

การเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็น
ของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร
Performance Comparison of Data Mining Techniques for Building Classification Models
of the Parent Toward Children who use Smart Phone

สุภัทสรุ สมเจตนา^{1*} และ จารี ทองคำ¹
Supatsara Somjettana^{1*} and Jaree Thongkam¹

¹กลุ่มสารสนเทศเชิงประยุกต์ ภาควิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม
อ.กันทรวิชัย จ.มหาสารคาม 44150

¹Applied Information Group, Department of Information Technology, Faculty of Informatics, Mahasarakham
University, Kantharawichai District, Maha Sarakham Province 44150

*Email: seemaprang135@gmail.com, jaree.thongkam@gmail.com

Received: 06 Jan 21

Revised: 04 Mar 21

Accepted: 26 Mar 21

บทคัดย่อ

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคในเหมืองข้อมูลในการสร้างแบบจำลองจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร งานวิจัยนี้ใช้ 6 เทคนิค ได้แก่ เทคนิคริบเปอร์ เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคเคเนียร์เรสเนเบอร์ และเทคนิคต้นไม้ป่าสุ่ม มาสร้างแบบจำลองความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร โดยข้อมูลนั้นถูกรวบรวมมาเฉพาะความคิดเห็นของผู้ปกครองที่มีลักษณะเป็นข้อความภาษาไทยบนเครือข่ายสังคมออนไลน์ผ่านเว็บไซต์พันทิป และเฟซบุ๊ก จำนวนทั้งหมด 1,925 ข้อความ ตัวแปรที่ใช้ในงานวิจัยนี้ผู้วิจัยได้เลือกเฉพาะคำวิเศษณ์ที่สามารถระบุความรู้สึกได้เป็นอย่างดี 10-โพล์ครอสวาไลเดชั่น ได้ถูกนำมาใช้ในการแบ่งกลุ่มข้อมูลชุดเรียนรู้และชุดทดสอบ รวมถึงวัดประสิทธิภาพของแบบจำลองด้วยโดยใช้ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล หลังจากทำการทดลองและวัดประสิทธิภาพของแบบจำลอง พบว่า เทคนิคป่าสุ่ม เป็นเทคนิคที่ดีที่สุดในการวิเคราะห์ความคิดเห็นสำหรับข้อมูลชุดนี้ โดยให้ค่าความแม่นยำ 89.62% ค่าความระลึก 78.38% และค่าความถ่วงดุล 83.55%

คำสำคัญ: การเปรียบเทียบประสิทธิภาพของแบบจำลอง แบบจำลองจำแนกความคิดเห็น เทคนิคเหมืองข้อมูล ความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร

Abstract

This research aimed to compare the efficacy of a data mining technique in modeling parental opinion on the use of their children's smartphones. This research investigated opinion mining, comparing the model performance of 6 techniques: the RIPPER, C4.5 decision tree, Naïve Bayes, Support Vector Machine, K-Nearest Neighbor, and Random Forest technique to analyze the opinions of parents on the use of smartphones of their children. Therefore, the data was collected from only the Thai texts of parents' opinions on the social platforms: Pantip and Facebook, a total of 1,925 messages. In this process, only adverbs were used to classify the state of the word, conveying both positive and negative sentiments well. The researcher applied the 10-fold cross-validation performance measurement technique to cluster the learning data and the testing data and measure the model performance using Precision, Recall, and F-measure. After testing and measuring the performance of the model, it could be concluded that the Random Forest technique was the best technique to analyze the opinions for this data set. Conclusively, this Random Forest technique provided a model performance score of 83.85% with Precision of 89.62%, Recall of 78.38%, and F-measure of 83.55%.

Keywords: Model performance comparison, Opinion classification model, Data mining techniques, Parents' Opinion of Your Child's Use of Smartphones

1. บทนำ

ในปัจจุบันเป็นที่ทราบกันดีอยู่แล้วว่าสมาร์ทโฟน ได้เข้ามาจับบทบาทมาก ทั้งด้านการทำงาน การติดต่อสื่อสารจนกลายเป็นส่วนหนึ่งในการใช้ชีวิตประจำวันไปแล้ว ไม่ว่าจะเป็นผู้สูงอายุ วัยทำงาน วัยรุ่น รวมไปถึงเด็กเล็ก ๆ ที่ปัจจุบันนี้ก็มี การใช้สมาร์ทโฟนมากขึ้น และการใช้ก็มีแนวโน้มเพิ่มสูงขึ้นเรื่อย ๆ [1] โดยเฉพาะในกลุ่มเด็ก ๆ ได้มีการใช้สมาร์ทโฟนกันสูงมาก [2] โดยขาดการความเอาใจใส่ของพ่อแม่ ดังมีผลการศึกษาพบว่าเด็กส่วนใหญ่ที่อาศัยอยู่กับพ่อแม่กว่าร้อยละ 80 ที่มีการใช้โทรศัพท์แบบสมาร์ทโฟนที่ตัวเองเป็นเจ้าของและมีการใช้อย่างอิสระ จำนวนชั่วโมงเฉลี่ยที่เล่นต่อวันอยู่ประมาณ 3 ชั่วโมง และมีพฤติกรรมติดเกมมากถึงร้อยละ 45 [3] ซึ่งพ่อแม่ในยุคปัจจุบันต้องทำงานไม่มีเวลาดูแล หรือฝากให้ปู่ย่าตายาย หรือญาติคนอื่น ๆ เลี้ยงให้ การให้เด็กใช้สมาร์ทโฟนทำให้เด็กสงบนิ่ง ส่งผลกระทบต่อพัฒนาการด้านทักษะการสื่อสาร ทักษะการแก้ปัญหา ทักษะการปฏิสัมพันธ์กับสังคม โดยตรงทั้งทางด้านสมองและสายตาแอลง [4] , [5] แต่บางครั้งการใช้สมาร์ทโฟนของเด็กสามารถช่วยเพิ่มพัฒนาการภาษา [6] ทำให้ผู้ปกครองเกิดการแลกเปลี่ยนความคิดเห็นในเฟซบุ๊กและเว็บไซต์พันธุทิพย์ เพื่อแนะนำการใช้สมาร์ทโฟนที่เป็นประโยชน์ให้แก่บุตร ซึ่งความคิดเห็นของผู้ปกครองจะเป็นเชิงบวกและเชิงลบ ในการวิเคราะห์ความคิดเห็นของผู้ปกครองที่มีต่อบุตรที่ใช้สมาร์ทโฟนในเชิงบวกและเชิงลบ เพื่อสนับสนุนการตัดสินใจให้แก่ผู้ปกครองในการเลี้ยงบุตรที่เหมาะสมกับยุคปัจจุบัน

เหมืองความคิดเห็น (Opinion Mining) [7] เป็นกระบวนการวิเคราะห์และจำแนกความคิดเห็นบนเครือข่ายสังคมออนไลน์ จากความคิดเห็นส่วนใหญ่ที่เกิดขึ้นจะเป็นไปในเชิงบวกและเชิงลบ นักวิจัยได้ใช้เทคนิคริปเปอร์ (RIPPER) เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 (Decision tree C4.5) เทคนิคนาอิวเบย์ (Naïve Bayes) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM Support Vector Machine) เทคนิคเคเนียร์เรสเนเบอร์ (KNN) และเทคนิคการสุ่มป่าไม้ (Random forest) ในการวิเคราะห์ความคิดเห็น เช่น นุชนาฏ ปิ่นเมือง และ จารี ทองคำ [8] , [9] , [10] ได้ทำการวิจัยเกี่ยวกับความคิดเห็นของคนไทยต่อสื่อออนไลน์ด้วยเทคนิค 5 เทคนิค อันได้แก่ เทคนิคนาอิวเบย์ เทคนิคจำแนกเชิงเส้น เทคนิคเคเนียร์เรสเนเบอร์ เทคนิคต้นไม้ตัดสินใจและเทคนิคต้นไม้ตัดสินใจแบบซี 4.5 พบว่าเทคนิคนาอิวเบย์ มีประสิทธิภาพในการจำแนกความคิดเห็นโดยมีค่าความถูกต้องมากถึงร้อยละ 93.88 และค่าความแม่นยำร้อยละ 94.02 ส่วน Gupta และคณะ ได้ทำเหมืองความคิดเห็นของลูกค้าที่มาใช้บริการโรงแรมและให้คะแนนการบริการ กลับพบว่า เทคนิคต้นไม้ตัดสินใจมีความ แม่นยำถึง 76.22 เปอร์เซ็นต์ และง่ายต่อการเข้าใจของพนักงานโรงแรม และผดุงนันอำไพและ จารี ทองคำ [11] ได้ทำการวิจัยเกี่ยวกับการตรวจจับการบุกรุกในระบบเครือข่ายด้วยเทคนิคการจำแนกในการทำเหมืองข้อมูลด้วย 4 เทคนิค คือ เทคนิคการตัดสินใจแบบตาราง, เทคนิคนาอิวเบย์, เทคนิคริปเปอร์และเทคนิค PART

decision list พบว่า แบบจำลองที่ใช้เทคนิคริปเปอร์ มีค่าความแม่นยำมากที่สุดคือ 99.00% และวัชรวิวรรณ จิตต์สกุล [12] ได้วิเคราะห์ความเสถียรของอัลกอริทึมการเชื่อมต่อกัน, เทคนิคต้นไม้ป่าสุ่ม, การถดถอยโลจิสติกแบบเบย์ และเทคนิคจำแนกเชิงเส้น ในการจำแนกข้อความแสดงความคิดเห็นเชิงบวกและเชิงลบในการให้บริการของเว็บไซต์ พบว่าอัลกอริทึมที่ให้ผลลัพธ์ดีที่สุด คือ อัลกอริทึมเทคนิคต้นไม้ป่าสุ่ม

ดังนั้นผู้วิจัยจึงมีแนวคิดที่จะพัฒนาแบบจำลองเพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร โดยการนำความคิดเห็นของผู้ปกครองบนเครือข่ายสังคมออนไลน์เว็บไซต์พันทิป และเฟซบุ๊ก ว่ามีความคิดเห็นไปในเชิงบวกหรือเชิงลบ โดยการเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลที่ใช้ในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็น ได้แก่ เทคนิคริปเปอร์ เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 เทคนิคนาอิวเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคเคเนียร์เรสเนเบอร์ และเทคนิคการสุ่มป่าไม้ แล้ววัดประสิทธิภาพของแบบจำลองด้วยค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล

2. วิธีดำเนินการวิจัย

ในการดำเนินการวิจัยนี้ได้ใช้กระบวนการในการทำเหมืองความคิดเห็นซึ่งมี 4 ขั้นตอนมาใช้ โดยเริ่มจากการรวบรวมข้อมูล การเตรียมข้อมูล การสร้างแบบจำลองเพื่อการจำแนก และการวัดประสิทธิภาพของแบบจำลอง

2.1 การรวบรวมข้อมูล

ในการรวบรวมข้อมูล ผู้วิจัยจะทำการเก็บรวบรวมข้อมูล ตั้งแต่วันที่ 1 มกราคม 2561 ถึงวันที่ 31 ธันวาคม 2562 โดยความคิดเห็นนั้นจะเป็นข้อความที่มีลักษณะเป็นภาษาไทยที่ได้มีการแสดงความคิดเห็นบนเครือข่ายสังคมออนไลน์ จากเว็บไซต์พันทิปและเฟซบุ๊กเท่านั้น ดังแสดงใน Figure 1 และ Figure 2 โดยทำการค้นหาข้อมูลซึ่งจะใช้ข้อความความคิดเห็นจากเพจหลักที่ได้รับความนิยมและน่าเชื่อถือที่ได้รับการจัดอันดับแนะนำ 10 เพจหมอดูเด็ก ให้ความรู้เรื่องการเลี้ยงลูก ที่พ่อกับแม่ต้องติดตาม เท่านั้น เช่น เพจเลี้ยงลูกนอกบ้าน, นายแพทย์ประเสริฐ ผลิตผลการพิมพ์, Dr.Pam book club, เลี้ยงลูกตามใจหมอ, เลี้ยงลูกให้เป็นคนปกติ, ชมรมจิตแพทย์เด็กและวัยรุ่นแห่งประเทศไทย, เข็นเด็กขึ้นภูเขา, สุธีรา เอื้อไพโรจน์กิจ, เลี้ยงลูกโตไปด้วยกันกับหมอ Hormone for Kids, หมอเสาวภา เลี้ยงลูกเชิงบวก และเพจสารพันปัญหาการเลี้ยงลูก โดยใช้คำสำคัญในการค้นหาเช่น เด็ก, สมาร์ทโฟน, เด็กเล่นมือถือ, การเล่นเกมของเด็ก, เด็กติดโทรศัพท์, ประโยชน์มือถือ เป็นต้นจากนั้นทำการดึงข้อมูลความคิดเห็น โดยในแต่ละเรคคอร์ดจะประกอบไปด้วยชื่อผู้ใช้งานหรือ ID ของผู้ใช้งานวันที่ และข้อความความคิดเห็นของผู้ใช้งาน เมื่อทำการเก็บข้อมูลความคิดเห็นออกมาแล้ว นำมาจัดเก็บใช้โปรแกรม Microsoft Excel ในการรวบรวมข้อมูล หรือแก้ไขข้อมูล ต่าง ๆ ได้ จำนวนทั้งหมด 1, 925 ข้อความ ดังแสดงใน Figure 3

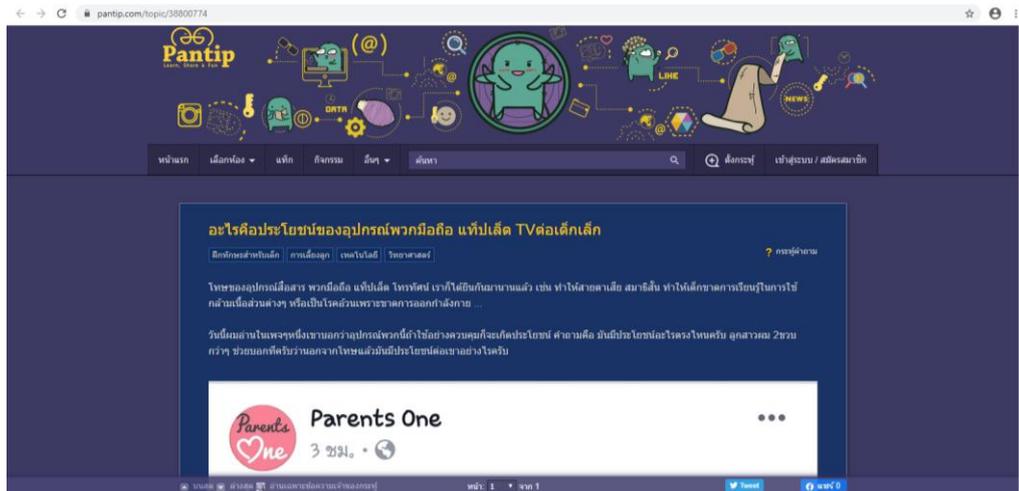


Figure 1 FROM URL: <https://pantip.com/topic/38800774>

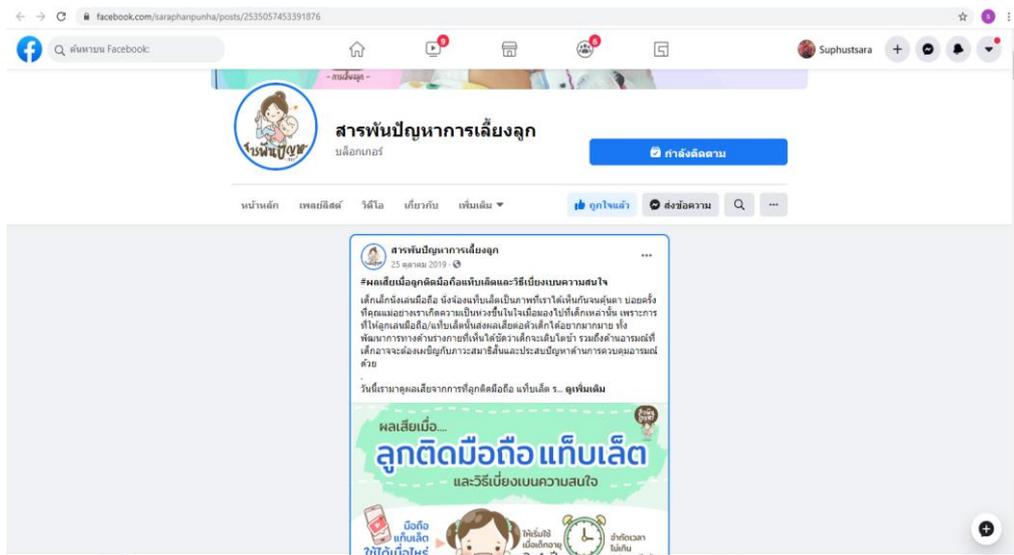


Figure 2 FROM URL: <https://www.facebook.com/saraphanpunha/posts/2535057453391876>

Date	Comment
25/4/62	มันก็ไม่ต่างกับ TV สมัยตอนคุณเด็กๆ แหละครับ ตอนเด็กๆ เราก็ดูมากกับการนั่งดู TV พ่อแม่ ก็ต้องคอยห้ามว่า ไม่ให้ดูเกิน
25/4/62	ความรู้
25/4/62	มันน่าสนใจว่าหนังสือต่างๆ เป็นธรรมชาติของตัวมันอยู่แล้ว จะทำให้เด็ก เรียนรู้อะไรต่างๆ ได้รวดเร็ว ด้านภาษาด้วยก็ได้ ให้
25/4/62	ประโยชน์มันก็มีบ้างครับ อย่างเช่น เราอยากให้ลูกน้อยรู้จักพวกสัตว์ต่างๆ เช่น โลมา แพนกวิน ช่าง ฯลฯ ใ้ครั้นจะว่า
25/4/62	จริงๆแล้ว คงต้องเรียกว่า ดาบ 2 คม ครับ ถ้าเรารู้จักใช้ให้เกิดประโยชน์ มันก็เกิด ประโยชน์
25/4/62	เห็นด้วยกับ คท.ที่6 ครับ...อยู่ที่พ่อแม่ต้องคอยสอนเขาด้วยว่า อันไหนดี หรืออันไหนไม่ดี ลูกสาวผมเริ่มให้เล่น ตอน อา:
25/4/62	ที่เห็นมา พ่อแม่ชี้แจงอยากพักผ่อนก็เอามือถือให้ลูกเล่น แล้วเด็กก็จะจมอยู่กับมือถือจนไม่มาทักพ่อแม่ แต่พอนานๆไ
25/4/62	เวลาลูกทำการบ้าน เราชี้แจงก็ให้ดู มือถือ พวกฝึกร้องเพลงเด็ก เกมส้อธิบายเรื่องการผสมสี แม่สี เกมส้อต่างๆ มีอยู่พวก
25/4/62	เรื่องการศึกษาค้นคว้ามากที่สุด
25/4/62	มันเชื่อม AI book และสามารถเปิดสารคดี รวมถึงคลิปและแอปต่างๆที่มีประโยชน์ต่อเด็กเอง แต่ต้องอยู่ในความดูแลของ

Figure 3 Examples of comments collected in Microsoft Excel

2.2 การเตรียมข้อมูล

ขั้นตอนการเตรียมข้อมูลนี้เป็นกระบวนการในการจัดทำข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ได้ ซึ่งมีขั้นตอนดังนี้

1) การแก้ไขคำผิด ในการแก้ไขคำผิดผู้วิจัยได้นำข้อมูลเข้าสู่โปรแกรมตารางคำนวณ Microsoft Excel และ

ได้ทำการแก้ไขคำผิดโดยการค้นหาคำที่พิมพ์ผิดบ่อย คำที่พิมพ์เป็นภาษาพูดและตัดสัญลักษณ์ต่าง ๆ ที่ไม่เกี่ยวข้องออกไป

2) จากนั้นจึงเข้าสู่กระบวนการตัดคำด้วยโปรแกรมภาษา Python ไลบรารีของ PyThaiNLP ดังแสดงใน Table 1

3) การจัดทำถ่วงคำ หลังผ่านกระบวนการตัดคำ กำจัดคำหยุดจะได้คำทั้งหมด 5, 409 คำ โดยการใช้พจนานุกรมไทยในการแยกประเภทของคำ ดังแสดงใน Table 2

Table 1 Wrapping examples

	Comment Cut
1	คุณพ่อ, คุณแม่, ดิฉัน, มื้อ, ค่ะ, แนะนำ, กรณีย์, ลูกสาว, เรา, ที่, ทำ, แล้ว, ได้ผล, นะคะ, เรา, สามี่, จะ, ไม่, เล่น, โทรศัพท์, เวลา, ที่อยู่, ลูก, ค่ะ, เรา, จะ, ตก ลง, กัน, ว่า, จะ, ไม่, ติด, โทรศัพท์, ทุกคน, แล้ว, ให้, ลูก, ใช้, ตามกำหนด, เวลา, ลูก, เรา, เรา, ปลอ่ย, ให้, เล่น, ได้, ชม., แล้ว, ...
2	หลาน, เรา, ก็, ติด, เกมส์, แต่, ไม่, ก้าวร้าว, บอก, อะไร, ก็, เชื้อ, รับผิดชอบ, เรื่อง, เรียน, เพราะ, แม่, เค้า, ไม่, ตามใจ, ทั้งหมด, ให้, รับผิดชอบ, ส่วน, ของ, ตัวเอง, ดูแลตัวเอง, ให้, ทำ, อะไร, เอง, และ, พ่อ, ดู, มาก, แต่, มี, ผล, ไม่, ดู, เพราะ, พ่อ, ดู, เมื่อ, ทำผิด, หนัก, เกรงใจ, ...
3	เอา, ง่าย ๆ, ผู้ใหญ่, ยัง, ติดงอมแงม, เด็ก, นี้, ยัง, ไม่, มี, วิจารณญาณ, เลย, ต้อง, คอย, คม, ดี
4	ที่, เห็น, มา, พ่อแม่, ชี้แจง, อายาก, พักผ่อน, ก็, เอา, มือถือ, ให้, ลูกเล่น, แล้ว, เด็ก, จะ, จมอยู่กับ, มือถือ, จน, ไม่, มา, กวน, พ่อแม่, แต่, พ่อ, นาน ๆ, ไป, กลายเป็น, ลูกคิด, มือถือ, ที่, พบ, บ่อย, คือ, ไม่, มี, ความอดทน, รอคอย, ไม่, มี, ระเบียบวินัย, ก้าวร้าว, สมาธิ, สั้น, ขาด, จินตนาการ, ...
5	ขอ, เข้ามา, ตอบ, บ้าง, นะคะ, มี, ลูกสาว, ค่ะ, แต่, เรา, ไม่, ได้, ห้าม, เรื่อง, จอ, นะคะ, ส่วนมาก, จะ, เป็น, ที่วี, ที่, ดู, เกือบ, ทุกวัน, ประมาณ, วัน, ละ, นาที, แต่, ไม่, ตายตัว, ค่ะ, มือถือ, นี้, ให้, นั่ง, ดู, บ้าง, แต่, ค่อนข้างน้อย, ถึง, แทบ, ไม่, มี, ไม่, ใช้, อะไร, หรือ, ลูก, ใช้, ไม่, เป็น, ค่ะ, ลูก,

Table 2 Example of specifying word types

Word	Word type
หยุดหยุด	ว.
แน่	ว.
แนบ	ก.
แนะ	ก.
ก็ดี	ว.
แบบ	น.
แก้ตัว	น.
แม่	สัน.
กีฬา	น.
เกร	ว.

4) การคัดเลือกคุณลักษณะ โดยเลือกเอาเฉพาะคำวิเศษณ์จำนวน 671 คำ มาระบุตามความหมายเชิงบวก เชิงลบ และเป็นกลาง การระบุจะพิจารณาเป็นคำ ๆ โดย

ไม่สนใจประโยค แล้วจะคงเหลือคำที่บ่งชี้ลักษณะคำ จำนวน 394 คำ โดยแบ่งคำออกเป็น 2 กลุ่ม คือ คำแทนคุณลักษณะเชิงบวกและคำแทนคุณลักษณะเชิงลบ ดังแสดงใน Table 3

Table 3 Examples of adverbs that state the feelings of positive and negative commentators in the word bag

Word	Word type	Weight
หยุดหยุด	ว.	-1
ก็ช่าง	ว.	-1
ก้าวร้าว	ว.	-1
งอมแงม	ว.	-1
กร้าว	ว.	-1
แถม	ว.	1
ก็ดี	ว.	1
ก็ได้	ว.	1
ทันสมัย	ว.	1
กล้า	ว.	1

5) การแทนค่าคุณลักษณะ เมื่อได้คำวิเศษณ์ที่ระบุความคิดเห็นเชิงบวกและเชิงลบแล้ว จากนั้นแทนค่าของจำนวนค่าคุณลักษณะที่ไม่ปรากฏเป็น 0 แทนค่าจำนวนค่าคุณลักษณะเชิงบวกที่ปรากฏเป็น 1 และแทนค่าจำนวนค่าคุณลักษณะเชิงลบที่ปรากฏเป็น -1 จะเห็นความถี่ของค่าคุณลักษณะที่ถูกจำแนกออกมาเป็นเชิงบวกและเชิงลบในแต่ละความคิดเห็น และทำการแบ่งความคิดเห็นเชิงบวกและเชิงลบ โดยใช้หลักเกณฑ์

ถ้าจำนวนความถี่คุณลักษณะเชิงบวกมากกว่า จะแทน ความคิดเห็นเชิงบวก ด้วย P

ถ้าจำนวนความถี่คุณลักษณะเชิงลบมากกว่า จะแทน ความคิดเห็นเชิงลบ ด้วย N

แต่ถ้าหากความถี่ของคุณลักษณะเชิงบวกและเชิงลบเท่ากัน จะแทน ด้วย B และข้อความนั้นจะถูกตัดออก ดังแสดงใน Table 4 แล้วจึงทำการลบแถวตั้งที่ข้อความ P และ N ออก ก่อนที่จะนำข้อมูลดังกล่าวไปสร้างแบบจำลอง

Table 4 Characterization example

Message	หุงตหจิด	กัซง	กักรั้ว	งอมแงม	กรั้ว	กัตี	กัใต้	พันสมัย	กล้า	...	P	N	Class
1	0	0	0	0	0	0	0	0	1	...	3	1	P
2	0	0	-1	0	0	1	0	0	0	...	1	0	P
3	0	0	0	-1	0	0	0	0	0	...	1	1	B
4	0	0	-1	0	0	0	0	0	0	...	0	1	N
5	-1	0	0	0	0	0	0	0	0	...	1	2	N

จากข้อความความคิดเห็นทั้งหมด จำนวน 1,925 ข้อความ พบว่ามีข้อความที่มีความคิดเห็นเชิงบวก จำนวน 528 ข้อความ ข้อความที่มีความคิดเห็นเชิงลบ จำนวน 809 ข้อความ และข้อความที่มีความคิดเห็นเป็นกลาง จำนวน 588 ข้อความ โดยจะเลือกเอาเพียงข้อความที่มีความคิดเห็นเชิงบวกและเชิงลบ เท่านั้นมาสร้างแบบจำลอง

2.3 การสร้างแบบจำลองเพื่อการจำแนกความคิดเห็น

การสร้างแบบจำลองเพื่อการจำแนกด้วยการใช้งานโปรแกรมเวก้า เวอร์ชัน 3.8.4 ในการเลือกเทคนิคมาใช้ผู้วิจัยได้เลือกอัลกอริทึมที่มีลักษณะพื้นฐานที่แตกต่างกัน 4 พื้นฐาน ได้แก่ ฐานกฎ ฐานต้นไม้ตัดสินใจ ฐานความน่าจะเป็นและฐานแห่งการเรียนรู้ และเทคนิคอื่นอีก 2 เทคนิคที่มีประสิทธิภาพสูงจากการวิจัยที่ได้ศึกษามา รวมเป็น 6 เทคนิค ดังนี้

เทคนิคริปเปอร์ [11] [13] เป็นเทคนิคที่พัฒนาจากเทคนิค IRIP สามารถสร้างกฎเองได้ โดยการเรียนรู้จากข้อมูลที่เตรียมไว้ให้กฎที่สร้างขึ้นจะอยู่ในรูปของ if then else ประกอบด้วย 2 เฟส คือ เฟสที่ 1 ระบุกฎเริ่มต้น เฟสที่ 2 ระบุค่า porst – process rule optimization โดยการเรียนรู้จากข้อมูลที่กำหนดคลาสไว้เรียบร้อยแล้ว ซึ่งแบ่งเป็น Growing set และ pruning set และจะหาค่าที่ดีที่สุด growing set ใน rule space อธิบายได้จาก BNF จากที่ได้ Growing set แล้ว จะทำการ pruning ข้อมูลทันที เพื่อให้ได้ผลออกมาแล้วเป็นที่พอใจมากที่สุด

เทคนิคต้นไม้ตัดสินใจ ซี4.5 เป็นเทคนิควิธีเหมือนข้อมูลที่เป็นที่นิยมกันมาก เนื่องจากเป็นวิธีการวิเคราะห์ที่ง่ายต่อการตีความหมาย ต้นไม้ตัดสินใจนั้นจะประกอบไปด้วย โหนด ซึ่งจะทำหน้าที่ในการแสดงคุณลักษณะที่ใช้สำหรับการ

ทดสอบข้อมูล กิ่งเป็นส่วนที่จะแสดงคุณสมบัติในโหนดที่ได้มีการแตกออกมา และใบจะแสดงกลุ่มหรือคลาสที่ได้มีการกำหนดเอาไว้ และในหาความสัมพันธ์ของแต่ละโหนดที่แสดงคุณลักษณะนั้นจะใช้ค่าเกนความรู้ (Information Gain) เพื่อหาความสัมพันธ์ของในแต่ละโหนดคุณลักษณะ โดยการใช้ค่าอัตราส่วนเกน (Gain Ratio) นี้จะช่วยลดจำนวนครั้งที่ใช้ในการทดสอบและทำให้ต้นไม้ตัดสินใจที่ได้ไม่มีความซับซ้อนมากเกินไปสามารถแสดงการคำนวณค่าเกนความรู้ได้

เทคนิคนาอิวเบย์ [12] มีพื้นฐานมาจากกฎของเบย์ โดยจะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละตัวกับตัวแปรตาม เพื่อสร้างเงื่อนไขความน่าจะเป็นของแต่ละความสัมพันธ์ กล่าวถึง ความน่าจะเป็นของเหตุการณ์ที่เกิดขึ้น (A) ถ้ามี เหตุการณ์อีกเหตุการณ์หนึ่งเกิดมาแล้ว (B) สามารถ เขียนให้อยู่ในรูปอย่างง่าย

เทคนิคซัพพอร์ตเวกเตอร์แมชชีน [12] เป็นเทคนิคเพื่อลดความผิดพลาดจากการทำนาย จัดเป็นเทคนิคที่ใช้ในการแก้ปัญหาทางด้านการรู้จำรูปแบบข้อมูลอาศัยหลักการจำแนกหมวดหมู่ข้อมูลด้วยการหาระนาบตัดสินใจและแบ่งข้อมูลออกเป็น 2 ส่วน โดยจะพยายามสร้างเส้นแบ่งตรงกึ่งกลางระหว่างกลุ่มให้มีระยะระหว่างขอบเขตทั้ง 2 กลุ่มมากที่สุด เพื่อหาระนาบการตัดสินใจในการแบ่งข้อมูล โดยใช้ฟังก์ชันแม่ปสำหรับย้ายข้อมูลจาก Input Space ไปยัง Feature Space ในลักษณะเชิงเส้น

เทคนิคเคเนียร์เรสเนเบอร์ [14] เป็นวิธีที่ใช้ในการจัดแบ่งคลาส โดยเทคนิคนี้จะตัดสินใจว่า คลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวมของจำนวนเงื่อนไข หรือกรณีต่าง ๆ สำหรับ

แต่ละคลาส และกำหนดเงื่อนไขใหม่ ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด ซึ่งขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุดนั้นเป็นเทคนิคที่ใช้ในการจำแนกกลุ่มข้อมูล โดยการจัดข้อมูลที่อยู่ใกล้กันให้เป็นกลุ่มเดียวกันซึ่งเทคนิคนี้จะทำให้ตัดสินใจได้ว่า คลาสไหนที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง โดยการตรวจสอบจำนวน K ซึ่งถ้าหากเงื่อนไขของการตัดสินใจมีความซับซ้อน วิธีนี้สามารถสร้างโมเดลที่มีประสิทธิภาพได้

เทคนิคการสุ่มป่าไม้ [15] เป็นเทคนิคการสุ่มเลือกข้อมูลและคุณลักษณะต้นไม้มัดตัดสินใจ ซึ่งถูกสร้างจากการนำข้อมูลไปสุ่มเลือกตัวอย่างแบบเลือกแล้วใส่กลับ แล้วนำมาสร้างเป็นต้นไม้ ซึ่งจะมีตัวอย่างส่วนหนึ่งที่ไม่ถูกเลือก ซึ่งข้อมูลส่วนนี้เรียกว่า Out-of-Bag (OOB) จะถูกนำมาใช้ในการทดสอบ วิธีการดังกล่าวนี้เรียกว่า Bagging ผลลัพธ์ที่ได้จะต่างจาก Decision Tree ในแต่ละต้นถูกนำมาคิดเป็นผลการโหวต ผลโหวตที่มากที่สุดจะใช้ระบุสถานะของคลาส เทคนิค Random Forest ไม่จำเป็นต้องมีข้อมูลทดสอบ เพื่อประมาณความผิดพลาดเพราะข้อมูล OOB นั้นถูกนำมาใช้ทดสอบ Decision Tree นั้นแล้ว

2.4 การวัดประสิทธิภาพของแบบจำลอง

ในการวัดประสิทธิภาพของแบบจำลองด้วยวิธี 10-โพลด์ครอสวาไลเดชัน จะทำการเลือกสุ่มข้อมูลออกเป็น K ชุดเท่า ๆ กัน จากนั้นจะทำการทดลองครั้งแรกด้วยข้อมูลชุดที่ 1 ซึ่งเป็นข้อมูลทดสอบและกำหนดให้ข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน และในการทดลองครั้งที่สองจะใช้ข้อมูลชุดที่ 2 เป็นชุดข้อมูลทดสอบและให้ข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน ทำจนกระทั่งข้อมูลทุกชุดข้อมูลได้ถูกนำมาเป็นชุดข้อมูลทดสอบทั้งหมด ซึ่งจำนวนในการทดสอบมีจำนวนเท่ากับ K ครั้ง โดยผลลัพธ์ที่ได้นั้นจะมากำหนดค่าเฉลี่ยความถูกต้องของการจำแนกข้อมูลในแต่ละรอบ โดยวิธีการทดสอบประสิทธิภาพแบบ 10-โพลด์ครอสวาไลเดชัน มีข้อเสียคือ จะต้องทำการเริ่มทดสอบใหม่โดยจะต้องทำทั้งหมด K รอบ

การวัดประสิทธิภาพการทำงานในแต่ละขั้นตอนวิธีสามารถวัดได้จากผลของการจำแนกกลุ่มข้อมูล และสามารถหาค่าความแม่นยำ ค่าความระสีก และค่าความถ่วงดุล โดย

ค่าความแม่นยำ (Precision) คือ การวัดความสามารถในการที่จะจัดเอกสารที่ไม่เกี่ยวข้องออกไปโดยที่ค่าความแม่นยำนั้นจะเป็นอัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและได้มีการถูกดึงออกมา เพื่อเทียบกับจำนวนเอกสารที่ถูกดึงออกมาทั้งหมด ดังสมการที่ 1

$$Precision = \frac{TP}{TP+FN} \quad (1)$$

ค่าความระสีก (Recall) คือ การวัดความสามารถของระบบในการดึงเอกสารที่เกี่ยวข้องออกมา โดยค่าความระสีกนั้นจะใช้อัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและได้

มีการถูกดึงออกมา เทียบกับจำนวนเอกสารที่เกี่ยวข้องทั้งหมด ซึ่งสามารถหาได้จากสมการ ดังสมการที่ 2

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

โดยที่ TP คือ จำนวนข้อมูลที่แบบจำลองพยากรณ์ได้อย่างถูกต้อง

TN คือ จำนวนข้อมูลที่แบบจำลองพยากรณ์ได้อย่างถูกต้อง

FP คือ จำนวนข้อมูลที่แบบจำลองพยากรณ์ได้ไม่ถูกต้อง

FN คือ จำนวนข้อมูลที่แบบจำลองพยากรณ์ได้ไม่ถูกต้อง

ค่าความถ่วงดุล (F-Measure) คือการวัดประสิทธิภาพโดยรวมของทั้งสองค่าระหว่างค่าความแม่นยำและค่าความระสีก พร้อมกันของโมเดล โดยพิจารณาแยกทีละคลาส ดังสมการที่ 3

$$F - Measure = \frac{2 * Precision + Recall}{Precision + Recall} \quad (3)$$

เมื่อ Precision คือ ค่าความแม่นยำของโมเดล

Recall คือ ค่าความระสีกของโมเดล

3. ผลการวิจัย

งานวิจัยนี้ได้ทำการวิเคราะห์ความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ตโฟนของบุตร โดยนำเทคนิคของการทำเหมืองข้อความมาใช้ในการวิเคราะห์ทั้งหมด 6 เทคนิค ได้แก่ เทคนิคริบเปอร์ เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 เทคนิคนาอ็พเบย์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคเคเนียร์ เรสเนเบอร์ และเทคนิคการสุ่มป่าไม้ ผ่านการใช้งานโปรแกรมเวก้า เวอร์ชัน 3.8.4 ในการทำการทดลอง โดยข้อมูลที่ใช้ในงานวิจัยนี้เก็บรวบรวมจากการนำความคิดเห็นของผู้ปกครองบนเครือข่ายสังคมออนไลน์เว็บไซต์พันธุทิพย์และเฟซบุ๊ก ตั้งแต่วันที่ 1 มกราคม 2561 ถึงวันที่ 31 ธันวาคม 2562 จำนวน 1,925 ชุดข้อมูล จากนั้นทำการค้นหาคุณลักษณะของแต่ละประโยค โดยใช้ค่าเพื่อทำการกำหนดคุณลักษณะของแต่ละประโยค โดยใช้ค่าเพื่อทำการกำหนดคุณลักษณะทั้งหมด 671 คำ และเลือกเอาเพียงคำบ่งชี้คุณลักษณะเชิงบวกและเชิงลบจำนวน 394 คำเท่านั้น สามารถแบ่งออกเป็นคำบ่งชี้คุณลักษณะเชิงบวก 192 คำ และคำบ่งชี้คุณลักษณะเชิงลบจำนวน 202 คำ ในการประเมินประสิทธิภาพของโมเดลที่ 6 โมเดล โดยใช้การทดสอบโมเดลด้วยวิธีการ 10-โพลด์ครอสวาไลเดชัน ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ จากการทดสอบสามารถแสดงผลการวิจัยออกเป็นการเปรียบเทียบประสิทธิภาพและรูปแบบของแบบจำลองที่เทคนิคเหมืองข้อมูลสร้างขึ้น

3.1 การเปรียบเทียบประสิทธิภาพของแบบจำลอง การเปรียบเทียบประสิทธิภาพของแบบจำลองในงานวิจัยนี้ได้

ใช้ค่าความแม่นยำ ค่าความระลึกและค่าความถูกต้อง สามารถ
แสดงได้ใน Figure 4 Figure 5 และ Figure 6

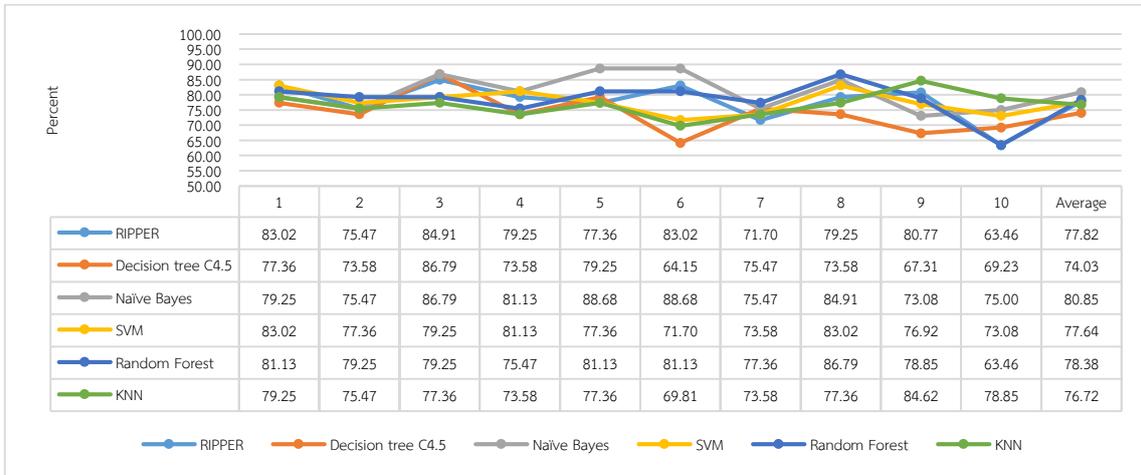


Figure 4 Precision

Figure 4 แสดงค่าความแม่นยำ ของเทคนิคริบเปอร์, เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5, เทคนิคนาอิวเบย์, เทคนิคซัพพอร์ตเวกเตอร์แมชชีน, เทคนิคเคเนียร์เรสเนเบอร์ และเทคนิคการสุ่มป่าไม้ ในการเปรียบเทียบประสิทธิภาพของเทคนิคในเหมืองข้อมูลในการสร้างแบบจำลองจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์โฟนของบุตร ผลปรากฏเทคนิคการสุ่มป่าไม้ให้ค่าความแม่นยำ สูงสุดที่ 89.62% เมื่อพิจารณาตามรอบแล้วจะเห็นได้ว่าในรอบที่ 1-6 มีค่าเฉลี่ยใกล้เคียงกัน อยู่ทีระหว่าง 87-93 % และในรอบที่ 7 ได้ลดลง

ต่ำสุดกว่าทุกรอบที่ 83.67% หลังจากลดลงต่ำสุดแล้วได้พุ่งสูงขึ้นที่สุดในรอบที่ 8 ที่ 95.83% แล้วค่อย ๆ ลดลงมาในรอบที่ 9 ที่ 95.35% จนมาในรอบสุดท้ายรอบที่ 10 ค่าได้ลดลงมาใกล้เคียงกับรอบที่ 1-6 จะเห็นได้ว่าเทคนิคการสุ่มป่าไม้ ที่ให้ค่าความแม่นยำเฉลี่ยสูงสุดที่89.62% รองลงมาคือเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ที่ 85.45% ส่วนเทคนิคที่ให้ค่าความแม่นยำ เฉลี่ยน้อยที่สุดคือ เทคนิคเคเนียร์เรสเนเบอร์ ที่ 74.61%

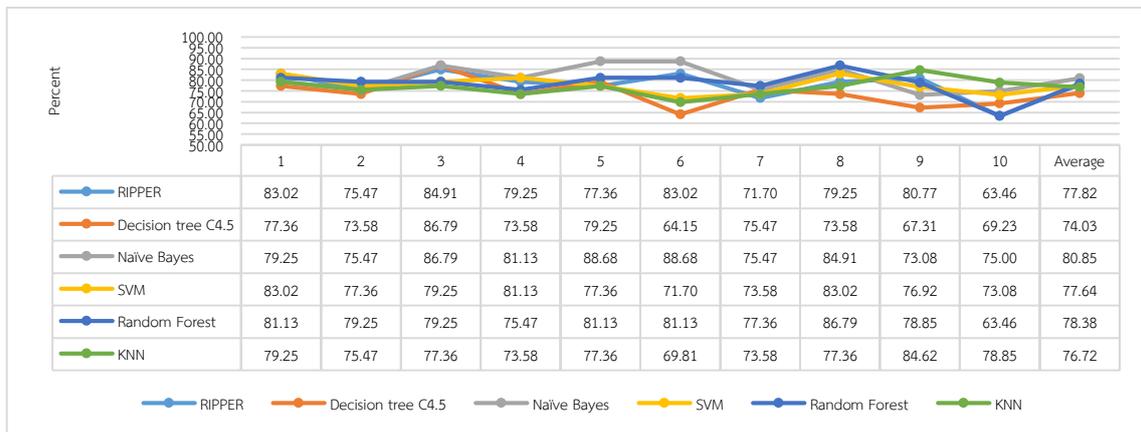


Figure 5 Recall

Figure 5 การแสดงค่าความระลึกของเทคนิคริบเปอร์, เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5, เทคนิคนาอิวเบย์, เทคนิคซัพพอร์ตเวกเตอร์แมชชีน, 4 เทคนิคเคเนียร์เรสเนเบอร์ และเทคนิคการสุ่มป่าไม้ ในการเปรียบเทียบประสิทธิภาพของเทคนิคในเหมืองข้อมูลในการสร้างแบบจำลองจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์โฟนของบุตร ผลปรากฏว่า เทคนิคนาอิวเบย์ ให้ค่าความระลึก เฉลี่ยสูงสุดที่ 80.85% เมื่อพิจารณาตามรอบแล้วจะเห็นได้ว่าในรอบที่ 1 และรอบที่ 2

มีค่าเฉลี่ยอยู่ที่ 75-79% และในรอบที่ 3 ได้สูงขึ้นมาจนเกือบสูงสุดที่ 86.79% แล้วลดลงในรอบที่ 4 และมาในรอบที่ 6-7 ได้พุ่งสูงขึ้นที่สุดที่ 88.68% ทั้ง 2 รอบ หลังจากพุ่งขึ้นสูงสุดแล้วก็ค่อย ๆ ลดลงมาในรอบที่ 8-10 จะเห็นได้ว่า เทคนิคนาอิวเบย์ ให้ค่าความระลึก เฉลี่ยสูงสุดที่ 80.85% รองลงมาคือเทคนิคการสุ่มป่าไม้ ที่ 78.38% ส่วนเทคนิคที่ให้ค่าความระลึกลดน้อยที่สุด คือ เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 ที่ 74.03%

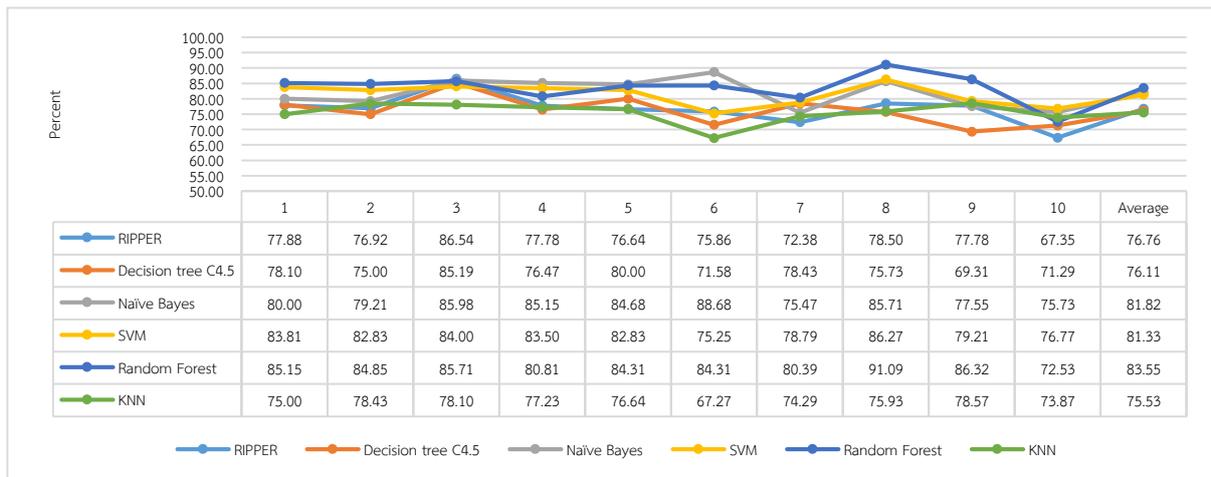


Figure 6 F-measure

Figure 6 แสดงค่าความถ่วงดุลของเทคนิคริปเปอร์, เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5, เทคนิคนาอิวเบย์, เทคนิคซัพพอร์ตเวกเตอร์แมชชีน, เทคนิคเคเนียร์เรสเนเบอร์ และเทคนิคการสุ่มป่าไม้ ในการเปรียบเทียบประสิทธิภาพของเทคนิคในเหมืองข้อมูลในการสร้างแบบจำลองจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์โฟนของบุตร ผลปรากฏว่า เทคนิคการสุ่มป่าไม้ ให้ค่าความถ่วงดุลสูงสุด เมื่อพิจารณาตามรอบแล้วจะเห็นได้ว่าในรอบที่ 1-7 มีค่าเฉลี่ยใกล้เคียงกัน อยู่ที่ระหว่าง 80-85% และมาพุ่งสูงขึ้นที่สุดในรอบที่ 8 ที่ 83.55% แล้วค่อยๆลดลงมาในรอบที่ 9 ที่ 86.32% จนมาในรอบสุดท้ายรอบที่ 10 ค่าได้ลดลงมาต่ำสุดกว่าทุกรอบ

จะเห็นได้ว่าเทคนิคการสุ่มป่าไม้ ให้ค่าความถ่วงดุลเฉลี่ยสูงสุดที่ 83.55% รองลงมาคือเทคนิคนาอิวเบย์ ที่ 81.82% ส่วนเทคนิคที่ให้ค่าความถ่วงดุลน้อยที่สุด คือ เทคนิคเคเนียร์เรสเนเบอร์ ที่ 75.53%

3.2 รูปแบบของแบบจำลอง

จากการสร้างแบบจำลอง สามารถที่จะแสดงรูปแบบจำลองได้จาก เทคนิคริปเปอร์ และเทคนิคต้นไม้ตัดสินใจแบบซี4.5 โดยที่แบบจำลองริปเปอร์ เป็นแบบจำลองที่มีรูปแบบเป็นเงื่อนไข จากการสร้างแบบจำลอง พบว่า เทคนิคริปเปอร์ สามารถสร้างกฎได้ทั้งหมด 11 กฎ ดังแสดงใน Figure 7

No.	Rule
Rule 1	(ไม่ >= 0) and (แต่ >= 0) and (อย่า >= 0) => class=p
Rule 2	(นี้ >= 1) and (กว่า >= 1) and (ดี >= 1) => class=p
Rule 3	(มาก >= 1) and (ไม่ >= -1) and (นี้ >= 1) => class=p
Rule 4	(ไม่ >= -1) and (นี้ >= 1) and (แต่ >= 0) and (แค่ >= 0) => class=p
Rule 5	(แต่ >= -1) and (มาก >= 2) and (เกือบจะ >= 0) and (ก้าวร้าว >= 0) => class=p
Rule 6	(ด้วย >= 1) and (ไม่ >= -2) and (แต่ >= 0) => class=p
Rule 7	(ไม่ >= -1) and (ควร >= 1) => class=p
Rule 8	(ใช่ >= 1) and (ไม่ >= -1) => class=p
Rule 9	(กว่า >= 1) and (แต่ >= 0) and (ไม่ >= -1) => class=p
Rule 10	(ไม่ได้ศัพท์ >= 1) and (แต่ >= -1) => class=p
Rule 11	=> class=n

Figure 7 Ripper model

Figure 7 แสดงรูปแบบของแบบจำลองริปเปอร์ โดยมีกฎทั้งหมด 11 กฎที่จัดอยู่ในหมวดความคิดเห็นเชิงบวก และหากได้แบบจำลองที่ไม่ตรงตามกฎ 11 กฎนี้ก็จะถูกจัดให้อยู่ในหมวดความคิดเห็นเชิงลบ ซึ่งจะขอยกตัวอย่างใน 3 เงื่อนไขแรกได้ดังนี้

กฎที่ 1 สามารถแปลได้ว่า ถ้าไม่มีคำว่า “ไม่” ถ้าไม่มีคำว่า “แต่” และถ้าไม่มีคำว่า “อย่า” ให้จัดอยู่ในกลุ่มความคิดเห็นเชิงบวก

กฎที่ 2 สามารถแปลได้ว่า ถ้ามีคำว่า “นี้” ถ้ามีคำว่า “กว่า” และ ถ้ามีคำว่า “ดี” ให้จัดอยู่ในกลุ่มความคิดเห็นเชิงบวก

กฎที่ 3 สามารถแปลได้ว่า ถ้ามีคำว่า “มาก” ถ้ามีคำว่า “ไม่” และถ้ามีคำว่า “นี้” ให้จัดอยู่ในกลุ่มความคิดเห็นเชิงบวก

แบบจำลองต้นไม้ตัดสินใจแบบ ซี4.5 เป็นแบบจำลองที่แสดงเงื่อนไขในลักษณะของต้นไม้ที่มีรากโหนดและความสัมพันธ์ จากการสร้างแบบจำลองพบว่าเทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 สามารถสร้างต้นไม้ที่มีขนาด 133 เงื่อนไข 67 เงื่อนไข สามารถแสดงตัวอย่างต้นไม้ 1 กิ่ง ดังแสดงใน Figure 8

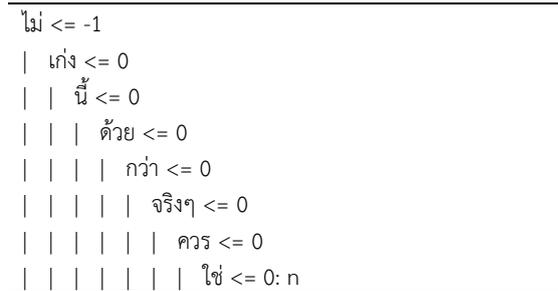


Figure 8 Decision Tree C 4.5 Model

Figure 8 สามารถแปลได้ว่า ถ้ามีคำว่า “ไม่” น้อยกว่าหรือเท่ากับ 1 คำ ไม่มีคำว่า “เก่ง” ไม่มีคำว่า “นี้” ไม่มีคำว่า “ด้วย” ไม่มีคำว่า “กว่า” ไม่มีคำว่า “จริงๆ” ไม่มีคำว่า “ควร” และถ้าไม่มีคำว่า “ใช่” ให้จัดอยู่ในกลุ่มความคิดเห็นเชิงลบ แต่เทคนิคการสุ่มป่าไม้จะเป็นต้นไม้หลายต้นที่เกิดจากการสุ่มข้อมูล แต่โปรแกรมเวก้าไม่สามารถแสดงแบบจำลองในรูปของต้นไม้หลายต้นได้ จึงไม่ได้แสดงผลของต้นไม้ป่าสุ่มได้ ส่วนแบบจำลองที่สร้างจากเทคนิคคานารี เทคนิคซัพพอร์ตเวกเตอร์แมชชีนและเทคนิคเคเนียร์เรสเนเบอร์ ซึ่งเป็นแบบจำลองที่ใช้สมการทางคณิตศาสตร์ในการจำแนก จึงไม่สามารถแสดงเป็นรูปแบบที่เข้าใจได้ง่ายดังแบบจำลองที่สร้างจากเทคนิคแบบจำลองริปเปอร์ และเทคนิคต้นไม้ตัดสินใจแบบ ซี4.5

4. วิจัยและสรุป

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลอง จากศึกษากระบวนการในการทำความเข้าใจของความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร ที่ถูกเขียนขึ้นด้วยภาษาไทย ซึ่งเก็บรวบรวมความคิดเห็นบนเครือข่ายสังคมออนไลน์ จากเว็บไซต์พันธุทิพย์และเฟซบุ๊ก ตั้งแต่วันที่ 1 มกราคม 2561 ถึงวันที่ 31 ธันวาคม 2562 จำนวนทั้งหมด 1,925 ข้อความ โดยงานวิจัยนี้ได้ทำการแบ่งคุณลักษณะออกเป็น 2 กลุ่ม คือ คุณลักษณะเชิงบวกและคุณลักษณะเชิงลบ โดยนำเอาคำวิเศษณ์จากความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร ซึ่งคำวิเศษณ์นี้สามารถแสดงถึงอารมณ์เชิงบวกและเชิงลบได้ดี จากนั้นได้นำเอาเทคนิควิธีการวิเคราะห์เหมือนข้อความมาทำการวิเคราะห์ข้อความความคิดเห็นจำนวน 6 เทคนิคด้วยกัน คือ

เทคนิคริปเปอร์, เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5, เทคนิคคานารี, เทคนิคซัพพอร์ตเวกเตอร์แมชชีน, เทคนิคเคเนียร์เรสเนเบอร์ และเทคนิคการสุ่มป่าไม้

สำหรับขั้นตอนวิธีการวัดประสิทธิภาพของทั้ง 6 โมเดล ได้ใช้หลักการ 10-โพลด์ครอสวาไลเดชัน ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ และวัดประสิทธิภาพของแบบจำลองด้วยค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล ผลการวิจัยพบว่า เทคนิคการสุ่มป่าไม้ เป็นเทคนิคที่ดีที่สุด ให้ค่าประสิทธิภาพของแบบจำลองที่ 87.885% โดยให้ค่าความแม่นยำ 89.62% ค่าความระลึก 78.38% และค่าความถ่วงดุล 83.55%

จากผลการวิจัยทำให้เห็นว่าเทคนิคการสุ่มป่าไม้ เหมาะกับชุดข้อมูลที่มีการจำแนกค่า เนื่องจากเทคนิคการสุ่มป่าไม้จะมีความได้เปรียบกว่าเทคนิคอื่น ๆ ที่ได้นำมาทดสอบ เพราะเทคนิคนี้มีความสามารถในการทำนายผลได้อย่างแม่นยำ โดยที่ไม่มีปัญหาในเรื่องของ overfitting ซึ่งสอดคล้องกับการวิจัยของ วัชรวิธรรม จิตต์สกุล [12] ได้วิเคราะห์ความเสถียรของอัลกอริทึม ในการจำแนกข้อความแสดงความคิดเห็นเชิงบวกและเชิงลบในการให้บริการของเว็บไซต์ พบว่าอัลกอริทึมที่ให้ผลลัพธ์ที่ดีที่สุด คือ เทคนิคการสุ่มป่าไม้ด้วยเช่นกัน ดังนั้นทำให้เห็นว่าเทคนิคการสุ่มป่าไม้ เหมาะที่จะใช้ในการวิเคราะห์ชุดข้อมูลความคิดเห็นได้ดีที่สุด ยิ่งไปกว่านั้นรูปแบบของแบบจำลองริปเปอร์และต้นไม้ตัดสินใจสามารถที่จะแปลผลได้ง่าย

5. กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณผู้ปกครองที่เข้าร่วมกันแสดงความคิดเห็นที่เป็นประโยชน์ และขอขอบคุณอาจารย์คณะวิทยการสารสนเทศ มหาวิทยาลัยมหาสารคามที่ให้คำปรึกษาและคำแนะนำต่าง ๆ เพื่อให้งานออกมาได้ดีและมีประสิทธิภาพมากที่สุด

6. References

[1] Fernández, A. 2011. Clinical Report: The impact of social media on children, adolescents and families, *Archivos de Pediatría del Uruguay*, 82(1): 31-32.

[2] Apirapee, S.T., Srirath, P. and Yanawuti, S. 2018. Screen Behaviour of Thai Toddlers Aged 0 – 3 years in Bangkok, *The journal of social communication innovation*, 6(2): 60-69. (in Thai)

[3] Apisit, N. and Supaporn, K. 2020. Factors related to Game Addiction among Prathom Suksa 4-6 Schoolchildren, *J. of Soc Sci & Hum*, vol. 46, no. 2, pp. 111-141 (in Thai)

- [4] Siritorn, Y. 2013. Gaming Addiction in Children: Screening and Solutions, **Journal of Nursing and Education**, 6(3): 16-24. *(in Thai)*
- [5] Sopana, J.C. and et al. 2018. A DANGER OF MOBILE TECHNOLOGY, **Research Journal Phranakhon Rajabhat: Science and Technology**, 13(1): 164-177. *(in Thai)*
- [6] Chat, C. and et al. 2019. The Learning Outcome of English Subject by Lesson One for Kids Application using Augmented Reality Technology, **Journal of Liberal Arts, Maejo University**, 7(2): 81-95. *(in Thai)*
- [7] Khan, F.H., Bashir, S. and U. Qamar. 2014. TOM: Twitter opinion mining framework using hybrid classification scheme, **Decision support systems**, vol. 57: 245-257.
- [8] Noochanat, P. and Jaree, T. 2018. Classifying Thai opinions on online media using text mining, **Journal of Science & Technology MSU**, 37(3). *(in Thai)*
- [9] Gupta, S. and et al. 2018. Opinion Mining for Hotel Rating through Reviews Using Decision Tree Classification Method, **International Journal of Advanced Research in Computer Science**, 9(2): 180.
- [10] Kanda, P. and Pramote, L. 2013. OPINION MINING FROM ONLINE SOCIAL NETWORKS, **Modern Management Journal**, 11(2): 11-20. *(in Thai)*
- [11] Phadung, N. and Jaree, T. 2019. Intrusion Detection Using Classification Techniques in Data Mining, **Journal of Information Technology Management and Innovation**, 6(2): 111-118. *(in Thai)*
- [12] Watchareewan, J. and Sunantha, S. 2017. TEXT CLASSIFICATION ANALYSIS BY STABILITY COMPARISON OF ALGORITHMS, **Sripatum Review of Science and Technology**, 9(1): 19-31. *(in Thai)*
- [13] Papidchaya, K. and Jaree, T. 2018. **Applying the Ensemble Technique for Improving Rule-based Models Performance**, M.Sc. Thesis, Mahasarakham University. *(in Thai)*
- [14] Joachims, T. 1998. Text categorization with Support Vector Machines: Learning with many relevant features, in *Machine Learning: ECML-98*, Berlin, Heidelberg, C. Nédellec and C. Rouveirol, Eds., : **Springer Berlin Heidelberg**, pp. 137-142.
- [15] sujatha, M. and Dr.Lavanya Devi, G. 2013. feature-selection-techniques-using-for-high-dimensional- data- in- machine- learning- IJERTV2IS90912.