การวิเคราะห์แนวโน้มการสื่อสารทางการเมืองระหว่างเจเนอเรชั่นบนทวิตเตอร์โดยใช้การวิเคราะห์ข้อมูลเชิงลึกด้วยการจำแนกหัวข้อและการแสดงผลด้วยแท็กคลาวด์

VISUALIZING POLITICAL COMMUNICATION TRENDS ACROSS GENERATIONS ON X (TWITTER): INSIGHTS THROUGH TOPIC MODELING AND WORD CLOUDS

กิตติศักดิ์พัฒนา ราเชาว์ดี

KITTISAKPATTANA RACHAODEE

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์

ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2568

KMITL-2025-SC-M-017-001

VISUALIZING POLITICAL COMMUNICATION TRENDS ACROSS GENERATIONS ON X (TWITTER): INSIGHTS THROUGH TOPIC MODELING AND WORD CLOUDS

KITTISAKPATTANA RACHAODEE

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE
IN DATA SCIENCE AND ANALYTICS
KMITL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2025

KMITL-2025-SC-M-017-001

หัวข้อการค้นคว้าอิสระ	การวิเคราะห์แนวโน้มการสื่อสารทางการเมืองระหว่างเจเนอเรชั่นบนทวิตเตอร์โดยใช้การวิเคราะห์ข้อมูลเชิงลึกด้วยการจำแนกหัวข้อและการแสดงผลด้วยแท็กคลาวด์
ชื่อนักศึกษา	กิตติศักดิ์พัฒนา ราเชาว์ดี
รหัสประจำตัว	63605095
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์) ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
พ.ศ.	2568
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	ผู้ช่วยศาสตราจารย์ ดร.ปัทมา เจริญพร

บทคัดย่อ

การศึกษานี้สำรวจความสนใจและความสำคัญของคำที่ใช้ใน Twitter (หรือ X) ของกลุ่มคนจากแต่ละเจเนอเรชัน ได้แก่ เจเนอเรชันเบบี้บูมเมอร์ (Baby Boomer), เจเนอเรชั่น เอ็กซ์ (Generation X), เจเนอเรชั่น วาย (Generation Y), และ เจเนอเรชั่น ซี (Generation Z) โดยใช้วิธีการจำแนกหัวข้อด้วย Latent Dirichlet Allocation (LDA) เพื่อค้นหาความสัมพันธ์และความสำคัญของคำในแต่ละกลุ่ม เนื่องจากผลการจำแนกหัวข้ออาจเข้าใจได้ยาก จึงใช้การแสดงผลด้วยเวิร์ดคลาวด์ (Word Cloud) เพื่อช่วยให้เข้าใจผลลัพธ์ของแต่ละเจเนอเรชันได้ชัดเจนขึ้น ผลการวิจัยพบลักษณะเฉพาะของแต่ละกลุ่ม เช่น เจเนอเรชั่นเบบี้บูมเมอร์ มักกล่าวถึงสื่อสิ่งพิมพ์ เว็บไซต์ข่าว และบุคคลทางการเมืองในไทยที่มีชื่อเสียง เจเนอเรชัน เอ็กซ์ เน้นบุคคลและปัญหาการเมืองท้องถิ่นในกรุงเทพฯ เจเนอเรชัน วายกล่าวถึงเหตุการณ์ทางการเมืองและสังคม ส่วนเจเนอเรชัน ซี มักตั้งคำถามเกี่ยวกับกิจกรรมทางการเมืองและสังคม การศึกษานี้สามารถนำไปใช้ได้ในหลากหลายภาษาและงาน เพื่อเข้าใจพฤติกรรมและความสนใจของแต่ละเจเนอเรชัน

คำสำคัญ : การสร้างโมเดลหัวข้อ, การประมวลผลภาษาธรรมชาติ, การสื่อสารทางการเมือง, แท็กคลาวด์, เวิร์ดคลาวด์

Independent Study Title Visualizing Political Communication Trends across Generations on X (Twitter): Insights through Topic Modeling and Word Clouds

Student Name Kittisakpattana Rachaodee

Student ID 63605095

Degree Master of Science (Data Science and Analytics)

KMITL-Digital Analytics and Intelligence Center

Year 2025

Independent Study Advisor Asst. Prof. Dr. Pattama Charoenporn

Abstract

This study examines the interests and significance of words on Twitter (or X) across different generational groups: Baby Boomers, Generation X, Generation Y, and Generation Z. Using Topic Modeling with Latent Dirichlet Allocation (LDA), the research explores relationships and word importance within each group. As the results of topic modeling are not always easy to interpret, we used word cloud visualization to help make sense of the results for each generation. The findings reveal distinct patterns: Baby Boomers frequently mention print media, news websites, and prominent Thai political figures; Generation X emphasizes individuals and local political issues in Bangkok; Generation Y discusses political and social events; and Generation Z uniquely questions political and social activities. This research methodology is applicable across languages and tasks, offering insights into generational behaviors and interests.

Keywords : Topic Modeling, Tag Cloud, Word Cloud, Natural Language Processing, Political Communication.

กิตติกรรมประกาศ

ข้าพเจ้าขอขอบคุณอาจารย์ที่ปรึกษา ผศ.ดร. ปัทมา เจริญพร ที่ให้คำแนะนำและสนับสนุนการทำวิจัยนี้เป็นอย่างดี และขอขอบคุณ อาจารย์ ผศ.ดร.กุลสวัสดิ์ จิตขจรวานิช ที่ได้ให้ข้อเสนอแนะและความรู้เพิ่มเติมที่เป็นประโยชน์ต่อการศึกษาในครั้งนี้ ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาของอาจารย์ทุกท่าน และขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณท่านคณาจารย์และวิทยากรทุกท่านที่มาถ่ายทอดความรู้ตลอดระยะเวลา 5 ปี ทำให้สามารถนำความรู้ต่างๆ ที่ได้เรียนมาใช้ประโยชน์ในการค้นคว้าอิสระนี้ได้อย่างเต็มที่ และความร่วมมือต่างๆ ของหลายท่าน ที่ให้การสนับสนุนผู้วิจัยตั้งแต่เริ่มต้นจนเสร็จสมบูรณ์ พร้อมทั้งขอบคุณเพื่อนๆ ทุกคนที่ให้ความช่วยเหลือทั้งในเรื่องการเรียน งานกลุ่ม การสอบ และคำแนะนำด้านเทคนิคการเขียนโปรแกรม และช่วยหาข้อมูลที่จะนำมาใช้ในการค้นคว้าอิสระให้สำเร็จลุล่วงไปด้วยดี

ขอขอบคุณครอบครัว บิดา และมารดา ที่ให้การเลี้ยงดูสนับสนุนทุกอย่าง และอบรมส่งเสริมด้านการศึกษาเป็นอย่างดีตลอดมา ข้าพเจ้าขอขอบคุณทุกท่านที่มีส่วนร่วมในการสนับสนุนการศึกษาครั้งนี้ ซึ่งทำให้ข้าพเจ้าได้ประสบความสำเร็จและบรรลุเป้าหมายของการวิจัยอย่างสำเร็จ

นายกิตติศักดิ์พัฒนา ราเชาว์ดี

สารบัญ

สารบัญ (ต่อ)

สารบัญรูป

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันนี้ ธุรกิจทั้งขนาดใหญ่และขนาดเล็กต่างหันมาใช้แพลตฟอร์มโซเชียลมีเดีย เช่น ทวิตเตอร์ (Twitter) หรือ เอกซ์ (X) ในการทำการตลาด เพราะเป็นวิธีที่มีประสิทธิภาพในการโฆษณาและขายสินค้า โดยใช้ต้นทุนต่ำหรือแทบไม่เสียค่าใช้จ่ายเลย เมื่อเทียบกับวิธีการตลาดแบบดั้งเดิม การตลาดบนโซเชียลมีเดียมีประสิทธิภาพสูงและได้ผลลัพธ์อย่างรวดเร็ว ในยุคนี้ธุรกิจใดที่ไม่ใช้โซเชียลมีเดียถือว่าล้าสมัย

การศึกษาวิจัยนี้มีจุดมุ่งหมายเพื่อวิเคราะห์ความสนใจของผู้ใช้ทวิตเตอร์ ตามเจเนอเรชัน โดยใช้เทคนิคการสร้างแบบจำลองหัวข้อ (Topic Modeling) และการแสดงผลแบบแท็กคลาวด์ (Tag Cloud) หรือ เวิร์ดคลาวด์ (Word Cloud) เพื่อสำรวจผลการวิจัย ในขอบเขตการวิจัยนี้ ผู้วิจัยใช้ชุดข้อมูลทวิตเตอร์ ที่เก็บผ่านทวิตเตอร์ เอพีไอ (Twitter API) เพื่อวิเคราะห์ความสนใจของผู้ใช้ตามเจเนอเรชัน โดยใช้ไพทอน (Python) ในการสร้างแบบจำลองหัวข้อและแท็กคลาวด์ (Tag Cloud) เมื่อวิเคราะห์ความสนใจของผู้ใช้ทวิตเตอร์ได้แล้ว จะสามารถแนะนำสินค้าและบริการที่ตรงกับความสนใจเหล่านี้ได้ ผู้ขายสามารถนำเสนอสินค้าและบริการให้ตรงกลุ่มเป้าหมายของผู้ใช้ทวิตเตอร์ ได้อย่างมีประสิทธิภาพยิ่งขึ้น

ดังนั้นการค้นคว้าอิสระนี้จึงจัดทำการศึกษาและวิเคราะห์แนวโน้มการสื่อสารทางการเมืองระหว่างเจเนอเรชั่นบนทวิตเตอร์โดยใช้เทคนิคการสร้างแบบจำลองหัวข้อ (Topic Modeling) และใช้เทคนิคการสร้างแท็กคลาวด์ เพื่อนำเสนอการอภิปรายผล

1.2 วัตถุประสงค์ของการค้นคว้าอิสระ

1) เพื่อศึกษาและวิเคราะห์แนวโน้มการสื่อสารทางการเมืองระหว่างเจเนอเรชั่นบนทวิตเตอร์

2) เพื่อการวิเคราะห์ข้อมูลเชิงลึกด้วยเทคนิคการจำแนกหัวข้อ

1.3 ขอบเขตของการค้นคว้าอิสระ

1) ใช้ข้อมูลทวิตบนทวิตเตอร์ผ่าน ทวิตเตอร์ เอพีไอ (Twitter API) มาทำการวิเคราะห์แนวโน้มการสื่อสารทางการเมืองระหว่างเจเนอเรชั่นบนทวิตเตอร์ โดยใช้คีย์เวิร์ดทางการเมือง ช่วงเวลาการเก็บข้อมูลที่เป็นการสื่อสารทางการเมือง ปี 2564 ถึงปี 2565

2) ภาษาที่ใช้ คือ ภาษาไพธอน (Python) โดยใช้สำหรับการค้นคว้าอิสระดังนี้

2.1) เพื่อสร้างแบบจำลองหัวข้อ (Topic Modeling) และสร้างการแสดงผลด้วยแท็กคลาวด์

2.2) เพื่อเก็บข้อมูลปีเกิด ผ่านการดึงข้อมูลจากเว็บไซต์ (Web scraping)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้ข้อมูลสำหรับแนะนำสินค้าที่ตรงกับความสนใจการทางเมือง ให้ตรงกลุ่มเป้าหมายของผู้ใช้ทวิตเตอร์ ได้อย่างมีประสิทธิภาพยิ่งขึ้น

2) ได้ข้อมูลสำหรับแนะนำบริการที่ตรงกับความสนใจการทางเมือง ให้ตรงกลุ่มเป้าหมายของผู้ใช้ทวิตเตอร์ ได้อย่างมีประสิทธิภาพยิ่งขึ้น

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้อง ซึ่งจะประกอบด้วยส่วนที่จะแสดงรายละเอียดเพิ่มเติมในหัวข้อ 2.1 ถึง 2.8

2.1 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) เป็นเทคนิควิธีการที่มีเป้าหมายเพื่อค้นหารูปแบบ และความสัมพันธ์ของข้อมูลที่ซ่อนอยู่จากข้อมูลจำนวนมหาศาลโดยอัตโนมัติ โดยใช้วิธีจากสถิติ การเรียนรู้ของเครื่อง การรู้จำ และหลักคณิตศาสตร์ โดยที่การทำเหมืองข้อมูลมีกระบวนการมาตรฐานในการวิเคราะห์ข้อมูลด้านการทำเหมืองข้อมูลที่พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือกันของ 3 บริษัท คือ DaimlerChrysler SPSS และ NCR กระบวนการทำงานนี้เรียกว่า “Cross-Industry Standard Process for Data Mining” หรือเรียกย่อว่า “CRISP-DM” โดยในกระบวนการ CRISP-DM นี้จะประกอบด้วย 6 ขั้นตอน ได้แก่

1. Business Understanding เน้นไปที่การทำความเข้าใจในงาน ระบุโอกาส และหาปัญหาที่จะเกิดขึ้นกับธุรกิจ กำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์หาความได้เปรียบทางการตลาดและแก้ไขปัญหาองค์กร ซึ่งต้องสามารถระบุผลลัพธ์ที่มีได้

2. Data Understanding ทำความเข้าใจข้อมูลโดยการรวบรวมข้อมูลที่เกี่ยวข้อง คัดเลือกให้เหลือเพียงข้อมูลที่มีความถูกต้องและสำคัญต่องานมาทำการวิเคราะห์

3. Data Preparation ทำการแปลงข้อมูล (Raw Data) ให้กลายเป็นข้อมูลที่สามารถนำมาช่วยในการวิเคราะห์ต่อไปได้ ขั้นตอนนี้จะใช้เวลามากที่สุดในทุกขั้นตอน เพราะคุณภาพของงานที่ได้จะดีเพียงใดขึ้นอยู่กับคุณภาพข้อมูลที่จัดเตรียมในขั้นนี้ การเตรียมข้อมูลประกอบด้วย การคัดเลือกข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล

4. Modeling การสร้างแบบจำลองเพื่อวิเคราะห์ข้อมูลที่ได้จากขั้นตอนที่ 3 พร้อมทดสอบผลลัพธ์แบบจำลองเพื่อให้ได้คำตอบที่ดีที่สุด บางครั้งอาจมีการย้อนกลับไปปรับการเตรียมข้อมูลเพื่อให้ได้แบบจำลองที่เหมาะสมที่สุด

5. Evaluation การประเมินผลลัพธ์ที่ได้ก่อนที่จะนำไปใช้จริง ว่าตรงกับวัตถุประสงค์หรือเป้าหมายที่ได้ตั้งไว้หรือมีความน่าเชื่อถือมากน้อยเพียงใด หากไม่ได้ผลลัพธ์ตามวัตถุประสงค์ต้องย้อนกลับไปปรับปรุงแก้ไขการดำเนินงานในขั้นตอนก่อนหน้า

6. Deployment การนำเอาข้อมูลที่เป็นผลลัพธ์จากทั้งหมด มาลองปฏิบัติจริงกับธุรกิจในองค์กร โดยแปลงแนวคิดที่มีให้เกิดเป็นสารสนเทศเพื่อให้ผู้บริหารหรือนักการตลาดเข้าใจสามารถนำไปใช้ประโยชน์ในทางธุรกิจได้จริง และติดตามประเมินผลที่ได้เพื่อนำกลับไปปรับปรุง Data

Mining ต่อเนื่องต่อไป ซึ่งการประเมินผลสามารถทำได้หลายทางเช่น วัดจากส่วนแบ่งของตลาด วัดจากปริมาณลูกค้า หรือ วัดจากกำไรสุทธิ เป็นต้น

2.2 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง (Machine Learning) เป็นการสอนอัลกอริทึมให้เรียนรู้ทำความเข้าใจและตัดสินใจได้ด้วยตัวเองจากข้อมูลที่ป้อนให้ และสามารถเรียนรู้ได้ด้วยตนเองเพื่อที่จะผลิตผลลัพธ์ที่แม่นยำออกมาได้

โดยที่การเรียนรู้ของเครื่อง (Machine Learning) นั้นเริ่มต้นมาตั้งแต่ปี ค.ศ.1950 เมื่อนักวิทยาศาสตร์คอมพิวเตอร์คิดหาวิธีสอนคอมพิวเตอร์ให้เล่นหมากฮอส จากนั้น เมื่อวิวัฒนาการทางเทคโนโลยี ทำให้ระบบการคำนวณค่าต่างๆของคอมพิวเตอร์เพิ่มขึ้น ทำให้คอมพิวเตอร์สามารถเข้าใจและจดจำรูปแบบของค่าต่างๆที่ซับซ้อนได้ แล้วจึงประยุกต์ไปสู่การคาดการณ์สถานการณ์ณ์รวมไปถึงการแก้ปัญหาด้วยตัวเอง อัลกอริทึมของ Machine Learning คือ การให้ ‘ชุดการสอน (teaching set)’ ของข้อมูล แล้วให้ใช้ข้อมูลดังกล่าวเพื่อตอบคำถาม ตัวอย่างเช่น คุณอาจเตรียมชุดการสอนเกี่ยวกับรูปภาพ ภาพบางส่วนคือ “นี่คือแมว” ภาพบางส่วน “นี่ไม่ใช่แมว” จากนั้นคุณสามารถแสดงภาพชุดใหม่ ๆ สอบถามคอมพิวเตอร์ แล้วคอมพิวเตอร์จะเริ่มแยกแยะได้ว่ารูปใดเป็นรูปแมว

2.3 การจัดกลุ่ม (Clustering)

การจัดกลุ่มข้อมูล เป็นส่วนหนึ่งของเทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการแบ่งกลุ่มข้อมูล โดยพิจารณาจากลักษณะที่คล้ายกันของข้อมูล โดยจะจัดกลุ่มข้อมูลที่มีลักษณะคล้ายกันมาอยู่ด้วยกัน และจะจัดกลุ่มข้อมูลที่มีลักษณะต่างกัน ให้อยู่คนละกลุ่มกัน

ขั้นตอนในการสร้างโมเดลด้วยวิธีการการจัดกลุ่มข้อมูล สามารถแบ่งขั้นตอนของการสร้างโมเดลออกเป็น 4 ขั้นตอน ดังนี้

1. กำหนดวิธีวัดความเหมือนหรือความต่าง วิธีที่นิยมใช้ เช่น ยูคลิเดียน (Euclidean Distance) โคไซน์ (Cosine) และ แมนฮัตตัน (Manhattan Distance) เป็นต้น

2. เลือกใช้อัลกอริทึมที่ใช้ในการจัดกลุ่มข้อมูล แบ่งได้เป็น 2 ประเภท 2.1 ประเภทที่มีการแบ่งกลุ่มอย่างชัดเจน (Hard clustering) เป็นเทคนิคที่แต่ละข้อมูลนั้นจะถูกจัดให้อยู่ในกลุ่มใดกลุ่มหนึ่งเท่านั้น ตัวอย่างเช่น เทคนิคการจัดกลุ่มข้อมูลแบบเคมีน (K-means clustering algorithm) 2.2 ประเภทที่มีการแบ่งกลุ่มแบบไม่ชัดเจน (Soft clustering) เป็นเทคนิคที่ข้อมูลสามารถอยู่ในหลายๆ กลุ่มได้ โดยขึ้นอยู่กับความน่าจะเป็นของตัวข้อมูล ตัวอย่างเช่น เทคนิคการจัดกลุ่มแบ่งกลุ่มข้อมูลแบบลำดับชั้น (Hierarchical clustering methods)

3. กำหนดจำนวนกลุ่มที่ต้องการ โดยที่อัลกอริทึมประเภทที่มีการแบ่งกลุ่มอย่างชัดเจนจำเป็นต้องกำหนดจำนวนกลุ่มที่ต้องการ แต่หากเป็นอัลกอริทึมประเภทที่มีการแบ่งกลุ่มแบบไม่ชัดเจน ไม่จำเป็นต้องกำหนดจำนวนกลุ่มก็ได้

4. ประเมินผล เนื่องการจัดกลุ่มข้อมูล เป็นส่วนหนึ่งของเทคนิคการเรียนรู้แบบไม่มีผู้สอน ซึ่งจะไม่สามารถวัดผลได้จากเปรียบเทียบหาความแม่นยำ เนื่องจากไม่มีผลลัพธ์ตั้งต้นให้เปรียบเทียบ แต่จะสามารถวัดผลได้ด้วยวิธีอื่น เช่น วัดจากความพึงใจในตัวโมเดล เป็นต้น

2.3.1 การจัดกลุ่มแบบเคมีน (K-mean Clustering)

การจัดกลุ่มข้อมูลแบบเคมีน เป็นวิธีการที่เป็นที่นิยม เนื่องจากมีขั้นตอนการทำงานที่ไม่ซับซ้อนและเข้าใจได้ง่าย โดยมีขั้นตอนการทำงานดังนี้

1. เริ่มต้นจากการกำหนดค่า K หรือจำนวนกลุ่มข้อมูลที่ต้องการ
2. สุ่มวางตำแหน่งจุดศูนย์กลางของข้อมูลแต่ละกลุ่มหรือเซ็นทรอยด์ (Centroid)
3. จากจุดศูนย์กลางแต่ละจุดจะมีการคำนวณระยะทางกับทุกข้อมูลในชุดข้อมูล ซึ่งคำนวณโดยวิธียูคลิเดียนดังได้อธิบายไว้แล้วในข้างต้น จากนั้นแต่ละข้อมูลจะถูกจัดอยู่ในกลุ่มของจุดศูนย์กลางที่มีระยะทางใกล้ที่สุดเท่านั้น
4. หลังจากทำการจัดกลุ่มข้อมูลใหม่แล้วทำการคำนวณค่าเฉลี่ยของสมาชิกในกลุ่ม เพื่อกำหนดเป็นจุดศูนย์กลางของกลุ่มข้อมูลใหม่
5. ทำซ้ำข้อ 3 ถึง 4 จนกระทั่งค่าจุดศูนย์กลางของกลุ่มข้อมูลใหม่ได้ค่าไม่ต่างหรือต่างเพียงเล็กน้อยจากค่าจุดศูนย์กลางรอบก่อนหน้า

รูปที่ 2.1 ขั้นตอนการจัดกลุ่มข้อมูลแบบเคมีน

2.3.2 การจัดกลุ่มแบบโครงสร้างลำดับชั้น (Hierarchical Clustering)

การจัดกลุ่มแบบโครงสร้างลำดับชั้น (Hierarchical Clustering) มี 2 ประเภทคือ 1.Agglomerative Clustering วิธีการจัดกลุ่มแบบโครงสร้างลำดับชั้นจากล่างขึ้นบน ในวิธีนี้จะจุดข้อมูลทั้งหมดเป็นกลุ่มและจะรวมจุดข้อมูลเข้าด้วยกันตามระยะห่างระหว่างกลุ่ม จะรวมจนกว่าข้อมูลทั้งหมดเป็นกลุ่มเดียวกัน 2.Divisive Clustering วิธีการจัดกลุ่มแบบโครงสร้างลำดับชั้นจากบนลงล่าง

ในวิธีนี้กลุ่มข้อมูลทั้งหมด และค่อยๆแตกกลุ่มเป็นกลุ่มขนาดเล็ก จะแตกกลุ่มจนกว่ากลุ่มข้อมูลมีเพียงจุดข้อมูลเดียว

2.4 การสร้างแบบจำลองหัวข้อ (Topic Modeling)

Topic modeling เป็นวิธีการทางสถิติที่ใช้ในการค้นพบ "หัวข้อ" ที่แฝงอยู่ในเอกสาร โดยวิธีนี้นิยมใช้ในการวิเคราะห์ข้อมูลเอกสารเพื่อค้นหาความสัมพันธ์ภายในของข้อมูลที่ไม่มีโครงสร้าง ซึ่งทำให้เห็นภาพรวมของข้อมูลภาษาได้อย่างชัดเจน นอกจากนี้ยังใช้ศึกษาและทำความเข้าใจข้อมูลในศาสตร์ต่าง ๆ ได้อีกด้วย

Topic modeling เป็นการเรียนรู้ด้วยเครื่องแบบไม่มีการชี้แนะ (unsupervised learning) โดยอาศัยการเกิดร่วมกันของคำในเอกสารเพื่อค้นหาความหมายที่ซ่อนอยู่ในข้อความ วิธีนี้สามารถประยุกต์ใช้กับงานประมวลผลภาษาต่าง ๆ เช่น การจัดประเภทเอกสารตามหัวข้อ และการสรุปคำสำคัญในหัวข้อ

“หัวข้อ” ที่ได้จากการทำ Topic modeling คือกลุ่มของคำที่มีความหมายใกล้เคียงกัน โดยคำนวณมาจากความถี่ของการปรากฏร่วมกันของคำในเอกสารต่าง ๆ อัลกอริทึมที่มักใช้ในงาน Topic modeling เช่น Latent Dirichlet Allocation (LDA) และ Latent Semantic Analysis (LSA) โดยหลักการของอัลกอริทึมเหล่านี้คือการสร้างเมทริกซ์เพื่อดูการกระจายตัวของคำในเอกสาร และสร้างเป็น Document-Term matrix

การสร้างเมทริกซ์ Document-Term จะช่วยในการวิเคราะห์ความเกี่ยวข้องระหว่างคำและหัวข้อในเอกสาร เช่น ถ้ามีเอกสาร d ชิ้น และมีจำนวนรูปศัพท์ n ศัพท์ จะสามารถสร้างเมทริกซ์ขนาด d x n ซึ่งเมทริกซ์นี้จะสามารถถูกแยกออกเป็นเมทริกซ์ขนาด d x k และ k x n โดยที่ k คือจำนวนหัวข้อที่มีในข้อมูล จากนั้นสามารถวิเคราะห์ได้ว่าคำใดเกี่ยวข้องกับหัวข้อใด และเอกสารแต่ละชิ้นมีความเกี่ยวข้องกับหัวข้อใดบ้าง

2.4.1 ทำไมจึงหากลุ่มคำในหัวข้อ

Topic modeling ไม่ต้องอาศัยความรู้ทางภาษาศาสตร์โดยตรงในการวิเคราะห์ข้อมูลภาษา โดยไม่จำเป็นต้องวิเคราะห์โครงสร้างประโยคหรือรู้ว่าอะไรเป็นประธานหรือกรรม เพียงมองข้อมูลภาษาเป็นกล่องใส่คำ (bag of words) ที่มีคำต่าง ๆ อยู่ในกล่องนั้น

ด้วยธรรมชาติของภาษา คำมีความหมาย เช่น เมื่อพูดถึงหัวข้อหนึ่ง ๆ เช่น การสร้างคลังข้อมูลภาษา คำเหล่านี้จะปรากฏซ้ำ ๆ ในตัวบท เช่น ตัวบท การสร้าง ตัวแทนภาษา ขนาด การกำกับข้อมูล เป็นต้น เอกสารหนึ่งมักมีหัวข้อไม่กี่หัวข้อ เอกสารที่กล่าวถึงหัวข้อต่างกัน คำที่ปรากฏในเอกสารก็จะต่างกัน ด้วยลักษณะเช่นนี้ เมื่อแจกแจงการปรากฏของคำเนื้อหาในเอกสารในรูปตาราง Document-Term matrix แต่ละ Document ก็มีชุดลำดับความถี่ของคำที่ไม่เหมือนกัน ชุดลำดับความถี่คำนี้ก็คือเว็กเตอร์ที่ใช้แทน Document นั้น

Document ที่คล้ายกันก็จะมีเว็กเตอร์ที่ใกล้เคียงกัน ในทางกลับกัน หากมองคำจากชุดลำดับความถี่ที่พบในแต่ละเอกสาร ก็จะได้เว็กเตอร์ที่ใช้แทนคำต่าง ๆ คำที่มีความหมายใกล้เคียงกันก็จะมีเว็กเตอร์ที่ใกล้กัน

ด้วยเหตุนี้ การทำ Topic modeling จึงสามารถหากลุ่มคำที่เป็นหัวข้อได้โดยไม่ต้องอาศัยความรู้ทางภาษาศาสตร์ แต่ใช้การวิเคราะห์จากความถี่ของคำที่ปรากฏร่วมกันในเอกสารแทน

2.5 การสร้างแท็กคลาวด์ (Tag Cloud)

Tag Cloud หรือที่รู้จักกันในชื่อ Word Cloud หรือ Weighted List คือการแสดงข้อมูลที่เป็นข้อความในรูปแบบภาพ ซึ่งมักใช้ในการออกแบบส่วนต่อประสานกับผู้ใช้ (UI) โดยแต่ละแท็กใน Tag Cloud จะแสดงด้วยขนาดตัวอักษร สี หรือสไตล์ที่แตกต่างกัน เพื่อเน้นความสำคัญ ความถี่ หรือความเกี่ยวข้องของแท็กภายในบริบทที่กำหนด

Tag Clouds เป็นเครื่องมือที่ใช้งานง่ายและรวดเร็วในการแสดงข้อมูลที่มีค่าจากชุดข้อมูลขนาดใหญ่ เช่น แนวโน้มคำหลัก หัวข้อยอดนิยม หรือการวิเคราะห์เนื้อหา วิธีนี้สามารถรวมเข้ากับแอปพลิเคชันประเภทต่าง ๆ ได้ ไม่ว่าจะเป็นแอปพลิเคชันบนเว็บ อุปกรณ์เคลื่อนที่ หรือแบ็กเอนด์ ช่วยให้ผู้ใช้สามารถโต้ตอบและสำรวจข้อมูลที่ซ่อนอยู่ได้อย่างมีประสิทธิภาพ

2.5.1 วิธีการทำงานของแท็กคลาวด์

Tag Clouds ทำงานโดยการวิเคราะห์ข้อมูลข้อความเพื่อแยกคำหรือวลีสำคัญ และกำหนดน้ำหนักให้กับแต่ละแท็กตามความถี่ของการเกิดขึ้น ความเกี่ยวข้องกับบริบทของแอปพลิเคชัน หรือตัวชี้วัดอื่น ๆ ข้อมูลแบบถ่วงน้ำหนักนี้จะถูกนำมาใช้เพื่อสร้างการจัดเรียงแท็กในลักษณะคลาวด์ที่ดึงดูดสายตา โดยขนาด สี หรือสไตล์ของแท็กจะสัมพันธ์กับน้ำหนักของแท็ก ยิ่งแท็กมีน้ำหนักมากเท่าใด แท็กก็จะยิ่งปรากฏเด่นชัดในคลาวด์มากขึ้นเท่านั้น ช่วยให้ผู้ใช้สามารถระบุและมุ่งเน้นไปที่คำหลักที่สำคัญในชุดข้อมูลได้อย่างรวดเร็ว

2.5.2 ประโยชน์ของแท็กคลาวด์

การวิจัยแสดงให้เห็นว่า Tag Clouds สามารถปรับปรุงประสิทธิภาพและความแม่นยำของการสำรวจข้อมูลได้อย่างมาก ทำให้ผู้ใช้สามารถแยกแยะข้อมูลเชิงลึกจากชุดข้อมูลขนาดใหญ่ได้อย่างรวดเร็ว จากการศึกษาด้านการโต้ตอบระหว่างมนุษย์กับคอมพิวเตอร์และการแสดงข้อมูลเป็นภาพ พบว่า Tag Clouds สามารถเพิ่มอัตราความสำเร็จของการค้นหาผู้ใช้ได้สูงสุดถึง 30% เมื่อเทียบกับอินเทอร์เฟซการค้นหาแบบเดิม และลดเวลาทำงานได้ถึง 25%

นอกจากนี้ Tag Clouds ยังช่วยเพิ่มความสามารถในการเรียกคืนและรับรู้ข้อมูล โดยผู้ใช้มีแนวโน้มที่จะเรียกคืนคำหลักที่แสดงใน Tag Cloud มากกว่าคำหลักที่แสดงในรูปแบบรายการแบบดั้งเดิมถึง 20%

โดยสรุป Tag Clouds เป็นองค์ประกอบ UI ที่มีคุณค่า ซึ่งมอบวิธีการโต้ตอบและดึงดูดสายตาให้ผู้ใช้สามารถสำรวจและเข้าใจชุดข้อมูลที่เป็นข้อความขนาดใหญ่ได้อย่างมีประสิทธิภาพ อีกทั้งยังช่วยปรับปรุงประสบการณ์ผู้ใช้ การแสดงข้อมูลเป็นภาพ และความสามารถในการวิเคราะห์ได้อย่างมาก

2.6 งานวิจัยที่เกี่ยวข้องกับการสร้างแบบจำลองหัวข้อและแท็กคลาวด์

Blei, Ng, และ Jordan (2003) ได้นำเสนอ LDA ซึ่งเป็นอัลกอริธึมในการจำลองหัวข้อที่ได้รับความนิยมมาก ช่วยให้สามารถค้นหาหัวข้อที่ซ่อนอยู่ในเอกสารจำนวนมากได้อย่างมีประสิทธิภาพ โดยการจำลองแต่ละเอกสารเป็นการผสมผสานของหัวข้อ LDA จึงช่วยเผยโครงสร้างของเนื้อหาในชุดข้อมูลที่ซับซ้อน ทำให้สามารถนำไปใช้ในงานจัดประเภทและสรุปเนื้อหาได้

Griffiths และ Steyvers (2004) แสดงให้เห็นว่า LDA มีประโยชน์ในการจัดระเบียบและตีความชุดข้อมูลในสาขาต่าง ๆ เช่น จิตวิทยาและพันธุศาสตร์ ซึ่งเป็นที่ต้องการในการทำความเข้าใจรูปแบบที่ซ่อนอยู่

ส่วน Pritchard, Stephens และ Donnelly (2000) ได้อภิปรายถึงการใช้ LDA ในการวิเคราะห์สถิติเพื่อหาความสัมพันธ์ของคำและสร้างกลุ่มคำที่มีความหมาย

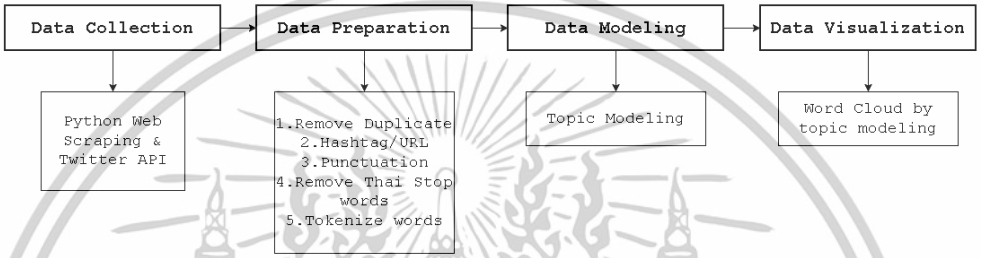
Hassan-Montero และ Herrero-Solana (2006) พบว่า Word Clouds (หรือ Tag Clouds) ช่วยให้ผู้ใช้เข้าใจแนวโน้มและหัวข้อสำคัญในเอกสารได้อย่างรวดเร็ว Word Clouds เป็นเครื่องมือแสดงภาพที่ช่วยให้เห็นภาพรวมของคำสำคัญได้อย่างรวดเร็ว

Sinclair และ Cardew-Hall (2008) แสดงให้เห็นว่า Word Clouds สามารถเน้นคำสำคัญและหัวข้อที่เกี่ยวข้องในบทความวิจัยได้อย่างมีประสิทธิภาพ ซึ่งช่วยเพิ่มความเข้าใจของผู้ใช้ในหัวข้อที่ซับซ้อนอย่างมาก

และ Rivadeneira และคณะ (2007) ยืนยันว่า Word Clouds ช่วยเพิ่มความเข้าใจโดยการเน้นคำที่ปรากฏบ่อย ๆ จึงทำให้เป็นเครื่องมือที่มีประสิทธิภาพในการแสดงหัวข้อจากผลลัพธ์ของ LDA

บทที่ 3
วิธีการดำเนินงาน

ในบทนี้จะอธิบายขั้นตอนต่าง ๆ ในการพัฒนาแนวทางการทำโมเดลหัวข้อเพื่อวิเคราะห์ความสนใจของผู้ใช้ Twitter โดยแบ่งตามเจเนอเรชั่น ขั้นตอนประกอบด้วย การเก็บข้อมูล การเตรียมข้อมูล การสร้างโมเดล และการนำเสนอข้อมูลตามที่แสดงดังรูปที่ 3.1



รูปที่ 3.1 ลำดับขั้นตอนการดำเนินงาน

3.1 การจัดเก็บข้อมูล (Data Collection)

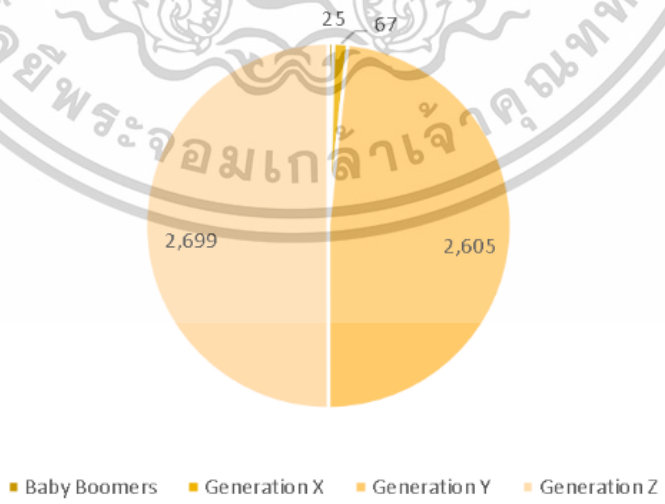
ผู้วิจัยเลือกใช้ Twitter เป็นแหล่งข้อมูลและใช้ Twitter API (เวอร์ชัน 1.0) เพื่อเก็บทวีต ข้อมูลที่เก็บรวมถึงคุณลักษณะดังนี้ รหัส (id) รหัสรูปแบบข้อความ (id_str) ชื่อบัญชีผู้ใช้งาน (name) ชื่อหน้าจอ (screen_name) ที่ตั้ง (location) ยูอาร์แอล (url) คำอธิบาย (description) การตรวจสอบยืนยันตัวตน (verified) จำนวนผู้ติดตาม (followers_count) จำนวนเพื่อน (friends_count) จำนวนรายการ สาธารณะที่ผู้ใช้รายนี้เป็นสมาชิก (listed_count) จำนวนรายการโปรด (favourites_count) จำนวนสถานะ (statuses_count) วันที่สร้างขึ้น (created_at) ทวิต (tweet) และวันที่ทวิต (tweet_created_at) แต่เนื่องจาก Twitter API ไม่ได้ให้ข้อมูลวันเกิด ซึ่งจำเป็นสำหรับการระบุเจเนอเรชั่นของผู้ใช้ เราจึงใช้การเก็บข้อมูลเว็บ (web scraping) ด้วย Python เพื่อเก็บข้อมูลวันเกิด (born) ที่แสดงในโปรไฟล์ของผู้ใช้ โดยเน้นไปที่ปีเกิด (born_year) ข้อมูลนี้มีความสำคัญในงานของเรา เพราะใช้ในการระบุเจเนอเรชั่นของผู้ใช้ในชุดข้อมูลของเราเท่านั้น ข้อมูลที่เก็บมาไม่สามารถย้อนกลับข้อมูล และระบุตัวตนได้ โดยใช้อัลกอริธึมการการเก็บรวบรวมข้อมูลโดยการดึงข้อมูลจากโปรไฟล์ผู้ใช้สาธารณะเพื่อระบุเจเนอเรชันของพวกเขาโดยใช้ทวิตเตอร์ เอพีไอ (Twitter API) แสดงในรูปที่ 3.2

Input: keyword search "`#politics`" (in Thai)
Output: Total tweet data files

1. Import necessary libraries/functions such as `tweepy`, `configparser`, `selenium`, `webdriver`, and `pandas`.
2. Authenticate access with Twitter API key and access token.
3. Specify keyword search as: "`#politics`".
4. Set the tweet number limit to 100 for retrieval at a time.
5. Initialize `total_tweets = {}`.
6. Retrieve a collection of 100 tweets at a time.
7. For each `tweet` in `tweets`:
8. Locate the user profile and retrieve the username (`u_web`).
9. Set `target = http://twitter.com/ + u_web`.
10. Access the `target` website.
11. Using `XPATH`, locate "`UserBirthdate`" and save as `born`.
12. Append `born` to the `tweet`.
13. Append `tweets` to `total_tweets`.
14. Save `total_tweets` to a file.

รูปที่ 3.2 อัลกอริธึมการเก็บรวบรวมข้อมูลโดยการดึงข้อมูลจากโปรไฟล์ผู้ใช้สาธารณะเพื่อระบุเจเนอเรชันของพวกเขาโดยใช้ทวิตเตอร์ เอพีไอ (Twitter API)

จากรูป 3.2 สามารถสรุปผลจากการเก็บ มีการเก็บทวีตทั้งหมด 13,249 ทวีต ซึ่งในจำนวนนี้สามารถระบุเจเนอเรชั่นของผู้ใช้ได้ 5,396 ทวีต ในขณะที่ไม่สามารถระบุได้ 7,853 ทวีต เนื่องจากไม่มีข้อมูลวันเกิดในโปรไฟล์ของผู้ใช้ การแบ่งทวีตตามเจเนอเรชั่นเป็นดังนี้ เจเนอเรชั่นเบบี้บูมเมอร์ (Baby Boomer) มีจำนวน 25 ทวีต, เจเนอเรชั่น เอ็กซ์ (Generation X) มีจำนวน 67 ทวีต, เจเนอเรชั่น วาย (Generation Y) มีจำนวน 2,605 ทวีต, และ เจเนอเรชั่น ซี (Generation Z) มีจำนวน 2,699 ทวีต ดังที่แสดงในรูปที่ 3.4



รูปที่ 3.3 กราฟแสดงจำนวนเจเนอเรชั่น

3.2 การเตรียมข้อมูล (Data Preparation)

การเตรียมข้อมูลประกอบด้วยการทำความสะอาดทวีตที่มีข้อมูลเจเนอเรชั่น ตัวอย่างของข้อมูลนี้แสดงในรูปที่ 3.4 (ส่วนที่ 1 คอลัมน์ ID ถึง friends_count) และ 3.5 ส่วนที่ 2 คอลัมน์ listed_count ถึง generation) กระบวนการทำความสะอาดข้อมูลประกอบด้วยขั้นตอนดังนี้ 1. ลบข้อมูลที่ซ้ำกัน 2. ลบแฮชแท็กและ URL 3. ลบคำหยุดภาษาไทย และ 4. ลบอีโมจิ

id	id_str	name	screen_name	location	url	description	verified	followers_count	friends_count
2915775614	2915775614	น้องน้ำตุยอยากไ	PPSEIKI			#ป๋อจ๋าน #monsta	FALSE	151	369
173841899	173841899	เกร์ทเตอะฟลุท	greattheflute		https://t.co/7L5v	ภก.เวทชวิน (เกร์	FALSE	2444	1595
2954242255	2954242255	คุณคนเดียว	tonyon_jsdmtw	-_-		อึหยังวะ?	FALSE	37	115
2951915689	2951915689	หัวใจพอร์นๆ	_superbug			exo nct	FALSE	73	861
3294031950	3294031950	dextorboyz เช็ค	ingmueng004			มีหวานเจี๊ยบแล้วค	FALSE	133	586
2951915689	2951915689	หัวใจพอร์นๆ	_superbug			exo nct	FALSE	73	861
2951915689	2951915689	หัวใจพอร์นๆ	_superbug			exo nct	FALSE	73	861
2954242255	2954242255	คุณคนเดียว	tonyon_jsdmtw	-_-		อึหยังวะ?	FALSE	37	115
2915775614	2915775614	น้องน้ำตุยอยากไ	PPSEIKI			#ป๋อจ๋าน #monsta	FALSE	151	368
2829563880	2829563880	I'm Gonnnnn.	portogus28			คุณคือรอยยิ้มของ	FALSE	37	373
340691370	340691370	yuè	pamnrm_			☺️𝑬𝑿𝑶♡	FALSE	439	653
2951915689	2951915689	หัวใจพอร์นๆ	_superbug			exo nct	FALSE	73	861
2951915689	2951915689	หัวใจพอร์นๆ	_superbug			exo nct	FALSE	73	861
3184386746	3184386746	ล่อบ	alphalobzter	รีดอนเท่น art+nsfw+blood+tw		intp 5w4 / feel m	FALSE	97	604
1435487227	1435487227	อย่าจำฉัน	smlttg	บนหลังหมาหน้าเซเว่น		선우경	FALSE	142	768
569612301	569612301	Don't_Cry | นักเ	DontTouch_Mee	ไม่ตอบแชท = อ่านนิยาย&หลับ		นกเขาขันยันรี	FALSE	146	1376
4915932680	4915932680	ink	LIENOXKIP86			cuz all of me love	FALSE	1753	340
103613505	103613505	庚希	care_hc	天上天下希澈	https://t.co/U6Q	LABEL	FALSE	369	246

รูปที่ 3.4 ตัวอย่างข้อมูลที่สามารถหาเจเนอเรชั่นได้ (ส่วนที่ 1 คอลัมน์ ID ถึง friends_count)

listed_count	favourites_count	statuses_count	created_at	tweet	tweet_created_at	born	born_year	Generation
2	21672	55382	2014-12-01	RT @fellerian: อย่าอวยชัชชาติมากไปเด้อ ว่า	2022-05-25	Born November 19, 1998	1998	Generation Z
15	3283	163426	2010-08-02	RT @ddough12: เอ็นดูชาว กทม. เค้าตั้งใจเลือ	2022-05-25	Born January 5, 1991	1991	Generation Y
0	3255	192466	2015-01-01	RT @mortorsorton: ขอบคุณ #ชัชชาติ ที่ทำให	2022-05-25	Born March 10, 1997	1997	Generation Z
1	1270	83516	2014-12-30	RT @fellerian: อย่าอวยชัชชาติมากไปเด้อ ว่า	2022-05-25	Born May 18, 1998	1998	Generation Z
10	55188	53060	2015-07-26	RT @SupersMint: @gorya_ilada แข็งแกร่งกว่	2022-05-25	Born 2002	2002	Generation Z
1	1270	83516	2014-12-30	RT @emmamadest: มึงจริง ๆ ชัชชาติเป็นผู้ว่าก	2022-05-25	Born May 18, 1998	1998	Generation Z
1	1270	83516	2014-12-30	RT @thecuriousdan2: เห็นบางกลุ่ม demand	2022-05-25	Born May 18, 1998	1998	Generation Z
0	3255	192466	2015-01-01	RT @juseyo_1427: ตอนแรกนึกว่าคุณชัชชาติแ	2022-05-25	Born March 10, 1997	1997	Generation Z
2	21672	55382	2014-12-01	RT @HuaLaaN_7824: เอาจริงๆ กูน่าไม่ออกอะ	2022-05-25	Born November 19, 1998	1998	Generation Z
1	1115	95209	2014-09-24	RT @fellerian: อย่าอวยชัชชาติมากไปเด้อ ว่า	2022-05-25	Born October 28, 1996	1996	Generation Y
8	25980	448459	2011-07-28	RT @juseyo_1427: ตอนแรกนึกว่าคุณชัชชาติแ	2022-05-25	Born 1999	1999	Generation Z
1	1270	83516	2014-12-30	RT @aorwiki: คุณเป้ามายืนหนังสือร้องเรียนเรื่อ	2022-05-25	Born May 18, 1998	1998	Generation Z
1	1270	83516	2014-12-30	RT @yamyummy: ชัชชาติคงไม่คุยกันตรงๆ แ	2022-05-25	Born May 18, 1998	1998	Generation Z
0	4647	64394	2015-05-03	RT @LittleBrightey1: ถ้าชัชชาติตุ๊บจริงมันต้อง	2022-11-17	Born 1998	1998	Generation Z
0	20047	113693	2013-05-17	RT @Gilg_kun: เราว่าชัชชาติก็ตอบได้ดีนะ เค้า	2022-11-17	Born 1999	1999	Generation Z
4	69454	272697	2012-05-03	RT @LittleBrightey1: แต่ใคร ถ้าชัชชาติตุ๊บนี่มี	2022-11-17	Born 1997	1997	Generation Z
2	2849	450749	2016-02-16	RT @LittleBrightey1: ถ้าชัชชาติตุ๊บจริงมันต้อง	2022-11-17	Born 1998	1998	Generation Z
27	6516	949579	2010-01-10	RT @HIPPOCRATES_i: มาถกเรื่องป้ายเลือกตั้ง	2022-11-17	Born 1992	1992	Generation Y

รูปที่ 3.5 ตัวอย่างข้อมูลที่สามารถหาเจเนอเรชั่นได้ (ส่วนที่ 2 คอลัมน์ listed_count ถึง generation)

จากนั้นนำข้อมูลข้อมูลที่สามารถหาเจเนอเรชั่นได้ ไปทำการสร้างโมเดลขึ้นหัวข้อถัดไป

3.3 การสร้างโมเดล (Data Modeling)

การสร้างโมเดลนี้เกี่ยวข้องกับการทำหัวข้อโมเดลลิ่งของข้อความภาษาไทย โดยใช้วิธี Latent Dirichlet Allocation (LDA) เพื่อค้นหาหัวข้อหรือธีมที่ซ่อนอยู่ในข้อมูล โมเดล LDA จะให้รายการคำและน้ำหนักที่สัมพันธ์กันของแต่ละคำ ในการทำหัวข้อโมเดลลิ่ง LDA เราต้องกำหนดจำนวนหัวข้อที่

ต้องการ สำหรับกรณีของเรา เรากำหนดไว้ที่ 10 หัวข้อ โดยกระบวนการทำงานของการทำหัวข้อโมเดลลิ่ง LDA สามารถอธิบายได้เป็นลำดับขั้นตอนดังนี้

1. นำเข้าไลบรารีที่จำเป็นใช้ไลบรารี

gensim ซึ่งเป็นเครื่องมือที่นิยมสำหรับการทำ Topic Modeling และ LdaModel ซึ่งเป็นอัลกอริทึมที่ใช้ในการระบุหัวข้อในชุดข้อมูล

2. การแปลงข้อความให้อยู่ในรูปแบบของ Token (Tokenization)

ข้อมูลทวีตทั้งหมด (total_tweets) จะถูกแปลงให้อยู่ในรูปแบบของรายการที่ประกอบด้วยคำ (list of tokens) ซึ่งในที่นี้เรียกว่า "texts" และกระบวนการนี้อาจรวมถึงการลบอักขระพิเศษ, การทำให้เป็นตัวพิมพ์เล็กทั้งหมด และการลบคำที่ไม่มีความหมาย (Stopwords Removal)

3. สร้างพจนานุกรม (Dictionary) สำหรับข้อความ

ใช้ gensim.corpora.Dictionary เพื่อสร้างพจนานุกรมที่ประกอบด้วยรายการคำศัพท์ทั้งหมดในชุดข้อมูลและแต่ละคำในพจนานุกรมจะได้รับการกำหนดดัชนีเฉพาะ

4. สร้างคลังข้อมูล (Corpus) ในรูปแบบของ Bag-of-Words (BoW)

ใช้แนวคิด Bag-of-Words (BoW) ซึ่งจะแปลงข้อความเป็นเวกเตอร์เชิงตัวเลข และการใช้ฟังก์ชัน doc2bow() ในการแปลงเอกสารแต่ละรายการให้เป็นเวกเตอร์ของความถี่ของคำศัพท์ในพจนานุกรม

5. รันแบบจำลองหัวข้อด้วย LDA (Latent Dirichlet Allocation - LDA)

ใช้โมเดล LdaModel โดยกำหนดพารามิเตอร์ ได้แก่

- Corpus Name คลังข้อมูลที่ใช้
- Dictionary พจนานุกรมของคำศัพท์
- Number of Topics จำนวนหัวข้อที่ต้องการให้โมเดลระบุ
- Number of Passes จำนวนรอบของการทำซ้ำเพื่อปรับปรุงความแม่นยำของโมเดล

6. แสดงผลหัวข้อที่ได้จากแบบจำลอง

แสดงรายการของหัวข้อที่โมเดลค้นพบ พร้อมคำศัพท์ที่เป็นตัวแทนของแต่ละหัวข้อ สามารถใช้การวัดค่าความสำคัญของคำ (Topic Coherence) เพื่อประเมินคุณภาพของหัวข้อ

7. บันทึกผลลัพธ์ที่ได้

สามารถบันทึกโมเดลและหัวข้อที่ได้รับเพื่อใช้วิเคราะห์ต่อไป หรือใช้ในการจำแนกข้อความใหม่ในอนาคต

โดยที่อัลกอริธึมที่อธิบายการสร้างและการทำงานของหัวข้อโมเดลลิ่ง LDA ได้แสดงอยู่ในรูปที่ 3.6

Input: Total tweet data files
Output: Results from topic modeling

1. Import necessary libraries/functions such as `gensim` and `LdaModel`.
2. Convert `total_tweets` into a list of tokens (named "`texts`").
3. Create a `dictionary` for `texts`.
4. Build a corpus using a bag-of-words (BoW) concept via the `doc2bow` function.
5. Run the topic model (LDA) by specifying parameters for the model: corpus name, dictionary, number of topics, and number of passes.
6. Display the resulting topics.
7. Save the resulting topics.

รูปที่ 3.6 อัลกอริธึมการวิเคราะห์หัวข้อด้วย LDA เพื่อนำหัวข้อออกมาจากกลุ่มข้อมูลทวีต

จากนั้นนำรายการคำและของแต่ละคำ ไปทำการนำเสนอข้อมูลในหัวข้อถัดไป

3.4 การนำเสนอข้อมูล (Data Visualization)

การแสดงผลข้อมูลประกอบด้วยการนำเสนอผลลัพธ์จากทวีตของแต่ละเจเนอเรชั่นในรูปแบบแท็กคลาวด์ แท็กคลาวด์เหล่านี้จะแสดงหัวข้อหรือธีมที่ระบุได้ โดยใช้ขนาดตัวอักษรที่แตกต่างกันเพื่อแสดงน้ำหนักของแต่ละหัวข้อ การนำเสนอรวมถึงแท็กคลาวด์ทั้งหมด 4 รูปแบบ ดังนี้ 1. เจเนอเรชั่น เบบี้ บูมเมอร์ (Baby Boomer), 2. เจเนอเรชั่น เอ็กซ์ (Generation X), 3. เจเนอเรชั่น วาย (Generation Y) และ 4. เจเนอเรชั่น ซี (Generation Z) โดยสามารถแสดงตัวอย่างผลลัพธ์ของแท็กคลาวด์ที่ได้จากการวิเคราะห์ แสดงในรูปที่ 3.7

รูปที่ 3.7 ตัวอย่างผลลัพธ์ในรูปแบบแท็กคลาวด์

บทที่ 4
ผลการดำเนินงาน

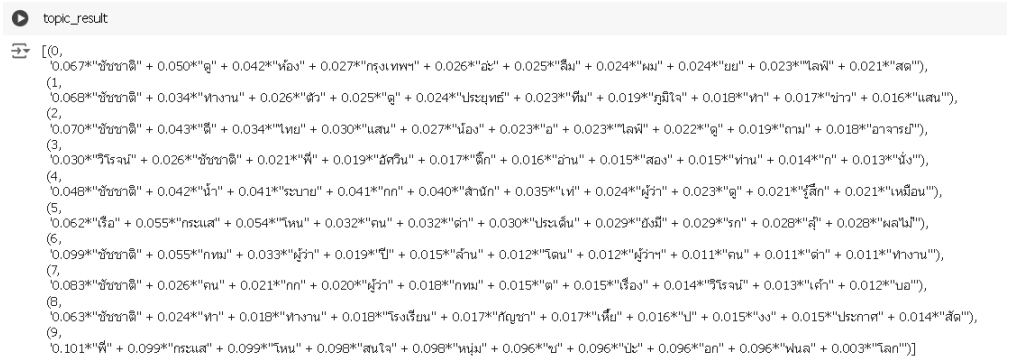
ในบทนี้ จะนำเสนอผลการศึกษา ซึ่งแบ่งออกเป็น 3 ส่วนย่อย ได้แก่ 1 การสร้างโมเดลการจัดประเภทหัวข้อ 2 ภาพที่ตีความแล้วในรูปแบบแท็กคลาวด์ และ 3 ผลการดำเนินงานที่ได้

4.1 ผลการดำเนินงานของการสร้างโมเดลการจัดประเภทหัวข้อ

โมเดลการจัดประเภทหัวข้อที่ได้ทั้งหมด 4 โมเดล หนึ่งโมเดลสำหรับแต่ละเจเนอเรชั่นจากการศึกษา ในแต่ละโมเดลการจัดประเภทหัวข้อต้องกำหนดจำนวนหัวข้อที่ต้องการ สำหรับกรณีของผู้วิจัยกำหนดไว้ที่ 10 หัวข้อ และในแต่ละแถวคือหนึ่งหัวข้อที่ประกอบด้วย 10 คำ

รูปที่ 4.1 ผลลัพธ์โมเดลสำหรับเจเนอเรชั่นเบบี้บูมเมอร์ (Baby Boomer)

รูปที่ 4.2 ผลลัพธ์โมเดลสำหรับเจเนอเรชั่น เอ็กซ์ (Generation X)



```
topic_result

[(0,
  '0.067*"ชัชชาติ" + 0.050*"ดู" + 0.042*"ห้อง" + 0.027*"กรุงเทพฯ" + 0.026*"อะ" + 0.025*"ลืม" + 0.024*"ผม" + 0.024*"ยย" + 0.023*"ไลฟ์" + 0.021*"สด"'),
 (1,
  '0.068*"ชัชชาติ" + 0.034*"ทำงาน" + 0.026*"ตัว" + 0.025*"ดู" + 0.024*"ประยุทธ์" + 0.023*"ทีม" + 0.019*"ภูมิใจ" + 0.018*"ทำ" + 0.017*"ข่าว" + 0.016*"แสน"'),
 (2,
  '0.070*"ชัชชาติ" + 0.043*"ดี" + 0.034*"ไทย" + 0.030*"แสน" + 0.027*"น้อง" + 0.023*"อ" + 0.023*"ไลฟ์" + 0.022*"ดู" + 0.019*"ถาม" + 0.018*"อาจารย์"'),
 (3,
  '0.030*"วิโรจน์" + 0.026*"ชัชชาติ" + 0.021*"ฟ้า" + 0.019*"อัศวิน" + 0.017*"ดึก" + 0.016*"อ่าน" + 0.015*"สอง" + 0.015*"ท่าน" + 0.014*"ก" + 0.013*"นั่ง"'),
 (4,
  '0.048*"ชัชชาติ" + 0.042*"น้ำ" + 0.041*"ระบาย" + 0.041*"กก" + 0.040*"สำนัก" + 0.035*"เท่" + 0.024*"ผู้ว่า" + 0.023*"ดู" + 0.021*"รู้สึก" + 0.021*"เหมือน"'),
 (5,
  '0.062*"เรือ" + 0.055*"กระแส" + 0.054*"โหน" + 0.032*"คน" + 0.032*"ด่า" + 0.030*"ประเด็น" + 0.029*"ยังมี" + 0.029*"รก" + 0.028*"สุ" + 0.028*"ผลไม้"'),
 (6,
  '0.099*"ชัชชาติ" + 0.055*"กทม" + 0.033*"ผู้ว่า" + 0.019*"ปี" + 0.015*"ล้าน" + 0.012*"โตน" + 0.012*"ผู้ว่าฯ" + 0.011*"คน" + 0.011*"ด่า" + 0.011*"ทำงาน"'),
 (7,
  '0.083*"ชัชชาติ" + 0.026*"คน" + 0.021*"กก" + 0.020*"ผู้ว่า" + 0.018*"กทม" + 0.015*"ต" + 0.015*"เรื่อง" + 0.014*"วิโรจน์" + 0.013*"เค้า" + 0.012*"บอ"'),
 (8,
  '0.063*"ชัชชาติ" + 0.024*"ทำ" + 0.018*"ทำงาน" + 0.018*"โรงเรียน" + 0.017*"กัญชา" + 0.017*"เหี้ย" + 0.016*"ป" + 0.015*"งง" + 0.015*"ประกาศ" + 0.014*"สัด"'),
 (9,
  '0.101*"ฟ้า" + 0.099*"กระแส" + 0.099*"โหน" + 0.098*"สนใจ" + 0.098*"หนุ่ม" + 0.096*"ข" + 0.096*"ปะ" + 0.096*"อก" + 0.096*"ฟนล" + 0.003*"โลก"')]
```

รูปที่ 4.3 ผลลัพธ์โมเดลสำหรับเจเนอเรชั่น วาย (Generation Y)

```
topic_result

[(0,
  '0.051*"คน" + 0.035*"ตึน" + 0.028*"ชัชชาติ" + 0.027*"เลือก" + 0.025*"โหน" + 0.025*"กระแส" + 0.023*"เต" + 0.022*"น" + 0.020*"ตัดผม" + 0.018*"ดี"'),
 (1,
  '0.107*"กระแส" + 0.107*"โหน" + 0.100*"ฟ้า" + 0.098*"หนุ่ม" + 0.096*"สนใจ" + 0.094*"ปะ" + 0.094*"อก" + 0.094*"ข" + 0.094*"ฟนล" + 0.006*"คลิป"'),
 (2,
  '0.082*"ชัชชาติ" + 0.028*"ประยุทธ์" + 0.019*"ทำงาน" + 0.018*"ปี" + 0.018*"กทม" + 0.015*"มหาชการ" + 0.014*"ทำ" + 0.014*"ออกมา" + 0.014*"เจอ" + 0.013*"ต้อนรับ"'),
 (3,
  '0.087*"ชัชชาติ" + 0.037*"ดู" + 0.030*"ดี" + 0.021*"ห้อง" + 0.020*"เรือ" + 0.020*"ฟ้า" + 0.019*"แสน" + 0.018*"น้อง" + 0.017*"ผม" + 0.016*"ขอบ"'),
 (4,
  '0.055*"ชัชชาติ" + 0.043*"เรือง" + 0.033*"คน" + 0.022*"บ้าน" + 0.021*"ล้าน" + 0.021*"กทม" + 0.018*"ปี" + 0.017*"ผู้ว่าฯ" + 0.016*"สลัม" + 0.015*"งบ"'),
 (5,
  '0.062*"ชัชชาติ" + 0.026*"เค้า" + 0.026*"ทำงาน" + 0.020*"คน" + 0.017*"วิโรจน์" + 0.016*"โตน" + 0.011*"เรื่อง" + 0.010*"ถาม" + 0.009*"เลือก" + 0.008*"ดู"'),
 (6,
  '0.066*"ชัชชาติ" + 0.041*"น้ำ" + 0.040*"ระบาย" + 0.040*"สำนัก" + 0.040*"กก" + 0.035*"เท่" + 0.027*"ดู" + 0.022*"ผู้ว่า" + 0.019*"เ" + 0.019*"หนัก"'),
 (7,
  '0.066*"ชัชชาติ" + 0.051*"กทม" + 0.040*"ผู้ว่า" + 0.026*"กก" + 0.021*"ต" + 0.020*"คน" + 0.019*"กรุงเทพฯ" + 0.017*"เลือกตั้ง" + 0.014*"หน้า" + 0.014*"กกต"'),
 (8,
  '0.077*"ชัชชาติ" + 0.035*"ดู" + 0.034*"งง" + 0.032*"อ" + 0.021*"ไทย" + 0.021*"กทม" + 0.020*"เรือง" + 0.019*"ผู้ว่า" + 0.018*"ไลฟ์" + 0.017*"ง"'),
 (9,
  '0.070*"ชัชชาติ" + 0.033*"เหี้ย" + 0.027*"สัด" + 0.023*"ไทย" + 0.022*"คน" + 0.021*"กทม" + 0.020*"ผู้ว่า" + 0.018*"ประเทศ" + 0.018*"ว่าการ" + 0.016*"หาเสียง"')]
```

รูปที่ 4.4 ผลลัพธ์โมเดลสำหรับเจเนอเรชั่น ซี (Generation Z)

4.2 ผลการดำเนินงานของการผลลัพธ์ที่แสดงในรูปแบบแท็กคลาวด์

จากขั้นตอนก่อนหน้านี้ ที่ได้สร้างแท็กคลาวด์สำหรับแต่ละเจเนอเรชั่นทั้งสี่ ได้แก่ เจเนอเรชั่น เบบี้บูมเมอร์ (Baby Boomer), เจเนอเรชั่น เอ็กซ์ (Generation X), เจเนอเรชั่น วาย (Generation Y) และ เจเนอเรชั่น ซี (Generation Z) โดยอิงจากคำและหัวข้อที่สร้างจากโมเดลของแต่ละเจเนอเรชั่น แท็กคลาวด์เหล่านี้มาจากผลลัพธ์ของโมเดลการจัดประเภทหัวข้อ โดยแท็กคลาวด์สำหรับแต่ละรุ่นแสดงอยู่ในรูปที่ 4.5 ถึง 4.8

รูปที่ 4.5 ผลการดำเนินงานของการแสดงผลในรูปแบบแท็กคลาวด์เจเนอเรชั่นเบบี้บูมเมอร์ (Baby Boomer)

รูปที่ 4.6 ผลการดำเนินงานของการแสดงผลในรูปแบบแท็กคลาวด์ เจเนอเรชั่น เอ็กซ์ (Generation X)

รูปที่ 4.7 ผลการดำเนินงานของการแสดงผลในรูปแบบแท็กคลาวด์ เจเนอเรชั่น วาย (Generation Y)

รูปที่ 4.8 ผลการดำเนินงานของการแสดงผลในรูปแบบแท็กคลาวด์ เจเนอเรชั่น ซี (Generation Z)

จากรูปแสดงผลการดำเนินงานของการแสดงผลในรูปแบบแท็กคลาวด์ ของแต่ละชุดข้อมูล เห็นได้ว่า

4.3 ผลการดำเนินงาน

ในส่วนนี้จะอภิปรายผลการศึกษาโดยใช้แท็กคลาวด์จากโมเดลการจัดประเภทหัวข้อของแต่ละเจเนอเรชั่น การอภิปรายเน้นถึงความแตกต่างและความคล้ายคลึงของผลลัพธ์ระหว่าง เจเนอเรชั่น เบบี้บูมเมอร์ (Baby Boomer), เจเนอเรชั่น เอ็กซ์ (Generation X), เจเนอเรชั่น วาย (Generation Y) และ เจเนอเรชั่น ซี (Generation Z) ตามที่แสดงในรูปที่ 4.2 ถึง 4.5

4.3.1 ความแตกต่างของการแสดงผลในแต่ละเจนเนเรชั่น

จากผลลัพธ์ในรูปที่ 4.2 ถึง 4.5 เราสามารถสังเกตเห็นความแตกต่างในความถี่และความสำคัญของคำที่ปรากฏในแท็กคลาวด์ของแต่ละเจเนอเรชั่น

เจเนอเรชั่นเบบี้บูมเมอร์ (Baby Boomer) มีคำที่ปรากฏบ่อยและโดดเด่นมักเกี่ยวข้องกับสื่อสิ่งพิมพ์และเว็บไซต์ข่าว รวมถึงชื่อบุคคลที่มีชื่อเสียงในแวดวงการเมืองไทย/สื่อ เช่น "สยามรัฐ" "ข่าว" และ "ประยุทธ์ (นายกรัฐมนตรีในช่วงเวลานั้น)" ซึ่งสะท้อนถึงการสื่อสารและการบริหารจัดการทางการเมืองผ่านหนังสือพิมพ์

เจเนอเรชั่น เอ็กซ์ (Generation X) มีคำสำคัญมักเกี่ยวข้องกับบุคคลและตำแหน่งทางการเมืองโดยเฉพาะในกรุงเทพฯ เช่น "ชัชชาติ" "BKK" และ "ประชาชน" บ่งบอกถึงความสนใจใน “การบริหารและการเมืองท้องถิ่น” ซึ่งแตกต่างจาก เจเนอเรชั่นเบบี้บูมเมอร์ (Baby Boomer) ที่ให้ความสนใจในระดับการเมืองระดับชาติ

เจเนอเรชั่น วาย (Generation Y) มีคำโดดเด่นในกลุ่มนี้มักเกี่ยวข้องกับการเมืองและเหตุการณ์ทางสังคม เช่น "ประเด็น" "กระแส" และ "แสนดี" แสดงถึงความสนใจและการมีส่วนร่วมในเรื่องการเมือง เหตุการณ์ปัจจุบัน และการเคลื่อนไหวต่าง ๆ

เจเนอเรชั่น ซี (Generation Z) มีคำที่ปรากฏบ่อยและไม่พบในรุ่นอื่น ๆ มักเกี่ยวข้องกับความอยากรู้อยากเห็นเกี่ยวกับการเมืองและเหตุการณ์ทางสังคม เช่น "หา" "ถาม" และ "สงสัย" ซึ่งแสดงถึงความกล้าหาญในการแสดงความสนใจและตั้งคำถามเกี่ยวกับเรื่องราวทางการเมือง

บทที่ 5
สรุปผลการดำเนินงานและข้อเสนอแนะ

ในบทนี้จะกล่าวถึงผลการดำเนินงานแบ่งออกเป็น ผลการดำเนินงานจากการสร้างโมเดลการจำแนกหัวข้อ ผลการดำเนินงานที่แสดงผลในรูปแบบแท็กคลาวด์ และการอภิปรายผลการดำเนินงาน

5.1 สรุปผลการดำเนินงาน

จากการดำเนินการค้นคว้าอิสระนี้ ผู้วิจัยหวังว่าจะสามารถเป็นแนวทางในการศึกษาและพัฒนาการค้นคว้าอิสระในอนาคต และเป็นประโยชน์ต่อการนำไปประยุกต์ใช้ในด้านต่างๆ ทั้งในทางธุรกิจ การตลาด และการพัฒนาสังคมต่อไป

สรุปการวิจัยนี้มีเป้าหมายเพื่อศึกษาและวิเคราะห์ข้อมูลเกี่ยวกับเจเนอเรชั่นต่าง ๆ ได้แก่ เจเนอเรชั่นเบบี้บูมเมอร์ (Baby Boomer), เจเนอเรชั่น เอ็กซ์ (Generation X), เจเนอเรชั่น วาย (Generation Y) และ เจเนอเรชั่น ซี (Generation Z) เพื่อเข้าใจความสนใจและความสำคัญของคำบางคำ โดยเริ่มจากการเก็บและคัดเลือกข้อมูลเกี่ยวกับเจเนอเรชั่นต่าง ๆ จาก Twitter ผ่าน Twitter API แล้วสร้างโมเดลการจัดประเภทหัวข้อเพื่อระบุความสัมพันธ์และความสำคัญของคำในแต่ละรุ่น แท็กคลาวด์สำหรับแต่ละรุ่นแสดงดังนี้

เจเนอเรชั่นเบบี้บูมเมอร์ (Baby Boomer) คำที่ปรากฏบ่อยและโดดเด่นมักเกี่ยวข้องกับสื่อสิ่งพิมพ์และเว็บไซต์ข่าว รวมถึงชื่อบุคคลสำคัญในแวดวงการเมืองไทย เช่น "สยามรัฐ" "ข่าว" และ "ประยุทธ์" ซึ่งสะท้อนถึงการบริหารและการสื่อสารทางการเมืองผ่านหนังสือพิมพ์

เจเนอเรชั่น เอ็กซ์ (Generation X) คำสำคัญมักเกี่ยวข้องกับบุคคลและตำแหน่งทางการเมืองในกรุงเทพฯ เช่น "ชัชชาติ" "กรุงเทพ" และ "ประชาชน" แสดงถึงความสนใจในด้านการปกครองและการเมืองท้องถิ่น

เจเนอเรชั่น วาย (Generation Y) คำโดดเด่นมักเกี่ยวข้องกับเหตุการณ์ทางการเมืองและสังคม เช่น "ประเด็น" "กระแส" และ "ดี" ซึ่งสะท้อนถึงความสนใจในเรื่องราวการเมืองและข่าวที่เป็นกระแส

เจเนอเรชั่น ซี (Generation Z) คำที่ปรากฏบ่อยและไม่พบในรุ่นอื่น ๆ มักเกี่ยวข้องกับการตั้งคำถามและการสำรวจเกี่ยวกับการเมืองและกิจกรรมทางสังคม เช่น "หา" "ถาม" และ "สงสัย"

สำหรับการวิจัยในอนาคตแนะนำให้เก็บรวบรวมข้อมูลจากแหล่งที่มีความหลากหลายและครอบคลุมมากขึ้น เพื่อให้ได้ความถูกต้องและสมบูรณ์ยิ่งขึ้น รวมถึงการปรับปรุงและพัฒนาระบบจำแนกหัวข้อ เพื่อให้รองรับการวิเคราะห์ข้อมูลที่ซับซ้อนยิ่งขึ้น นอกจากนี้ ในด้านการตลาดและธุรกิจเราแนะนำให้ใช้การวิเคราะห์คำสำคัญของแต่ละรุ่น เพื่อวางแผนกลยุทธ์การตลาดและการสื่อสารที่เจาะจงกลุ่มเป้าหมายได้ดีขึ้น อีกทั้งยังแนะนำให้นำผลการวิจัยไปพัฒนาผลิตภัณฑ์และบริการที่ตรง

กับความต้องการและความสนใจของแต่ละรุ่น สุดท้าย สำหรับการพัฒนาสังคมและนโยบายสาธารณะ เราแนะนำให้นำข้อมูลนี้ไปใช้ในการวางแผนและพัฒนานโยบายที่ตอบโจทย์ความต้องการของแต่ละรุ่น เพื่อให้เกิดความสมดุลและความยั่งยืนในการพัฒนาสังคม

เอกสารอ้างอิง

Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications–A decade review from 2000 to 2011. Expert systems with applications, 39(12), 11303-11311.

Rivo, E., de la Fuente, J., Rivo, Á., García-Fontán, E., Cañizares, M. Á., & Gil, P. (2012). Cross-Industry Standard Process for data mining is applicable to the lung cancer surgery domain, improving decision making as well as knowledge and quality management. Clinical and Translational Oncology, 14, 73-79.

Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9(1), 381-386.

Rokach, L., & Maimon, O. (2005). Clustering methods. Data mining and knowledge discovery handbook, 321-352.

Burkardt, J. (2009). K-means clustering. Virginia Tech, Advanced Research Computing, Interdisciplinary Center for Applied Mathematics.

Kaser, O., & Lemire, D. (2007). Tag-cloud drawing Algorithms for cloud visualization. arXiv preprint cs/0703109.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

ประวัติผู้เขียน

ชื่อ	นายกิตติศักดิ์พัฒนา ราเชาว์ดี
วัน เดือน ปี เกิด	13 กันยายน พ.ศ.2539
ที่อยู่ปัจจุบัน	บ้านเลขที่ 189 พุทธมณฑลสาย 2 แขวงศาลาธรรมสพน์ เขตทวีวัฒนา กรุงเทพมหานคร 10170
ประวัติการศึกษา	(2561) วิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ เกรดเฉลี่ย 2.81 (สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง)
ทุนการศึกษาที่ได้รับ	ไม่มี
ผลงานทางวิชาการ	กิตติศักดิ์พัฒนา ราเชาว์ดี. (2561). การวิเคราะห์พฤติกรรมผู้ใช้งานบัตรเครดิต [วิทยาศาสตรบัณฑิต]. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.