

การสร้างแบบจำลองหัวข้อ (Topic Modeling) และวิเคราะห์ความรู้สึก (Sentiment Analysis) ของผู้ใช้บริการศูนย์ซื้อขายสินทรัพย์ดิจิทัล (Digital Asset Exchange) กรณีศึกษาเพจเฟซบุ๊กศูนย์ซื้อขายสินทรัพย์ดิจิทัล ในประเทศไทย

วออัปสร บุปผา^{1*} และ เอกรัฐ รัชฎาญจน์²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาทำความเข้าใจพฤติกรรมและความต้องการของผู้ใช้บริการศูนย์ซื้อขายสินทรัพย์ดิจิทัล (Digital Asset Exchange) ในโซเชียลมีเดีย กรณีศึกษาเพจเฟซบุ๊กศูนย์ซื้อขายสินทรัพย์ดิจิทัล ในประเทศไทย โดยใช้ข้อมูลความคิดเห็นของผู้ใช้บริการในเพจเฟซบุ๊กศูนย์ซื้อขายสินทรัพย์ดิจิทัล ได้แก่ บริษัท บิทคับ ออนไลน์ จำกัด (Bitkub), บริษัท สตางค์ คอร์ปอเรชั่น จำกัด (Satang Pro), บริษัท ซิปเม็กซ์ จำกัด (Zipmex) และบริษัท อப்பิต เอ็กซ์เชนจ์ (ประเทศไทย) จำกัด (Upbit) ตั้งแต่วันที่ 1 มกราคม 2564 – 30 เมษายน 2564 ซึ่งได้รับความอนุเคราะห์ใช้เครื่องมือ ZOCIAL EYE จากบริษัท ไวซ์ไซท์ (ประเทศไทย) จำกัด ในการเก็บรวบรวมข้อมูลที่เกี่ยวข้อง โดยผู้วิจัยสนใจศึกษา 6 แง่มุมที่เป็นปัญหาที่นักลงทุนเป็นกังวล ได้แก่ 1) ความปลอดภัย (Security), 2) ค่าธรรมเนียม (Trading Fees), 3) บริษัท (Company), 4) การบริการ (Customer Service), 5) ระบบ และแอปพลิเคชัน (System & Application), และ 6) การสมัคร การลงทะเบียน และการยืนยันตัวตน (Register & KYC) โดยใช้โมเดล Universal Sentence Encoder (USE) ในการหา Topic Modeling เพื่อจำแนกแ่งมุมปัญหาดังกล่าวจากข้อความของผู้ใช้บริการ ในส่วนของการวิเคราะห์ความรู้สึก (Sentiment Analysis) ผู้วิจัยได้สร้างตัวแบบในการประเมินความถูกต้องเพื่อทำการเปรียบเทียบ คือ 1) Support Vector Machine (SVM), 2) Naïve Bayes, 3) Logistic Regression, และ 4) Random Forest การผลประเมินภาพรวมของโมเดล ตัวแบบ Random Forest ให้ประสิทธิภาพดีที่สุด เท่ากับ 72% ในแ่งมุมของบริษัท และในการประเมินเฉพาะความคิดเห็นเชิงลบ (Negative) ตัวแบบ Logistic Regression ให้ประสิทธิภาพดีที่สุด เท่ากับ 72% ในแ่งมุมของบริษัทเช่นกัน และสิ่งที่พบจากการวิเคราะห์ความรู้สึกแต่แ่งมุมของผู้ใช้บริการ คือ ศูนย์ซื้อขายสินทรัพย์ดิจิทัลควรให้ความสำคัญกับเรื่องระบบ และแอปพลิเคชัน (System & Application) เป็นอย่างแรกเพื่อสร้างความพึงพอใจให้แก่ลูกค้า

คำสำคัญ: การสร้างแบบจำลองหัวข้อ, การวิเคราะห์ความรู้สึก, ศูนย์ซื้อขายสินทรัพย์ดิจิทัล, ยูนิเวอร์แซล เซนเทนซ์ เอ็นโค้ดเดอร์

¹ Corresponding Author. E-mail: woraupsom@gmail.com

^{1,2} คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

Topic Modeling and Sentiment Analysis for Digital Asset Exchange: A Case Study of Digital Asset Exchange Facebook page in Thailand

Woraupsorn Bubpha^{1*} and Ekarat Rattagan²

Abstract

This research aims to understand the behavior and needs of digital asset exchange users on social media. A case study of digital asset exchange Facebook page in Thailand by using conversation (comments) on the Facebook page of Bitkub Online Company Limited (Bitkub), Satang Corporation Company Limited (Satang Pro), Zipmex Company Limited (Zipmex), and Upbit Exchange (Thailand) Company Limited (Upbit) from January 1, 2021 - April 30, 2021. The secondary data support by ZOCIAL EYE tool from Wisersight (Thailand) Company Limited. The researcher is interested in 6 aspects that are the investor problems are 1) Security, 2) Trading Fees, 3) Company, 4) Customer Service, 5) System and Application, and 6) Registration and KYC using Universal Sentence Encoder (USE) model to find Topic Modeling to classify the problem from the user messages. In part of Sentiment Analysis, the assessment model for comparison is 1) Support Vector Machine (SVM), 2) Naïve Bayes, 3) Logistic Regression, and 4) Random Forest. The result for the overall model showed Random Forest f1-score was the best performance with 72% from the company aspect and the result for the negative model showed Logistic Regression f1-score was the best performance with 72% in terms of the company as well. Finally, the digital asset exchange should focus on the development and improvement of systems and applications as the priority to create satisfaction to customers.

Keywords: Topic Modeling, Sentiment Analysis, Digital Asset Exchange, Universal Sentence Encoder (USE)

* Corresponding Author. E-mail: woraupsorn@gmail.com

^{1,2} Faculty of Applied Statistics, National Institute of Development Administration

1. บทนำ

1.1 ที่มาและความสำคัญ

ปัจจุบัน การซื้อขายสกุลเงินดิจิทัล หรือ Cryptocurrency เป็นที่นิยมเพิ่มมากขึ้นอย่างต่อเนื่อง โดยเฉพาะบิตคอยน์ (Bitcoin) ซึ่งเป็นสกุลเงินดิจิทัลที่ได้รับความนิยมสูงสุด มูลค่าตลาดของ Cryptocurrency ในปัจจุบันมีมูลค่าสูงถึง 324.7 พันล้านดอลลาร์ และมีจำนวนเหรียญเกิดขึ้นมาแล้วมากถึง 7,812 สกุลทั่วโลก (John Wanguba, 2021) สกุลเงินดิจิทัลสามารถใช้ซื้อขายแลกเปลี่ยนสินค้าและบริการ รวมถึงการลงทุนเพื่อเก็งกำไรจากความผันผวนของราคา โดยการใช้เทคโนโลยีที่เรียกว่า บล็อกเชน (blockchain) ในการทำธุรกรรม ระบบบล็อกเชนจะบันทึกการทำธุรกรรมทางการเงินต่างๆ ระหว่างบุคคล แบบการกระจายศูนย์ (decentralize) โดยไม่จำเป็นต้องมีตัวกลางในการดำเนินการ จึงมีความโปร่งใส น่าเชื่อถือ ปลอดภัย และมีความถูกต้องของข้อมูล (โคกิน ถนอมเพชรสง่า, 2561)

การซื้อขายสกุลเงินดิจิทัลมีหลากหลายช่องทาง ซึ่งช่องทางที่เป็นที่นิยม คือการซื้อขายผ่านศูนย์ซื้อขายสินทรัพย์ดิจิทัล (Digital Asset Exchange) ในประเทศไทยมีผู้ประกอบการธุรกิจที่ได้รับใบอนุญาตให้เป็นศูนย์ซื้อขายสินทรัพย์ดิจิทัลจำนวน 7 บริษัทที่ยังคงดำเนินกิจการ ได้แก่ บริษัท บิทคับ ออนไลน์ จำกัด (Bitkub), บริษัท สตางค์ คอร์ปอเรชั่น จำกัด (Satang Pro), บริษัท ซิปเม็กซ์ จำกัด (Zipmex), บริษัท หัวบี (ประเทศไทย) จำกัด (Huobi), บริษัท อีอาร์เอ็กซ์ จำกัด (ERX), บริษัท อัปบิต เอ็กซ์เชนจ์ (ประเทศไทย) จำกัด (Upbit) และ บริษัท จีเอ็มไอ-แซด.คอม คริปโทโนมิกส์ (ประเทศไทย) จำกัด (Z.comEX) (สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์, 2564)

จากการสำรวจของ Statista ในปี 2018 พบว่าปัญหาที่นักลงทุนได้รับผลกระทบมากที่สุด เมื่อทำการซื้อขายผ่านศูนย์ซื้อขายสินทรัพย์ดิจิทัล คือ ความปลอดภัย (40%) ค่าธรรมเนียมที่สูง (37%) บริษัทขาดสภาพคล่อง (36%) การบริการลูกค้า (33%) รวมไปถึงการใช้งานบนแพลตฟอร์ม (21%) ดังนั้น ทางผู้จัดทำจึงสนใจที่จะศึกษาประเด็นปัญหาเหล่านี้ โดยการสร้างแบบจำลองหัวข้อ (Topic Modeling) พร้อมทั้งวิเคราะห์ความรู้สึก (Sentiment Analysis) ของผู้ใช้บริการศูนย์ซื้อขายสินทรัพย์ดิจิทัลในประเทศไทย โดยการใช้ข้อมูลคำสนทนาในเพจเฟซบุ๊ก (Facebook Page) ของศูนย์ซื้อขายสินทรัพย์ดิจิทัล คือ บริษัท บิทคับ ออนไลน์ จำกัด (Bitkub), บริษัท สตางค์ คอร์ปอเรชั่น จำกัด (Satang Pro), บริษัท ซิปเม็กซ์ จำกัด (Zipmex) และบริษัท อัปบิต เอ็กซ์เชนจ์ (ประเทศไทย) จำกัด (Upbit) เพื่อวิเคราะห์ประเด็นปัญหา และเป็นแนวทางในการสร้างนโยบายในการแก้ไขปัญหาในอนาคต

1.2 วัตถุประสงค์งานวิจัย

การศึกษานี้ มีจุดประสงค์เพื่อศึกษาบทสนทนาของผู้ใช้บริการในสื่อสังคมออนไลน์ดังนี้

1.2.1 ศึกษาและทำความเข้าใจถึงพฤติกรรมและความต้องการของผู้ใช้บริการศูนย์ซื้อขายสินทรัพย์ดิจิทัล ในประเทศไทย

1.2.2 วิเคราะห์ความรู้สึก (Sentiment Analysis) ของผู้ใช้บริการศูนย์ซื้อขายสินทรัพย์ดิจิทัล ในประเทศไทย

1.2.3 เป็นแนวทางให้ศูนย์ซื้อขายสินทรัพย์ดิจิทัลในประเทศไทยเข้าใจความต้องการ ปัญหาของผู้ใช้บริการ และสามารถนำไป พัฒนา แก้ไข ปรับปรุงให้ดีขึ้นในอนาคต

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ความรู้เบื้องต้นเกี่ยวกับสกุลเงินดิจิทัล (Cryptocurrency) และศูนย์ซื้อขายสินทรัพย์ดิจิทัล (Digital Asset Exchange) (ธนาคารแห่งประเทศไทย, 2563)

สกุลเงินดิจิทัล หรือ คริปโทเคอร์เรนซี (Cryptocurrency) หมายถึง หน่วยข้อมูลอิเล็กทรอนิกส์ซึ่งถูกสร้างขึ้นบนระบบหรือเครือข่ายอิเล็กทรอนิกส์ โดยมีความประสงค์ที่จะใช้เป็นสื่อกลางในการแลกเปลี่ยนเพื่อให้ได้มาซึ่ง

สินค้า บริการ หรือสิทธิ์อื่นใด หรือแลกเปลี่ยนระหว่างสินทรัพย์ดิจิทัล และให้หมายความรวมถึงหน่วยข้อมูลอิเล็กทรอนิกส์อื่นใด ตามที่คณะกรรมการ ก.ล.ต. ประกาศกำหนด และตามที่จะอาจจะมีการแก้ไขเปลี่ยนแปลงภายใต้กฎหมายดังกล่าว

บล็อกเชน (Blockchain) หมายถึง เทคโนโลยีการประมวลผลและจัดเก็บข้อมูลแบบกระจายศูนย์ Distributed Ledger Technology (DLT) เป็นรูปแบบการบันทึกข้อมูลที่ใช้หลักการ Cryptography ร่วมกับกลไก Consensus ทำให้ข้อมูลที่ถูกรับบันทึกไปแล้วสามารถเปลี่ยนแปลงหรือแก้ไขได้ยาก เพิ่มความถูกต้องเชื่อถือได้ของข้อมูล

ศูนย์ซื้อขายสินทรัพย์ดิจิทัล (Digital Asset Exchange) หมายถึง ศูนย์กลางหรือเครือข่ายใด ๆ ที่จัดให้มีขึ้นเพื่อการซื้อขายหรือแลกเปลี่ยนสินทรัพย์ดิจิทัล โดยการจับคู่หรือหาคู่สัญญาให้ หรือการจัดระบบหรืออำนวยความสะดวกให้ผู้ซึ่งประสงค์จะซื้อขายหรือแลกเปลี่ยนสินทรัพย์ดิจิทัลสามารถทำความตกลงหรือจับคู่กันได้ โดยกระทำเป็นทางคำปกติ แต่ทั้งนี้ไม่รวมศูนย์กลางหรือเครือข่ายในลักษณะที่คณะกรรมการ ก.ล.ต. ประกาศกำหนด

2.1.2 ผู้ประกอบธุรกิจที่ได้รับใบอนุญาตศูนย์ซื้อขายสินทรัพย์ดิจิทัล (แสดงเฉพาะบริษัทที่ใช้ในงานวิจัย ข้อมูล ณ วันที่ 10 กรกฎาคม 2564)

ตารางที่ 1 ผู้ประกอบธุรกิจที่ได้รับใบอนุญาตศูนย์ซื้อขายสินทรัพย์ดิจิทัล

ศูนย์ซื้อขายสินทรัพย์ดิจิทัล	บริษัท บิทคับออนไลน์ จำกัด (Bitkub)	บริษัท สตางค์ คอร์ปอเรชั่น จำกัด (Satang Pro)	บริษัท ซิปเม็กซ์ จำกัด (Zipmex)	บริษัท อัปบิต เอ็กซ์เชนจ์ (ประเทศไทย) จำกัด (Upbit)
ปีที่ก่อตั้ง	2561	2560	2561	2560
ผู้ก่อตั้ง/ผู้ร่วมทุน	คุณสกลกรรย์ สระเกวี	คุณปรามินทร์ อินโซม	คุณมาคัส ลิม และ ดร.เอกกลภ ยิ้มวิไล	คุณสมโภชน์ อาหุนัย, คุณปริษา ไพระภักทรกุล และคุณชัชวาลย์ เจียรนนท์
ทุนจดทะเบียนเริ่มต้น (ล้านบาท)	290	91.8	62	80
ค่าธรรมเนียมในการซื้อขายต่อครั้ง	0.25%	0.20%	0% - 0.20%	0.15% - 0.25%

2.1.3 กระบวนการจัดการข้อมูลภาษาไทย และโมเดลทางภาษา

2.1.3.1 การประมวลผลภาษาธรรมชาติ (Natural language processing: NLP)

การประมวลผลภาษาธรรมชาติ หรือ Natural language processing (NLP) เป็นวิทยาการแขนงหนึ่งที่ช่วยให้คอมพิวเตอร์สามารถเข้าใจ ตีความ และใช้งานภาษามนุษย์ได้ ความสามารถนี้เข้ามาช่วยธุรกิจได้เป็นอย่างดี โดยปัจจุบันนี้ ชาวโซเชียลอาจเห็นการนำวิทยาการนี้มาช่วยแบรนด์ตอบคำถามอัตโนมัติผ่าน Chatbot หรือการเฝ้าระวังคำสนทนาเชิงลบบนโซเชียลมีเดียเพื่อให้ฝ่ายที่เกี่ยวข้องเข้าไปช่วยแก้ปัญหาก่อนที่จะประเด็นั้นจะลุกลามอย่างไรก็ตาม ความท้าทายของการประมวลผลภาษาธรรมชาติที่เป็นภาษาไทย คือ ความกำกวมของภาษาไทย รวมถึงศัพท์ใหม่ที่เกิดขึ้นบนโซเชียลมีเดียอีกด้วย (ณิชาภา รัตนปรีชาเวช, 2562)

2.1.3.2 Universal Sentence Encoder (USE)

Universal Sentence Encoder (USE) เป็นโมเดลทางภาษาของ google โดยเป็นโมเดลที่ได้รับการเทรนเรียบร้อยแล้ว (Pre-trained) วัตถุประสงค์หลักเพื่อใช้ในการจัดกลุ่มคำ หรือประโยคที่เหมือนกัน ใช้ได้ทั้งภาษา

เดียวกัน หรือต่างภาษา โมเดลรองรับภาษามากถึง 16 ภาษา รวมทั้งภาษาไทย หลักการทำงานของ USE คือ การนำประโยค หรือคำมาแปลงให้เป็นตัวเลข และจัดกลุ่มจากค่าตัวเลข ถ้าประโยค หรือคำใด ที่มีความหมายใกล้เคียงกัน ตัวเลขที่แปลงเหล่านั้นก็จะมีค่าใกล้เคียงกัน และจัดอยู่ในกลุ่มเดียวกัน โมเดลสามารถจัดกลุ่มได้ทั้งภาษาเดียวกัน หรือต่างภาษากัน หรือสามารถใช้เป็นเครื่องมือในการแปลภาษาได้ (Pined Laohapiengsa, 2020)

2.1.3.3 การวิเคราะห์ความรู้สึก (Sentiment Analysis)

การวิเคราะห์ความรู้สึก (Sentiment Analysis) เป็นการนำการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) มาประยุกต์ใช้เพื่อมุ่งเน้นการวิเคราะห์ และตรวจสอบความรู้สึกได้อย่างอัตโนมัติ โดยมีข้อมูลเป็นข้อความแสดงความคิดเห็นของผู้ใช้สินค้าหรือบริการ ที่ได้แสดงความคิดเห็นไว้ในที่ต่าง ๆ ทั้งในกระดานสนทนา (Web board) หรือในสื่อสังคมออนไลน์ เช่น Facebook, Twitter โดยนำข้อมูลมาวิเคราะห์และจำแนกประเภทของความรู้สึก เช่น ความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) หรือความรู้สึกเป็นกลาง (Neutral) (นันทศักดิ์ สุทธิเลิศ, 2560)

2.1.3.4 เหมือนข้อความ

เหมือนข้อความหรือการค้นหาคำหรือข้อความ (Knowledge-Discovery in Text - KDT) คือ กระบวนการที่กระทำกับข้อความ (โดยส่วนใหญ่จะใช้กับข้อความที่มีจำนวนมาก) เพื่อค้นหารูปแบบแนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อความนั้นโดยใช้หลักการค้นคืนสารสนเทศ (Information retrieval), เหมือนข้อมูล (Data mining), การเรียนรู้ของเครื่อง (Machine learning), สถิติ (Statistics) และหลักการประมวลผลทางภาษาธรรมชาติ (Computational linguistics) (วิสุตา เทศเมือง, 2560)

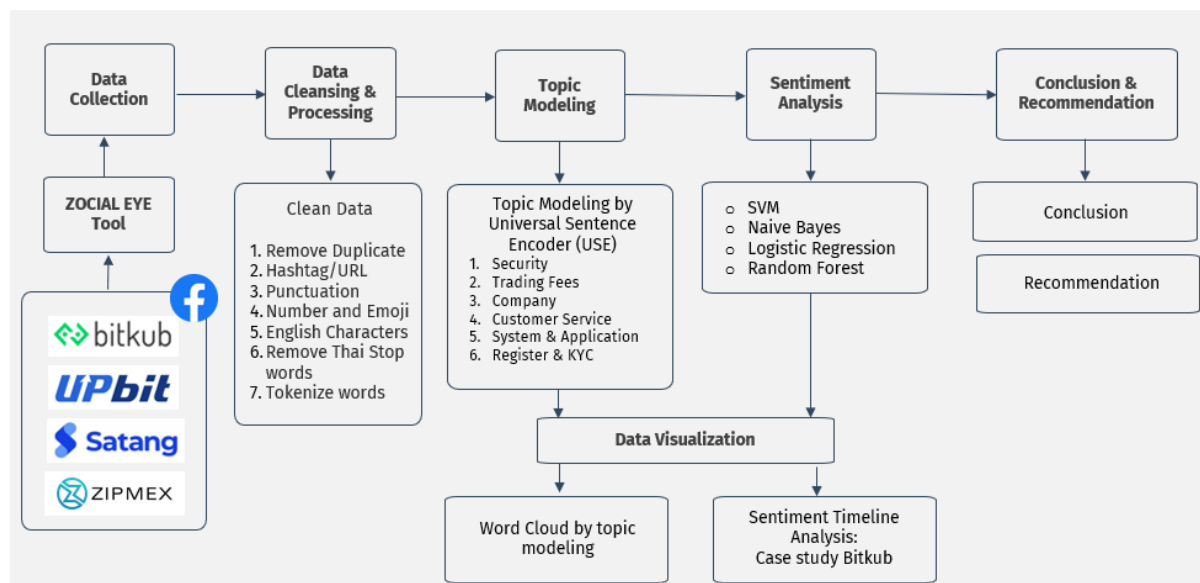
2.2 งานวิจัยที่เกี่ยวข้อง

วิสุตา เทศเมือง (2560) ได้ทำการศึกษาเรื่องการวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิวสินค้าออนไลน์ โดยใช้ขั้นตอนวิธีซอฟต์แวร์แมชชีน นำเสนอวิธีการจำแนกความคิดเห็นโดยการวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิวสินค้าออนไลน์ ด้านการบริการห้องพัก โรงแรม รีสอร์ท จาก Agoda Thailand และ Twitter Thailand จำนวน 2,890 ข้อมูล จากการทดลองพบว่าคุณลักษณะที่ดีที่สุดคือ ซอฟต์แวร์แมชชีน รองลงมาเป็นนาอูฟเบย์ , ต้นไม้ตัดสินใจ และเคเน็ยเรสเนเบอร์ ตามลำดับ

ณิชาภา รัตนปริชาเวช (2562) ได้ทำการศึกษาเรื่องความสัมพันธ์ระหว่างประเด็นต่าง ๆ ที่ถูกพูดถึงในโซเชียลมีเดีย กรณีศึกษา เว็บบอร์ดกลุ่มแม่และเด็ก โดยใช้ข้อมูลความคิดเห็นทั้งหมดที่เกิดขึ้นในเว็บบอร์ด Pantip.com ห้องชานเรือน ผลการวิจัยพบว่า ผู้บริโภคกลุ่มแม่และเด็กส่วนใหญ่พูดคุยกันในประเด็นเรื่อง ค่าใช้จ่ายของโรงเรียน, ความกังวลต่อการตั้งครรภ์, และค่าใช้จ่ายของโรงพยาบาล ที่ Jaccard Similarity 0.2, 0.19, 0.19 ตามลำดับ ทั้งยังมีการแสดงความเครียดและกังวลกับทุกเรื่อง ที่ Normalized Degree Centrality สูงสุด 0.648 โดยเฉพาะสถานการณ์ ท้องไม่พร้อม

สุชีมา แสงจันทร์ (2563) ได้ทำการศึกษาเรื่องการวิเคราะห์ความรู้สึกแต่ละแง่มุมของผู้ใช้บริการส่งอาหารในประเทศไทย ผู้วิจัยได้สร้างตัวแบบสำหรับการวิเคราะห์ความรู้สึกแต่ละแง่มุมของผู้ใช้บริการส่งอาหารในประเทศไทย โดยสร้างตัวแบบหลัก 2 ตัวแบบ คือ ตัวแบบสำหรับการระบุแง่มุม ซึ่งตัวแบบซอฟต์แวร์แมชชีนให้ประสิทธิภาพที่ดีที่สุด และตัวแบบสำหรับการระบุความรู้สึกซึ่งตัวแบบซอฟต์แวร์แมชชีน และนาอูฟเบย์ ให้ประสิทธิภาพดีต่างกันในแต่ละแง่มุม

3. วิธีการดำเนินงานวิจัย



แผนภาพที่ 1 แผนผังวิธีการดำเนินงานวิจัย

3.1 แหล่งที่มาของข้อมูล

ผู้วิจัยเลือกแหล่งที่มาของข้อมูลแบบเจาะจง (Purposive Sampling) คือ บทสนทนาทั้งหมดที่เกิดขึ้นในเพจเฟซบุ๊กของศูนย์ซื้อขายสินทรัพย์ดิจิทัล ได้แก่ เพจ Bitkub, Satang pro, Zipmex และ Upbit เนื่องจากเป็นเพจที่ได้รับความนิยม กำหนดจากยอดผู้ติดตามเพจเกิน 10,000 คน โดยข้อมูลบทสนทนาประกอบไปด้วยข้อมูลที่แอดมินเพจและผู้ให้บริการได้โพสต์ (Post), แสดงความคิดเห็น (Comment) และตอบกลับ (Reply) ดังนั้นจึงเลือกใช้เพียงข้อมูลบทสนทนาของผู้ให้บริการเท่านั้น เพื่อให้ทราบถึงประเด็นปัญหา และความรู้สึกที่แท้จริง นอกจากนี้ข้อมูลที่ได้รับยังประกอบไปด้วยวันที่ และเวลากำกับ (Timestamp) จึงสามารถรับรู้ได้ว่าเป็นประเด็น ณ ขณะใด (Real Time) และจากการตรวจสอบข้อมูล พบว่าบริษัท Bitkub มีข้อมูลที่ครบถ้วน และมากเพียงพอ จึงนำข้อมูลของบริษัท Bitkub เป็นกรณีศึกษาในการนำเสนอ Sentiment Timeline Analysis

3.2 ขั้นตอนการดำเนินงานวิจัย

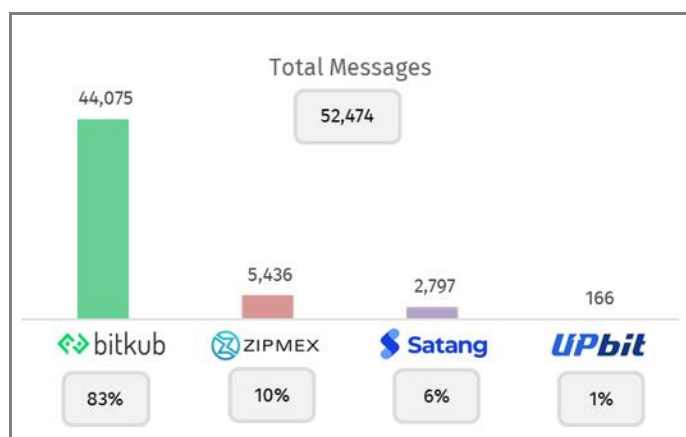
ขั้นตอนการดำเนินงานวิจัย แบ่งเป็น 6 ขั้นตอน ได้แก่ การเก็บรวบรวมข้อมูล, การเตรียมข้อมูล และทำความสะอาดข้อมูล, การจำแนกหัวข้อ (Topic Modeling), การวิเคราะห์ความรู้สึก (Sentiment Analysis) พร้อมประเมินตัวแบบ, การนำเสนอข้อมูล (Visualization) และการสรุปผล โดยมีรายละเอียดดังต่อไปนี้

3.2.1 การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในงานวิจัย เป็นข้อมูลแบบทุติยภูมิ (Secondary Data) จากเครื่องมือ ZOCIAL EYE ของบริษัท ไวซ์ไซท์ (ประเทศไทย) จำกัด ซึ่งเก็บข้อมูลคำสนทนาบนเฟซบุ๊กเพจของศูนย์ซื้อขายสินทรัพย์ดิจิทัล ได้แก่ เพจ Bitkub, Satang pro, Zipmex และ Upbit ตั้งแต่วันที่ 1 มกราคม 2564 – 30 เมษายน 2564 รวม 4 เดือน ซึ่งเป็นข้อมูลตัวอักษรภาษาไทย

3.2.2 การเตรียมข้อมูล และทำความสะอาดข้อมูล

ข้อความทั้งหมด 57,888 ข้อความ บทสนทนาเฉพาะผู้ให้บริการ มีข้อความทั้งหมด 52,474 ข้อความ แบ่งเป็นเพจ Bitkub 44,075 ข้อความ คิดเป็น 83% เพจ Zipmex 5,436 ข้อความ คิดเป็น 10% เพจ Satang Pro 2,797 ข้อความ คิดเป็น 6% และเพจ Upbit 166 ข้อความ คิดเป็น 1%



แผนภาพที่ 2 สัดส่วนข้อมูล

เนื่องจากชนิดของข้อมูลอยู่ในรูปแบบข้อความ (Text) การเตรียมข้อมูลจึงจำเป็นต้องใช้เทคนิค Natural Language Processing (NLP) ในการประมวลผลผ่านโปรแกรมไพทอน (Python) ดังนี้

ผู้วิจัยทำความสะอาดข้อมูล เพื่อตรวจสอบและแก้ไขข้อมูลที่ไม่เกี่ยวข้องออกไปจากชุดข้อมูล ได้แก่ การลบข้อความซ้ำ ลบ Hashtag/URL ลบเครื่องหมายต่างๆ (Punctuation) ลบตัวเลข ลบอีโมจิ ลบอักขรภาษาอังกฤษ เพราะจะใช้เฉพาะอักขรภาษาไทยเท่านั้น ลบคำ Stop word ของภาษาไทย และทำการตัดคำ (Tokenization) เก็บไว้ใน Custom Dictionary รวมถึงตรวจสอบ Label ด้านความรู้สึก ถ้าข้อความแสดงความชื่นชอบ ชื่นชมบริษัท เช่น ให้กำลังใจ, สนับสนุน, ดีมาก, เร็วมาก จะ Label เป็นความรู้สึกเชิงบวก (Positive) ถ้าข้อความแสดงความคำตำหนิ ร้องเรียน แสดงความไม่พอใจ เช่น แย่, ช้า, แพง, ล่ม จะ Label เป็นความรู้สึกเชิงลบ (Negative) แต่ถ้าข้อความไม่ได้แสดงความรู้สึกชัดเจน หรือเป็นคำถาม จะ Label เป็นความรู้สึกกลาง (Neutral)

Account	Message	Post time	Sentimen	original_text	Review_tokenized
UNKNOWN	มาตรฐานแบบ	2021-01-03 1	Negative	มาตรฐานแบบนี้ยังเก็บค่าธรรมเนียมแพงกว่าเว็บอื่นแถมแถมจะ มาตรฐาน,ฐาน,แพง,แถม,ปรับปรอง,ส่งหน้า	
UNKNOWN	ค่าธรรมเนียม	2021-01-03 1	Negative	ค่าธรรมเนียมถอนแพงไปนะเราว่า เมื่อเทียบกับค่ายอื่น ยื่นยื่นแล้ว คถอน,แพง,เทียบ,ค่าย,ตัวตน,ก้,ทันการ,เอาเลย	
UNKNOWN	ค้ทออก	2021-01-04 1	Negative	ค้ทออก เกือบไม่ทันเพราะระบบล่มกว่าจะกดได้ กดได้แล้วตัวเลข ค้ท,ล้ม,กด,กด,ตัวเลข,แอฟ,เสียหาย,รีบผิดชอบ,แพง,คืน,มานะ	
UNKNOWN	ค่าธรรมเนียม	2021-01-04 2	Negative	ค่าธรรมเนียมในการโอนbtค้ทออกไปแพงไปหน่อยหรือครับตอนนี้ btc,แพง	
UNKNOWN	https://www	2021-01-05 0	Negative	แนะนำ ครับเหรียญไทย ไม่ล้ม ไม่มีล้ง ที่สำคัญค่าธรรมเนียมๆ แนะนำ,เสียหาย,ล้ม,ล้ง,ได้เงิน,คืน,กด,สมัคร,ขอ,สงสัย,สอบถาม	
UNKNOWN	ขอทราบ	2021-01-05 1	Neutral	ขอทราบวันที่จะได้เครดิตค่าธรรมเนียมครับ	ทราบ,เครดิต
UNKNOWN	แอฟไม่	2021-01-06 2	Negative	แอฟไม่อัปเดตทันโลก ค้าง ค่าธรรมเนียมแพงมาก	แอฟ,อัปเดต,ทันโลก,ค้าง,แพง

แผนภาพที่ 3 ผลลัพธ์จาก Data Cleansing and Processing

3.2.3 การจำแนกหัวข้อ (Topic Modeling)

การจำแนกหัวข้อ (Topic Modeling) จากข้อความภาษาไทย โดยใช้โมเดล Universal Sentence Encoder (USE) ทางผู้วิจัยได้ทำการจำแนกหัวข้อที่เป็นปัญหาที่ผู้ใช้บริการเป็นกังวลจากข้อมูลอ้างอิงข้างต้น ประกอบด้วย 6 แง่มุม คือ ความปลอดภัย (Security), ค่าธรรมเนียม (Trading Fees), บริษัท (Company), การบริการ (Customer Service), ระบบ และแอปพลิเคชัน (System & Application), การสมัคร การลงทะเบียน และการยืนยันตัวตน (Register & KYC) และมีการทำความสะอาดข้อมูลเพิ่มเติม เพื่อให้ได้ Topic Modeling ตรงกับประเด็นที่ต้องการจะศึกษา

Account	Message	Post tim	Sentiment	original_text	Review_tokenized	category
UNKNOWN	มาตรฐานแบบ	2021-01-03	1 Negative	มาตรฐานแบบที่ยังเก็บค่าธรรมเนียมแพงกว่าเว็บอื่นและครี มาตรา,ฐาน,แพง,แยะ,ปรับปรุง,ล่วงหน้า		ค่าธรรมเนียม
UNKNOWN	ค่าธรรมเนียมถ	2021-01-03	1 Negative	ค่าธรรมเนียมถอนแพงไปนะเราว่า เมื่อเทียบกับค่ายอื่น ยินดี ถอน,แพง,เทียบ,ค่าย,ตัวตน,กั,ทันการ,เอาเลย		ค่าธรรมเนียม
UNKNOWN	คัทออก เก็บ	2021-01-04	1 Negative	คัทออก เก็บไม่ทันเพราะระบบล่มกว่าจะกดได้ กดได้แล้วดี คัท,ล้ม,กด,กด,ตัวเลข,แอฟ,เสียหาย,รับคัทชอม,แพง,คืน,มาจะ		ค่าธรรมเนียม
UNKNOWN	ค่าธรรมเนียมใ	2021-01-04	2 Negative	ค่าธรรมเนียมในการโอนbtออกไม่แพงไปหน่อยหรือคริบ btc,แพง		ค่าธรรมเนียม
UNKNOWN	https://www	2021-01-05	0 Negative	แนะนำ คริบเหรียญไทย ไม่ล้ม ไม่มีสิ่ง ที่สำคัญค่าธรรมเนียม แะนะนำ,เสียหาย,ล้ม,ส่ง,ไต่เงิน,คืน,กด,สมัคร,ขอ,สงสัย,สอบถาม		ค่าธรรมเนียม
UNKNOWN	ขอทราบวันที่	2021-01-05	1 Neutral	ขอทราบวันที่จะได้เครดิตค่าธรรมเนียมคริบ	ทราบ,เครดิต	ค่าธรรมเนียม
UNKNOWN	แอฟไม่อัปเดต	2021-01-06	2 Negative	แอฟไม่อัปเดตทั่วโลก ค้าง ค่าธรรมเนียมแพงมาก	แอฟ,อัปเดต,พันโลก,ค้าง,แพง	ค่าธรรมเนียม

แผนภาพที่ 4 แสดงผลลัพธ์จาก Universal Sentence Encoder (USE)

3.2.4 การวิเคราะห์ความรู้สึก (Sentiment Analysis)

นำข้อมูลจากการจำแนกหัวข้อ (Topic Modeling) ในแต่ละแ่งมุม นำมาวิเคราะห์ความรู้สึก แบ่งข้อมูล เป็น Train data set 70% และ Test data set 30% และจำแนกข้อมูลโดยใช้ตัวแบบในการทำนายว่าแต่ละข้อความ เป็นความรู้สึกด้านใด ได้แก่ 1) Support Vector Machine (SVM), 2) Naïve Bayes, 3) Logistic Regression และ 4) Random Forest

ในส่วนการประเมินผลประสิทธิภาพการจำแนกข้อมูลในงานวิจัย จะประกอบไปด้วยค่าความแม่นยำ (Precision) ค่าเรียกคืน (Recall) ค่าความถูกต้อง (Accuracy) และการวัดประสิทธิภาพโดยรวม (F-measure) ผู้วิจัยทำการประเมินประสิทธิภาพโมเดลโดยใช้ ประสิทธิภาพโดยรวม (F-measure) จะแบ่งเป็น การประเมินภาพรวมของโมเดล กับการประเมินประสิทธิภาพเฉพาะความคิดเห็นเชิงลบ (Negative) ของโมเดล เนื่องจากผู้วิจัยต้องการศึกษา แ่งมุมที่เกี่ยวข้องกับปัญหาของนักลงทุนที่เป็นกังวล ดังนั้นการประเมินประสิทธิภาพความคิดเห็นเชิงลบ (Negative) จึงมีความสำคัญ

3.2.5 การนำเสนอข้อมูล (Visualization) และวิเคราะห์ข้อมูล

การนำเสนอ และวิเคราะห์ข้อมูล แบ่งเป็น 3 รูปแบบ คือ

3.2.5.1 Word Cloud หลังจากการจำแนกหัวข้อ (Topic Modeling) จากข้อความภาษาไทย โดยใช้โมเดล Universal Sentence Encoder (USE) จึงนำผลที่ได้ไปแสดงในรูปแบบ Word Cloud เพื่อแสดงภาพรวมของคำสำคัญที่ปรากฏขึ้นมากที่สุดในแต่ละแ่งมุม

3.2.5.2 Sentiment Timeline Analysis โดยใช้ข้อมูลของบริษัท Bitkub เป็นกรณีศึกษา เพื่อบอกความสัมพันธ์ระหว่างประเด็นปัญหาที่ช่วงเวลาต่างๆ

3.2.5.3 การวิเคราะห์แบบสรุปอุปนัย (Analytical Induction) เนื่องจากข้อความต่าง ๆ บนโซเชียลมีเดียจำเป็นต้องได้รับการตีความเพื่อทำความเข้าใจถึงบริบทของบทสนทนา จึงใช้การวิเคราะห์แบบสรุปอุปนัย กล่าวคือตีความสร้างข้อสรุปข้อมูลจากปรากฏการณ์ที่มองเห็น

3.2.6 การสรุปผล และข้อเสนอแนะ

สรุปประเด็นสำคัญของการจำแนกหัวข้อ (Topic Modeling), การวิเคราะห์ความรู้สึก (Sentiment Analysis) และ Sentiment Timeline Analysis กรณีศึกษาข้อมูลของบริษัท Bitkub พร้อมข้อเสนอแนะ

4. ผลการดำเนินงานวิจัย

ผลการดำเนินงานวิจัย สามารถสรุปได้ดังนี้

4.1 ผลลัพธ์จากการหา Topic Modeling โดยใช้โมเดล Universal Sentence Encoder (USE)

4.1.1 ผลลัพธ์จากการหาหัวข้อ (Topic Modeling) โดยใช้โมเดล Universal Sentence Encoder (USE) แสดงความถี่ของคำที่ปรากฏมากที่สุด 10 อันดับ ในแต่ละแ่งมุม และนำมาแบ่งตามความรู้สึกเชิงบวก และความรู้สึกเชิงลบจากความถี่ของคำที่ปรากฏมากที่สุด 5 อันดับ สรุปได้ตารางที่ 2

ตารางที่ 2 ผลลัพธ์คำที่ปรากฏมากที่สุด 10 อันดับในแต่ละแง่มุม

แง่มุม	คำที่ปรากฏมากที่สุด 10 อันดับ	ความรู้สึก (Sentiment) คำที่ปรากฏมากที่สุด 5 อันดับ	
		ความรู้สึกเชิงบวก (Positive)	ความรู้สึกเชิงลบ (Negative)
ความปลอดภัย	ความเสี่ยง, เสียหาย, ปัญหา, การลงทุน ความเสี่ยง, ล่ม, ปลอดภัย, การลงทุน, เว็บไซต์, กลัว, มั่นใจ	ปัญหา, ความเสี่ยง, เสียหาย, มั่นใจ, แข็งแกร่ง	ความเสี่ยง, ล่ม, เสียหาย, เว็บไซต์, ปัญหา
ค่าธรรมเนียม	ชดเชย, ลด, เสียหาย, ดี, แพง, เสียโอกาส, ล้า, รับผิดชอบ, เกร็ด, ลดลง	ชดเชย, ลด, ดี, ดีมาก, กำลังใจ	ชดเชย, ลด, แพง, ล่ม, เสียหาย
บริษัท	เสียหาย, กำลังใจ, ดี, ปัญหา, แยก, ปรับปรุง, เว็บไซต์, ชดเชย, พัฒนา, เสียโอกาส	กำลังใจ, ปัญหา, ทีม, พัฒนา, ดี	เสียหาย, ดี, รับผิดชอบ, ล่ม, ปรับปรุง
การบริการ	ปัญหา, ปรับปรุง, ดี, ดูแล, หาย, เสียโอกาส, กำลังใจ, ล่ม, แยก, เมล	ดีมาก, ดี, กำลังใจ, ปัญหา, ดูแล	ปัญหา, ปรับปรุง, ล่ม, เสียโอกาส
ระบบ และแอปพลิเคชัน	ล่ม, ปรับปรุง, ดี, หาย, ค้าง, พัฒนา, เสียโอกาส, ล่วงหน้า, รองรับ, แก้	กำลังใจ, ดี, ปรับปรุง, พัฒนา, ดีมาก	ล่ม, ปรับปรุง, ค้าง, ดี, หาย
การสมัครลงทะเบียน และยืนยันตัวตน	รับสมัคร, ตรวจสอบ, ปัญหา, ดี, อนุมัติ, เปิดรับ, ทราบ, ลงทุน, ปรับปรุง, เพื่อน	กำลังใจ, สมัคร, ดี, KYC, เต็ม	ล่ม, ย้าย, ปัญหา, ตรวจสอบ, ปรับปรุง

แง่มุมความปลอดภัย คำที่ปรากฏมากที่สุด คือ คำว่าความเสี่ยง เนื่องจากการซื้อขายสกุลเงินดิจิทัลผ่านศูนย์ซื้อขายสินทรัพย์ดิจิทัล (Digital Asset Exchange) ถึงแม้ว่าศูนย์ซื้อขายที่ได้รับอนุญาต จะได้เตรียมระบบความปลอดภัยไซเบอร์ (cyber security) ไว้เป็นอย่างดีแล้ว แต่ก็ยังมีความเสี่ยงจากการสูญหาย และโอกาสที่จะถูกโจรกรรม (hack) ได้ ดังนั้นการซื้อขายสกุลเงินดิจิทัล จึงควรศึกษาระบบการเก็บรักษาเงินดิจิทัลอย่างปลอดภัย และถูกวิธี เช่น การเก็บเงินดิจิทัลใน hardware wallet มีลักษณะเป็นอุปกรณ์ (device) ซึ่งจะมีกุญแจส่วนตัว (private key) จะถูกเก็บไว้ในอุปกรณ์ในลักษณะออฟไลน์ ทำให้มีความปลอดภัยจากการโจรกรรมทางไซเบอร์สูง

แง่มุมค่าธรรมเนียม คำที่ปรากฏมากที่สุด คือ คำว่าชดเชย การซื้อขายผ่านศูนย์ซื้อขาย จะมีการเก็บค่าธรรมเนียมในการซื้อขาย ซึ่งค่าธรรมเนียมจะแตกต่างกันตามนโยบายของแต่ละบริษัท ดังนั้น จากการศึกษาพบว่า ชดเชย ปรากฏมากที่สุด เนื่องจากในช่วงเวลาที่เก็บข้อมูล เกิดปัญหาเกี่ยวกับระบบการซื้อขาย บริษัทจึงได้มีการชดเชยค่าธรรมเนียมให้แก่ลูกค้า ในขณะที่เดียวกันลูกค้าก็มีการร้องเรียนให้ชดเชยความเสียหายจากปัญหาดังกล่าว โดยสามารถแสดงได้จากความรู้สึกเชิงบวก และความรู้สึกเชิงลบ ซึ่งจะพบคำว่า ชดเชย ปรากฏทั้ง 2 ความรู้สึก โดยความรู้สึกเชิงบวก ตัวอย่างประโยคที่ปรากฏ เช่น ได้ค่าชดเชยแล้วครับ ขอขอบคุณมากครับ และความรู้สึกเชิงลบ ตัวอย่างประโยคที่ปรากฏ เช่น ค่าธรรมเนียมก็แพงกว่าชาวบ้าน ผมเสียหายจะชดเชยอย่างไรครับ

ผลลัพธ์แสดงในตารางที่ 3 เมื่อพิจารณาค่า F1-Score พบว่า ตัวแบบ Random Forest มีประสิทธิภาพดีกว่าตัวแบบอื่นๆ ด้วยค่า F1-Score เท่ากับ 72% ในแง่มุมมองของบริษัท และหากพิจารณาเฉพาะแง่มุมมองบริษัท พบว่า ค่า F1-Score จะมีค่าที่ดีที่สุด เมื่อเทียบกับแง่มุมมองอื่นๆ ในทุกตัวแบบ ได้แก่ Logistic Regression ค่า F1-Score เท่ากับ 68%, Support Vector Machine (SVM) ค่า F1-Score เท่ากับ 67% และ Naïve Bayes ค่า F1-Score เท่ากับ 66%

ตารางที่ 3 ผลลัพธ์การประเมินประสิทธิภาพโดยภาพรวมของโมเดล

ตัวแบบ	แง่มุมมอง	Precision	Recall	F1-Score
Support Vector Machine (SVM)	ความปลอดภัย	38%	39%	38%
	ค่าธรรมเนียม	63%	52%	55%
	บริษัท	68%	66%	67%
	การบริการ	54%	48%	49%
	ระบบ และแอปพลิเคชัน	63%	65%	64%
	การสมัคร ลงทะเบียน และยืนยันตัวตน	59%	51%	54%
Naïve Bayes	ความปลอดภัย	43%	50%	41%
	ค่าธรรมเนียม	43%	48%	39%
	บริษัท	68%	65%	66%
	การบริการ	43%	49%	41%
	ระบบ และแอปพลิเคชัน	42%	47%	40%
	การสมัคร ลงทะเบียน และยืนยันตัวตน	41%	48%	37%
Logistic Regression	ความปลอดภัย	45%	53%	44%
	ค่าธรรมเนียม	41%	35%	36%
	บริษัท	68%	68%	68%
	การบริการ	48%	53%	49%
	ระบบ และแอปพลิเคชัน	47%	55%	47%
	การสมัคร ลงทะเบียน และยืนยันตัวตน	43%	50%	42%
Random Forest	ความปลอดภัย	47%	46%	46%
	ค่าธรรมเนียม	45%	45%	45%
	บริษัท	73%	72%	72%
	การบริการ	57%	52%	53%
	ระบบ และแอปพลิเคชัน	53%	55%	54%
	การสมัคร ลงทะเบียน และยืนยันตัวตน	51%	45%	46%

4.2.2 ประเมินประสิทธิภาพเฉพาะความคิดเห็นเชิงลบ (Negative)

ผลลัพธ์ประเมินประสิทธิภาพเฉพาะความคิดเห็นเชิงลบ (Negative) ของโมเดล ผลลัพธ์แสดงในตารางที่ 4 เมื่อพิจารณาค่า F1-Score พบว่า ตัวแบบ Logistic Regression มีประสิทธิภาพดีกว่าตัวแบบอื่นๆ ด้วยค่า F1-Score เท่ากับ 72% ในแง่มุมมองบริษัท

ตารางที่ 4 ผลลัพธ์การประเมินประสิทธิภาพเฉพาะความคิดเห็นเชิงลบ (Negative)

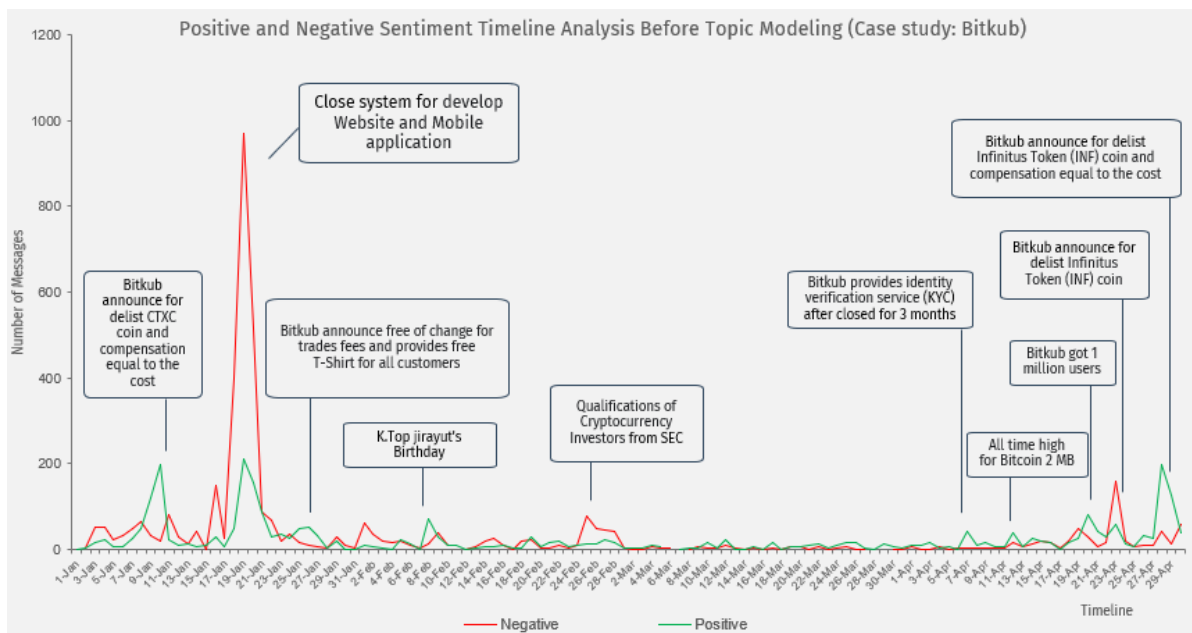
ความรู้สึกเชิงลบ (Negative)	แง่มุม	Precision	Recall	F1-Score
Support Vector Machine (SVM)	ความปลอดภัย	46%	40%	43%
	ค่าธรรมเนียม	53%	56%	54%
	บริษัท	60%	72%	66%
	การบริการ	61%	38%	47%
	ระบบ และแอปพลิเคชัน	67%	70%	68%
	การสมัคร ลงทะเบียน และยืนยันตัวตน	51%	44%	47%
Naïve Bayes	ความปลอดภัย	47%	49%	48%
	ค่าธรรมเนียม	42%	52%	46%
	บริษัท	65%	71%	68%
	การบริการ	46%	59%	51%
	ระบบ และแอปพลิเคชัน	59%	52%	55%
	การสมัคร ลงทะเบียน และยืนยันตัวตน	28%	53%	37%
Logistic Regression	ความปลอดภัย	49%	60%	54%
	ค่าธรรมเนียม	44%	44%	44%
	บริษัท	75%	69%	72%
	การบริการ	48%	47%	47%
	ระบบ และแอปพลิเคชัน	63%	54%	58%
	การสมัคร ลงทะเบียน และยืนยันตัวตน	27%	47%	35%
Random Forest	ความปลอดภัย	55%	51%	57%
	ค่าธรรมเนียม	58%	56%	57%
	บริษัท	72%	66%	68%
	การบริการ	64%	40%	49%
	ระบบ และแอปพลิเคชัน	65%	59%	62%
	การสมัคร ลงทะเบียน และยืนยันตัวตน	40%	33%	36%

4.3 การนำเสนอข้อมูลในรูปแบบของ Sentiment Timeline Analysis

การนำเสนอข้อมูลในรูปแบบของ Sentiment Timeline Analysis เพื่อเข้าใจความรู้สึกเชิงบวก และความรู้สึกเชิงลบของผู้ใช้บริการในแต่ละช่วงเวลา โดยใช้ข้อมูลของบริษัท Bitkub เป็นกรณีศึกษา

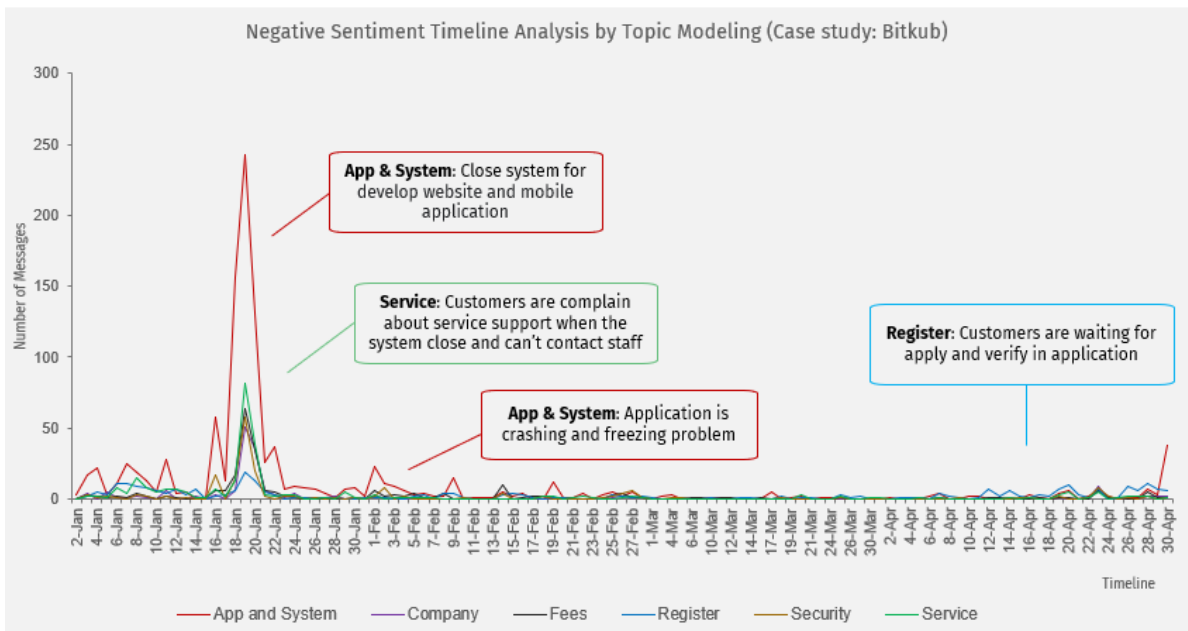
จากแผนภาพที่ 8 แสดง Positive and Negative Sentiment Timeline Analysis ก่อนทำการจำแนกหัวข้อ (Topic Modeling) ของเฟสบุ๊กเพจ Bitkub เนื่องจากต้องการทราบประเด็นปัญหาของผู้ใช้บริการ จึงแสดงเฉพาะความรู้สึกเชิงบวก และความรู้สึกเชิงลบ ในระยะเวลาตั้งแต่วันที่ 1 มกราคม 2564 ถึง 30 เมษายน 2564 พบว่าจากระบบการซื้อขายในช่วงวันที่ 2, 3 และ 16 มกราคม 2564 ไม่สามารถให้บริการลูกค้าได้ตามปกติ และมีการหยุดชะงักของระบบ จึงมีลูกค้าร้องเรียนเป็นจำนวนมาก ทำให้ทางคณะกรรมการ ก.ล.ต. มีมติให้บริษัทดำเนินการแก้ไขระบบงานที่เป็นประเด็นปัญหาให้แล้วเสร็จภายใน 5 วัน ทำให้ในช่วงวันที่ 18-19 มกราคม 2564 ได้ปิดปรับปรุงแอปพลิเคชัน และเว็บไซต์ชั่วคราว และกลับมาเปิดให้บริการวันที่ 20 มกราคม 2564 เวลา 15.00 น. หากพิจารณาจากข้อความที่เป็นความรู้สึก จะพบว่าผู้ใช้บริการมีข้อความที่เป็นความรู้สึกเชิงลบสูงมากในช่วงเวลาดังกล่าว และในวันที่ 25-31 มกราคม 2564 บริษัทได้ชดเชยโดยการให้ลูกค้าทุกคนชื้อขายสกุลเงินดิจิทัล โดยไม่เสียค่าธรรมเนียม และมีกิจกรรมแจกสื่อให้กับลูกค้า ทำให้ช่วงเวลาดังกล่าวมีข้อความที่เป็นความรู้สึกเชิงบวกมากกว่าความรู้สึกเชิงลบ หรือหาก

พิจารณาช่วงวันที่ 8 กุมภาพันธ์ 2564 เป็นวันเกิดของคุณท็อป จิรายุส ทรัพย์ศรีโสภา ประธานเจ้าหน้าที่บริหารของบริษัท หรือช่วงวันที่ 12 เมษายน 2564 ราคาของบิตคอยน์ (Bitcoin) ได้ทำจุดสูงสุดใหม่ที่ราคา 2 ล้านบาทเป็นครั้งแรก จึงทำให้ข้อความที่เป็นความรู้สึกเชิงบวกมากกว่าความรู้สึกเชิงลบ ดังนั้นจาก Sentiment Timeline Analysis พบว่า ข้อความที่เป็นความรู้สึกเชิงบวก หรือความรู้สึกเชิงลบไม่ได้เกิดจากข้อความที่เกี่ยวกับปัญหา การร้องเรียน หรือการชื่นชมเพียงอย่างเดียว แต่มักจะเกี่ยวข้องกับปัจจัยอื่นๆ ที่เป็นประเด็น หรือเป็นกระแสในโลกโซเชียลมีเดียในขณะนั้น และการโพสต์ข้อความของบริษัทก็เป็นแรงจูงใจในการแสดงความคิดเห็นในเชิงบวก หรือเชิงลบเช่นกัน (Influencer) ดังนั้น การวิเคราะห์ความรู้สึก (Sentiment Analysis) โดยไม่ได้คัดกรองข้อมูลที่เกี่ยวข้อง จึงไม่สามารถเข้าใจ และเข้าถึงปัญหาที่ผู้ใช้บริการต้องการจะสื่อสารได้อย่างมีประสิทธิภาพ



แผนภาพที่ 8 Positive and Negative Sentiment Timeline Analysis ก่อนทำการจำแนกหัวข้อ (Topic Modeling) ของเฟสบุ๊คเพจ Bitkub

ดังนั้นทางผู้วิจัยจึงใช้ประโยชน์จากการการจำแนกหัวข้อ (Topic Modeling) ทั้ง 6 แ่งมุม มานำเสนอในรูปแบบเดิม แต่จะเลือกเฉพาะข้อความที่มีความรู้สึกลบ เพื่อให้เข้าใจ และเข้าถึงปัญหาของผู้ใช้บริการ จากแผนภาพที่ 9 แสดง Negative Sentiment Timeline Analysis หลังจากทำการจำแนกหัวข้อ (Topic Modeling) ของเฟสบุ๊คเพจ Bitkub พบว่าสอดคล้องกับ Sentiment Timeline Analysis ภาพรวมที่ยังไม่ได้จำแนกหัวข้อ (Topic Modeling) โดยสามารถระบุ และเจาะจงได้อย่างชัดเจนว่า ในช่วงเวลาที่มีการความรู้สึกเชิงลบสูงนั้นเกิดจากปัญหาเรื่องใด โดยพบว่า ในช่วงวันที่ 18-20 มกราคม 2564 มีการความรู้สึกเชิงลบสูง เกิดจากปัญหาเรื่องระบบ และแอปพลิเคชัน ตามมาด้วยเรื่องการบริการ เนื่องจากในช่วงเวลาที่บริษัททำการปิดปรับปรุงระบบ ผู้ใช้บริการไม่สามารถติดต่อเจ้าหน้าที่เพื่อขอคำแนะนำได้ และพบว่าผู้ใช้บริการมีการร้องเรียนปัญหาเรื่องระบบ และแอปพลิเคชันมาอย่างต่อเนื่องตั้งแต่วันที่ 1 มกราคม จนถึงช่วงต้นเดือนมีนาคม แสดงให้เห็นว่า ถึงแม้ว่าบริษัทจะมีการปิดปรับปรุง และพัฒนาระบบให้ดีขึ้น แต่ก็ยังมีข้อความที่ผู้ใช้บริการแสดงความรู้สึกเชิงลบเกี่ยวกับปัญหาดังกล่าว จึงเป็นสิ่งที่บริษัทควรตระหนัก และพัฒนาระบบให้มีประสิทธิภาพ ในส่วนเรื่องการสมัคร การลงทะเบียน และการยืนยันตัวตนมีข้อความแสดงความรู้สึกเชิงลบเช่นกัน เนื่องจากผู้ใช้บริการยังคงรอคอยที่จะสมัคร และรอการอนุมัติเพื่อใช้งาน ดังนั้น การจำแนกหัวข้อ (Topic Modeling) สามารถเข้าใจ และเข้าถึงปัญหาของผู้ใช้บริการได้มากขึ้น และบริษัทสามารถแก้ไขปัญหาได้ตรงจุด เมื่อเกิดการร้องเรียนของผู้ใช้บริการ



แผนภาพที่ 9 Negative Sentiment Timeline Analysis หลังจากทำการจำแนกหัวข้อ (Topic Modeling) ของเฟสบุ๊คเพจ Bitkub

5. สรุปผลการดำเนินงานวิจัย

5.1 สรุปผลการจำแนกหัวข้อ (Topic Modeling)

การจำแนกหัวข้อที่เป็นปัญหาที่ผู้ใช้บริการเป็นกังวล ประกอบด้วย 6 แง่มุม คือ ความปลอดภัย, ค่าธรรมเนียม, บริษัท, การบริการ, ระบบ และแอปพลิเคชัน, การสมัคร ลงทะเบียน และยืนยันตัวตน สรุปจากคำที่ปรากฏมากที่สุด และคำที่ปรากฏเหมือนกันในแต่ละแง่มุม สรุปได้ดังนี้

แง่มุมความปลอดภัย ผู้ใช้บริการพูดถึง เรื่องความเสี่ยง ความเสียหาย และปัญหามากที่สุด แง่มุมค่าธรรมเนียม ผู้ใช้บริการพูดถึง การชดเชยค่าธรรมเนียม ลดค่าธรรมเนียม และความเสียหายมากที่สุด แง่มุมบริษัท ผู้ใช้บริการพูดถึง ความเสียหาย และให้กำลังใจมากที่สุด แง่มุมการบริการ และแง่มุมระบบ และแอปพลิเคชัน ผู้ใช้บริการพูดถึงเหมือนกันคือ เรื่องการปรับปรุง ระบบล่ม และเสียโอกาส แง่มุมสุดท้าย คือ แง่มุมการสมัครลงทะเบียน และยืนยันตัวตน ผู้ใช้บริการพูดถึง การรับสมัคร ตรวจสอบ และปัญหามากที่สุด

5.2 สรุปผลตัวแบบการระบุความรู้สึก (Sentiment Analysis)

5.2.1 ประเมินประสิทธิภาพโดยภาพรวมของโมเดล

ตัวแบบ Random Forest มีประสิทธิภาพดีกว่าตัวแบบอื่นๆ ด้วยค่า F1-Score เท่ากับ 72% ในแง่มุมของบริษัท

5.2.2 ประเมินประสิทธิภาพเฉพาะความคิดเห็นเชิงลบ (Negative)

ตัวแบบ Logistic Regression มีประสิทธิภาพดีกว่าตัวแบบอื่นๆ ด้วยค่า F1-Score เท่ากับ 72% ในแง่มุมบริษัท

5.3 สรุปผลวิเคราะห์ Sentiment Timeline Analysis

จาก Sentiment Timeline Analysis กรณีศึกษาบริษัท Bitkub ก่อนทำการจำแนกหัวข้อ (Topic Modeling) พบว่าข้อความที่เป็นความรู้สึกเชิงบวก หรือความรู้สึกเชิงลบไม่ได้เกิดจากข้อความที่เกี่ยวกับปัญหา การร้องเรียน หรือ การชื่นชมเพียงอย่างเดียว แต่มักจะเกี่ยวข้องกับปัจจัยอื่นๆ ที่เป็นประเด็น หรือเป็นกระแสในโซเชียลมีเดียใน

ขณะนั้น และการโพสต์ข้อความของบริษัทก็เป็นแรงจูงใจในการแสดงความคิดเห็นในเชิงบวก หรือเชิงลบเช่นกัน (Influencer) ดังนั้น การวิเคราะห์ความรู้สึก (Sentiment Analysis) โดยไม่ได้คัดกรองข้อมูลที่เกี่ยวข้อง จึงไม่สามารถเข้าใจ และเข้าถึงปัญหาที่ผู้ใช้บริการต้องการจะสื่อสารได้อย่างมีประสิทธิภาพ และหลังจากทำการจำแนกหัวข้อ (Topic Modeling) ทั้ง 6 แง่มุม สามารถระบุ และเจาะจงได้อย่างชัดเจนว่า ในช่วงเวลาที่มีการความรู้สึกเชิงลบสูง นั้นเกิดจากปัญหาเรื่องระบบ และแอปพลิเคชัน ทางบริษัทควรปรับปรุง และพัฒนาให้มีประสิทธิภาพมากยิ่งขึ้น เพื่อรองรับลูกค้าใหม่ที่ต้องการจะเข้ามาใช้บริการ ดังนั้น การจำแนกหัวข้อ (Topic Modeling) สามารถเข้าใจ และเข้าถึงปัญหาของผู้ใช้บริการได้มากขึ้น และบริษัทสามารถแก้ไขปัญหาได้ตรงจุด เมื่อเกิดการร้องเรียนของผู้ใช้บริการ

5.4 ข้อเสนอแนะ (Recommendation)

5.4.1. การติดตามข้อมูลแบบ Realtime โดยใช้ข้อมูลจากโซเชียลมีเดีย จำเป็นต้องคัดกรองเรื่องที่เป็นประเด็น หรือเป็นกระแสออกไปด้วย เนื่องจากข้อมูลบนโซเชียลมีเดียเป็นการพูดถึง ณ ขณะที่พวกเขาารู้สึก ดังนั้น เพื่อทราบถึงความต้องการ และปัญหาของผู้ใช้บริการอย่างแท้จริง จึงจำเป็นต้องจัดการคัดกรองข้อมูลที่ไม่ได้เกี่ยวข้องออกไป

5.4.2. สามารถศึกษาเพิ่มเติมประเด็นอื่นๆ จากโซเชียลมีเดีย นอกเหนือจากเรื่องปัญหา เช่น สกุลเงินดิจิทัลที่เป็นที่นิยมในขณะนั้น เป็นต้น

5.4.3. สามารถนำไปต่อยอดไปยังการวิจัยเชิงสำรวจ เช่น ทำแบบสอบถามออนไลน์ เพื่อให้ประเด็นที่ศึกษามีความน่าเชื่อถือยิ่งขึ้น

5.4.4. สามารถเพิ่มเติมขอบเขตการศึกษาในโซเชียลมีเดียในพื้นที่อื่นๆ เช่น พันทิป (Pantip), ทวิตเตอร์ (Twitter) เป็นต้น

6. เอกสารอ้างอิง

เทคซอส ทีม. (2564). เปิดตัว **Upbit Thailand** แพลตฟอร์มเทรดคริปโต โดยทีมผู้บริหารบริษัทใหญ่. สืบค้น 10 กรกฎาคม 2564 จาก <https://techsauce.co/news/upbit-thailand-cryptocurrency>

ไศภิน ถนอมเพชรสง่า. (2561). ปัจจัยที่มีผลต่อราคาสกุลเงินดิจิทัล (**Cryptocurrency**). มหาวิทยาลัยหอการค้าไทย.

ณิชภาภา รัตนปรีชาเวช. (2562). ความสัมพันธ์ระหว่างประเด็นต่าง ๆ ที่ถูกพูดถึงในโซเชียลมีเดียกรณีศึกษา เว็บไซต์บอร์ดกลุ่มแม่และเด็ก. สถาบันบัณฑิตพัฒนบริหารศาสตร์.

ทรินิตี้ดี. (2564). ความเสี่ยงและคำเตือนในการลงทุนสินทรัพย์ดิจิทัล. สืบค้น 10 กรกฎาคม 2564 จาก <https://www.trinitythai.com/en/Service/Digitalasset/14>

ธนาคารแห่งประเทศไทย. (2563). แนวปฏิบัติการใช้เทคโนโลยีบล็อกเชน (**Blockchain**) ในการให้บริการทางการเงิน. ธนาคารแห่งประเทศไทย.

นันท์ภักดิ์ สุทธิเลิศ. (2560). การวิเคราะห์ความรู้สึกผู้โดยสารที่ใช้บริการสายการบินของบริษัทในประเทศไทย สหรัฐอเมริกา. คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา.

ประชาชาติธุรกิจ. (2564). กระดานเทรดบิตคอยน์ '**Satang Pro**' ยืนยันระบบเสถียร พร้อมรับธุรกรรม. สืบค้น 10 กรกฎาคม 2564 จาก <https://www.prachachat.net/finance/news-597231>

วิรัชดา เทศเมือง. (2560). การวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิวสินค้าออนไลน์ โดยใช้ขั้นตอนวิธีชัพพอร์ตเวกเตอร์แมทซ์. มหาวิทยาลัยศรีปทุม.

ลงทุนแบบมนุษย์เงินเดือน. (2564). 5 เว็บไซต์ "Crypto" มีใบอนุญาตในไทย เว็บไหนน่าใช้บ้าง?. สืบค้น 5 กรกฎาคม 2564 จาก <https://salaryinvestor.com/guide/investment-invest-guide/5-trade-crypto/>

- สมศักดิ์ ศรีสุวรรณ. (2563). การวิเคราะห์เหมืองความคิดเห็นโดยใช้เทคนิคการสกัดคำ. คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏลำปาง.
- สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์. (2564). กฎเกณฑ์. สืบค้น 9 กรกฎาคม 2564 จาก <https://www.sec.or.th/TH/pages/lawandregulations/digitalassetbusiness.aspx>
- สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์. (2564). สิทธิทรัพย์ดิจิทัล. สืบค้น 9 กรกฎาคม 2564 จาก <https://www.sec.or.th/digitalasset>
- สิริรามย์. (2564). ย่อหรือ 'Bitkub' เว็บเทรดบิตคอยน์เบอร์ 1 ของไทย ก่อน ก.ล.ต. ลงดาบสั่งแก็ระบบใน 5 วัน. สืบค้น 12 กรกฎาคม 2564 จาก <https://workpointtoday.com/bitkub-trading-bitcoin-5-day-01/>
- สุธีมา แสงจันทร์. (2563). การวิเคราะห์ความรู้สึกแต่ละแง่มุมของผู้ใช้บริการส่งอาหารในประเทศไทย. สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- Bitkub. (2021). **Founders of Bitkub**. cited 2021 July 10 Retrieved from <https://www.bitkub.com/content/founders>
- Bodrunova, S. S., Orekhov, A. V., Blekanov, I. S., Lyudkevich, N. S., & Tarasov, N. A. (2020). **Topic Detection Based on Sentence Embeddings and Agglomerative Clustering with Markov Moment**. Future Internet, 12(9), 144. <https://doi.org/10.3390/fi12090144>
- Jain, A., Tripathi, S., Dwivedi, H. D., & Saxena, P. (2018). **Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis**. 2018 Eleventh International Conference on Contemporary Computing (IC3), 1–7. <https://doi.org/10.1109/IC3.2018.8530659>
- John Wanguba. (2021). **How Many Cryptocurrencies Are There In 2021?**. cited 2021 July 10 Retrieved from <https://e-cryptonews.com/how-many-cryptocurrencies-are-there-in-2021/>
- Pined Laohapiengsa. (2020). ตรวจสอบความถูกต้องของการแปล ด้วย **Universal Sentence Encoder**. cited 2021 June 25 Retrieved from <https://medium.com/airesearch-in-th/-universal-sentence-encoder-ad3c156c2e9b>
- Satangcorp. (2021). **About Us**. cited 2021 July 10 Retrieved from <https://satangcorp.com/en/about/company-profile>
- Statista. (2021). **What are the biggest problems that cryptocurrency traders see in currently available exchanges?**. cited 2021 January 20 Retrieved from <https://www.statista.com/statistics/596462/biggest-problems-cryptoexchanges/>
- Supalerk. (2021). **Categorize Free Text Using Universal Sentence Encoder, Step-by-Step with Python Code**. cited 2021 June 25 Retrieved from <https://spps-supalerk.medium.com/categorize-free-text-using-universal-sentence-encoder-step-by-step-with-python-code-4acf1be42591>
- Techsauce Team. (2021). เปิดตัว **Upbit Thailand** แพลตฟอร์มเทรดคริปโต โดยทีมผู้บริหารบริษัทใหญ่. cited 2021 July 10 Retrieved from <https://techsauce.co/news/upbit-thailand-cryptocurrency>
- Yang, Y., Cer, D., Ahmad, A., Guo, M., Law, J., Constant, N., Abrego, G. H., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2019). **Multilingual Universal Sentence Encoder for Semantic Retrieval**. ArXiv:1907.04307 [Cs]. <http://arxiv.org/abs/1907.04307>
- Zipmex. (2021). **About us**. cited 2021 July 10 Retrieved from <https://zipmex.com/th/about-us/>