนำมาใช้และมีวิธีการลงโปรแกรม ดังนี้

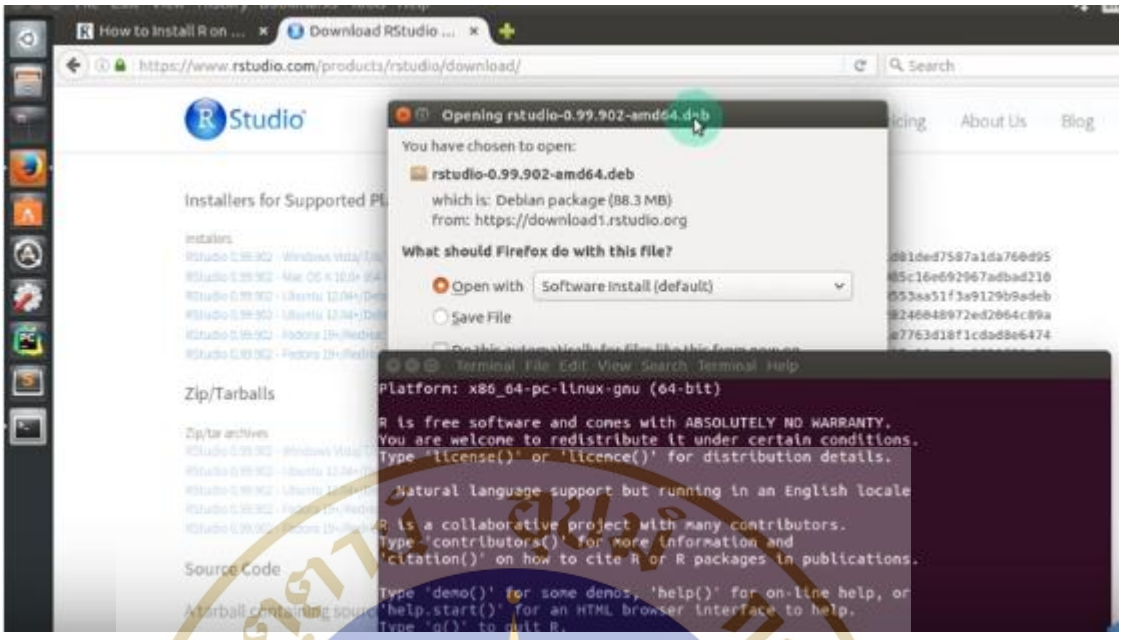
1. ลงโปรแกรม R ใน ระบบปฏิบัติการ Ubuntu ตามตัวอย่าง Youtube การใช้โปรแกรม R: การติดตั้ง R และ RStudio บน Ubuntu 16.04 (Install R and RStudio on Ubuntu) ของคณะพาณิชยศาสตร์และการบัญชีจุฬาลงกรณ์มหาวิทยาลัย ซึ่งเป็นสิ่งที่ทำให้สามารถใช้โปรแกรมR ในระบบปฏิบัติการ Ubuntu ได้ โดยการติดตั้ง สามารถดูรายละเอียดเพิ่มเติมได้จาก R-bloggers ดังนี้

หน้าตัวอย่างรายละเอียดการติดตั้งโปรแกรมR ในUbuntuจากเว็บไซต์ R-bloggers

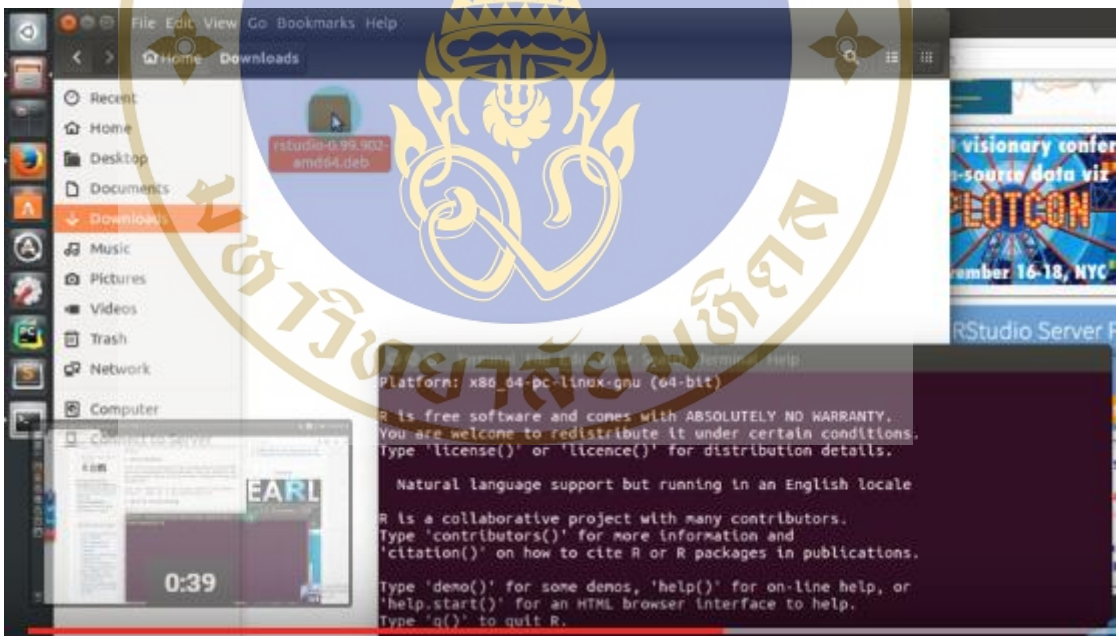
หน้าตัวอย่างรายละเอียดการติดตั้งโปรแกรมR ในUbuntuจากเว็บไซต์ R-bloggers



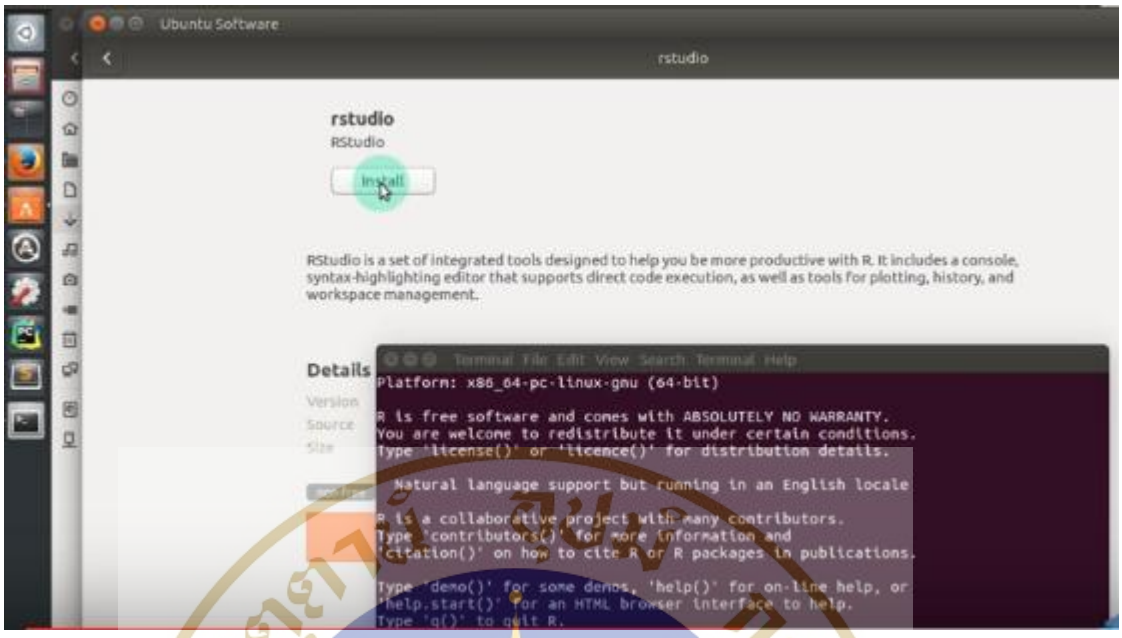
2. Download RStudio จากเว็บไซต์ https://www.rstudio.com/products/rstudio/download/ และตรวจสอบ Version ว่าเป็นโปรแกรม R Versionล่าสุด โดยเลือก Download RStudio-Ubuntu แบบ 64bits จากเว็บไซต์ RStudio ตามรูป



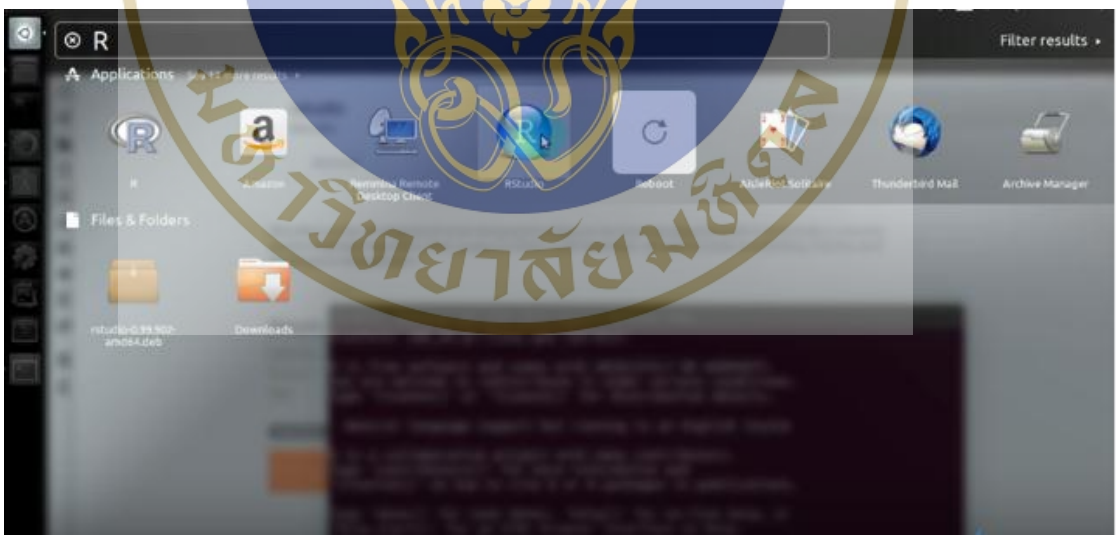
หลังจากที่ทำการ Download Program RStudioเสร็จสิ้นแล้วจะได้ไฟล์ที่อยู่ใน Folder rstudio-0.99.902-amd64.deb ตามรูปตัวอย่างไฟล์ที่ได้จากการ Download RStudito



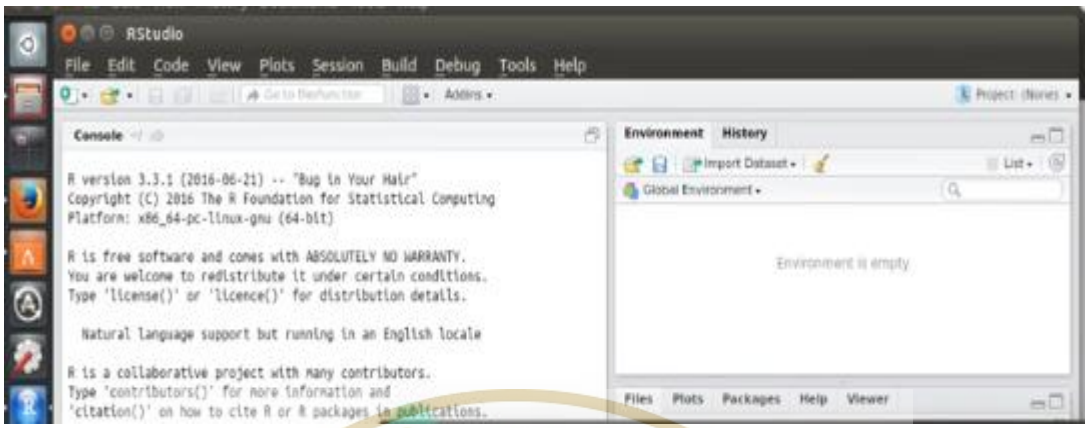
ทำการ Install โปแกรม โดยทำการดับเบิ้ลคลิก (Double click)ไปที่ไฟล์ข้างต้น จากนั้นกด Install โปรแกรมเพื่อทำการเริ่มติดตั้งโปรแกรม ตามภาพแสดงการ install file RStudio ที่ได้จากการdownload เพื่อติดตั้งโปรแกรม



หลังจากที่ทำการติดตั้งเสร็จสมบูรณ์แล้ว สามารถทดสอบโดยการเรียกใช้โปรแกรม RStudio จากการSearch ในแถบคำสั่งดังรูป และสามารถเริ่มใช้งานได้ ตามภาพแสดงการเรียกใช้โปรแกรม RStudio ผ่านระบบปฏิบัติการ Ubuntu

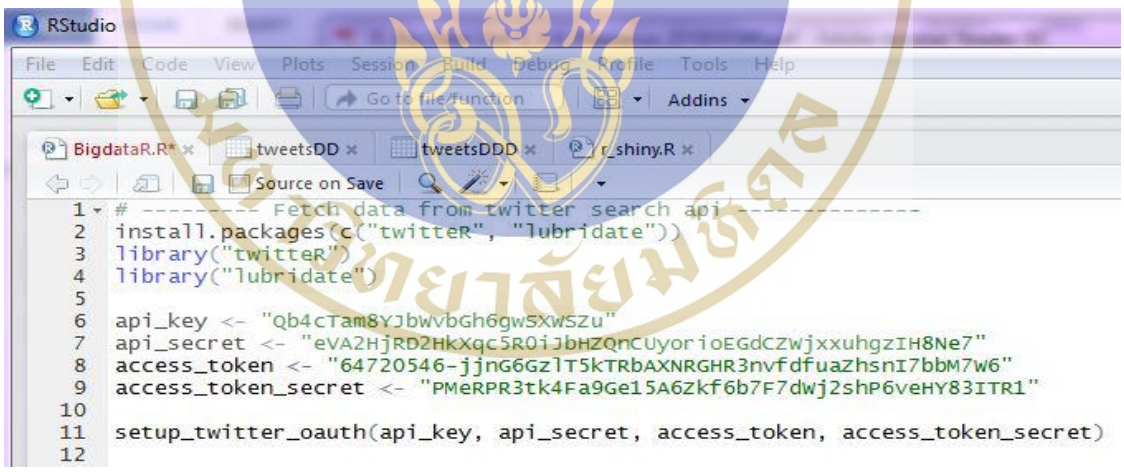


ตัวอย่างการเริ่มใช้งาน RStudio ผ่านระบบปฏิบัติการ Ubuntu



3. การดึงข้อมูลจาก API Twitter ด้วย RStudio

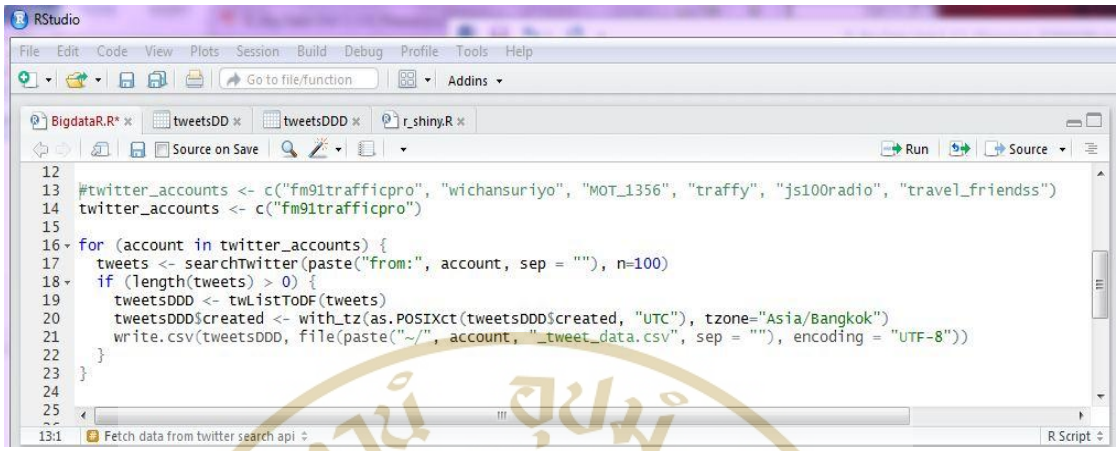
สำหรับขั้นตอนนี้ต้องมีการเรียกใช้ library twitteR และมีการกรอกข้อมูล api_key, api_secret, access_token และ access_token_secret ตามที่ได้สมัครไว้ในส่วนของ API Twitter ข้างต้น เพื่อที่ให้ R Programสามารถเชื่อมต่อกับ Twitterเตรียมเก็บข้อมูลการเกิดอุบัติเหตุ และเรียกใช้ข้อมูลได้ ตามภาพ



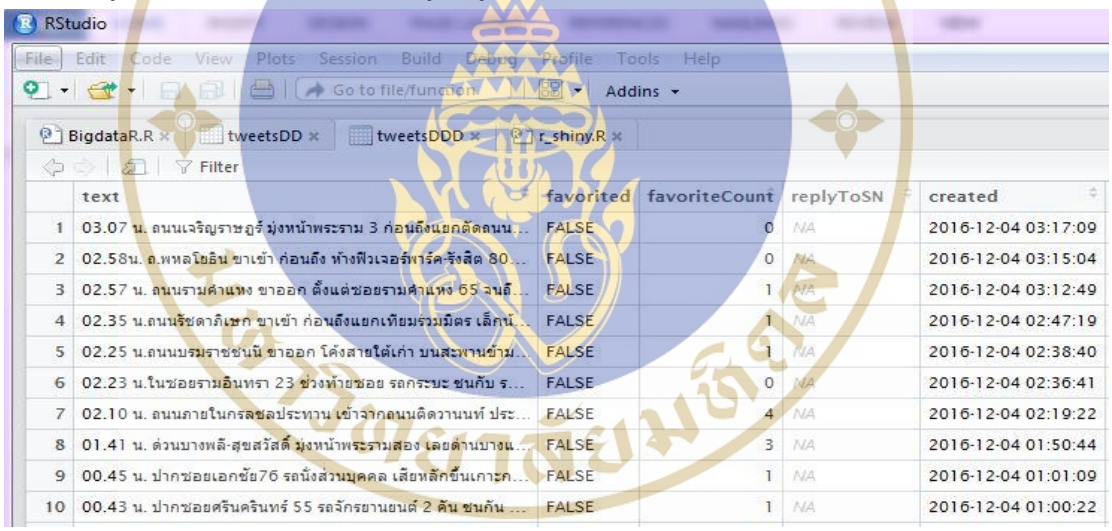
การปรับค่าให้อ่านภาษาไทยได้ดียิ่งขึ้น

หลังจากที่มีการเชื่อมต่อ R Program กับ Twitter แล้วนั้น ได้ทำการเก็บข้อมูลการจราจรและอุบัติเหตุจาก fm91trafficpro, js100radio, MOT_1356, traffy, travel_friendss, wichansuriyo จากทวิตเตอร์ทั้งหมดที่มีการทวิตเกี่ยวกับการจราจรและอุบัติเหตุบนถนน โดยเริ่มจากการเลือก fm91trafficpro มาเป็นตัวอย่างกลุ่มแรกในการเก็บข้อมูล 100 ตัวอย่าง ทั้งนี้มีการแปลงค่า

encoding = “UTF-8” เพื่อให้สามารถอ่านภาษาไทยได้ดียิ่งขึ้น จากนั้นทำการบันทึกเป็นfile CSV จากคำสั่ง write.csv



เมื่อได้ข้อมูลมาแล้ว สามารถเรียกไฟล์ข้างต้นเปิดดูข้อมูลดังกล่าวด้วยคำสั่ง Read.csv จะได้ข้อมูลตามตัวอย่างภาพการเรียกดูข้อมูลจาก Fm91 จากfile CSV



	text	favorited	favoriteCount	replyToSN	created
1	03.07 น. ถนนเจริญราษฎร์ มุ่งหน้าพระราม 3 ก่อนถึงแยกตัดถนน...	FALSE	0	NA	2016-12-04 03:17:09
2	02.58น. ถ.พหลโยธิน ขาเข้า ก่อนถึง ห้างฟิวเจอร์พาร์ค-รังสิต 80...	FALSE	0	NA	2016-12-04 03:15:04
3	02.57 น. ถนนรามคำแหง ขาออก ตั้งแต่ซอยรามคำแหง 65 จนถึ...	FALSE	1	NA	2016-12-04 03:12:49
4	02.35 น.ถนนรัชดาภิเษก ขาเข้า ก่อนถึงแยกเทียมร่วมมิตร เล็กน้...	FALSE	1	NA	2016-12-04 02:47:19
5	02.25 น.ถนนบรมราชชนนี ขาออก โค้งสายใต้เก่า บนสะพานข้าม...	FALSE	1	NA	2016-12-04 02:38:40
6	02.23 น.ในซอยรามอินทรา 23 ช่วงท้ายซอย รถกระบะ ชนกับ ร...	FALSE	0	NA	2016-12-04 02:36:41
7	02.10 น. ถนนภายในกรุงชุลประทาน เข้าจากถนนติวานนท์ ประ...	FALSE	4	NA	2016-12-04 02:19:22
8	01.41 น. ด่วนบางพลี-สุขสวัสดิ์ มุ่งหน้าพระรามสอง เลยด่านบางแ...	FALSE	3	NA	2016-12-04 01:50:44
9	00.45 น. ปากซอยเอกชัย76 รถนั่งส่วนบุคคล เสียหลักขึ้นเกาะก...	FALSE	1	NA	2016-12-04 01:01:09
10	00.43 น. ปากซอยศรีนครินทร์ 55 รถจักรยานยนต์ 2 คัน ชนกัน ...	FALSE	1	NA	2016-12-04 01:00:22

4. การนำเสนอกราฟบนแผนที่ในส่วนของสถานที่ที่เกิดอุบัติเหตุ (Location)

ในส่วนนี้จะเป็นการนำเสนอการเกิดอุบัติเหตุในสถานที่ต่างๆ ทั้งหมด 50 เขต ในกรุงเทพมหานคร โดยการ plot ลงในแผนที่ตามโปรแกรม R Package ของShiny ซึ่งแผนที่นี้จะสามารถ Zoom in หรือ Zoom out เขตที่ต้องการดูการแสดงผล โดยผู้จัดทำได้ใส่เป็นจุดวงกลมสีม่วงตามรัศมีเขตต่างๆที่มีการเกิดอุบัติเหตุ ขนาดจุดเล็กแสดงถึงการเกิดอุบัติเหตุน้อย และขนาดจุดใหญ่แสดงถึงจำนวนการเกิดอุบัติเหตุมาก นอกจากนั้นเมื่อนำเม้าท์ไปชี้ที่จุดวงกลมจะมีการแสดงค่า

จำนวนการเกิดเหตุในภาพด้วยเพื่อความชัดเจนในการแสดงผลมากยิ่งขึ้น ตามภาพแสดงคำสั่งโปรแกรมเพื่อให้แสดงผลค่าการเกิดอุบัตุเหตุในแผนที่