

ระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดีย
ด้วยซัพพอร์ตเวกเตอร์แมชชีน
Sentiment Analysis System of Thai Video on Social Media
Using Support Vector Machine

ปรีตาวรรณ เกษเมธีการุณ^{1*}
พยุ่ง มีสัจ²
สุมิตรา นวลมีศรี³

¹นักศึกษา สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
E-mail: preedawan.k@email.kmutnb.ac.th

²อาจารย์ สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
E-mail: phayung.m@it.kmutnb.ac.th

³อาจารย์ สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยราชภัฏสวนสุนันทา
E-mail: sumitra.nu@ssru.ac.th

บทคัดย่อ : การวิจัยนี้ได้พัฒนาระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซัพพอร์ตเวกเตอร์แมชชีนบนพื้นฐานการวิเคราะห์ความรู้สึกของผู้บริโภคจากวิดีโอบนโซเชียลมีเดียเพื่อเป็นแนวทางในการตัดสินใจในการซื้อของผู้บริโภคและการจัดการข้อมูลของภาคธุรกิจต่อวิดีโอบนสื่อสังคมออนไลน์ทั้งเชิงบวกและเชิงลบโดยใช้เทคนิคการเรียนรู้ของเครื่อง งานวิจัยได้ทำการสกัดข้อความจากวิดีโอ จากนั้นข้อมูลจะผ่านกระบวนการตัดคำ และส่งไปเข้าสู่กระบวนการสกัดคุณลักษณะโดย Term Frequency-Inverse Document Frequency (TF-IDF) และเข้าสู่ขั้นตอนการจำแนกประเภทด้วยซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) จากการทดลองพบว่าระบบมีประสิทธิภาพดีทั้งด้านความถูกต้องอยู่ที่ร้อยละ 98 เมื่อเทียบกับผู้เชี่ยวชาญจำนวน 15 คน และทำการประเมินประสิทธิภาพด้านความพึงพอใจของผู้ใช้งานจำนวน 60 คน ต่อระบบ มีค่าเฉลี่ยเท่ากับ 4.79 ส่วนเบี่ยงเบนมาตรฐานที่ 0.41 กล่าวได้ว่าระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซัพพอร์ตเวกเตอร์แมชชีนที่พัฒนาขึ้นสามารถวิเคราะห์ความรู้สึกจากวิดีโอภาษาไทยบนโซเชียลมีเดียได้อย่างมีประสิทธิภาพ

คำสำคัญ : การวิเคราะห์ความรู้สึก สื่อสังคมออนไลน์ ซัพพอร์ตเวกเตอร์แมชชีน

ABSTRACT : This research developed a system for analyzing Thai videos published on social media based on sentimental analysis method. It deployed a technique called support vector machine to provide guideline for customers to make decision on their goods purchase. It also provides guideline for business enterprise to manage enterprise information regarding videos on social media both positive and negative perspectives based on machine learning technique. First, this system extracts a dialog from video and featured it to text by using a technique called Term Frequency-Inversed Document Frequency (TF-IDF). Then, each text will processed by using a technique called Support Vector Machine (SVM) to attain the result. Experimental survey shows that this system achieved 98% of accuracy compared with 15 experts. It gained 4.79 of satisfaction with 0.41 of Standard Deviation from 60 users. Finally, It is apparently shown that this system: Sentiment Analysis System of Thai Video on Social Media Using Support Vector Machine is capable of analyzing sentiment from Thai videos at very high efficiency.

Keywords : Sentiment Analysis, Social Networks, Support Vector Machine

1. บทนำ

สื่อสังคมออนไลน์เป็นช่องทางการสื่อสารที่นิยมใช้กันอย่างแพร่หลายในการแลกเปลี่ยนข้อมูล ความคิดเห็น ทศนคติต่างๆ ผู้ใช้สามารถแบ่งปันข้อมูล เช่น ข้อความ รูปภาพ เสียง และวิดีโอ เป็นต้น ลูกค้าจะมีการแบ่งปันประสบการณ์การใช้ผลิตภัณฑ์และบริการต่างๆ ได้ทันทีผ่านสื่อสังคมออนไลน์ที่ทำหน้าที่เป็นเสมือนตัวกลางเผยแพร่ข้อมูลออกไปด้วยอุปกรณ์ที่มีอยู่ เช่น สมาร์ทโฟน แท็บเล็ต และคอมพิวเตอร์ ไปยังคนอื่นๆ ได้อย่างรวดเร็วผ่านสื่อสังคมออนไลน์ที่ได้รับความนิยมได้แก่ Facebook, Twitter และ YouTube เป็นต้น [1] การแสดงความคิดเห็นหรือที่นิยมเรียกว่าการรีวิวสินค้ามักจะนำเสนอเป็นวิดีโอผ่านสื่อสังคมออนไลน์ YouTube โดยเฉพาะสินค้าด้านความงามที่มีการแข่งขัน

ทางการตลาดที่สูงมาก วิดีโอที่ลูกค้าแบ่งปันจากประสบการณ์การใช้ถือเป็นข้อมูลที่มีความสำคัญต่อผู้บริโภคคนอื่นๆ และต่อภาคธุรกิจเป็นอย่างมาก ส่วนผู้บริโภคคนอื่นๆ สามารถทราบข้อมูลผลิตภัณฑ์ รูปร่าง และความคิดเห็นของลูกค้าที่เคยใช้ผลิตภัณฑ์นั้นมาก่อนผ่านวิดีโอเพื่อประกอบการตัดสินใจในการซื้อผลิตภัณฑ์หรือบริการนั้นๆ ซึ่งถือเป็นข้อมูลที่มีผลต่อการตัดสินใจซื้อได้เป็นอย่างดี และส่วนภาคธุรกิจหากธุรกิจสามารถรวบรวมและทราบข้อมูลความคิดเห็นของลูกค้าที่มีต่อผลิตภัณฑ์และบริการได้ย่อมทราบข้อมูลในการปรับปรุงสินค้า การคาดการณ์ยอดขาย และปิดช่องโหว่ที่อาจส่งความเสียหายในความคิดเห็นเชิงลบของลูกค้าที่จะแบ่งปันด้วยวิดีโอผ่านสื่อสังคมออนไลน์ได้

การวิเคราะห์ความรู้สึก (Sentiment Analysis) เป็นการนำการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) มาประยุกต์ใช้เพื่อมุ่งเน้นการวิเคราะห์และตรวจสอบความรู้สึก (Opinion) ได้อย่างอัตโนมัติ ซึ่งในปัจจุบันได้มีงานวิจัยที่มีการวิเคราะห์ความรู้สึกจากข้อความ เช่น ความคิดเห็น สัญลักษณ์ทางอารมณ์ (Emoticon) จากเสียง และภาพจากวิดีโอออนไลน์ในภาษาอังกฤษและภาษาสเปน การวิเคราะห์ความรู้สึกในงานวิจัยภาษาไทยพบว่าส่วนใหญ่เป็นการนำเฉพาะข้อความ (Text) ที่มีการเขียนหรือโพสต์เอาไว้มาใช้วิเคราะห์ความรู้สึกของผู้บริโภคในหัวข้อเรื่องที่น่าสนใจ อาทิ สินค้า บริการ และการเมือง [2]

จากการศึกษาการวิเคราะห์ความรู้สึกบนสื่อสังคมออนไลน์ที่ใช้งานในปัจจุบันพบว่า มีข้อจำกัดในเรื่องภาษาไทยและแหล่งข้อมูลที่ยังไม่มีการวิเคราะห์จากวิดีโอบนสื่อสังคมออนไลน์ที่เป็นภาษาไทย ซึ่งความคิดเห็นเชิงบวกและเชิงลบที่มีผลโดยตรงต่อการตัดสินใจซื้อสินค้าของผู้บริโภคและต่อภาคธุรกิจ ดังนั้นงานวิจัยนี้จะมุ่งเน้นพัฒนาระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนสื่อสังคมออนไลน์ YouTube ภาษาไทย เพื่อวิเคราะห์ความคิดเห็นของลูกค้าที่มีผลต่อผลิตภัณฑ์เครื่องสำอาง ด้วยกระบวนการสกัดข้อความจากวิดีโอบนสื่อสังคมออนไลน์ ตัดคำ สกัดคำ สกัดคุณลักษณะ โดยแต่ละความคิดเห็นจะถูกจำแนกเป็น 2 ประเภท ได้แก่ เชิงบวกและเชิงลบ ข้อมูลที่ผ่านกระบวนการแล้วจะพร้อมเข้าสู่การจำแนกด้วยเทคนิค SVM เพื่อการจำแนกการวิเคราะห์ความรู้สึกบนสื่อสังคมออนไลน์ให้มีประสิทธิภาพดีที่สุด

2. งานวิจัยที่เกี่ยวข้อง

การวิเคราะห์ความรู้สึกบนสื่อสังคมออนไลน์ส่วนใหญ่จะอาศัยวิธีการเรียนรู้ของเครื่อง (Machine Learning) ด้วยการจำแนกประเภท เช่น Naïve Bayes, Decision Tree, Maximum Entropy และ Support Vector Machine (SVM) เป็นต้น งานวิจัยของ Lima และ Castro (2013) ได้นำ Naïve Bayes มาพัฒนาระบบการวิเคราะห์ความรู้สึกจากข้อความ Twitter ต่อรายการทีวีทางช่องบราซิลเลียน มีค่าความถูกต้องที่ 91.94% [3] Isah, Neagu และ Trundle (2014) นำ Naïve Bayes มาใช้ในจำแนกข้อมูลจาก Facebook และ Twitter เกี่ยวกับการวิเคราะห์ผลิตภัณฑ์ประเภทยาและเครื่องสำอาง ด้วยการจำแนกเป็น 2 ประเภท ได้แก่ Positive และ Negative ผลการวิจัยระบุว่า Naïve Bayes ให้ค่าความถูกต้องที่ 83% เมื่อเปรียบเทียบกับวิธีการจำแนกโดยผู้เชี่ยวชาญ [4] Preedawon และ Sumitra (2015) ได้นำเสนอกระบวนการวิเคราะห์ความรู้สึกด้วยเทคนิค Naïve Bayes โดยใช้ข้อมูลเสียงจากวิดีโอที่นำมาจาก YouTube ผลการทดสอบให้ค่าความแม่นยำที่ดีที่สุดที่ระดับ 87.70% [5]

การวิเคราะห์ความรู้สึกด้วยเทคนิค SVM งานวิจัยของ Naradhipa และ Purwarianti (2011) ได้ทำการจำแนกความรู้สึกบนสื่อสังคมออนไลน์ในภาษาอินโดนีเซีย โดยแบ่งออกเป็น 4 กลุ่ม ได้แก่ Positive, Negative, Neutral และ Question ซึ่งผลวิจัยสนับสนุน SVM มีค่าความถูกต้อง 83.5% [6] Binali, Wu และ Potdar (2009) นำ SVM มาทำการจำแนกอารมณ์พื้นฐาน ผลการวิจัยมีค่าความถูกต้อง 96.43% [7] Rosas, Mihalcea และ Morency

(2013) ได้จำแนกความรู้สึกจากภาษาสเปน จากวิดีโอที่แสดงความคิดเห็นในหัวข้อทั่ว ๆ ไปจาก YouTube จากข้อความ เสียง และภาพแสดงท่าทางร่วมกัน ด้วย SVM พบว่าให้ค่าความแม่นยำที่น่าพอใจที่ระดับ 75% [8]

งานวิจัยได้ทำการเปรียบเทียบประสิทธิภาพในการวิเคราะห์ความรู้สึกด้วยการจำแนกระหว่าง Naïve Bayes, Decision Tree, Maximum Entropy และ Support Vector Machine (SVM) เป็นต้น Pongthanoo และคณะ (2012) ได้นำเสนอโมเดลการวิเคราะห์ความพึงพอใจของลูกค้าจากข้อความคำแนะนำโดยการทำให้เหมือนความคิดเห็นและเปรียบเทียบผลการสร้างโมเดล Decision Tree และ Naïve Bayes พบว่าเทคนิคต้นไม้ตัดสินใจให้ผลความถูกต้องดีกว่าที่ค่า 95.50% [9] Akaich, Dhouioui และ Perez (2013) ทำการเปรียบเทียบการจำแนก Naïve Bayes และ SVM ความรู้สึกจากภาษาตุนิเซียบน Facebook ในประเทศตุนิเซีย ทดสอบโดยด้วยชุดข้อมูลเดียวกัน ซึ่งมีการแบ่งข้อมูลฝึกหัดไว้ที่ 60% ส่วนอีก 40% เป็นชุดข้อมูลทดสอบ ผลการวิจัยพบว่า SVM สามารถให้ประสิทธิภาพความถูกต้องดีกว่า Naïve Bayes [10] Kaushik, Sangwan และ Hansen (2013) ได้จำแนกความรู้สึกโดยสกัดเสียงจากวิดีโอที่นำมาจาก YouTube ด้วยภาษาอังกฤษและสร้างแบบจำลองความรู้สึกและระบบรู้จำเสียงอัตโนมัติ โดยใช้ข้อมูลที่แสดงความคิดเห็นในด้านต่างๆ เช่น วิจารณ์สินค้า ภาพยนตร์ และประเด็นทางสังคม เป็นต้น จากผู้แสดงความคิดเห็นทั้งชายและหญิง โดยใช้เทคนิค Maximum Entropy สร้างแบบจำลองความรู้สึกและทดสอบระบบด้วยเทคนิคออฟเบส

(Naïve Bayes : NB), ซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine : SVM) และ Maximum Entropy จากนั้นเปรียบเทียบค่าความแม่นยำระหว่างเทคนิคที่ใช้สร้างแบบจำลองความรู้สึกกับเทคนิคมาตรฐาน โดยเทคนิคที่ใช้สร้างแบบจำลองให้ค่าความแม่นยำที่ดีสูงกว่า 80% [11] อติเทพ และรัฐสิทธิ์ (2013) พัฒนาระบบการประมาณอารมณ์จากความคิดเห็นภาษาไทยโดยรวบรวมความคิดเห็นจากเว็บไซต์บริการข่าว บริการवाईต์แบงก์ และบริการวิจารณ์สินค้าจำนวน 6,000 ความคิดเห็น โดยมีผู้เชี่ยวชาญจำแนกความคิดเห็นตามกลุ่ม อารมณ์ 6 อารมณ์ ได้แก่ กลุ่ม Love, Joy, Surprise, Sadness, Fear และ Anger กลุ่มละ 1,000 ความคิดเห็น ทำการเปรียบเทียบความถูกต้องจาก 3 เทคนิค Naïve Bayes, SVM และ Decision tree ผลการวิจัยพบว่า SVM ให้ค่าความถูกต้องสูงสุดที่ 69.15 [12] Brynielsson และคณะ (2014) ได้การจัดกลุ่มอารมณ์ของผู้ใช้ Twitter เกี่ยวกับ Hashtag พายุเฮอริเคนแซนดี้ (Sandy Hurricane) จำแนกเป็น 4 ประเภท ได้แก่ Positive, Anger, Fear และอื่นๆ พบว่า SVM มีค่าถูกต้องที่ 75.9% ในขณะที่ Naïve Bayes มีค่าความถูกต้องที่ 69.1% [13] Sarakit และคณะ (2015) ทำการจำแนกอารมณ์ในเว็บบอร์ด YouTube ที่เกี่ยวข้องกับการแบ่งปันวิดีโอคลิป โดยข้อมูลประกอบไปด้วยกลุ่ม Thai Music Video (MV) จำนวน 2,771 ชุดข้อมูล และกลุ่ม Thai Commercial Advertisement Video (AD) จำนวน 3,077 ชุดข้อมูลซึ่งจำแนกตามทฤษฎีของ Ekman's Six Basic Emotions ได้แก่ Anger, Disgust, Fear Happiness, Sadness และ Surprise โดยเปรียบเทียบ

ประสิทธิภาพ Multinomial Naïve Bayes (MNB), Decision Tree และ SVM ผลการวิจัยพบว่า การจำแนกความรู้สึกของความคิดเห็นในชุดข้อมูล AD ค่าความถูกต้อง SVM สูงสุดที่ค่า 76.14% และการจำแนกความรู้สึกของความคิดเห็นในชุดข้อมูล MV ค่าความถูกต้อง MNB สูงสุดที่ค่า 84.48% [14]

จากการศึกษางานวิจัยที่เกี่ยวข้อง การวิเคราะห์ความรู้สึกบนสื่อสังคมออนไลน์ ทั้งข้อความและวิดีโอพบว่า การวิเคราะห์ความรู้สึกด้วย SVM ให้ประสิทธิภาพที่ดีที่สุด SVM เป็นกระบวนการสอนเครื่องแบบมีผู้สอน (Supervise Learning) เพื่อให้สามารถสร้างตัวจัดประเภทข้อมูล (Classifier) ให้สามารถจำแนกได้ดีกับตัวอย่างที่ไม่รู้จัก (Unknown Dataset) นำมาใช้เพื่อหาระนาบการตัดสินใจ ในการแบ่งข้อมูลออกเป็นสองส่วน โดยพยายามสร้างเส้นแบ่งตรงกึ่งกลางระหว่างกลุ่มให้มีระยะห่างระหว่างขอบเขตของทั้งสองกลุ่มให้มากที่สุด SVM นำมาใช้สำหรับข้อมูลที่มีมิติของข้อมูลสูงด้วย SVM เป็นข่ายงานประสาทเทียมที่มีหน่วยที่จำลองลักษณะของเซลล์ประสาท เพื่อจำแนกข้อมูลนำเข้า (Input Space) ให้อยู่ในรูปแบบของข้อมูลที่มีมิติสูง (High Dimension Dataset) บนพื้นที่ข้อมูลคุณลักษณะ (Feature Space) โดยใช้ฟังก์ชันในการปรับรูปแบบข้อมูลที่เรียกว่า ฟังก์ชันเคอร์เนล (Kernel Function) ซึ่งสามารถดั่งกล่าวช่วยให้การสร้างตัวจัดประเภทข้อมูลด้วยสมการกำลังสอง (Quadratic Equation) บนพื้นที่ข้อมูลคุณลักษณะเป็นไปได้ง่ายขึ้น และมีความชัดเจนในการจัดประเภทมากยิ่งขึ้นในการแยกแยะประเภทของชุดข้อมูลแต่ละประเภทออกจากกันอย่างชัดเจน ซึ่งเส้นที่เหมาะสม

ดังกล่าวถูกเรียกว่า ระนาบแบ่งเขตข้อมูลที่เหมาะสม (The Optimal Separating Hyper plane)

กำหนดให้ $(x_1, y_1), \dots, (x_n, y_n)$ เป็นตัวอย่างที่ใช้สำหรับการสอน n คือจำนวนข้อมูลตัวอย่าง m คือจำนวนมิติข้อมูลเข้า และ y คือผลลัพธ์ที่มีค่า +1 หรือ -1 [15] ดังสมการที่ 1

$$(x_1, y_1), \dots, (x_n, y_n) \text{ เมื่อ } x \in R^m, y \in \{+1, -1\} \quad (1)$$

เมื่อ i มีค่าตั้งแต่ 1 ถึง n

สำหรับปัญหาเชิงเส้นที่มีมิติของข้อมูลมีขนาดสูงถูกแบ่งเป็น 2 กลุ่ม โดยระนาบตัดสินใจ [15] คำนวณดังสมการที่ 2

$$(w^*x) + b = 0 \quad (2)$$

เมื่อ w คือค่าน้ำหนัก และ b คือ ค่า bias [15] แสดงดังสมการที่ 3 และสมการที่ 4 เพื่อการจำแนกข้อมูล

$$(w^*x) + b > 0 \text{ if } y_i = +1 \quad (3)$$

$$(w^*x) + b < 0 \text{ if } y_i = -1 \quad (4)$$

SVM สามารถแบ่งกลุ่มข้อมูลได้ด้วยเส้นตรงและไม่เป็นเส้นตรง และยังรองรับข้อมูลที่มีคุณลักษณะจำนวนมาก เนื่องจากใช้การแทนข้อมูลแบบเวกเตอร์และพิจารณาเส้นแบ่งกลุ่มข้อมูลจากเวกเตอร์ซัพพอร์ต แต่ต้องมีการทดลองเพื่อปรับค่าพารามิเตอร์ที่เหมาะสม โดยต้องการหาเส้นแบ่งที่สามารถแบ่งข้อมูล 2 กลุ่มออก

จากกันให้ได้ทั้งหมด แต่ด้วยอาจมีข้อมูลที่มีความแตกต่างออกไปจากกลุ่มมาก จึงมีการกำหนดความสามารถในการยอมรับความผิดพลาดของการจัดกลุ่มได้ ด้วยค่าพารามิเตอร์ C (Cost Parameter) หากมีการกำหนดค่าพารามิเตอร์ C ที่มีค่าสูง หมายถึง กำหนดให้ยอมรับความผิดพลาดน้อย สำหรับชุดข้อมูลที่ไม่สามารถแบ่งกลุ่มข้อมูลได้ด้วยเส้นตรง SVM มีฟังก์ชันเคอร์เนล (Kernel Function) ในการแปลงข้อมูลนำเข้าให้เป็นฟีเจอร์สเปซ (Feature Space) เพื่อให้สามารถแบ่งข้อมูลที่ไม่สามารถแบ่งได้ด้วยเส้นตรง เช่น ฟังก์ชัน Polynominal ที่มีพารามิเตอร์ค่าดีกรี (Degree) สำหรับการปรับค่าความโค้งของเส้นแบ่งข้อมูลให้เหมาะสมกับข้อมูล และฟังก์ชัน RBF (Radial Basic) ที่มีการปรับความกว้างด้วยพารามิเตอร์แกมมา (Gamma) ให้เส้นแบ่งเหมาะสมกับข้อมูล

3. วิธีดำเนินการวิจัย

การวิจัยระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียไทยด้วยซอฟต์แวร์แมชชีน มีวิธีดำเนินการวิจัยดังนี้

3.1 การรวบรวมข้อมูล

คณะผู้วิจัยดำเนินการรวบรวมข้อมูลจากทฤษฎี งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความรู้สึกบนสื่อสังคมออนไลน์ และรวบรวมข้อมูลวิดีโอบทวิจารณ์จาก Youtube จำนวน 500 วิดีโอ ซึ่งแต่ละวิดีโอมีความยาวอยู่ระหว่าง 2-5 นาที

3.2 การออกแบบระบบ

การออกแบบระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีน โดยระบบประกอบไปด้วยส่วน

การสกัดวิดีโอเป็นข้อความ ส่วนการตัดคำจากข้อความที่ได้จากการสกัดจากวิดีโอ ส่วนการสกัดคุณลักษณะของคำโดยอาศัยน้ำหนักของคำ นำมาคำนวณค่า Term Frequency (TF) และ Inverse Documents Frequency (IDF) และการวิเคราะห์ความรู้สึกเชิงบวกหรือเชิงลบ โดยทำการเปรียบเทียบจากฐานข้อมูล และส่วนการแสดงผลการวิเคราะห์ความรู้สึกเชิงบวกหรือเชิงลบ ฐานข้อมูลที่เก็บคุณลักษณะของคำว่าเป็นเชิงบวกหรือเชิงลบจะมีการปรับปรุงด้วยวิธีการจำแนกคุณลักษณะด้วย SVM แล้วจะนำค่าผ่านการจำแนกคุณลักษณะไปปรับปรุงและจัดเก็บในฐานข้อมูลเพื่อให้ฐานข้อมูลมีความสามารถในการจำแนกความรู้สึกได้ดียิ่งขึ้น การออกแบบระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีน แสดงดังภาพที่ 1

3.3 การพัฒนาระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียไทยด้วยซอฟต์แวร์แมชชีน

จากขั้นตอนการออกแบบระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีน จะทำการพัฒนาเป็นระบบโดยเขียนด้วยภาษา PHP และ ฐานข้อมูล MySQL โดยในระบบประกอบไปด้วยส่วนต่างๆ ดังนี้

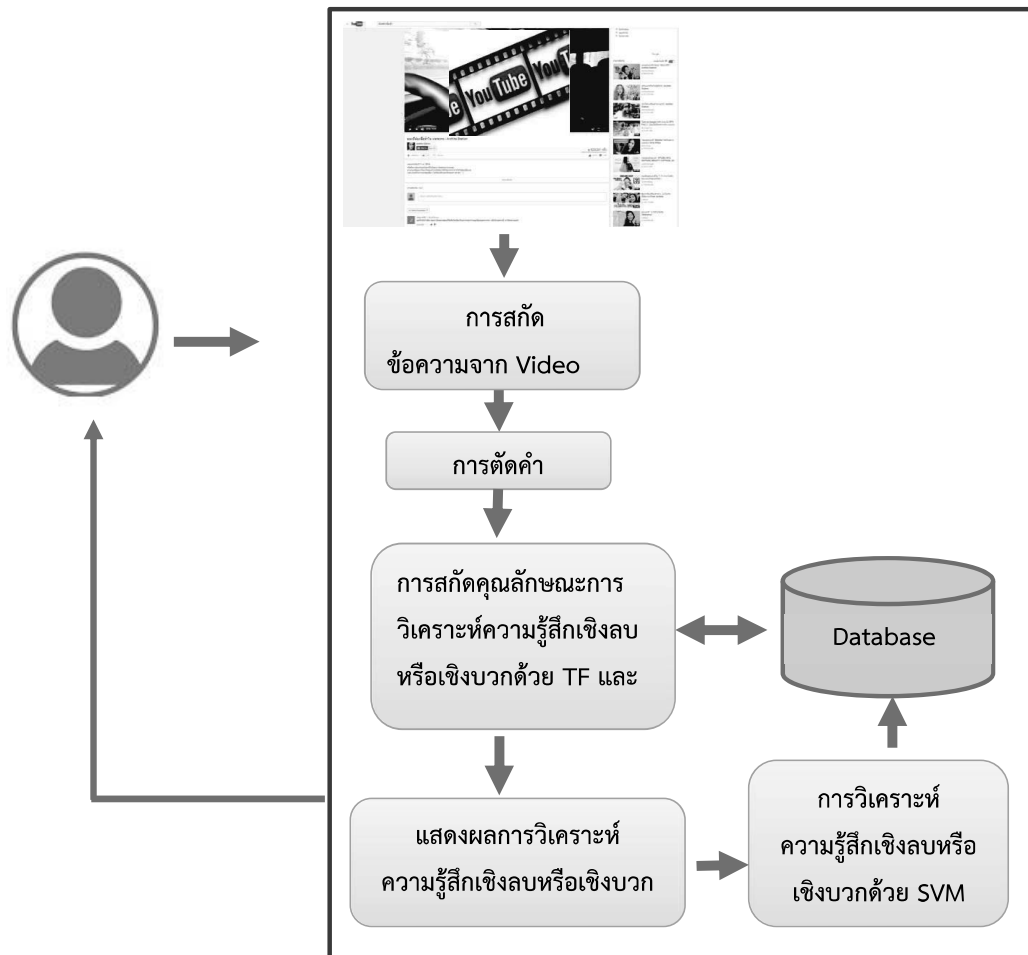
1) การสกัดข้อความจากวิดีโอ การพัฒนาระบบจะทำการเชื่อมโยงเพื่อเรียกใช้งาน Web Speech API ของกูเกิลโดยผู้ใช้งานนำไฟล์วิดีโอหรือไฟล์ลิงค์เข้าสู่ Web Speech API ซึ่ง จะทำการสกัดออกมาเป็นข้อความ แสดงดังภาพที่ 2 ซึ่งจากขั้นตอนนี้ผู้ใช้สามารถจัดเก็บข้อความที่สกัดได้จากวิดีโอมาเก็บไว้เป็นไฟล์นามสกุล .txt

2) การตัดคำ จากข้อความที่ได้จากการการสกัดข้อความจากวิดีโอซึ่งจัดเก็บไว้ในลักษณะไฟล์นามสกุล .txt การพัฒนาระบบจะเขียนโปรแกรมเรียกใช้ฟังก์ชันการตัดคำ thsplitlib เพื่อทำการตัดคำไว้ในระบบ ผู้ใช้สามารถนำเข้าไฟล์นามสกุล .txt แสดงดังภาพที่ 3 ระบบจะแสดงข้อความที่นำเข้าระบบ จากนั้นผู้ใช้กดปุ่มแยกคำ ระบบจะทำการตัดคำ แสดงดังภาพที่ 4

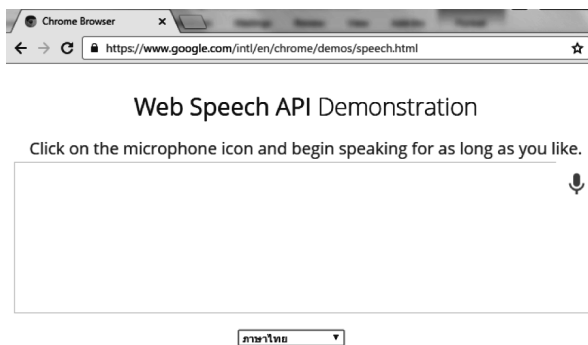
3) การสกัดคุณลักษณะด้วย TF และ IDF และการวิเคราะห์ความรู้สึกเชิงลบหรือเชิงบวก ระบบมีการพัฒนาส่วนการสกัดคุณลักษณะโดยอาศัยน้ำหนักของคำนำมาคำนวณค่า TF และ IDF เพื่อวิเคราะห์ความถี่ของคำที่ปรากฏในเอกสาร และทำการพัฒนาระบบฐานข้อมูลคำเชิงความหมายในเชิงบวกและเชิงลบ จะทำการวิเคราะห์คำจากโปรแกรมตัดคำ thsplitlib ซึ่งมีคำ 76,314 คำ [16] โดยผู้เชี่ยวชาญด้านภาษาจะทำการวิเคราะห์คำว่าคำใดมีความหมายเชิงบวกหรือเชิงลบและจะนำมาจัดเก็บเป็นฐานข้อมูลเพื่อทำการเปรียบเทียบคำที่ผ่านจากขั้นตอนการตัดคำและทำการวิเคราะห์โดยเปรียบเทียบกับฐานข้อมูลที่ได้จัดทำไว้ว่าคำนั้นเป็นเชิงบวกหรือเชิงลบและคำนวณค่า TF และ IDF แสดงดังภาพที่ 5

4) การแสดงผลการวิเคราะห์ความรู้สึกเชิงลบหรือเชิงบวก การพัฒนาระบบจะพัฒนาส่วนการสกัดคุณลักษณะโดยอาศัยน้ำหนักของคำนำมาคำนวณค่า TF และ IDF เพื่อวิเคราะห์ความถี่ของคำที่ปรากฏในเอกสาร และทำการพัฒนาระบบฐานข้อมูลคำเชิงความหมายในเชิงบวกและเชิงลบ จะทำการวิเคราะห์คำจากโปรแกรมตัดคำ thsplitlib ซึ่งมีคำ 76,314 คำ โดยผู้เชี่ยวชาญด้านภาษาจะทำการวิเคราะห์คำว่าคำใดมีความหมายเชิงบวกหรือเชิงลบและจะนำมาจัดเก็บเป็นฐานข้อมูลเพื่อทำการเปรียบเทียบคำที่ผ่านจากขั้นตอนการตัดคำและทำการวิเคราะห์โดยเปรียบเทียบกับฐานข้อมูลที่ได้จัดทำไว้ว่าคำนั้นเป็นเชิงบวกหรือเชิงลบและคำนวณค่า TF และ IDF แสดงดังภาพที่ 6 ถึงภาพที่ 8

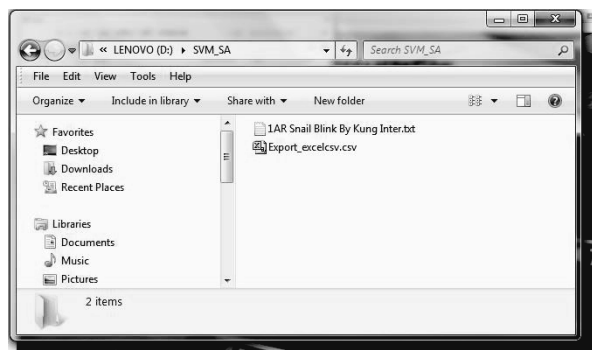
5) การจำแนกคุณลักษณะด้วย SVM ระบบมีการพัฒนาการจำแนกคุณลักษณะด้วย SVM และทำการจัดเก็บคำที่มีคุณลักษณะที่ดีที่สามารถเป็นตัวจำแนกความคิดเห็นเชิงบวกหรือเชิงลบไว้ในฐานข้อมูลเพื่อปรับปรุงฐานข้อมูลให้สามารถวิเคราะห์ได้ดี ในการจำแนกคุณลักษณะด้วย SVM เวกเตอร์คุณลักษณะที่ได้นำไปฝึกฝนตัวจำแนกประเภทด้วยการใช้ SVM จากข้อมูลในการทดสอบจำนวน 500 ตัวอย่าง



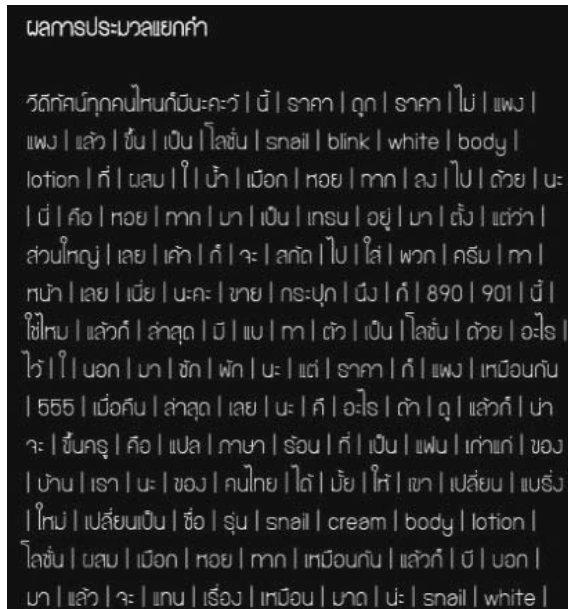
ภาพที่ 1 การออกแบบระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนเซิร์ฟเวอร์ด้วยซอฟต์แวร์แมชชีน



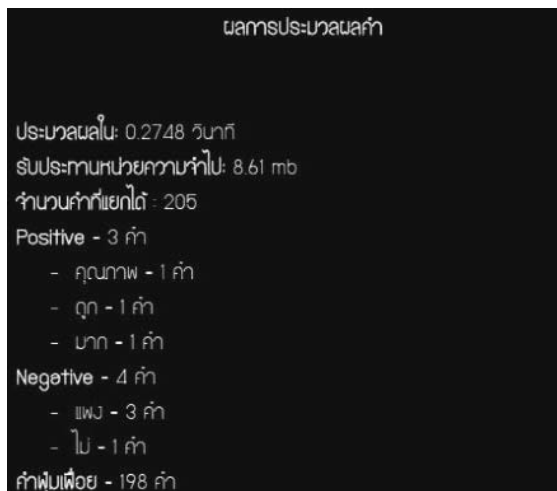
ภาพที่ 2 การสกัดข้อความจากวิดีโอผ่าน
Web Speech API



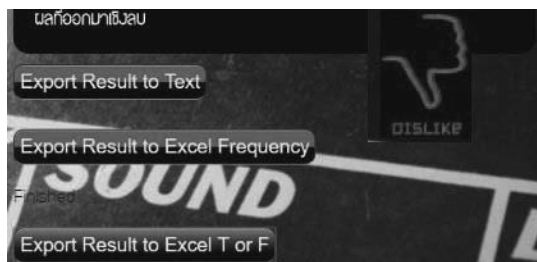
ภาพที่ 3 การนำเข้าไฟล์นามสกุล .txt



ภาพที่ 4 ผลการตัดคำ



ภาพที่ 5 ผลการสกัดคุณลักษณะด้วย TF และ ID และการวิเคราะห์ความรู้สึกเชิงลบหรือเชิงบวก



ภาพที่ 6 ผลการวิเคราะห์เชิงบวกหรือเชิงลบ



ภาพที่ 7 ผลการนำออกข้อมูลการสกัดคุณลักษณะเป็นไฟล์



ภาพที่ 8 ผลการนำออกข้อมูลเป็นไฟล์เพื่อนำไปใช้ในการจำแนกคุณลักษณะด้วย SVM

โดยข้อมูลจะถูกแบ่งเป็นชุดฝึกฝนและชุดทดสอบ ใช้ค่า Split 65% เป็นการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งเปอร์เซ็นต์ เลือกข้อมูลแบบสุ่มออกมา 35 ชุด เพื่อใช้เป็นชุดฝึกฝนและใช้ข้อมูลอีก 65 ชุดที่เหลือในการทดสอบ สำหรับพารามิเตอร์ของ SVM ที่ถูกปรับทำให้ได้รับค่าความถูกต้องสูงสุดที่พบในการทดลอง คือที่ 96.57% ได้แก่ ค่า C = 50000 ค่า Kernel = RBF และ ค่า gamma = 0.1 เพื่อกำหนดความสามารถในการพยากรณ์ข้อมูลใหม่ที่ไม่เคยพบ และสามารถจำแนกข้อมูลนั้นได้ถูกต้อง

3.4 การประเมินประสิทธิภาพระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียไทยด้วยซอฟต์แวร์เวกเตอร์แมชชีน

1) การประเมินประสิทธิภาพด้วยวิธีการที่วัดประสิทธิภาพของการจำแนกข้อมูลด้านความถูกต้อง (Accuracy) ซึ่งคำนวณดังสมการที่ 5

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN}) \quad (5)$$

TP = ความคิดเห็นเชิงบวกและจำแนกว่าเป็นเชิงบวก

FP = ความคิดเห็นเชิงบวกและจำแนกว่าเป็นเชิงลบ

TN = ความคิดเห็นเชิงลบและจำแนกว่าเป็นเชิงลบ

FN = ความคิดเห็นเชิงลบและจำแนกว่าเป็นเชิงบวก

การประเมินประสิทธิภาพระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีนด้วยค่าความถูกต้องจะทำการประเมินโดยผู้เชี่ยวชาญที่มีความรู้ด้านคอมพิวเตอร์จำนวน 15 คน เพื่อประเมินว่าระบบสามารถวิเคราะห์ความรู้สึกจากวิดีโอออกมาเป็นเชิงบวกหรือเชิงลบได้ถูกต้องพบว่าระบบมีความถูกต้องในการจำแนกคิดเป็น 98%

2) การประเมินประสิทธิภาพความพึงพอใจต่อระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีน คณะผู้วิจัยจัดทำแบบสอบถามความพึงพอใจเพื่อนำไปให้ผู้ใช้งานจำนวน 60 คน ซึ่งเป็นนักศึกษาระดับปริญญาตรี โดยการคัดเลือกกลุ่มตัวอย่างได้จากการอาสาสมัคร โดยทำการประเมินความพึงพอใจการใช้งานระบบ

การวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีน จากนั้นนำข้อมูลมาวิเคราะห์หาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเพื่อประเมินผลประสิทธิภาพด้านความพึงพอใจต่อระบบ ซึ่งมีเกณฑ์การให้คะแนนของแบบประเมินความพึงพอใจตามเกณฑ์ของลิเคิร์ต (Likert Scale) [18] ดังตารางที่ 1

ตารางที่ 1 เกณฑ์การให้คะแนนของแบบประเมินความพึงพอใจ

เกณฑ์การให้คะแนน		ความหมาย
เชิงปริมาณ	เชิงคุณภาพ	
4.51-5.00	มากที่สุด	ผู้มีความพึงพอใจมากที่สุด
3.51-4.50	มาก	ผู้มีความพึงพอใจมาก
2.51-3.50	ปานกลาง	ผู้มีความพึงพอใจปานกลาง
1.51-2.00	น้อย	ผู้มีความพึงพอใจน้อย
1.00-1.50	น้อยที่สุด	ผู้มีความพึงพอใจน้อยที่สุด

4. ผลการวิจัย

การจำแนกความคิดเห็นสำหรับระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีน โดยทำการทดสอบจากวิดีโอบน YouTube ของผู้เชี่ยวชาญที่ให้ความคิดเห็นเกี่ยวกับผลิตภัณฑ์เครื่องสำอางจำนวน 500 วิดีโอ การประเมินประสิทธิภาพด้านความถูกต้องโดยผู้เชี่ยวชาญที่มีความรู้ด้านคอมพิวเตอร์จำนวน 15 คน ว่าสามารถจำแนกความคิดเห็นเชิงบวกเชิงลบได้ถูกต้องคิดเป็น 98% จากนั้นจัดให้มีการใช้งานระบบและทำการประเมินประสิทธิภาพด้านความพึงพอใจระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีน โดยจัดทำแบบสอบถามความพึงพอใจเพื่อนำ

ไปให้ผู้ใช้งานระบบจำนวน 60 คน ซึ่งเป็นนักศึกษา
ระดับปริญญาตรี ทำการประเมินความพึงพอใจ
ในระบบการวิเคราะห์ความรู้สึกจากวิดีโอบน
โซเชียลมีเดียด้วยซอฟต์แวร์แมชชีน ผล
การประเมินประสิทธิภาพด้านความพึงพอใจ
ระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียล
มีเดียด้วยซอฟต์แวร์แมชชีน พบว่าผล
การประเมินความพึงพอใจมีค่าเฉลี่ยเลขคณิต
เท่ากับ 4.79 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ
0.41 แสดงให้เห็นว่าผู้ใช้งานมีความพึงพอใจระบบ
การวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดีย
ด้วยซอฟต์แวร์แมชชีน อยู่ในระดับมาก
ที่สุด แสดงดังตารางที่ 2

ตารางที่ 2 ผลการประเมินความพึงพอใจระบบการ
วิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วย
ซอฟต์แวร์แมชชีน

รายการประเมิน	ค่าเฉลี่ย เลขคณิต	ส่วน เบี่ยงเบน มาตรฐาน	ระดับความ พึงพอใจ
ระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีนสามารถใช้งานได้ง่าย	4.78	0.42	มากที่สุด (95.67%)
ระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีนสามารถแสดงผลข้อมูลได้ถูกต้อง	4.78	0.42	มากที่สุด (95.67%)
ระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีนสามารถจำแนกข้อมูลได้	4.73	0.45	มากที่สุด (94.67%)

รายการประเมิน	ค่าเฉลี่ย เลขคณิต	ส่วน เบี่ยงเบน มาตรฐาน	ระดับความ พึงพอใจ
ระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีนสามารถนำเข้าสู่ข้อมูลได้อย่างถูกต้อง	4.78	0.42	มากที่สุด (95.67%)
ระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีนสามารถวิเคราะห์ความคิดเห็นจากวิดีโอได้	4.87	0.34	มากที่สุด (95.67%)
สรุป	4.79	0.41	มากที่สุด (95.80%)

5. สรุปและอภิปรายผลการวิจัย

งานวิจัยนี้ได้พัฒนาระบบการวิเคราะห์
ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์
แมชชีน โดยจัดเก็บความคิดเห็นเกี่ยวกับ
ผลิตภัณฑ์เครื่องสำอางจากวิดีโอบนสื่อสังคม
ออนไลน์ YouTube โดยผู้เชี่ยวชาญจะจำแนก
ความคิดเห็นเป็น 2 กลุ่ม ได้แก่ เชิงบวกและเชิง
ลบ การเตรียมข้อมูลก่อนการจำแนกจะทำการ
สกัดข้อความจากวิดีโอบนสื่อสังคมออนไลน์
YouTube จากนั้น ทำการตัดคำ และใช้ SVM
ในการจำแนกความคิดเห็นเป็น 2 ประเภท
ดังกล่าวข้างต้น และระบบจะแสดงผลการ
วิเคราะห์ความรู้สึกไปยังผู้ใช้ จากการประเมิน
ความถูกต้องของระบบมีความถูกต้อง 98%
อัตราความผิดพลาดในการจำแนกคิดเป็น 2%
เมื่อเทียบกับการจำแนกโดยผู้เชี่ยวชาญ และผล
ประเมินความพึงพอใจของผู้ใช้ต่อระบบที่พัฒนา
ขึ้น จำนวน 60 คน พบว่าค่าเฉลี่ยเท่ากับ 4.79
และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.41 คิดเป็น

95.80% แสดงว่าผู้ใช้มีความพึงพอใจระบบการวิเคราะห์ความรู้สึกจากวิดีโอบนโซเชียลมีเดียด้วยซอฟต์แวร์แมชชีนในระดับมากที่สุด ซึ่งผลการวิจัยในการวิเคราะห์ความรู้สึกด้วยเทคนิค SVM มีค่าความถูกต้องที่สอดคล้องกับงานวิจัย Naradhipa และ Purwarianti (2011), Binali, Wu และ Potdar (2009), Rosas, Mihalcea และ Morency (2013), Akaich, Dhouioui และ Perez (2013), Kaushik, Sangwan และ Hansen (2013), อติเทพ และ รัฐสิทธิ์ (2013) และ Brynielsson และคณะ (2014)

การวิเคราะห์ความรู้สึกบนโซเชียลมีเดียทั้งในส่วนข้อความ ภาพ เสียง และวิดีโอ ยังคงมีการวิเคราะห์อย่างต่อเนื่อง ปัจจุบันโซเชียลมีเดียเป็นที่นิยมของผู้ใช้ ตลอดจนผู้ใช้สามารถแสดงความคิดเห็นและบอกเล่าเรื่องราวไปยังผู้อื่นได้อย่างง่ายซึ่งคำบอกเล่าเหล่านั้นส่งผลต่อการตัดสินใจของผู้อื่นทั้งทางตรงและทางอ้อม และข้อมูลดังกล่าวสามารถแพร่กระจายไปได้อย่างรวดเร็ว ในทางที่เป็นประโยชน์ต่อผู้บริโภค อาจช่วยเตือนในการบริโภคสินค้า เป็นข้อมูลในการพัฒนาสินค้าแก่ภาคธุรกิจ และในทางกลับกันเป็นข้อมูลเชิงลบที่ส่งผลต่อภาคธุรกิจได้หรือเป็นเพียงข้อมูลชวนเชื่อเพื่อการขายสินค้า การใช้เทคโนโลยีเพื่อเข้ามาช่วยในการวิเคราะห์ความรู้สึกจึงยังคงมีการวิเคราะห์อย่างแพร่หลายเพื่อการประมวลผลที่รวดเร็วและสามารถแสดงผลข้อมูลเชิงบวกและเชิงลบได้อย่างทันทั่วทั้งที่ เพื่อช่วยให้ภาคธุรกิจสามารถรับมือกับปัญหาที่อาจจะส่งผลกระทบได้

6. กิตติกรรมประกาศ

ขอขอบคุณมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ และขอขอบคุณมหาวิทยาลัยราชภัฏสวนสุนันทาที่อำนวยความสะดวกด้านการวิจัย

บรรณานุกรม

- (1) สุพัตรา วิริยะวิสุทธิสกุล, 2559. ระบบแจ้งเตือนโซเชียลมีเดียไทยสำหรับธุรกิจด้วยซอฟต์แวร์แมชชีน. วารสารปัญญาทัศน์, 8 : 223-233.
- (2) Neethu M.S., and Rajasree R., 2013. Sentiment Analysis in Twitter using Machine Learning Techniques. International Conference Computing, communications and Networking Technologies (ICCCNT), 4-6July 2013.
- (3) Lima A.C.E.S. and Castro deL.N., 2013. Automatic Sentiment Analysis of Twitter Messages. Forth International Conference on Computational Aspects of Social Networks, 52-57.
- (4) Isah H., Neagu D. and Trundle P., 2014. Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis. 14th UK Workshop on Computational Intelligence, 19 June 2015.
- (5) Kadmateekarun, P. and Nuanmeesri, S., 2015. Automatic Sentiment Analysis from Opinion of Thais Speech Audio. International Conference on Science and Technology (TICST), 25-26 May 2015.
- (6) Naradhipa, A. R. and Purwarianti, A., 2011. Sentiment Classification for Indonesian Message in Social Media. Proceedings of the 2011 International Conference on Electrical Engineering and Informatics (IEEE), 1-4.

- (7) Haji H. B., Chen W., and Vidyasagar P., 2009. A New Significant Area: Emotion Detection in E-learning using Opinion Mining Techniques, International Conference on Digital Ecosystems and Technologies, (IEEE, DEST '09), 1-3 June 2009.
- (8) Rosas, V.P., Mihalcea, R., and Morency, L.-P., 2013. Multimodal Sentiment Analysis of Spanish Online Videos. IEEE Intelligent Systems.
- (9) Pongthano P., et al., 2012. Customer Satisfaction Analysis from Comment Text using Opinion Mining. International Conference on Knowledge and Smart Technologies (KST-2012), 7-8 July 2012, 53-60.
- (10) Akaichi J., Dhouioui Z., and Perez M.J.L.H., 2013. Text Mining Facebook Status Updates for Sentiment Classification. International Conference on System Theory, Control and Computing (ICSTCC 17th), 11-13 October 2013.
- (11) Kaushik L., Sangwan A. and Hansen J.H.L., 2013. Sentiment Extraction from Natural Audio Streams. International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP 2013), 26-31 May 2013, 8485-8489.
- (12) อติเทพ ไชยสาร และ รัฐสิทธิ์ สุขะหุต, 2013. การประมาณอารมณ์จากความคิดเห็นภาษาไทยโดยใช้เทคนิคการเรียนรู้ของเครื่อง. The 9th National Conference on Computing and Information Technology, 9-10 May 2013; 260-266.
- (13) Brynielsson J., et al., 2014. Emotion Classification of Social Media Posts for Estimating People's Reactions to Communicated Alert Messages During Crises. Security Informatics, 3(1),
- (14) Sarakit, P., et al., 2015. Classifying Emotion in Thai Youtube Comments. International Conference of Information and Communication Technology for Embedded Systems (6th IC-ICTES IEEE), 1-5.
- (15) นิเวศ จิระวิฑิตชัย, ปริญญา สงวนลัทธิ และพญง มีสัจ. 2554. การพัฒนาประสิทธิภาพการจัดหมวดหมู่เอกสารภาษาไทยแบบอัตโนมัติ. วารสารพัฒนบริหารศาสตร์, 51(3): 188-205.
- (16) Thai Split Library : THsplitlib 3.0 (Computer software). (Online) <http://www.alogik.com/thsplitlib/>
- (17) Viriyavisuthisakul S., et al., 2015. A Comparison of Similarity Measures for Online Social Media Thai Text Classification. International Conference on Electrical Engineering/Electronics, Computer Telecommunications and Information Technology (12th ECTI-CON IEEE), 1-6.
- (18) ราตรี นันทสุนก, 2554. การวิจัยในชั้นเรียนและการวิจัยพัฒนาการเรียนการสอน. กรุงเทพฯ: จุฑทองจำกัด.