



CS 2566/2

วิจัยและพัฒนาระบบที่สามารถวิเคราะห์ความรู้สึกและความคิดเห็น

จากบทวิจารณ์สินค้า

Research and Development of Sentiment Analysis System from
Products Reviews.

โดย

643020045-6 นายณัฐวัฒน์ หมายบุญ

643021257-6 นายอภิสิทธิ์ น้อยหว่า

อาจารย์ที่ปรึกษา : รศ.ดร.อุรฉัตร โคแก้ว

ตำแหน่ง รองศาสตราจารย์

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์

มหาวิทยาลัยขอนแก่น

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชา 313 762 Research Methodology

ภาคเรียนที่ 2 ปีการศึกษา 2567

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์

มหาวิทยาลัยขอนแก่น

(เดือน มีนาคม พ.ศ. 2567)

CS 2566/2

วิจัยและพัฒนาระบบที่สามารถวิเคราะห์ความรู้สึกและความคิดเห็น

จากบทวิจารณ์สินค้า

Research and Development of Sentiment Analysis System from
Products Reviews.

โดย

643020045-6 นายณัฐวัฒน์ หมายบุญ
643021257-6 นายอภิสิทธิ์ น้อยหว่า

อาจารย์ที่ปรึกษา : รศ.ดร.อุรฉัตร โคแก้ว

ตำแหน่ง รองศาสตราจารย์

สาขาวิชาวิทยาการคอมพิวเตอร์วิทยาลัยการคอมพิวเตอร์

มหาวิทยาลัยขอนแก่น

ชื่อ นายณัฐวัฒน์ หมายบุญ รหัสประจำตัว 643020045-6

Mr.Nattawat Maiboon

นายอภิสิทธิ์ น้อยหว่า รหัสประจำตัว 643021257-6

Mr.Aphisit Noiwa

นักศึกษาระดับปริญญาตรี หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

อาจารย์ที่ปรึกษาสัมมนา รศ.ดร.อุรฉัตร โคแก้ว

Project Advisor Assoc. Prof.Urachart Kokaew

1. ชื่อหัวข้อสัมมนา

ภาษาไทย	วิจัยและพัฒนาระบบที่สามารถวิเคราะห์ความรู้สึกและความคิดเห็นจากบทวิจารณ์สินค้า
ภาษาอังกฤษ	Research and Development of Sentiment Analysis System from Products Reviews.

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์ความรู้สึกช่วยให้ธุรกิจต่างๆ ปรับปรุงการนำเสนอสินค้าของตนได้โดยการเรียนรู้ว่าสิ่งใดใช้ได้ผลและสิ่งใดใช้ไม่ได้ผล สามารถวิเคราะห์ความคิดเห็นเกี่ยวกับบทวิจารณ์ออนไลน์ เพื่อรับข้อมูลเชิงลึกเกี่ยวกับคุณลักษณะเฉพาะของสินค้า การวิเคราะห์ความคิดเห็นและความรู้สึก (Sentiment analysis) เพื่อวิเคราะห์และตรวจสอบความรู้สึกหรืออารมณ์ในข้อความ โดยเก็บข้อมูลคอมเมนต์สินค้าภาษาไทยรวมทั้งหมด 4,500 ความคิดเห็นและทำการแยกเป็น 3 ส่วนคือ 1,500 ความคิดเห็น Lazada หลังจากนั้นจะทำการจำแนกความคิดเห็นว่าเป็นข้อความในด้านบวก (Positive) , ด้านลบ (Negative) , ค่ากลาง (Neutrally)

ผลการวิจัยพบว่า เทคนิคที่นำมาใช้ในวิเคราะห์ความคิดเห็นและความรู้สึกประกอบไปด้วย 4 โมเดล ได้แก่ Navie Bayes, Support Vector Machine, Feed Forward Neural Network, และ BERT โดยเลือกจากงานวิจัยที่เกี่ยวข้องที่มีการใช้งานมากที่สุดและผลการทดสอบ เปรียบเทียบการวิเคราะห์ความคิดเห็นและความรู้สึกจากบทวิจารณ์สินค้า พบว่า โมเดลของ Bert ได้ผลลัพธ์การทำนายความคิดเห็นภาษาไทยอยู่ที่ 0.710 และเป็นวิธีที่ดีที่สุดสำหรับการวิเคราะห์ความรู้สึกและความคิดเห็น ทั้งนี้เมื่อพิจารณาผลลัพธ์การทำนายความคิดเห็นภาษาไทยแต่ละ โมเดลจากมากไปน้อยพบว่า Bert ได้เท่ากับ 0.710, Feed Forward Neural Network ได้เท่ากับ 0.704, Support Vector Machine ได้เท่ากับ 0.691, Navie Bayes ได้เท่ากับ 0.64 ตามลำดับ

กิตติกรรมประกาศ

ในงานวิจัยฉบับนี้ สำเร็จลุล่วงได้อย่างสมบูรณ์ด้วยความกรุณาอย่างยิ่งจาก รศ.ดร.อุรฉัตร โคแก้ว ที่ได้สละเวลาอันมีค่าแก่ผู้วิจัย เพื่อให้คำปรึกษาและแนะนำตลอดจน ตรวจทานแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่เป็นอย่างยิ่ง จนงานวิจัยฉบับนี้สำเร็จสมบูรณ์ ลุล่วงได้ด้วยดีคณะผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้ จากใจจริง

สุดท้ายนี้ผู้วิจัยหวังว่างานวิจัยฉบับนี้คงเป็นประโยชน์สำหรับหน่วยงานที่เกี่ยวข้อง และผู้ที่สนใจศึกษาต่อไป

ผู้ทำวิจัย

นายณัฐวัฒน์ หมายบุญ

นายอภิสิทธิ์ น้อยหว่า

สารบัญ		หน้า
บทคัดย่อ		ก
กิตติกรรมประกาศ		ข
สารบัญตาราง		ง
สารบัญภาพ		จ
บทที่ 1	บทนำ	1
	1.ความเป็นมาและความสำคัญของปัญหา	1
	2.วัตถุประสงค์	1
	3.เป้าหมายและขอบเข	1
	4.ประโยชน์ที่ได้รับ	1
บทที่ 2	งานวิจัยและทฤษฎีที่เกี่ยวข้อง	2
	1.ทฤษฎีที่เกี่ยวข้อง	2
	2.งานวิจัย	3
บทที่ 3	วิธีการดำเนินการวิจัย	5
บทที่ 4	ผลของการวิจัย	8
บทที่ 5	สรุปผล อภิปรายและข้อเสนอแนะ	12
	สรุปผล	12
	อภิปรายผล	12
	ข้อจำกัดของระบบ	13
	ปัญหาอุปสรรค	13
	แนวทางการแก้ไข	13
	ข้อเสนอแนะ	13
บรรณานุกรม		14
ภาคผนวก	เครื่องมือ วิธีการเก็บข้อมูล ผลลัพธ์การทดสอบ	15
ประวัติย่อของผู้ทำวิจัย		22-23

สารบัญตาราง

	หน้า
ตารางที่ 1	7
แผนและระยะเวลาดำเนินการ	

สารบัญรูปภาพ

		หน้า
รูปภาพที่ 1	Synthesis Matric	5
รูปภาพที่ 2	ขั้นตอนการออกแบบระบบโดยใช้โมเดล	6
รูปภาพที่ 3	แสดงจำนวนข้อความ	8-11
รูปภาพที่ 4	แสดงจำนวนและร้อยละของข้อความตามที่จำแนกเอาไว้ หลังจากนำไปฝึกกับโมเดล	8
รูปภาพที่ 5	แสดงจำนวนและร้อยละของข้อความตามที่จำแนกเอาไว้ หลังจากนำไปฝึกกับโมเดล	9
รูปภาพที่ 6	แสดงจำนวนและร้อยละของข้อความตามที่จำแนกเอาไว้ หลังจากนำไปฝึกกับโมเดล	10
รูปภาพที่ 7	แสดงจำนวนและร้อยละของข้อความตามที่จำแนกเอาไว้ หลังจากนำไปฝึกกับโมเดล	11
รูปภาพที่ 8	Anaconda Navigator	16
รูปภาพที่ 9	JupyterLab	16
รูปภาพที่ 10	Canva	17
รูปภาพที่ 11	Microsoft Word	17
รูปภาพที่ 12	เก็บ Comment จาก Lazada	18
รูปภาพที่ 13	เก็บ Comment จาก Lazada	18
รูปภาพที่ 14	เก็บ Commend จาก Lazada	19
รูปภาพที่ 15	ผลลัพธ์การทดสอบของ Bert	19
รูปภาพที่ 16	ผลลัพธ์การทดสอบของ Naïve Bayes	20
รูปภาพที่ 17	ผลลัพธ์การทดสอบของ Support Vector Machine	20
รูปภาพที่ 18	รูปภาพที่ 18 ผลลัพธ์การทดสอบของ Feed Forward Neural Network	21

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัญหาที่เกี่ยวข้องกับการทราบความคิดเห็นและอารมณ์จากข้อความ (Sentiment Analysis) คือ เรื่องที่มีความสำคัญอย่างมากเนื่องจากปัญหา ความกำกวมของภาษา ภาษาไทยที่มีโครงสร้างประโยคที่แตกต่างจากภาษาอังกฤษ ในการวิเคราะห์ข้อมูลและเข้าใจความรู้สึกของผู้ใช้หรือลูกค้าต่อสินค้าหรือบริการต่าง ๆ โดยเฉพาะในสถานการณ์ธุรกิจและการตลาด เช่น การวิเคราะห์ความพึงพอใจของลูกค้าต่อสินค้า เพื่อปรับปรุงคุณภาพหรือวิธีการให้บริการให้ดียิ่งขึ้น

การพัฒนาระบบที่สามารถวิเคราะห์ความคิดเห็นและอารมณ์จากข้อความเพื่อใช้ในการปรับปรุงสินค้าหรือบริการของธุรกิจเป็นเรื่องที่มีความสำคัญอย่างมากในสถานการณ์ธุรกิจที่แข่งขันสูง การใช้เทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) และ การเรียนรู้ของเครื่อง (Machine Learning) สามารถช่วยในการวิเคราะห์ข้อมูลที่ซับซ้อนและความรู้สึกและอารมณ์ที่หลากหลายในข้อความได้อย่างมีประสิทธิภาพ ทำให้ธุรกิจสามารถตอบโจทย์ต่อความต้องการและความคาดหวังของลูกค้าได้อย่างมีประสิทธิภาพและรวดเร็วขึ้น และทำให้สามารถเพิ่มประสิทธิภาพในด้านการตลาดและบริการลูกค้าได้ดีขึ้นในอนาคต

วัตถุประสงค์

- 1.เพื่อสำรวจความพึงพอใจของบทวิจารณ์สินค้า และนำมาปรับปรุงสินค้าต่อไป
- 2.เพื่อวิเคราะห์ความคิดเห็นบนร้านค้าออนไลน์ และจำแนกความคิดเห็นเป็นประเภทต่าง ๆ
- 3.เพื่อศึกษาความแม่นยำของโมเดลในการทำนายความคิดเห็นและอารมณ์จากข้อความ

เป้าหมายและขอบเขต

- 1. เนื่องจากความซับซ้อนของภาษาอาจทำให้เกิดข้อผิดพลาดในการวิเคราะห์ความรู้สึกหรือความคิดเห็นจากบทวิจารณ์สินค้าออนไลน์
- 2. ระบบฐานข้อมูลรองรับการเก็บชุดข้อมูลความรู้สึกหรือความคิดเห็นจากบทวิจารณ์สินค้าออนไลน์เป็นภาษาไทย โดยที่ข้อมูลจะอยู่ในช่วงปี 2020-2023
- 3. ระบบสามารถจำแนกความรู้สึกหรือความคิดเห็นจากบทวิจารณ์สินค้าจากแพลตฟอร์มร้านค้าออนไลน์ได้อย่างแม่นยำ

ประโยชน์ที่คาดว่าจะได้รับ

- 1. เข้าใจความรู้สึกและความคิดเห็นของคนและช่วยให้เข้าใจความต้องการของลูกค้าความคิดเห็นของคนทั่วไป หรือกระแสสังคม
- 2. นำไปใช้ในการปรับปรุงสินค้าและบริการให้ตรงกับความต้องการของลูกค้ามากขึ้น
- 3. พัฒนากลยุทธ์การตลาด เพื่อนำไปใช้ในการวางแผนการตลาดและการสื่อสารให้มีประสิทธิภาพมากขึ้น

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้อง

การเรียนรู้เชิงลึก (Deep learning) Deep Learning คือวิธีการเรียนรู้แบบอัตโนมัติด้วยการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ (Neurons) โดยนำระบบโครงข่ายประสาท (Neural Network) มาซ้อนกัน หลายชั้น (Layer) และทำการเรียนรู้ข้อมูลตัวอย่าง ซึ่งข้อมูล ดังกล่าวจะถูกนำไปใช้ในการตรวจจับรูปแบบ (Pattern) หรือจัดหมวดหมู่ข้อมูล (Classify the Data)

การเรียนรู้ของเครื่อง (Machine learning) Machine Learning คือ “การทำให้ระบบคอมพิวเตอร์สามารถเรียนรู้ได้ด้วยตนเองโดยใช้ข้อมูล” Machine Learning เป็น subset ของ AI จุดประสงค์คือเพื่อใช้ในการสร้างแอปพลิเคชันที่มีประสิทธิภาพมากกว่ามนุษย์ในการทำงานบางประเภท โดยการทำให้อัตโนมัติ สามารถพัฒนา และเรียนรู้ได้ด้วยตนเอง

เครื่องมือทำนายเวกเตอร์ (Support Vector Machine) Support Vector Machine (SVM) เป็นอัลกอริทึมที่สามารถนำมาช่วยแก้ปัญหาการจำแนกข้อมูล ใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูล โดยอาศัยหลักการของการหาสมมติฐานของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกต้องเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูลได้ดีที่สุด

เบสเนียนบายส์ (Naïve Bayes) Naive Bayes Classification เป็นหนึ่งใน Classification Model ใช้ในการแบ่งกลุ่มหรือหาเหตุการณ์ที่จะเกิดขึ้นโดยการอิงทฤษฎีความน่าจะเป็นของ Bayes หรือ Bayesian ซึ่ง Target ของโมเดลจะมีความคล้ายคลึงกับ Logistic Regression ว่าจะเกิดเหตุการณ์นั้นหรือไม่โดยจะเพิ่มโอกาสในการเกิดเหตุการณ์เข้าไปด้วย โดยมักจะใช้ในการวิเคราะห์ข้อมูลที่มีความต่อเนื่องของเหตุการณ์ (Dependent Event) เช่น โอกาสในการเกิดโรคในกลุ่มประชากรที่เราสนใจ

การประมวลผลภาษาธรรมชาติ (Natural Language Processing) Natural Language Processing (NLP) คือสาขาของปัญญาประดิษฐ์ที่มุ่งเน้นการประมวลผลและเข้าใจภาษาธรรมชาติของมนุษย์โดยใช้คอมพิวเตอร์และเทคโนโลยี การพัฒนา NLP มีวัตถุประสงค์เพื่อให้เครื่องจักรสามารถเข้าใจและแปลงข้อมูลภาษามนุษย์ให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถประมวลผลได้, รวมถึงสร้างปฏิสัมพันธ์ระหว่างมนุษย์กับคอมพิวเตอร์ผ่านภาษาธรรมชาติ

TF-IDF (Term Frequency-Inverse Document Frequency) เป็นเทคนิคที่ใช้ในการประมวลผลข้อความเพื่อจัดลำดับความสำคัญของคำศัพท์ในเอกสาร โดยคำนวณความถี่ของคำในเอกสาร (TF) และความหนาแน่นของคำในคลังข้อมูล (IDF) และนำมาคูณกัน เพื่อให้ได้ค่า TF-IDF ซึ่งช่วยในการจัดลำดับคำที่มีความสำคัญสูงในเอกสารนั้น ๆ หรือเอกสารที่เกี่ยวข้องกับคำนั้น ๆ มากที่สุด

โครงข่ายประสาทแบบป้อนไปหน้า (Feed Forward Neural Network) ถือเป็นโมเดลที่มีโครงสร้างที่เรียบง่ายที่สุด เพราะว่า การดำเนินการของข้อมูลจะเป็นไปในทิศทางเดียว ก็คือ รับข้อมูลจาก input layer แล้วส่งต่อไปยัง hidden layer เลือดยๆ จนกระทั่งถึง output layer ก็จะหยุด

เบิร์ต (BERT) BERT ย่อมาจาก Bidirectional Encoder Representations from Transformers ถูกพัฒนาโดย Google ก็คือ Artificial Intelligent (AI) ของ Google Algorithm เวอร์ชันใหม่ล่าสุด เหมือนเครื่องมือที่ทำให้สามารถเรียนรู้ภาษาของมนุษย์ Natural Language Processing (NLP) ได้ดีขึ้น สามารถทำความเข้าใจภาษามนุษย์ได้และเรียนรู้แบบ Deep Learning จะช่วยเพิ่มทักษะให้กับระบบ Search Engine ได้ดี นั่นหมายความว่ามันจะช่วยให้ผู้ใช้งานค้นหาควิเวิร์ดได้ดีขึ้นและสร้าง Organic Traffic ได้ตรงใจผู้ใช้งานมากที่สุด

งานวิจัย

X. Yifan และ R. Yong (2021) นำเสนอเกี่ยวกับอัลกอริธึมการทำซ้ำที่อัปเดตด้วยตนเองสำหรับการวิเคราะห์ความรู้สึกของข้อความจากประเมินสินค้าโภคภัณฑ์ โดยที่อัลกอริทึมจะแก้ไขปัญหาความตรงกันระหว่างการประเมินและการให้คะแนน โดย Dataset ได้มาจากแพลตฟอร์มช้อปปิ้งออนไลน์ โดยเฉพาะ JD, Sunnig และ Taobao ในการวิเคราะห์ความรู้สึกของข้อความประเมินสินค้าโดย Dataset ที่ได้มา ได้ทำการทำความสะอาดข้อมูลเรียบร้อยแล้ว โดยจะใช้วิธีการจำแนกแบบ Bayesian classification ซึ่งคำนวณความน่าจะเป็นก่อนหน้าและความน่าจะเป็นภายหลังเพื่อหลีกเลี่ยงอคติเชิงอัตวิสัยและความเหมาะสมมากเกินไป และได้ความแม่นยำของการวิเคราะห์อารมณ์คิดเป็นร้อยละ 98.003 (Sunning) และ 98.931 (Taobao) ตามลำดับ

V. R. Welgamage และคณะ (2020) นำเสนอการวิเคราะห์อารมณ์หรือการขุดเจาะความคิดเห็น (Sentiment Analysis หรือ Opinion mining) โดยรวบรวมข้อมูลจาก Amazon ที่เกี่ยวกับมุมมองของความรู้สึก และความคิดเห็นของผู้ใช้แต่ละรายเกี่ยวกับสินค้านั้นๆ โดยนำข้อมูลมาจากเว็บไซต์ Amazon ทั้งหมด 142.8 ล้านข้อมูลหลังจากนั้นทำการทำความสะอาดข้อมูลให้เหลืออยู่ทั้งหมด 80% เพื่อนำไปใช้ฝึกโมเดล Random Forest, Multinomial Naïve Bayes, Complement Naïve Bayes, Bernoulli Naïve Bayes จึงได้ความแม่นยำของการวิเคราะห์อารมณ์คิดเป็นร้อยละ 83.28 82.69 82.28 75.75 ตามลำดับ

M. A. M. Salem และ A. Y. A. Maghari (2022) นำเสนอการวิเคราะห์ความรู้สึกหรือขุดความคิดเห็นและอารมณ์เพื่ออนุมานแนวโน้มและความประทับใจที่แสดงในข้อมูลที่วิเคราะห์และจัดประเภทเป็นเชิงบวกหรือเชิงลบ โดยการนำข้อมูลมาจาก Kaggle จำนวน 82815 ข้อมูลแล้วนำไปคัดกรองข้อมูลหลังจากนั้นจะทำการแบ่ง Positive, Neutral, Negative, Total ของ Dataset เพื่อทำไปฝึกกับโมเดล Naïve Bayes, SVM, Maximum Entropy, Decision Tree, KNN ทำให้ได้ความแม่นยำในการวิเคราะห์อารมณ์จากความคิดเห็น คิดเป็นร้อยละ 82.1 80.7 82.7 73.6 76.8 ตามลำดับ

K. Purwandari และคณะ (2022) นำเสนอการจำแนกประเภทความรู้สึกของสินค้ายาของอินโดนีเซียเพื่อจัดประเภทข้อมูล Twitter โดยอัตโนมัติเพื่อระบุความรู้สึกหรือความคิดเห็นเชิงบวกหรือเชิงลบ โดยการดึงข้อมูลจาก Twitter ด้วยการใช้ Key-Word ในการค้นหาคำหลังจากนั้นทำการเก็บไว้ใน DataSet หลังจากนั้นทำการทำความสะอาดข้อมูลและทำการแยก Positive และ Negative เพื่อนำไปฝึกกับโมเดล Naïve Bayes, Support Vector Machine, Logistic Regression, Decision Tree, Random Forest ทำให้ผลลัพธ์ในการจำแนกเป็นร้อยละ 0.87 0.89 0.85 0.92 0.90 ตามลำดับ

R. Man และ K. Lin (2021) นำเสนออัลกอริธึมการวิเคราะห์ความรู้สึกที่รวม BERT และ Convolutional Neural Network (CNN) เพื่อการดึงข้อมูลและจำแนกคุณลักษณะ โดยบทความนี้เสนออัลกอริธึมการวิเคราะห์ความรู้สึกที่รวม BERT และ Convolutional Neural Network เพื่อการดึงและจำแนกคุณลักษณะการใช้ BERT และ Convolutional Neural Network เป็นแบบจำลองการแยกคุณลักษณะทำให้ได้ผลลัพธ์ที่ดีที่สุด โดยมีค่า F1 อยู่ที่ย้อยละ 0.903 เมื่อเทียบกับโมเดลอื่น word2vec-svm 0.812, word2vec-cnn 0.850, word2vec-Att-cnn 0.872 ตามลำดับ

S. A. Rahin, T. Hasib และ M. Hassan(2022) นำเสนอวิธีการปรับปรุงประสิทธิภาพการวิเคราะห์ความรู้สึกตามแง่มุม Aspect-based sentiment analysis บนชุดข้อมูลขนาดเล็กโดยการเปรียบเทียบ เครือข่ายประสาทที่เกิดขึ้นซ้ำ (Recurrent Neural Network) และพีดไปข้างหน้า (Feedforward Neural Network) และรวมข้อมูลอินพุตเพิ่มเติมที่สร้างขึ้นโดยใช้เครื่องมือ Neural Network ที่พร้อมใช้งาน และการทดลองแสดงให้เห็นว่าการใช้เวกเตอร์ร่วมกับข้อความที่มีป้ายกำกับจะช่วยให้เพิ่มความแม่นยำในการทำนายความรู้สึกตามแง่มุมสำหรับบทวิจารณ์การแปลงโทเค็นประโยคให้เป็นเวกเตอร์คำที่สร้างขึ้นและป้อนเข้าไปในเครือข่ายประสาทเทียม ให้ได้ผลลัพธ์เช่นกัน และการใช้แท็ก Part of speech ซึ่งมีข้อมูลเชิงความหมาย พร้อมด้วยเวกเตอร์คำ ช่วยเพิ่มประสิทธิภาพการทำงานของโครงข่ายประสาทเทียมใน ABSA ให้ดียิ่งขึ้น

วิธีการดำเนินการวิจัย

1.กำหนดหัวข้องานวิจัย

1.1 เก็บรวบรวมข้อมูลเนื้อหาจากแหล่งข้อมูลต่าง ๆ และทำ Synthesis Matric รูปภาพที่ 1

Author & Date	Propose	Method	Sample	Result/Finding	Weakness	Strong
P. Yuvraj, S. K. Dilip, (2023)	ประเมินความรู้สึกจากทวิ วิจารณ์	โดยใช้โมเดล PAM,AbSA,RNN และ CNN ในการจัดประเภท	Dataset จากทวิวิจารณ์ของ Amazon มากกว่า 14,000 ความคิดเห็น	วิจารณ์ คิดเป็นร้อยละ 90.6 91.7 92.32 93.52 94.91 ตามลำดับ	จำกัดในการปรับเข้ากับ สถานการณ์และความ แม่นยำทางภาษา	เข้าใจความรู้สึกของผู้บริโภค ได้ตรงประการ
X. Yifan, R. Yong, (2023)	การวิเคราะห์ความรู้สึกของ ข้อความจากประเมินสินค้า โลกทัศน์	โดยใช้อัลกอริทึมที่จะแก้ไข ปัญหาความตรงกันระหว่าง การประเมินและการให้ คะแนน	Dataset จากร้านค้าออนไลน์ JD,Sunning และ Taobao	คิดเป็นร้อยละ 98.003 (Sunning) และ 98.931 (Taobao) ตามลำดับ	แหล่งข้อมูลที่นำมาไม่ระบุ โดยชัดเจนถึงข้อมูลของ โมเดลวิเคราะห์ความรู้สึกที่ ถูกนำเสนอในงานวิจัยนี้	โมเดลวิเคราะห์ความรู้สึก ที่นำเสนอ ด้วยเทคนิคโครง ข่ายประสาทเทียมมี ประสิทธิภาพ โมเดลนี้ สามารถทำนายความรู้สึก ของรีวิว ผลิตภัณฑ์ที่ขาย ออนไลน์ได้ดี มีความแม่นยำ สูงถึง 99.17% ขึ้นไป.
S. N. Shailendra, S. Twinkle, (2020)	สำหรับการวิเคราะห์ความรู้สึก ของการวิจารณ์ผลิตภัณฑ์	โดยใช้อัลกอริทึม Random Forest การจำแนก	Dataset จากเว็บไซต์ e-commerce	ผลิตภัณฑ์หนึ่งหรือไม่ที่ ผลิตภัณฑ์หนึ่งหรือไม่ที่	การให้คะแนนดาวจากลูกค้า ไม่ทุกครั้งสอดคล้องกับรีวิวที่ เขียนโดยลูกค้า, ทำให้ยาก ต่อการวัดความรู้สึกที่แม่นยำ ตามคะแนนเท่านั้น	ระบบให้ผลลัพธ์ในรูปแบบ สั้นๆบอกความสัมพันธ์หรือไม่ดี, ลดความจำเป็นในการ อ่านรีวิวทั้งหมดเพื่อวิเคราะห์ ผลิตภัณฑ์
V. R. Welgamage, et al., (2020)	การวิเคราะห์อารมณ์หรือการ จุดเจาะความคิดเห็น	โดยใช้โมเดลคือ Random Forest, Multinomial Naive Bayes, Complement Naive Bayes, Bernoulli Naive Bayes	Dataset จากเว็บไซต์ Amazon 142,8 ล้านความ คิดเห็น	คิดเป็นร้อยละ 83.28 82.69 82.28 75.75 ตามลำดับ	อาจใช้เวลาในการฝึก คอมพิวเตอร์ให้ทำงาน เข้าใจ	การวิเคราะห์ความรู้สึก สามารถจัดประเภทรีวิว ลูกค้าเป็นกลุ่มบวก, กลุ่ม กลาง, และกลุ่มลบ, ช่วยใน การวิเคราะห์ข้อมูลปริมาณ มาก
M. A. M. Salem, A. Y. A. Maghari, (2022)	การวิเคราะห์ความรู้สึกหรือ ชุดความคิดเห็นและอารมณ์ เพื่ออนุมานแนวโน้มและ ความประทับใจที่แสดงใน ข้อมูลทวิวิเคราะห์และจัด ประเภทเป็นเชิงบวกหรือเชิง ลบ	โดยทำการแบ่ง Positive และ Negative ของ Dataset เพื่อนำไปฝึกกับโมเดล Naive Bayes, SVM, Maximum Entropy, Decision Tree, KNN Naive Bayes, SVM, Maximum Entropy, KNN	Dataset จาก kaggle จำนวน 82815 ข้อมูล	คิดเป็นร้อยละ 82.1 80.7 82.7 73.6 76.8 ตามลำดับ	ความยากลำบากในการหา เว็บไซต์ที่เกี่ยวข้องและสรุป ข้อมูล	สามารถจำแนกความรู้สึกที่ ถูกแสดงในรีวิวสินค้าบนเว็บไซต์ และสาขาอื่น ๆ ได้อย่าง แม่นยำ
K. Purwandari, et al., (2022)	การจำแนกประเภทความรู้สึก ของผลิตภัณฑ์ขายของ ออนไลน์เพื่อจัดประเภท ข้อมูล Twitter โดยอัตโนมัติ เพื่อระบุความรู้สึกหรือความ คิดเห็นเชิงบวกหรือเชิงลบ	ทำการทำความสะอาดข้อมูล และทำการแยก Positive และ Negative เพื่อนำไปฝึก กับโมเดล Naive Bayes, Support Vector Machine, Logistic Regression, Decision Tree, Random Forest	Dataset จาก Twitter	จำแนกเป็นร้อยละ 0.87 0.89 0.85 0.92 0.90 ตาม ลำดับ	โมเดลต้นไม้ตัดสินใจ (Decision Tree) และ Random Forest ที่ใช้สำหรับ การจำแนกรูปลักษณ์แสดง ประสิทธิภาพที่ดีกว่าในการ ทำนายความรู้สึกบวก	การวิเคราะห์ความรู้สึก ทั้งหมดมีแก่นแท้ที่แสดง ความรู้สึกแรงในทิศทางเดียว จึงสามารถให้อรรถาธิบาย เกี่ยวกับความคิดเห็นและ ความชื่นชอบของผู้บริโภค ต่อผลิตภัณฑ์

K. Purwandari, et al., (2022)	การจำแนกประเภทความรู้สึก ของผลิตภัณฑ์ขายของ ออนไลน์เพื่อจัดประเภท ข้อมูล Twitter โดยอัตโนมัติ เพื่อระบุความรู้สึกหรือความ คิดเห็นเชิงบวกหรือเชิงลบ	ทำการทำความสะอาดข้อมูล และทำการแยก Positive และ Negative เพื่อนำไปฝึก กับโมเดล Naive Bayes, Support Vector Machine, Logistic Regression, Decision Tree, Random Forest	Dataset จาก Twitter	จำแนกเป็นร้อยละ 0.87 0.89 0.85 0.92 0.90 ตาม ลำดับ	โมเดลต้นไม้ตัดสินใจ (Decision Tree) และ Random Forest ที่ใช้สำหรับ การจำแนกรูปลักษณ์แสดง ประสิทธิภาพที่ดีกว่าในการ ทำนายความรู้สึกบวก	การวิเคราะห์ความรู้สึก ทั้งหมดมีแก่นแท้ที่แสดง ความรู้สึกแรงในทิศทางเดียว จึงสามารถให้อรรถาธิบาย เกี่ยวกับความคิดเห็นและ ความชื่นชอบของผู้บริโภค ต่อผลิตภัณฑ์
S. A. Rahin, T. Hasib, M. Hassan(2022)	วัตถุประสงค์ของการวิจัยคือ เพื่อปรับปรุงประสิทธิภาพ ของการวิเคราะห์ความรู้สึก ตามแง่มุม (ABSA) บนชุด ข้อมูลขนาดเล็กโดยการ เปรียบเทียบ Recurrent networkและFeed forward neural	เปรียบเทียบโครงข่าย ประสาท เทียมRecurrentและ แบบFeed forward สำหรับ การวิเคราะห์ความรู้สึกตาม แง่มุม (ABSA) และพบว่า โครงข่ายประสาทเทียมแบบ ฝึกไปข้างหน้าทำงานได้ดี กว่า พวกเขาตรวจสอบชุดข้อมูล เพิ่มเติมที่สร้างขึ้นโดยใช้ เครื่องมือ NLP ต่างๆ ที่ พร้อมใช้งาน เช่น แยกส่วน ของคำพูด (POS) และการ พินิจคำ	Amazon Review DataและLaptop train Data	recurrent network ค่าPrecision Aspect count 79.7%,Aspects 54.32%,Polarities 59.8% และ Feed-forward network ค่าPrecision Aspect count 94.84%,Aspects 69.61 %,Polarities 64.17%	จำนวนบทวิจารณ์ที่สำหรับ ด้วยประเภทการฝึกมี จำนวนน้อยมาก และมีค่า overfitting เยอะเกินไป	การวิเคราะห์ความรู้สึกตาม แง่มุม (ABSA) มุ่งหวังที่จะ ทำความเข้าใจความคิดเห็น ของผู้ใช้ในมิติที่ลึกซึ้งขึ้น โดยเฉพาะ โดยให้อรรถาธิบาย ลักษณะของคำและคำบริบท โครงข่ายประสาทเทียม รวม ถึงโครงข่ายประสาทเทียม แบบfeed forwardและrecurrent ถูก นำมาใช้เพื่อจัดการกับงาน ABSA เพื่อปรับปรุงประสิทธิภาพ ของโครงข่ายประสาทเทียม ใน ABSA
R. Man, K. Lin (2021)	การวิเคราะห์ความรู้สึกต่อ เหตุการณ์ความคิดเห็น สาธารณะ	โดยใช้แบบจำลอง BERT สำหรับการดึงข้อมูล คุณลักษณะ และ CNN สำหรับการจำแนกประเภท	Dataset จาก Weibo	คิดเป็นร้อยละ 0.903	ประสิทธิภาพของอัลกอริทึม อาจแตกต่างกันไปขึ้นอยู่กับ ชุดข้อมูลและบริบทเฉพาะ เนื่องจากผลการทดลองขึ้น อยู่กับชุดข้อมูลสาธารณะ	เข้าใจลักษณะเฉพาะของ ข้อมูลได้ดีขึ้น อัลกอริทึมได้รับการแนะนำ สูง ค่าrecall และค่า F1 โดย ค่า F1 สูงถึง 0.903

รูปภาพที่ 1 Synthesis Matric

1.2 จัดเตรียมข้อมูลจากบทวิจารณ์สินค้าจากแพลตฟอร์มร้านค้าออนไลน์ Lazada โดยที่ Positive จะเก็บที่ 4 – 5 ดาว ส่วน Neutrally จะเก็บที่ 3 ดาว และ Negative จะเก็บที่ 1 – 2 ดาว

1.3 ทำการกำหนด Label ของคำจากข้อมูลที่คัดแยกไว้เพื่อนำเข้าโมเดลสำหรับทดสอบ

1.4 สรุปผลลัพธ์ที่ได้หลังจากการทดสอบโมเดล

1.5 รวบรวมผลการทดสอบโมเดล เพื่อใช้ในการพัฒนา

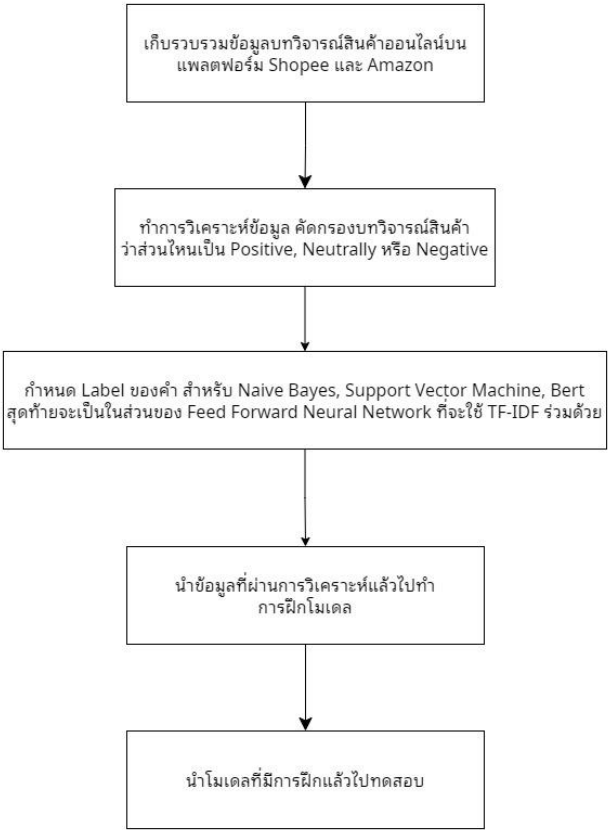
1.6 พัฒนาโมเดลในการวิเคราะห์ความรู้สึกและความคิดเห็น

1.7 ทดสอบโมเดลและประเมินผลลัพธ์ที่ได้

1.8 วิเคราะห์โมเดลและปรับปรุงโมเดลให้มีประสิทธิภาพ

1.9 นำโมเดลไปใช้งาน

งานวิจัยนี้แบ่งขั้นตอนในการออกแบบโมเดลโดยใช้ Feed Forward Neural Network ,Navie Bayes, Support Vector Machine และ BERT ดังรูปภาพที่ 2



รูปภาพที่ 2 ขั้นตอนการออกแบบระบบโดยใช้โมเดล

ข้อมูลบทวิจารณ์สินค้าที่นำเข้า

ข้อมูลที่นำเข้าเป็นภาษาไทยรวมทั้งหมด 4,500 ความคิดเห็นและทำการแยกเป็น 3 ส่วน คือ 1,500 ความคิดเห็นและจะต้องสามารถทำการแยกความคิดเห็นนั้นได้ว่าเป็น Positive, Neutrally หรือ Negative

การคัดกรองบทวิจารณ์สินค้า

ทำการแยก Positive, Neutrally หรือ Negative ของบทวิจารณ์สินค้า หลังจากนั้นทำความสะอาดข้อมูลด้วยการลบอีโมจิและคำซ้ำต่างๆ

กำหนดค่าตัวเลขให้กับคำ

ในการทำการกำหนดค่าตัวเลขให้กับคำจะใช้ Label ในส่วนของ Naïve Bayes, Support Vector Machine, Bert, สุดท้ายจะเป็นในส่วนของ Feed Forward Neural Network ที่จะใช้ TF-IDF ร่วมด้วย

การฝึกโมเดล

ฝึกโมเดลด้วย Naïve Bayes, Support Vector Machine, Bert และ Feed Forward Neural Network โดยจะนำข้อมูลที่ทำการคัดกรองแล้วไปใส่ในโมเดล

แผนและระยะเวลาดำเนินการ

การดำเนินงาน	สัปดาห์/เดือน																			
	ปี 2566/2567																			
	พฤศจิกายน				ธันวาคม				มกราคม				กุมภาพันธ์				มีนาคม			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.กำหนดหัวข้องานวิจัย																				
2.ศึกษาอ่านบทความจากแหล่งข้อมูลต่างๆ																				
3.ใช้ตารางสังเคราะห์																				
4.เขียนเอกสารอ้างอิง																				
5.ร่างเค้าโครงและปรับปรุง																				
6.เขียนโครงงานและตรวจสอบความเรียบร้อย																				
7.เก็บข้อมูล																				
8.ทดลองเขียนโมเดลและปรับปรุง																				
9.นำเสนอโครงงาน																				

ตารางที่ 1 แผนและระยะเวลาดำเนินการ

ผลของการวิจัย

การศึกษาวิจัยครั้งนี้ผู้วิจัยได้ศึกษา”วิจัยและพัฒนาระบบที่สามารถวิเคราะห์ความรู้สึกและความคิดเห็นจากบทวิจารณ์สินค้า”โดยการศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาความคิดเห็นของร้านค้าออนไลน์ ศึกษาการปรับปรุงสินค้าร้านค้าออนไลน์จากความคิดเห็น ศึกษาการจำแนกความคิดเห็น ศึกษาความแม่นยำของโมเดลที่ได้จากการวิเคราะห์จากความคิดเห็น จากวิธีการดำเนินการวิจัย ผู้วิจัยสามารถวิเคราะห์ดำเนินการตามลำดับหัวข้อต่อไปนี้

- 1. ฝึกด้วยโมเดล Naïve Bayes
- 2. ฝึกด้วยโมเดล Support Vector Machine
- 3. ฝึกด้วยโมเดล Bert
- 4. ฝึกด้วยโมเดล Feed Forward Neural Network

1.ฝึกด้วยโมเดล Naïve Bayes

โมเดล Naïve Bayes ได้ข้อมูลมาจาก Lazada ทั้งหมด 4500 ข้อความ แบ่งเป็น 3 ส่วนดังนี้

Positive,Neutrally และ Negative

ส่วนที่ 1 ความคิดเห็นจากร้านค้าออนไลน์

	User	Message	Sentiment
289	*****765	เสื้อตัวเล็กและผ้าบางมากๆ	Neutrally
819	0***8	บางไปนิดคะ	Neutrally
758	ภกมล แ	สั่งสีขาวได้สีส้มแปด ปวดหัว ร้านค้าตอบกลับด...	Negative
696	จันทนา ใ	สินค้าเล็กมากๆๆใส่ไม่ได้	Negative
986	*****644	เนื้อผ้าไม่โอเคไม่ผ่าน โทรมมาก	Negative
...
24	0***5	ส่งสินค้าเร็วมากๆ รองเท้านุ่ม ใส่ดีสมราคาละ	Positive
554	Thanakorn N.	เนื่องงานด้านออกดูโอเค แต่ข้างในไม่เนียน กระเป...	Neutrally
528	ร***.	กระเพาใบใหญ่ดี มีรอยบนนิดหน่อย รอยเหมือนรูปใน...	Neutrally
32	สุชาติ จ.	การเคลื่อนไหวที่เยี่ยมสงขและราบรื่น ดีไซน์ที่ดู...	Positive
1177	WORAPON S.	คุณภาพสมราคา ไม้ได้ดีมาก	Neutrally

รูปภาพที่ 3 แสดงจำนวนข้อความ

<pre>from sklearn.metrics import classification_report print(classification_report(test_labels,pred))</pre>				
	precision	recall	f1-score	support
0	0.65	0.66	0.66	150
1	0.79	0.33	0.46	150
2	0.60	0.95	0.74	150
accuracy			0.64	450
macro avg	0.68	0.64	0.62	450
weighted avg	0.68	0.64	0.62	450

รูปภาพที่ 4 แสดงจำนวนและร้อยละของข้อความตามที่จำแนกเอาไว้ หลังจากนำไปฝึกกับโมเดล

2.ฝึกด้วยโมเดล Support Vector Machine

โมเดล Support Vector Machine ได้ข้อมูลมาจาก Lazada ทั้งหมด 4500 ข้อความ แบ่งเป็น 3 ส่วนดังนี้ Positive,Neutrally และ Negative

ส่วนที่ 1 ความคิดเห็นจากร้านค้าออนไลน์

	User	Message	Sentiment
289	*****765	เสื้อตัวเล็กและผ่าบางมากๆๆ	Neutrally
819	0***8	บางไปนิดค่ะ	Neutrally
758	ภาภมล แ	สั่งสีขาวได้สีส้มแปด ปวดหัว ร้านค้าตอบกลับด...	Negative
696	จันทนา ใ	สินค้าเล็กมากๆๆใส่ไม่ได้	Negative
986	*****644	เนื้อผ้าไม่โอเคไม่ผ่าน โทรมมาก	Negative
...
24	0***5	ส่งสินค้าเร็วมากๆๆ รองเท้านุ่ม ใส่ดีสมราคาค่ะ	Positive
554	Thanakorn N.	เนื้องานด้านนอกดูโอเค แต่ข้างในไม่เนียน กระเป๋...	Neutrally
528	ร***.	กระเพาะใบใหญ่ดี มีรอยบนนิดหน่อย รอยเหมือนจรูปใน...	Neutrally
32	สุชาติ จ.	การเคลื่อนไหวที่เจียบสงบนและราบรื่น ดีไหมที่ดู...	Positive
1177	WORAPON S.	คุณภาพสมราคา ไม่ได้ดีมมาก	Neutrally

รูปภาพที่ 3 แสดงจำนวนข้อความ

<pre># Suppress warnings warnings.filterwarnings("ignore", category=UserWarning) # Print classification report print(classification_report(y_test, predictions))</pre>				
	precision	recall	f1-score	support
0	0.74	0.61	0.67	150
1	0.65	0.67	0.66	150
2	0.70	0.79	0.74	150
accuracy			0.69	450
macro avg	0.69	0.69	0.69	450
weighted avg	0.69	0.69	0.69	450

รูปภาพที่ 5 แสดงจำนวนและร้อยละของข้อความตามที่จำแนกเอาไว้ หลังจากนำไปฝึกกับโมเดล

3.ฝึกด้วยโมเดล Bert

โมเดล Bert ได้ข้อมูลมาจาก Lazada ทั้งหมด 4500 ข้อความ แบ่งเป็น 3 ส่วนดังนี้
Positive,Neutrally และ Negative

ส่วนที่ 1 ความคิดเห็นจากร้านค้าออนไลน์

	User	Message	Sentiment
289	*****765	เสื้อตัวเล็กและผ้าบางมากๆ	Neutrally
819	0***8	บางไปนิดค่ะ	Neutrally
758	ภกมล แ	สั่งสีขาวได้สีส้มเปิด ปวดหัว ร้านค้าตอบกลับด...	Negative
696	จันทนา ใ	สินค้าเล็กมากกกกกกกใส่ไม่ได้	Negative
986	*****644	เนื้อผ้าไม่โอเคไม่ผ่าน โทรมมาก	Negative
...
24	0***5	ส่งสินค้าเร็วมากกกๆ รองเท้านุ่ม ใส่ดีสมราคาค่ะ	Positive
554	Thanakorn N.	เนื้องานด้านนอกดูโอเค แต่ข้างในไม่เนียน กระเป๋...	Neutrally
528	ร***.	กระเพาะใบใหญ่ดี มีรอยบนนิดหน่อย รอยเหมือนๆรูปใน...	Neutrally
32	สุชาติ จ.	การเคลื่อนไหวที่เจ็บสงบนและราบรื่น ดีไหมที่ดู...	Positive
1177	WORAPON S.	คุณภาพสมราคา ไม่ได้ดีมาก	Neutrally

รูปภาพที่ 3 แสดงจำนวนข้อความ

Epoch 10/10				
Validation Accuracy: 0.7089				
	precision	recall	f1-score	support
0	0.70	0.66	0.68	150
1	0.66	0.65	0.65	150
2	0.76	0.82	0.79	150
accuracy			0.71	450
macro avg			0.71	450
weighted avg			0.71	450

รูปภาพที่ 6 แสดงจำนวนและร้อยละของข้อความตามที่จำแนกเอาไว้ หลังจากนำไปฝึกกับโมเดล

4.ฝึกด้วยโมเดล Feed Forward Neural Network

โมเดล Feed Forward Neural Network ได้ข้อมูลมาจาก Lazada ทั้งหมด 4500 ข้อความ แบ่งเป็น 3 ส่วนดังนี้ Positive,Neutrally และ Negative

ส่วนที่ 1 ความคิดเห็นจากร้านค้าออนไลน์

	User	Message	Sentiment
289	*****765	เสื้อตัวเล็กและผ้าบางมากๆ	Neutrally
819	0***8	บางไปนิดค่ะ	Neutrally
758	ภกมล แ	สั่งสีขาวได้สีส้มแปด ปวดหัว ร้านค้าตอบกลับด...	Negative
696	จันทนา ใ	สินค้าเล็กมากๆใส่ไม่ได้	Negative
986	*****644	เนื้อผ้าไม่โอเคไม่ผ่าน โทรมมาก	Negative
...
24	0***5	ส่งสินค้าเร็วมากๆ รองเท้านุ่ม ใส่ดีสมราคาค่ะ	Positive
554	Thanakorn N.	เนื้องานด้านออกอุโอเค แต่ข้างในไม่เนียน กระเป๋...	Neutrally
528	ร***.	กระเพาะใบใหญ่ดี มีรอยบนนิดหน่อย รอยเหมือนรูปใน...	Neutrally
32	สุชาติ จ.	การเคลื่อนไหวที่เจ็บสงบนและราบรื่น ดีไหมที่ดู...	Positive
1177	WORAPON S.	คุณภาพสมราคา ไม่ได้ดีมาก	Neutrally

รูปภาพที่ 3 แสดงจำนวนข้อความ

```
#ลองทำนายจากชุดทดสอบ
tfidfX_test = tfidfX_test.to(device)
Y_test = Y_test.to(device)

# Forward pass to get output
model.eval()
test_prediction = model.forward(tfidfX_test.float())
test_prediction = torch.squeeze(test_prediction)

#Calculate accuracy on test set
test_accuracy = calculate_accuracy(Y_test,test_prediction)

print("Test Accuracy:",round(test_accuracy.item(),4), "\n")

Test Accuracy: 0.7044
```

รูปภาพที่ 7 แสดงจำนวนและร้อยละของข้อความตามที่จำแนกเอาไว้ หลังจากนำไปฝึกกับโมเดล

สรุปผล อภิปรายและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์การวิจัย คือ เพื่อสำรวจความพึงพอใจของบทวิจารณ์สินค้า และนำมาปรับปรุงสินค้าต่อไป วิเคราะห์ความคิดเห็นบนร้านค้าออนไลน์ และจำแนกความคิดเห็นเป็นประเภทต่าง ๆ และศึกษาความแม่นยำของโมเดลในการทำนายความคิดเห็นและอารมณ์จากข้อความในบทนี้เป็นการนำเสนอ การสรุปกระบวนการวิจัยโดยรวมผลการวิจัยโดยสรุป การอภิปรายผลการวิจัย ข้อจำกัดของระบบ ปัญหาอุปสรรค และแนวทางการแก้ไข และข้อเสนอแนะ ซึ่งมีรายละเอียดดังต่อไปนี้

สรุปผล

ข้อมูลที่ใช้ในการศึกษาในครั้งนี้ คือ ความคิดเห็นจากบทวิจารณ์สินค้าออนไลน์ของ Lazada ซึ่งวิธีการเก็บคือ Positive จะเก็บที่ 4 – 5 ดาว ส่วน Neutrally จะเก็บที่ 3 ดาว และ Negative จะเก็บที่ 1 – 2 ดาว ทำให้ได้ข้อมูลความคิดเห็นภาษาไทยทั้งหมด 4,500 ความคิดเห็นแบ่งออกเป็น 3 ส่วนเป็น Positive, Neutrally และ Negative โดยแต่ละส่วนจะมีความคิดเห็นที่ใช้ในการฝึกทั้งหมด 1,500 ความคิดเห็น หลังจากนั้นจะผ่านกระบวนการ การคัดกรองบทวิจารณ์สินค้า กำหนดค่าตัวเลขให้กับคำ เพื่อให้ได้ข้อมูลที่ใช้สำหรับการฝึกโมเดล เทคนิคที่นำมาใช้ในวิเคราะห์ความคิดและความรู้สึกประกอบไปด้วย 4 โมเดลได้แก่ Navie Bayes, Support Vector Machine, Feed Forward Neural Network, และ BERT หลังการฝึกและทดสอบโมเดลพบว่า โมเดลของ Feed Forward Neural Network ได้ผลลัพธ์การทำนายความคิดเห็นภาษาไทยอยู่ที่ร้อยละ 0.695 และเป็นวิธีที่ดีที่สุดสำหรับการวิเคราะห์ความรู้สึกและความคิดเห็น ทั้งนี้เมื่อพิจารณาผลลัพธ์การทำนายความคิดเห็นภาษาไทยแต่ละ โมเดลจากมากไปน้อยพบว่า Feed Forward Neural Network ได้ร้อยละเท่ากับ 0.69, Support Vector Machine ได้ร้อยละเท่ากับ 0.691, Navie Bayes ได้ร้อยละเท่ากับ 0.64, Bert ได้ร้อยละเท่ากับ 0.60 ตามลำดับ

อภิปรายผล

ผู้วิจัยจะนำเสนอการอภิปรายผลการวิจัยตามวัตถุประสงค์ของการวิจัย

1.เพื่อสำรวจความพึงพอใจของบทวิจารณ์สินค้า และนำมาปรับปรุงสินค้าต่อไป

งานวิจัยนี้พบว่า การสำรวจความพึงพอใจของบทวิจารณ์สินค้า บริษัทหรือธุรกิจสามารถเข้าใจได้ว่าสินค้าของตนได้รับความนิยมอย่างไรและมีปัญหาหรือข้อเสียอะไรที่ควรปรับปรุง เพื่อเพิ่มมูลค่าทางการตลาดได้ดียิ่งขึ้น

2.เพื่อวิเคราะห์ความคิดเห็นบนร้านค้าออนไลน์ และจำแนกความคิดเห็นเป็นประเภทต่าง ๆ

งานวิจัยนี้พบว่า การจำแนกความคิดเห็นเป็นประเภทต่างๆ ช่วยให้ร้านค้าออนไลน์สามารถเข้าใจความคิดเห็นของลูกค้าต่อสินค้าหรือบริการของพวกเขาได้ดียิ่งขึ้น นอกจากนี้ยังช่วยในการจำแนกประเภทของความคิดเห็น เช่น ดี ปานกลาง แย่

3.เพื่อศึกษาความแม่นยำของโมเดลในการทำนายความคิดเห็นและอารมณ์จากข้อความ

งานวิจัยนี้พบว่า เพื่อศึกษาความแม่นยำของโมเดลจะช่วยให้เรารู้ว่าโมเดลที่ใช้ในการวิเคราะห์อารมณ์จากความคิดเห็น นั้นมีประสิทธิภาพอย่างไรในการทำนายความคิดเห็นและอารมณ์จากข้อความ ซึ่งสามารถนำไปพัฒนาและปรับปรุงโมเดลเพื่อให้มีประสิทธิภาพมากยิ่งขึ้นในอนาคต

ข้อจำกัดของระบบ

- 1.จำแนกความคิดเห็นของ Positive, Neutrally หรือ Negative ได้ทั้งหมด 3 ประเภท
- 2.สามารถจำแนกความคิดเห็นภาษาไทยเท่านั้น

ปัญหาอุปสรรค และแนวทางการแก้ไข

ปัญหาอุปสรรค

- 1.ข้อมูลมีไม่มากพอสำหรับการฝึกโมเดลเพื่อให้ได้ประสิทธิภาพเพิ่มขึ้น
- 2.การตัดคำที่ยังไม่สมบูรณ์ทำให้เกิดข้อผิดพลาดในการฝึกโมเดลได้
- 3.การคัดกรองคำส่วนของ Positive, Neutrally หรือ Negative ให้มีความละเอียด

แนวทางการแก้ไข

- 1.ทำการเก็บรวบรวมข้อมูลจากแหล่งต่างๆ ที่มีคนสร้างเอาไว้และนำมารวมกับที่ตัวเองเก็บเอาไว้
- 2.เปลี่ยนวิธีการตัดคำใหม่ เพื่อให้ได้ความหมายที่ชัดเจน
- 3.ทำการจำแนกประเภท Positive, Neutrally หรือ Negative เป็นส่วนย่อยๆ เพื่อให้มีความละเอียดในส่วนของการจำแนกประเภท

ข้อเสนอแนะ

- 1.เพิ่มความแม่นยำของโมเดลให้มากขึ้น
- 2.การทำความสะอาดข้อมูลให้ดียิ่งขึ้น

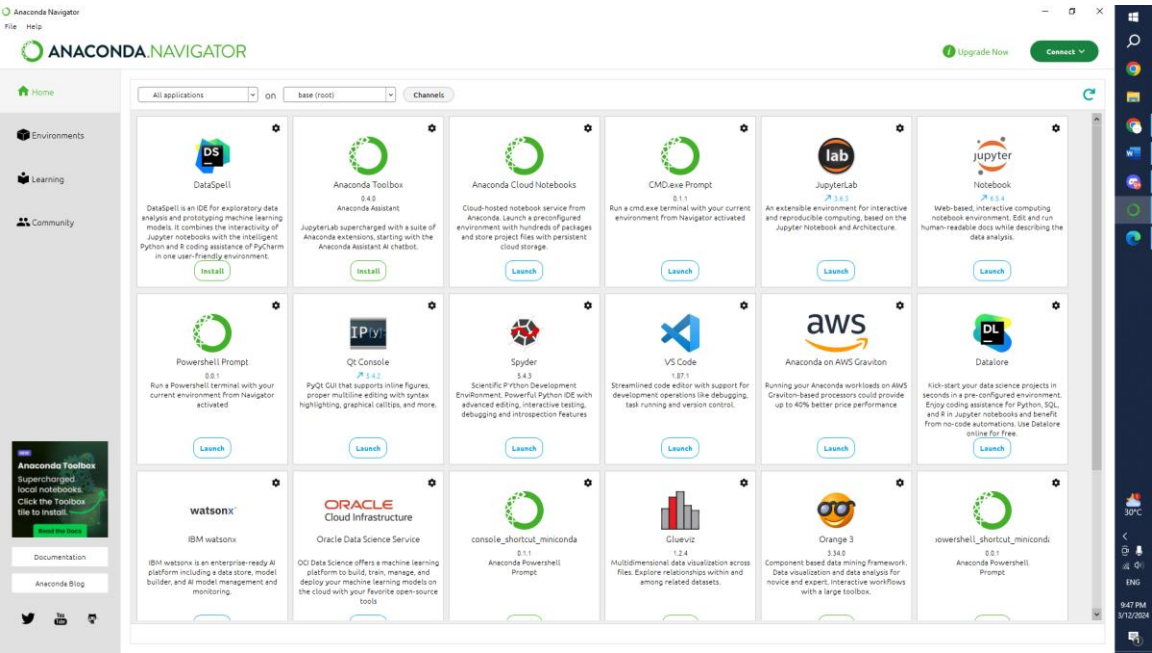
บรรณานุกรม

1. Adisorn Kaewchansilp. (2023). รู้จัก BERT AI อัลกอริทึมใหม่ของกูเกิ้ล, สืบค้นเมื่อ 27 พฤศจิกายน 2566. จาก <https://thewhitemarketing.com/bert-algorithm/>
2. Davoy. (2023). Natural Language Processing คืออะไร?, สืบค้นเมื่อ 27 พฤศจิกายน 2566. จาก <https://davoy.tech/th/natural-language-processing-คืออะไร/>
3. Cloud Ace. (2019). AI, Machine Learning (ML) คืออะไร? ทำความรู้จักกับบริการ AI และ ML จาก Google Cloud, สืบค้นเมื่อ 27 พฤศจิกายน 2566 จาก <https://cloud-ace.co.th/blogs/o0v9a6-ai-machine-learning-ml-ai-ml-goog>
4. K. Purwandari, N. H. Jeremy, P. Arisaputra, S. Philip และ D. Suhartono (2022). Twitter-based Sentiment Analysis for Indonesian Drug Products using Supervised Feature Engineering. 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI).(pp.221-226). Retrieved October 7,2023 from <https://ieeexplore.ieee.org/document/10053058>
5. nessesence. (2018). Deep Learning คืออะไร?, สืบค้นเมื่อ 3 มีนาคม 2567. จาก [https://www.thaiprogrammer.org/2018/12/deep-learning-คืออะไร/#:~:text=Feed-forward%20neural%20networks%20ถือ,\(loop\)%20เกิดขึ้นเลย%20](https://www.thaiprogrammer.org/2018/12/deep-learning-คืออะไร/#:~:text=Feed-forward%20neural%20networks%20ถือ,(loop)%20เกิดขึ้นเลย%20)
6. Salem, M. A. M., & Maghari, A. Y. A. (2020). Sentiment Analysis of Mobile Phone Products Reviews Using Classification Algorithms. 2020 International Conference on Promising Electronic Technologies (ICPET). (pp.84-88). Retrieved October 7,2023 from <https://ieeexplore.ieee.org/document/9319284>
7. Welgamage, V. R., Senarathne, U. A. C., Madhubhashani, N. H. A. C., Liyanage, T. C., & Asanka, P. P. G. D. (2022). Overall and Feature Level Sentiment Analysis of Amazon Product Reviews Using Machine Learning Techniques and Web-Based Chrome Plugin. 2022 International Research Conference on Smart Computing and Systems Engineering (SCSE). (pp.205-210). Retrieved October 7,2023 from <https://ieeexplore.ieee.org/document/9905125>
8. Y. Xu และ Y. Ren (2021). Y. Xu และ Y. Ren (2021). Research on Sentiment Analysis Model of Online Shopping Product Evaluation Based on Machine Learning. 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). (pp. 482-486). Retrieved October 7,2023 from <https://ieeexplore.ieee.org/document/9498066>
9. R. Man and K. Lin. (2021). Sentiment Analysis Algorithm Based on BERT and Convolutional Neural Network. 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2021, pp. 769-772, doi: 10.1109/IPEC51340.2021.9421110. Retrieved March 3,2024 from <https://ieeexplore.ieee.org/document/9421110>
10. S. A. Rahin, T. Hasib and M. Hassan. (2022). Aspect-Based Sentiment Analysis Using SemEval and Amazon Datasets. 2022 Fifth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU), Riyadh, Saudi Arabia, 2022, pp. 85-90, doi: 10.1109/WiDS-PSU54548.2022.00029. Retrieved March 3,2024 from <https://ieeexplore.ieee.org/document/9764870>

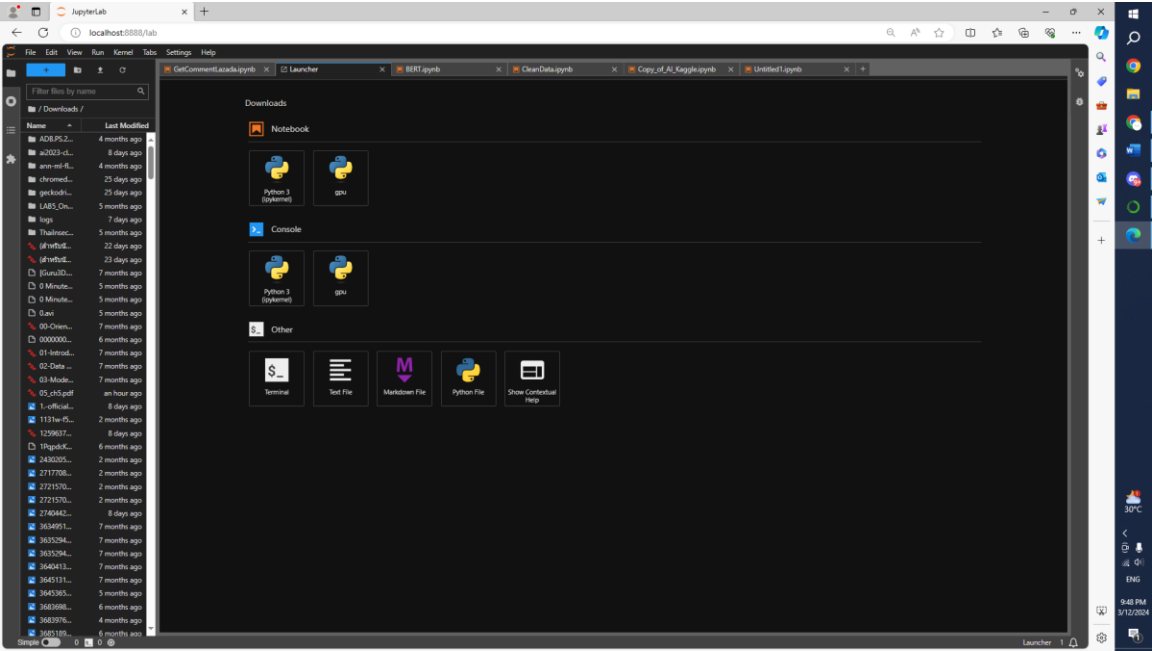
ภาคผนวก

เครื่องมือ วิธีการเก็บข้อมูล และผลลัพธ์การทดสอบ

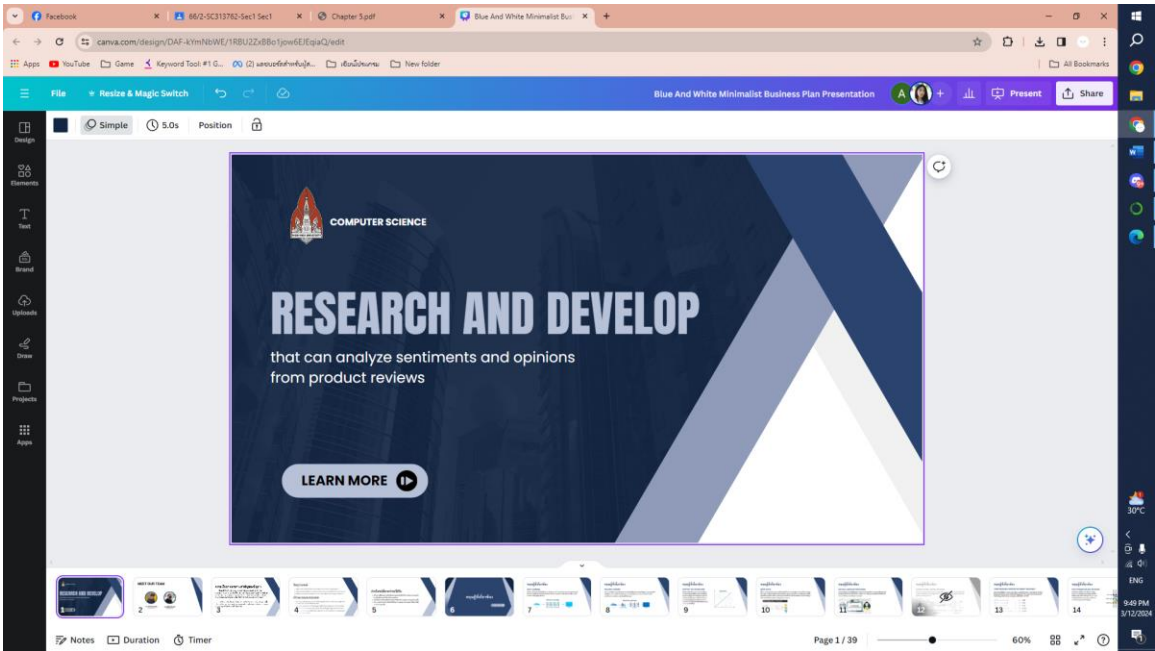
เครื่องมือ



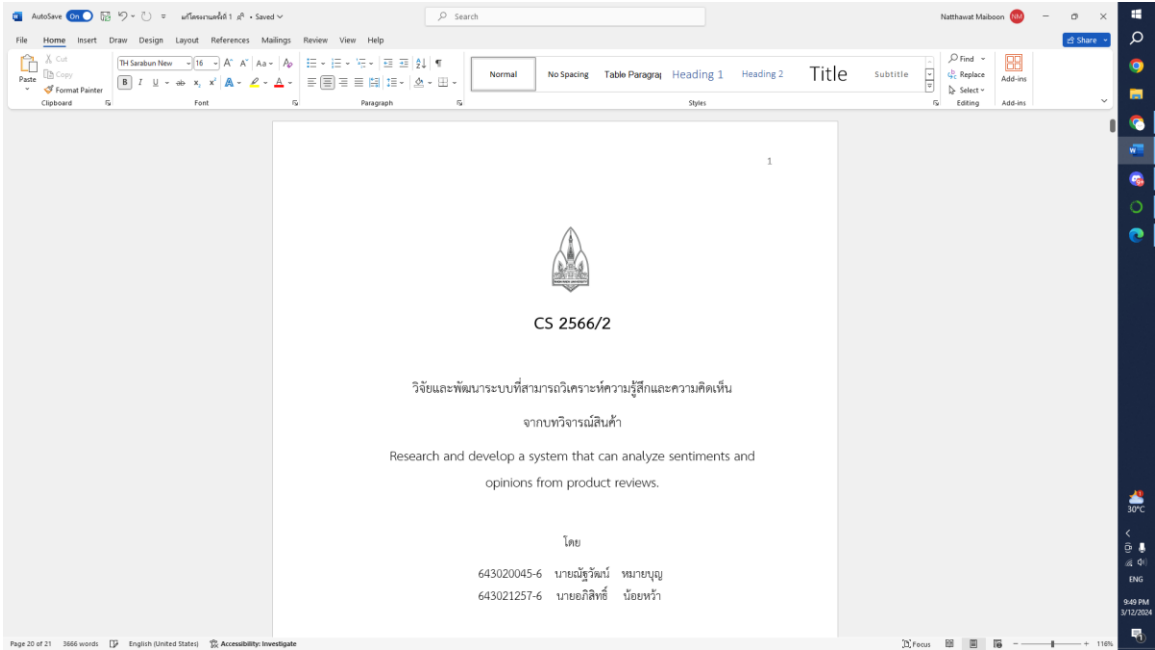
รูปภาพที่ 8 Anaconda Navigator



รูปภาพที่ 9 JupyterLab

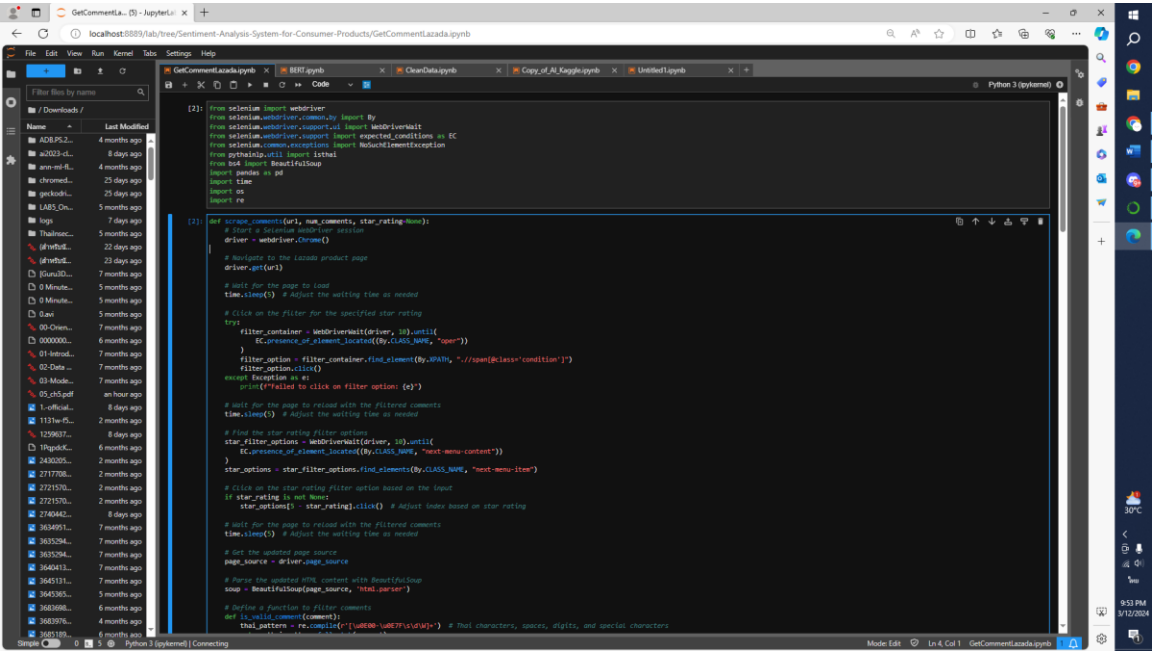


รูปภาพที่ 10 Canva

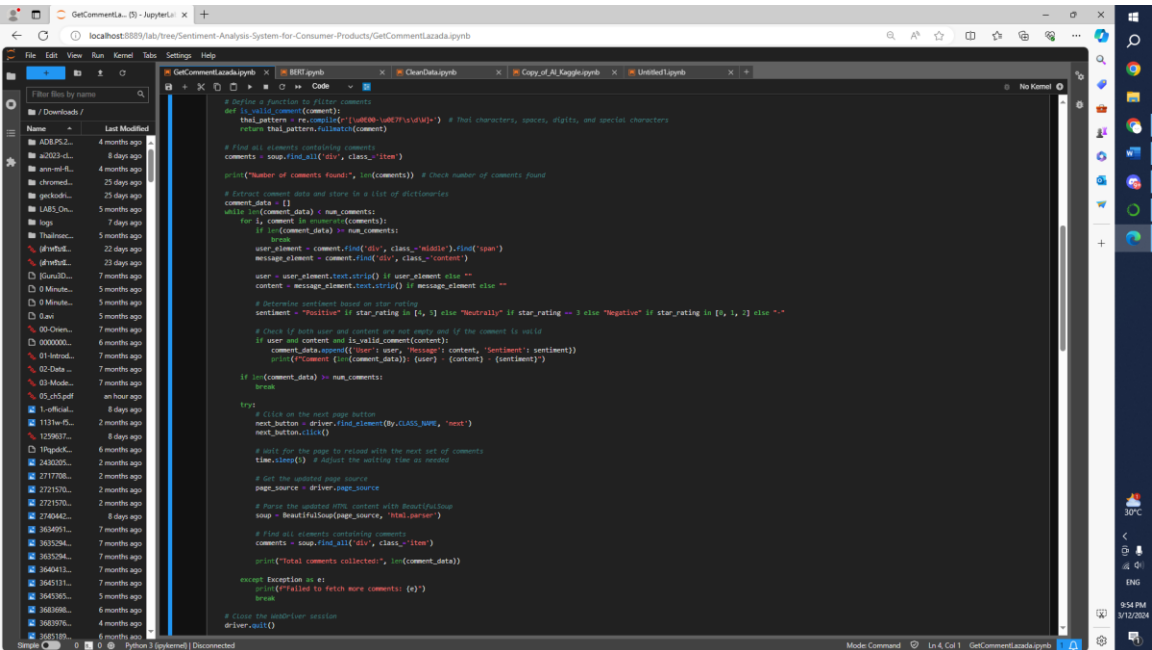


รูปภาพที่ 11 Microsoft Word

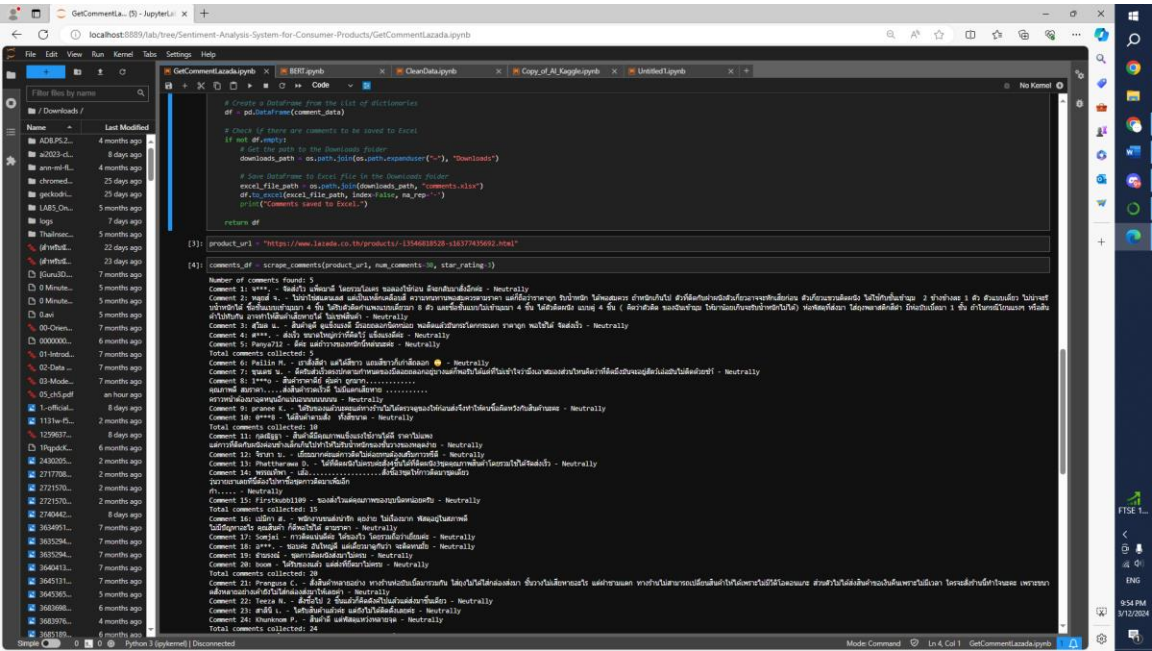
วิธีการเก็บข้อมูล



รูปภาพที่ 12 เก็บ Comment จาก Lazada

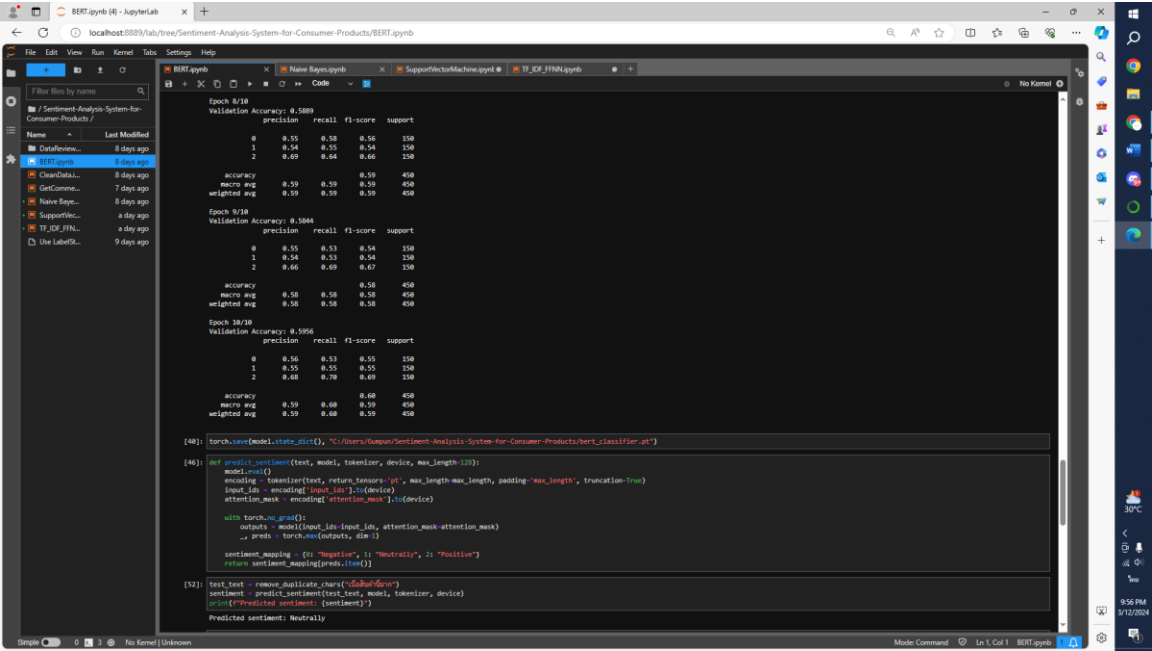


รูปภาพที่ 13 เก็บ Comment จาก Lazada

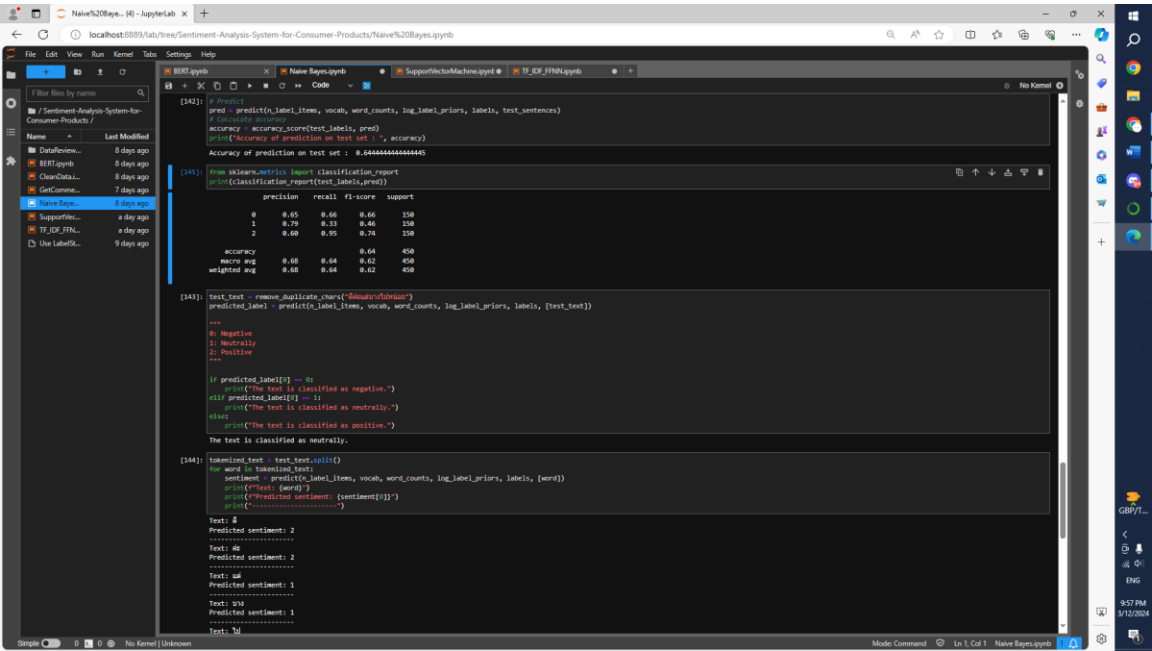


รูปภาพที่ 14 เก็บ Comment จาก Lazada

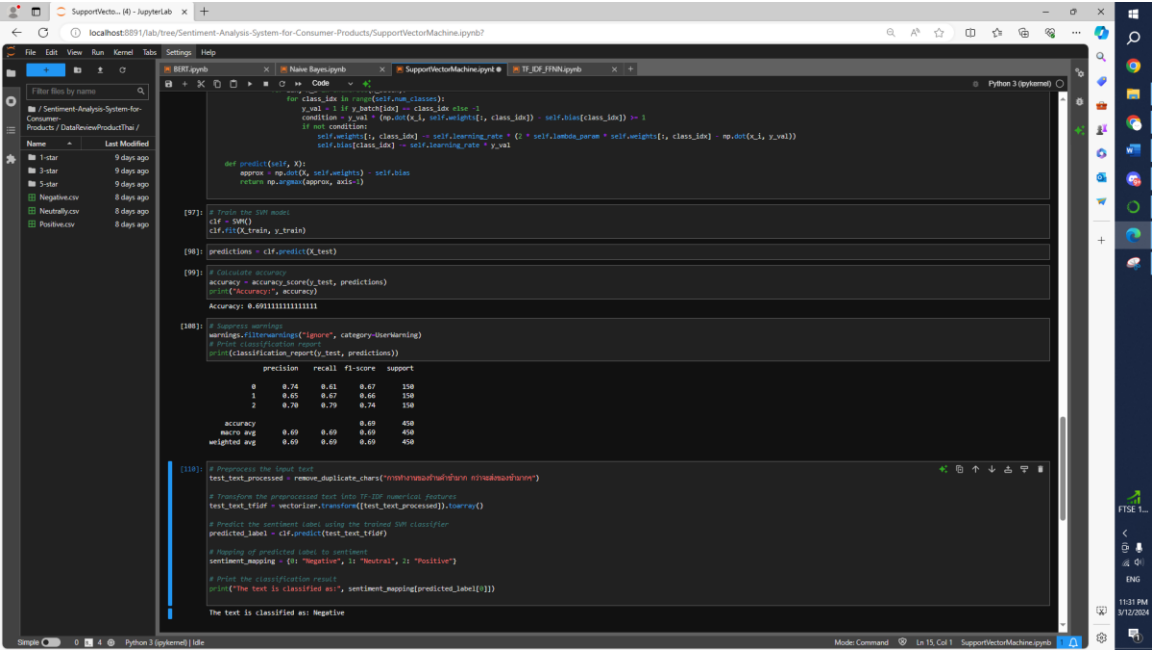
ผลลัพธ์การทดสอบ



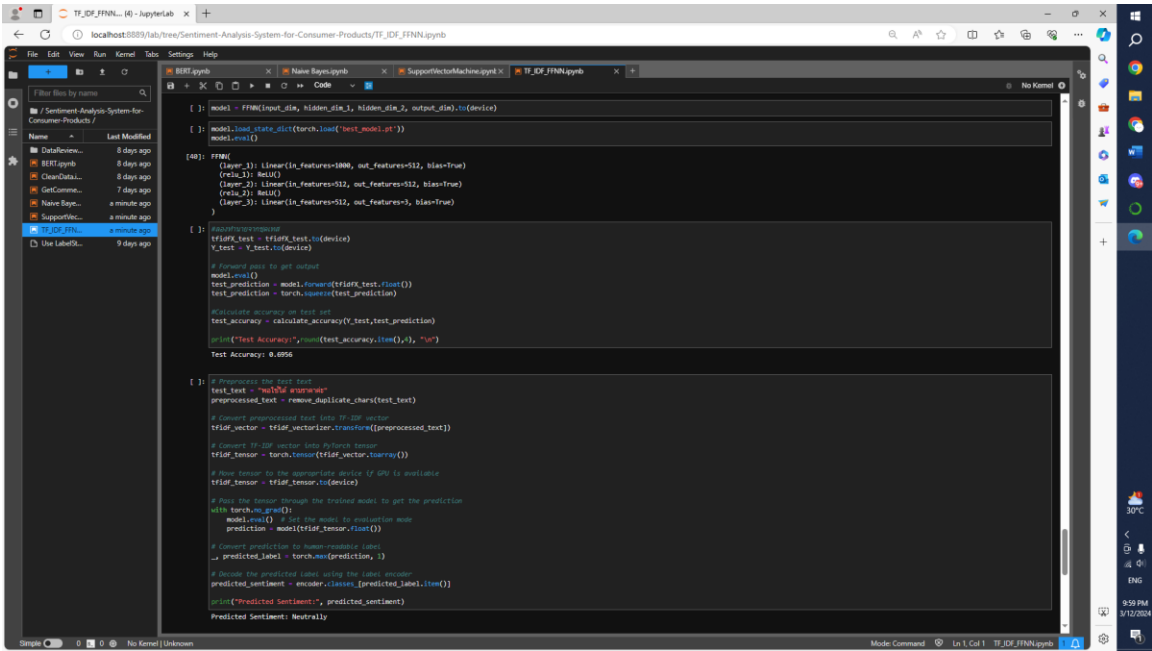
รูปภาพที่ 15 ผลลัพธ์การทดสอบของ Bert



รูปภาพที่ 16 ผลลัพธ์การทดสอบของ Naïve Bayes



รูปภาพที่ 17 ผลลัพธ์การทดสอบของ Support Vector Machine



รูปภาพที่ 18 ผลลัพธ์การทดสอบของ Feed Forward Neural Network

ประวัติย่อของผู้ทำวิจัย

ชื่อ สกุล	นายณัฐวัฒน์ หมายบุญ
วัน เดือน ปีเกิด	1 มิถุนายน พ.ศ. 2546
สถานที่อยู่ปัจจุบัน	62 ซ.4-5 ถ.โคลัมโบ ต.ศิลา อ.เมืองขอนแก่น ขอนแก่น 40000
ตำแหน่ง	นักศึกษา
สถานที่ทำงาน	มหาวิทยาลัยขอนแก่น อาคารวิทยวิภาส (SC09)
ประวัติการศึกษา	
	พ.ศ.2564 หลักสูตรวิทยาศาสตรบัณฑิต
	สาขาวิชาวิทยาการคอมพิวเตอร์

ชื่อ สกุล	นายอภิสิทธิ์ น้อยหว่า
วัน เดือน ปีเกิด	12 กรกฎาคม พ.ศ. 2545
สถานที่อยู่ปัจจุบัน	123 หมู่ 16 ต.ในเมือง อ.เมือง จ.ขอนแก่น ขอนแก่น 40000
ตำแหน่ง	นักศึกษา
สถานที่ทำงาน	มหาวิทยาลัยขอนแก่น อาคารวิทยวิภาส (SC09)
ประวัติการศึกษา	พ.ศ.2564 หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์