



การวิเคราะห์ข้อมูลโควิด-19

สู่ความยั่งยืนในธุรกิจท่องเที่ยว



ต่อ อาทอนันต์แล้ว

พค.ดร.นัทวัฒน์ ลีละวัฒน์

คณะวิศวกรรมศาสตร์ ภาควิชาวิศวกรรมมหาวิทาลัย
หน่วยปฏิบัติการวิจัยระบบสารสนเทศการศึกษาระดับปริญญาโทและ
ความเสียบ คณะวิศวกรรมศาสตร์ ภาควิชาวิศวกรรมมหาวิทาลัย
<http://natt.leelawat.com>

นอกจากนี้ยังมีงานวิจัยที่ได้ใช้ Google Earth Engine เป็นเครื่องมือในการวิเคราะห์ค่า NO₂ ภายในพื้นที่ศึกษา เช่น จังหวัดอยุธยา (Kodaka et al., 2021) และจังหวัดฮอกไกโด ในช่วงก่อนและหลังการเกิดการระบาดของโรค COVID-19 โดยพบว่าเมื่อการเดินทางของประชาชนลดลงนั้น จำนวนของก๊าซดังกล่าวที่ถูกตรวจจับได้โดยดาวเทียมนั้นจะลดลงตามไปด้วยเช่นกัน

อีกทั้งยังมีการศึกษาในประเด็นของการเดินทางของวัยรุ่น โดยมีผลสำรวจออกมาว่า หลังจากช่วงการระบาดของ COVID-19 นั้น วัยรุ่นมีความต้องการที่จะเดินทางเพื่อการท่องเที่ยวมากขึ้นอย่างไม่เคยเป็นมาก่อน ดังนั้นจึงมีการวิเคราะห์ต่อมาว่าหากกลุ่มวัยรุ่นนั้นต้องการที่จะท่องเที่ยว สถานที่ และกิจกรรมแบบใดที่จะตอบโจทย์ หรือดึงดูดนักท่องเที่ยววัยรุ่นเหล่านั้นได้ ซึ่งจะเป็นผลการศึกษาที่สามารถช่วยธุรกิจการเดินทางให้พัฒนาได้ตอบโจทย์ความต้องการ และดึงดูดลูกค้าได้มากขึ้นเช่นกัน

พื้นฐานข้อมูลกวิตเตอร์สำหรับการวิเคราะห์ข้อมูล

การบรรยายในช่วงต่อมาโดย อาจารย์ ดร. Jing Tang อาจารย์ประจำหลักสูตรวิศวกรรมหุ่นยนต์ และปัญญาประดิษฐ์ สำนักบริหารหลักสูตรวิศวกรรมนานาชาติ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และอาจารย์ประจำหลักสูตรสาขาวิชาการจัดการความเสี่ยง และภัยพิบัติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย ได้บรรยายในหัวข้อพื้นฐานข้อมูลทวิตเตอร์ (Twitter) สำหรับการวิเคราะห์ข้อมูล ซึ่งจะเชื่อมโยงกับกระบวนของ COVID-19 โดยการบรรยายนั้น จะแบ่งออกเป็นสองส่วนใหญ่ ๆ คือ เหตุผลว่าทำไมการศึกษาวัยรุ่นจึงควรเลือกใช้ข้อมูลจากทวิตเตอร์ รวมถึงผลของการวิจัยที่มีออกมา ซึ่งจะมีการบรรยายต่อเนื่องในประเด็นนี้โดยคุณกัณฑ์ แสงทับทิม นักศึกษาระดับปริญญาเอก ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ประเด็นแรกที่อาจารย์ ดร. Jing Tang กล่าวถึงคือการอธิบายถึง Big Data ซึ่งเป็นสิ่งที่ผู้คนในปัจจุบันให้ความสนใจ สำหรับการเก็บข้อมูลที่มีจำนวนมากไว้ด้วยกันเพื่อนำมาวิเคราะห์นั้น ไม่ว่าจะเก็บข้อมูลจากการเก็บแบบสอบถาม การทดสอบ หรือแม้กระทั่งโพสต์ต่าง ๆ จากสื่อสังคมออนไลน์ เช่น ทวิตเตอร์ ล้วนนับรวมในส่วนนี้ด้วยเช่นกัน ทั้งนี้ Big Data นั้น ก็ยังสามารถแบ่งได้เป็นสองลักษณะ คือข้อมูลที่มีการสร้างขึ้นโดยเครื่องมือ

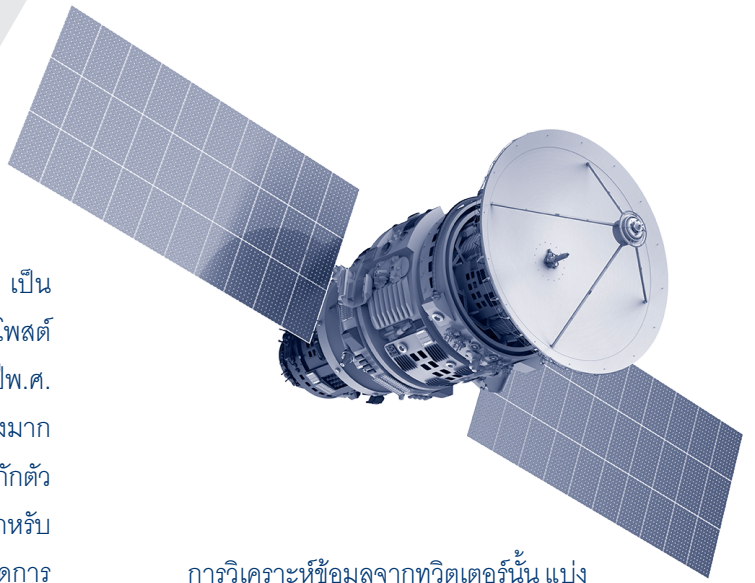
(เช่น ข้อมูลจากดาวเทียม เป็นต้น) และข้อมูลที่สร้างขึ้นโดยมนุษย์ (เช่น ข้อมูลจากสื่อสังคมออนไลน์ เป็นต้น)

ในปัจจุบันนั้นสื่อสังคมออนไลน์อย่างทวิตเตอร์นั้น เป็นช่องทางที่ได้รับความนิยมเป็นอย่างมาก โดยในทุก ๆ วันจะมีโพสต์จำนวน 400 – 500 ล้านโพสต์ที่เกิดขึ้นในทวิตเตอร์ และจากปีพ.ศ. 2562 – 2563 นั้น จำนวนของโพสต์เหล่านี้เพิ่มขึ้นเป็นอย่างมาก เนื่องจากการระบาดของ COVID-19 ที่ทุกคนจำเป็นที่จะต้องกักตัวอยู่ที่บ้าน ยอดในการใช้สื่อสังคมออนไลน์จึงสูงขึ้นเช่นกัน สำหรับในประเทศไทยนั้นการใช้ Facebook คิดเป็น 60% ของยอดการใช้สื่อสังคมออนไลน์ทั้งหมด และทวิตเตอร์นับเป็นอันดับที่สองอยู่ที่ประมาณ 20% ก่อนที่จะเป็น Youtube และ Instagram ซึ่งยอดการใช้ทวิตเตอร์นั้นก็เพิ่มขึ้นในทุก ๆ ปีเช่นกัน

หากแต่โดยทั่วไปแล้วการใช้ทวิตเตอร์นั้น เหมาะสมในการใช้เพื่อวิเคราะห์ข้อมูลในงานศึกษานี้มากกว่า เพราะมีการเกาะกลุ่มของผู้ที่มีความสนใจในประเด็นเดียวกัน และสามารถ



ติดตามประเด็นต่าง ๆ ได้มากกว่าการใช้ Facebook และมีความ Real-time ที่มากกว่า นอกจากนี้ผู้ใช้มีพฤติกรรมในการติดตามประเด็น และข้อมูลมากกว่าด้วย ดังนั้นการใช้ข้อมูลจากสื่อสังคมออนไลน์เหล่านี้ในการศึกษาพฤติกรรม และความคิดเห็นเกี่ยวกับอุตสาหกรรมการท่องเที่ยวของผู้ใช้งาน และการใช้ประโยชน์จาก Hashtag (#) ที่จะรวบรวมข้อมูลที่เกี่ยวข้องของโพสต์ที่มีความเชื่อมโยงกันไว้ในกลุ่มเดียวกันได้ ทำให้การเก็บข้อมูลสามารถทำได้ง่ายขึ้น เช่น การค้นหาข้อมูลที่เกี่ยวข้องกับการกล่าวถึงเกี่ยวกับ Phuket Sandbox เป็นต้น



การวิเคราะห์ข้อมูลจากทวิตเตอร์นั้น แบ่งออกได้เป็นสองแบบ กล่าวคือ การวิเคราะห์โดยไม่ได้กล่าวถึงเนื้อหาในโพสต์ แต่ให้ความสนใจเกี่ยวกับรายละเอียดรอบ ๆ นั้น เช่น ยอดการติดตาม ยอดการ Retweet ยอดการ Like ของโพสต์นั้น ๆ และแบบที่สองคือการวิเคราะห์เนื้อหา ซึ่งให้ความสำคัญกับเนื้อหาในโพสต์เหล่านั้นมากกว่า สำหรับการศึกษานี้เนื้อหาของโพสต์ในทวิตเตอร์นั้น ยังสามารถถูกแยกย่อยตามการเก็บข้อมูล และความสนใจได้เป็นอีกสามกลุ่ม ได้แก่ (1) Sentiment Analysis ที่จะวิเคราะห์ถึงทัศนคติหรือความเห็นของผู้ใช้งานว่ามีความพึงพอใจมากแค่ไหน หรือมีความรู้สึกอย่างไร (2) Topic Modelling ซึ่งวิเคราะห์เกี่ยวกับหัวข้อที่เกี่ยวข้อง และหัวข้อที่ได้รับความนิยม รวมถึงความเกี่ยวข้องของคำภายในแต่ละโพสต์ และ (3) Name Identity Recognition ซึ่งศึกษา และค้นหาจากชื่อของคน สถานที่ หรือกิจกรรม หรือประเทศที่ได้มีการกล่าวถึงในโพสต์

สำหรับตัวอย่างของงานวิจัยนั้น มีการศึกษาด้วยการใช้ Content Analysis ที่มีการศึกษาตั้งแต่ในช่วงของการเริ่มต้นของ COVID-19 โดยพิจารณาศึกษาการใช้ชื่อเรียกต่าง ๆ ที่เกิดขึ้นของโรค COVID-19 นี้ เพราะในขณะนั้นยังไม่ได้มีการกำหนดชื่ออย่างเป็นทางการ โดยหน่วยงานระดับนานาชาติขึ้นมาใช้อย่างแพร่หลาย และต่อเนื่องไปยังการเก็บข้อมูลจากชื่อเหล่านั้นที่ได้ทำการวิเคราะห์ออกมาในช่วงต้นว่า ในแต่ละเวลานั้นมีการกล่าวถึงการระบาดของโรคนี้มากหรือน้อยเท่าใด โดยการศึกษามุ่งเน้นทวิตในภาษาอังกฤษ ภาษาจีน และภาษาญี่ปุ่น หลังจากเก็บข้อมูลแล้ว จึงวิเคราะห์หาความเชื่อมโยงเหล่านั้นออกมาเพื่อหาหัวข้อที่ได้รับความนิยม เช่น การพบผู้ติดเชื้อในประเทศ หรือการกล่าวถึงคลัสเตอร์ต่าง ๆ ที่เกิดขึ้น (Leelawat et al., 2021)