การวิเคราะห์ความคิดเห็นต่อเกมมือถือพับจีด้วยเหมืองข้อความ

Opinion analysis on PlayerUnknown's Battlegrounds (PUBG) mobile games using text mining

วสวัตติ์ อินทร์แปลง[1*], จารี ทองคำ[2]
Wossawat Inplang[1*], Jaree Thongkam[2]

Received: 16 January 2020 ; Revised: 10 April 2020 ; Accepted: 19 May 2020

บทคัดย่อ

เหมืองข้อความเป็นกระบวนการวิเคราะห์ข้อมูลตัวอักษรเพื่อสกัดข้อมูลที่เป็นประโยชน์จากแหล่งข้อมูล ปัจจุบันเทคนิคในการจำแนกเหมืองข้อความมีหลายวิธี งานวิจัยนี้มีวัตถุประสงค์เพื่อค้นหาเทคนิคการจำแนก จาก 5 เทคนิคที่มีประสิทธิภาพ คือ เทคนิค Naïve Bayes เทคนิค Support Vector Machine (SVM) เทคนิค K-Nearest Neighbor เทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิค Random Forest โดยเก็บรวบรวมข้อมูลความคิดเห็นต่อเกมมือถือผับจีจำนวน 3,798 ข้อความ ในกระบวนการคัดเลือกคำบ่งชี้เพื่อใช้ในการแยกคุณลักษณะได้เลือกใช้คำวิเศษณ์ และคำสแลงบางคำที่ความหมายของคำเป็นคำวิเศษณ์เพื่อทำการแยกคุณลักษณะเชิงบวกและเชิงลบ ผลการศึกษาพบว่ามีความไม่สมดุลของคลาสในข้อมูล โดยมีจำนวนคลาสหนึ่งมากกว่าอีกคลาสหนึ่งเป็นจำนวนมาก ผู้วิจัยจึงได้ทำการแก้ปัญหาโดยการปรับความสมดุลของข้อมูลด้วยวิธี SMOTE (Synthetic Minority Over-sampling Technique) และใช้หลักการ 10-fold cross validation ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ และวัดประสิทธิภาพการจำแนกของแบบจำลองด้วยค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถูกต้อง (Accuracy) เมื่อทำการทดสอบและวัดประสิทธิภาพของโมเดลพบว่า เทคนิค K-Nearest Neighbor ให้ผลดีที่สุดในการวิเคราะห์ความคิดเห็น โดยให้ค่าความแม่นยำ 99.75% ค่าความระลึก 100% และค่าความถูกต้อง 99.87%

คำสำคัญ: เหมืองข้อความ ความคิดเห็น เกมมือถือ ข้อมูลไม่สมดุล การสุ่มเพิ่มตัวอย่างกลุ่มน้อย

Abstract

Text mining is one of the most effective data analysis processes using alphabetic methods. Currently, text mining techniques are classified in a variety of ways. This research aims to find the most effective of 5 techniques that were Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbor C4.5, and Random Forest. The data collected, in total of 3,798 messages, were all made by the viewers. The categorization process divided the data into 2 groups: positive character and negative character. Interestingly, the process has only indicated selection of adverbs and slangs as a core division to produce positive and negative characters. After analyzing the data, two problems were found class imbalanced. SMOTE were used for filtering and to increase the minority class. 10-fold cross validation was applied to segment the data into training and testing sets. Moreover, precision recall and accuracy are used as the criteria for selecting the most effective model. The results showed that the K-Nearest Neighbor produced greatest accuracy in categorizing the messages with a precision of 99.75% recall of 100% and accuracy score of 99.87%.

Keywords: Text mining, Opinion, Mobile game, Imbalanced data, SMOTE

---

[1] นิสิตปริญญาโท, หน่วยวิจัยสารสนเทศประยุกต์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม อำเภอกันทรวิชัย จังหวัดมหาสารคาม 44150,
[2] อาจารย์ที่ปรึกษา, หน่วยวิจัยสารสนเทศประยุกต์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม อำเภอกันทรวิชัย จังหวัดมหาสารคาม 44150,
[1] Master degree student, Applied Informatics Research Unit, Faculty of Informatics, Mahasarakham University, Kantharawichai District, Maha Sarakham 44150, Thailand.
[2] Advisor, Applied Informatics Research Unit, Faculty of Informatics, Mahasarakham University, Kantharawichai District, Maha Sarakham 44150, Thailand.
* Corresponding author ; Email address: wossawat.inplang@gmail.com

บทนำ

ปัจจุบันโทรศัพท์มือถือทำให้คนทุกเพศทุกวัยสามารถเข้าถึงเกมได้ง่ายและมากขึ้น การที่เกมถูกผสมผสานเข้ากับเทคโนโลยีต่างๆ ทำให้เกมบนมือถือได้รับความนิยมและมีความน่าสนใจมากขึ้น ซึ่งอุตสาหกรรมเกมเป็นอุตสาหกรรมหนึ่งที่มีมูลค่าทางเศรษฐกิจสูง ทำให้ผู้ผลิตและพัฒนาเกมต้องแข่งขันกันพัฒนาเกมมือถือออกมาให้มีคุณภาพและตรงกับความต้องการของผู้เล่นอยู่เสมอ PlayerUnknown's Battlegrounds: Mobile (PUBG Mobile) หรือเรียกสั้นๆ ว่า ผับจี โมบาย เป็นเกมมือถือที่ดาวน์โหลดได้ฟรี ซึ่งได้รับความนิยมอย่างมากในปัจจุบัน[1] แม้ว่าเกมมือถือผับจีจะสามารถดาวน์โหลดได้ฟรี แต่ถ้าหากผู้เล่นต้องการซื้ออุปกรณ์หรือไอเทมภายในเกมเพิ่มเติม ก็จะต้องเสียค่าใช้จ่ายผ่านการเติมเงินเข้าระบบเพื่อนำมาซื้อไอเทมนั้น การพัฒนาเกมให้มีคุณภาพ ไม่มีข้อบกพร่อง (Bug) และตรงกับความต้องการของผู้เล่นจึงเป็นสิ่งจำเป็นอย่างยิ่งของผู้ผลิตและพัฒนาเกม ซึ่งแหล่งที่มาสำคัญเพื่อช่วยระบุคุณลักษณะของเกมคือความคิดเห็น (Opinion) ที่มาจากผู้เล่น ซึ่งผู้เล่นจะกล่าวถึงความพึงพอใจที่มีต่อเกม รวมถึงปัญหาที่ผู้เล่นพบเจอจากการเล่นเกม และความคิดเห็นของผู้เล่นที่มีต่อเกมมือถือผับจียังเป็นตัวแปรสำคัญที่มีผลต่อผู้เล่นเกมอื่นๆ ในการตัดสินใจเลือกดาวน์โหลดเกมมือถือผับจีมาเล่น หรือตัดสินใจเติมเงินเข้าระบบเพื่อซื้อไอเทมภายในเกม[2] หากผู้ผลิตและพัฒนาเกมสามารถนำความคิดเห็นของผู้เล่นมาสร้างความพึงพอใจหรือนำมาปรับปรุงคุณภาพของเกมด้วยการสกัดข้อมูลจากความคิดเห็นของผู้เล่นที่มีต่อเกมได้ ก็จะสามารถพัฒนาเกมให้มีคุณภาพที่ดียิ่งขึ้น เป็นไปตามความต้องการของผู้เล่น และยังเป็นอีกช่องทางหนึ่งในการเข้าถึงผู้เล่น อีกทั้งยังสามารถแย่งชิงความได้เปรียบในการแข่งขันกับคู่แข่งทางการตลาดอีกด้วย อย่างไรก็ตาม เนื่องจากข้อความความคิดเห็นของผู้เล่นเกมมือถือผับจีนั้นมีจำนวนมาก และมีลักษณะเป็นภาษาธรรมชาติ (Natural Language) ทำให้ข้อความความคิดเห็นมีความแตกต่างกัน ซึ่งมีการใช้ภาษาที่ไม่ถูกต้องตามหลักไวยากรณ์ และมักไม่มีการสรุปว่าประโยคที่ผู้เล่นกล่าวถึงนั้นมีความคิดเห็นเป็นเชิงบวกหรือเชิงลบ ซึ่งหากต้องการทราบต้องเสียเวลาในการจำแนกเป็นอย่างมาก

การประมวลผลภาษาธรรมชาติ คือการแปลความจากภาษาธรรมชาติที่มนุษย์ใช้สื่อสารกันให้อยู่ในรูปแบบที่เป็นโครงสร้าง (Structured Data) ที่เครื่องคอมพิวเตอร์สามารถเข้าใจได้ 2 แนวทาง[3] คือ แนวทางการศึกษาและทำความเข้าใจกับโครงสร้างทางภาษาศาสตร์ และอีกแนวทางคืออาศัยความรู้ด้านปัญญาประดิษฐ์โดยการแทนความรู้ (Knowledge Representation) ด้วยคลังคำ (Corpus) ซึ่งการวิเคราะห์ความคิดเห็นที่เป็นภาษาไทยเป็นเรื่องท้าทาย เนื่องจากข้อความแสดงความคิดเห็นส่วนใหญ่ที่อยู่บนอินเทอร์เน็ตนิยมใช้ภาษาที่มีโครงสร้างประโยคที่ไม่แน่นอน (Unstructured Data) หรือเป็นภาษาธรรมชาติ ไม่ถูกต้องตามหลักไวยากรณ์ทางภาษา ซึ่งภาษาไทยมีรูปแบบการเขียนคำยาวต่อกัน ไม่มีการเว้นวรรคระหว่างคำดังเช่นภาษาอังกฤษ ทำให้ยากต่อการวิเคราะห์ ทำให้ข้อความความคิดเห็นที่เป็นภาษาไทยนั้นจะต้องตัดประโยคออกเป็นคำก่อน

การวิเคราะห์เหมืองข้อความ (Text Mining) ซึ่งเป็นกระบวนการเพื่อสกัดเอาความรู้จากภาษาธรรมชาติที่มีลักษณะของข้อมูลแบบไม่มีโครงสร้าง อีกทั้งไม่มีการกำหนดรูปแบบไว้ล่วงหน้า ในการทำเหมืองข้อความมีนักวิจัยหลายท่านนำเทคนิคเหมืองข้อมูลมาประยุกต์ในการวิเคราะห์ความคิดเห็น ซึ่งเป็นอีกกระบวนการของการวิเคราะห์เหมืองข้อความ[4] โดยการนำเอาความคิดเห็นมาทำการวิเคราะห์เพื่อให้ทราบถึงความพึงพอใจที่มีต่อสิ่งนั้นๆ และการสกัดคำตามคุณลักษณะที่แตกต่างกันก็อาจได้มาซึ่งประโยชน์เพื่อการพิจารณาวิเคราะห์ที่หลากหลายและแม่นยำมากขึ้น เช่น ประพัฒน์ พรมน้ำอ่างและคณะ[5] ได้ทำการแบ่งกลุ่มข้อความจากข้อความรีวิว โดยใช้เทคนิคเหมืองข้อมูล ซึ่งประกอบด้วยเทคนิค SVM เทคนิค Decision Tree เทคนิค k-NN และเทคนิค Naïve Bayes จากการทดลองพบว่าโดยเทคนิค SVM ได้ค่าความถูกต้องสูงที่สุดอยู่ที่ 86.26% ส่วนพัชรนิกานต์ พงษ์ธนู และคณะ[6] นำเสนอการวิเคราะห์เหมืองข้อความจากการเก็บข้อมูลการแสดงความพึงพอใจของลูกค้าบนเว็บไซต์เพื่อหาแนวทางในการปรับปรุงการบริการของเว็บไซต์ผู้ให้บริการโรงแรมให้มีประสิทธิภาพมากขึ้น โดยใช้เทคนิควิธีต้นไม้ตัดสินใจ ID3 และเทคนิค Naive Bayes จากการทดลองพบว่าเทคนิคต้นไม้ตัดสินใจให้ค่าเฉลี่ยมากที่สุดที่ 95.50% K. Gowtham Reddy และ Jagadeesh Gopal[7] ได้นำเสนอการวิเคราะห์ความคิดเห็นเกี่ยวกับบทวิจารณ์เกมที่อยู่บน Twitter ด้วยเทคนิคการเรียนรู้ของเครื่อง โดยนำความคิดเห็นมาจำแนกว่าความคิดเห็นแต่ละความคิดเห็นเป็นความคิดเห็นเชิงบวกหรือเชิงลบ โดยเก็บข้อมูลบน Twitter จำนวน 1,200 ข้อความ มาจำแนกด้วยเทคนิค Support vector machine, Naïve Bayes และ Maximum Entropy ซึ่งจากผลการวิจัยพบว่าเทคนิค Maximum Entropy ให้ค่าความถูกต้องมากที่สุดที่ 90% Ha-Na Kang, Hye-Ryeon Yong, และ Hyun-Seok Hwang[8] ได้ศึกษาการวิเคราะห์ข้อมูลรีวิวเกมออนไลน์บนชุมชน STEAM โดยใช้เทคนิคเหมืองมูลมาจำแนกว่ารีวิวใดมีประโยชน์บ้าง ซึ่งใช้ข้อมูลรีวิวเกมจำนวน 79,437 ข้อความจากทั้งหมด 11 เกม โดยใช้เทคนิคต้นไม้ตัดสินใจ CART เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks) และใช้หลักการการวัดประสิทธิภาพแบบ 10-fold cross validation จากผลการทดลองพบว่าเทคนิคต้นไม้ตัดสินใจ

CART มีประสิทธิภาพดีกว่า ส่วน Rohan Bais, Pasal Odek และ Seyla Ou[9] ได้ศึกษาและทดสอบการจำแนกความคิดเห็นของรีวิวเกมบน STEAM ซึ่งใช้ข้อมูลจำนวน 5,000 ข้อความจากเกมทั้งหมด 100 เกม โดยแบ่งความคิดเห็นเป็น 2 กลุ่มคือ ความคิดเห็นเชิงบวกและเชิงลบ โดยใช้เทคนิค Naïve Bayes เทคนิค SVM และเทคนิค Logistic Regression และใช้หลักการการวัดประสิทธิภาพแบบ 10-fold cross validation โดยแบ่งเป็นชุดข้อมูลฝึก 70% ชุดข้อมูลทดสอบ 30% จากผลการทดลองพบว่าเทคนิค SVM ให้ความถูกต้องมากที่สุดที่ 93.50% แต่เทคนิคเหล่านี้ไม่สามารถสร้างแบบจำลองที่มีประสิทธิภาพได้ถ้ามีความไม่สมดุลของข้อมูล[10] โดยมีนักวิจัยหลายท่านได้ทำการแก้ปัญหาโดยใช้เทคนิคการสังเคราะห์ข้อมูลเพิ่มของข้อมูล SMOTE (Synthetic Minority Oversampling Technique:) เช่น งานวิจัยของ เชาวนันท์ โสโท[11] ได้สร้างแบบจำลองการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกด้วยโครงข่ายประสาทเทียม และใช้ SMOTE มาทำการปรับสมดุลของข้อมูลแล้วทำการสร้างแบบจำลอง ผลการทดลองโครงข่ายประสาทเทียมที่มีการปรับความสมดุลของข้อมูลด้วยวิธีการ SMOTE มีประสิทธิภาพในการทำนายที่ดีกว่าวิธีอื่นโดยมีค่า ความถูกต้องเท่ากับ 81.70% ค่าความไวเท่ากับ 94.47% และค่าความจำเพาะเท่ากับ 55.47%

ดังนั้น งานวิจัยนี้จึงได้ทำการพัฒนาแบบจำลองในการวิเคราะห์ความคิดเห็นต่อเกมมือถือผับจีด้วยเหมืองข้อความ โดยการนำข้อความความคิดเห็นของผู้เล่นเกมมือถือผับจีมาจำแนกเป็น 2 กลุ่ม คือ ความคิดเห็นเชิงบวกและความคิดเห็นเชิงลบ และปรับความไม่สมดุลของข้อมูลด้วยเทคนิค SMOTE เปรียบเทียบเทคนิคที่ใช้ในการสร้างแบบจำลองด้วยเทคนิค Naïve Bayes เทคนิค SVM (Support Vector Machine) เทคนิค K-Nearest Neighbor เทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิค Random Forest ในการวัดประสิทธิภาพของแบบจำลองผู้วิจัยจะใช้หลักการ 10-fold cross validation ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ และวัดประสิทธิภาพของแบบจำลองและเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยค่า ความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถูกต้อง (Accuracy)

ทฤษฎีและเทคนิคที่เกี่ยวข้อง

1. ความหมายของเกม

เกม[12] หมายถึงการเล่นที่ผู้เล่นอยู่ภายใต้เงื่อนไขการควบคุมและมีจุดเริ่มต้นและจุดจบที่ชัดเจน ในเงื่อนไขที่จำกัดและอิสระ การเล่นเกมเป็นการมุ่งใช้ความสามารถในการแสดงความเหนือกว่าเพื่อเอาชนะหรือแข่งขันบนเงื่อนไขข้อจำกัดต่าง ๆ

2. PlayerUnknown's: Battlegrounds (PUBG)

เป็นเกมเอาชีวิตรอด (Survival Game) แบบออนไลน์หลายผู้เล่น (Online Multiplayer) จัดอยู่ในประเภทเกมแอคชั่น (Action Game) พัฒนาและเผยแพร่โดย PUBG Corporation ซึ่งเป็นบริษัทย่อยของ Bluehole บริษัทวิดีโอเกมของเกาหลีใต้ เกมดังกล่าวมีพื้นฐานมาจาก mods ก่อนหน้านี้ที่สร้างขึ้นโดย Brendan Greene ผู้ใช้ชื่อในโลกอินเทอร์เน็ตว่า "PlayerUnknown" ซึ่งได้แรงบันดาลใจจากภาพยนตร์เรื่อง Battle Royale ของญี่ปุ่นปี 2000 และขยายเป็นเกมเดี่ยว (Standalone) ภายใต้ทิศทางการสร้างสรรค์ของ Greene ในเกม PUBG จะมีผู้เล่นถึง 100 คนกระโดดลงมาจากเครื่องบินพร้อมร่มชูชีพลงบนแผนที่ที่เป็นเกาะร้าง และวิ่งหาอาวุธและอุปกรณ์เพื่อฆ่าคนอื่นในขณะที่ต้องหลีกเลี่ยงไม่ให้ตัวเองถูกฆ่าตาย พื้นที่ปลอดภัย (Safe Zone) ที่มีอยู่ในแผนที่ของเกมจะลดขนาดลงเรื่อย ๆ เมื่อเวลาผ่านไป ซึ่งจะบีบผู้เล่นที่รอดชีวิตมาสู่พื้นที่ที่จำกัดมากขึ้นเพื่อบังคับให้เผชิญหน้ากัน ผู้เล่นหรือทีมใดอยู่รอดเป็นทีมสุดท้ายก็คือผู้ชนะ[13]

3. เทคนิค Naïve Bayes

มีพื้นฐานมาจากกฎของเบย์[14] เป็นทฤษฎีทางด้านสถิติโดยนำความน่าจะเป็นมาใช้ประเมินความไม่แน่นอนให้เป็นตัวเลข กล่าวถึงความน่าจะเป็นของเหตุการณ์ที่เกิดขึ้น (A) ถ้ามีเหตุการณ์อีกเหตุการณ์หนึ่งเกิดมาแล้ว (B) สามารถเขียนให้อยู่ในรูปอย่างง่าย ดังสมการที่ 1

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

P(A|B) คือ ความน่าจะเป็นที่เหตุการณ์ A จะเกิดขึ้นถ้าเหตุการณ์ B เกิดขึ้นแล้ว

P(B|A) คือ ความน่าจะเป็นที่เหตุการณ์ B จะเกิดขึ้นถ้าเหตุการณ์ A เกิดขึ้นแล้ว

P(A) คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ A

P(B) คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ B

4. เทคนิค Support Vector Machines

เป็นเทคนิคที่สามารถนำมาช่วยแก้ปัญหาการจำแนกข้อมูล ใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูล[15] โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้โดยเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูลได้ดีที่สุด แนวความคิดของเทคนิควิธี SVM นั้นเกิดจากการที่นำค่าของกลุ่มข้อมูลมาวางลงในฟีเจอร์สเปซ จากนั้นจึงหาเส้นที่ใช้แบ่งข้อมูล

ทั้งสองออกจากกัน โดยจะทำการสร้างเส้นแบ่งที่เป็นเส้นตรงขึ้นมา เพื่อให้ทราบว่าเส้นตรงที่แบ่งกลุ่มสองกลุ่มออกจากกันนั้น เส้นใดเป็นเส้นที่ดีที่สุดสำหรับ SVM นั้นเดิมได้มีการนำมาใช้กับข้อมูลที่เป็นเชิงเส้นแต่ในความเป็นจริงแล้วข้อมูลที่นำมาใช้ในระบบการสอนให้ระบบเรียนรู้ส่วนใหญ่มักเป็นข้อมูลแบบไม่เป็นเชิงเส้น ซึ่งสามารถแก้ปัญหาดังกล่าวด้วยการนำ Kernal Function มาใช้ การจำแนกข้อมูลบนระนาบหลายมิติจะใช้ส่วนการเลือกที่มีความเหมาะสมที่สุดเรียกว่า โครงสร้างในการคัดเลือกซึ่งโครงสร้างในการคัดเลือกมาจากข้อมูลที่สอนให้ระบบเรียนรู้จำนวนเซตของโครงสร้างที่ใช้อธิบายในกรณีหนึ่ง เรียกว่า เวกเตอร์ ดังนั้นจุดมุ่งหมายของตัวแบบ SVM คือ แบ่งแยกกลุ่มของเวกเตอร์ในกรณีนี้ด้วยหนึ่งกลุ่มของตัวแปรของเป้าหมายที่อยู่ข้างหนึ่งของระนาบและกรณีของกลุ่มอื่นที่อยู่ทางระนาบต่างกัน ซึ่งเวกเตอร์ที่อยู่ข้างระนาบหลายมิติทั้งหมดเรียกว่า ซัพพอร์ตเวกเตอร์ ซึ่งวิธีการนี้เหมาะสำหรับข้อมูลที่มีมิติของข้อมูลสูง

5. เทคนิค K-Nearest Neighbor

เป็นวิธีการในการจัดแบ่งคลาส โดยเทคนิคนี้จะตัดสินใจว่า คลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง หลักการของวิธีการนี้[16] จะจำแนกประเภทข้อมูลโดยขึ้นกับข้อมูลที่มีคุณสมบัติใกล้เคียงกันมากที่สุด k ตัวจากข้อมูลบนชุดข้อมูลตัวอย่าง ทำงานโดยขึ้นกับระยะทางน้อยสุดจากสมาชิกใหม่หรือข้อมูลที่ป้อนถาม (Input Query Instance) กับข้อมูลตัวอย่างฝึกฝนจะคำนวณหาเพื่อนบ้านที่ใกล้ที่สุด k ตัว หลังจากนั้นจะรวบรวมสมาชิกที่ใกล้เคียงที่สุด k ตัวแล้วเลือกคลาสที่สมาชิกส่วนใหญ่ที่ในกลุ่ม k ดังกล่าวสังกัดอยู่มากที่สุดให้กับสมาชิกใหม่

6. เทคนิคต้นไม้ตัดสินใจ C4.5

C4.5[15] เป็นเทคนิคในการสร้างต้นไม้ การตัดสินใจพัฒนาโดย J. Ross Quinlan ในปี 1993 โดยนำเอา ID3 มาปรับปรุงให้มีความสามารถมากขึ้นใช้วิธีการ Information Gain เพิ่มเติมการจัดการกับข้อมูล, ตัวเลข, ข้อมูลที่ขาดไปและไม่สมบูรณ์ และการ Prune ด้วยการแทนกิ่ง (Branch) ที่ไม่ช่วยในการตัดสินใจด้วย Leaf Node ที่ตัดสินใจได้ดีกว่า การแบ่งของ tree ในการทำงานขั้นตอนแรกคล้ายกับการทำงานด้วย ID3 คือต้องหา Info และ Gain ออกมาก่อน

7. เทคนิค Random Forest

เป็นเทคนิคการสุ่มเลือกใช้ข้อมูลและคุณลักษณะ Decision Tree[17] ซึ่งถูกสร้างจากการนำข้อมูลไปสุ่มเลือกตัวอย่างแบบเลือกแล้วใส่กลับ(Sampling with Replacement) แล้วนำมาสร้างเป็น Tree ซึ่งจะมีตัวอย่างส่วนหนึ่งที่ไม่ถูกเลือก ซึ่งข้อมูลส่วนนี้เรียกว่าOut-of-Bag (OOB) จะถูกนำมาใช้ในการทดสอบ Decision Tree วิธีการดังกล่าวนี้เรียกว่า Bagging ผลลัพธ์ที่ได้อย่างอิสระจาก Decision Tree ในแต่ละต้นถูกนำมาคิดเป็นผลการโหวต ผลโหวตที่มากที่สุดจะใช้ระบุสถานะของคลาส เทคนิค Random Forest ไม่จำเป็นต้องมีข้อมูลทดสอบ เพื่อประมาณความผิดพลาดเพราะข้อมูล OOB นั้นถูกนำมาใช้ทดสอบ Decision Tree นั้นแล้ว

8. Synthetic Minority Over-sampling Technique

Synthetic Minority Over-sampling Technique (SMOTE)[10] เป็นเทคนิคในการสุ่มตัวอย่างของคลาสที่น้อย เพื่อแก้ปัญหาชุดข้อมูลที่มีคลาสไม่สมดุล ซึ่งข้อมูลมีจำนวนตัวอย่างแตกต่างกันมากในแต่ละคลาส เมื่อทำการจำแนกประเภท จะทำให้มีการเรียนรู้แต่ข้อมูลกลุ่มที่มาก ผลที่ได้ก็จะจำแนกไปในข้อมูลที่มีกลุ่มมาก วิธี SMOTE เป็นวิธีการเพิ่มจำนวนข้อมูลประเภทที่มีข้อมูลน้อย ให้เพิ่มปริมาณข้อมูลใกล้เคียงกับประเภทที่มีมากที่สุด โดยสุ่มค่าขึ้นมาหนึ่งค่า และหาระยะห่างระหว่างค่าที่เลือกกับทุกๆ ค่า แล้วเลือกค่าที่ใกล้เคียงที่สุด เช่น กำหนดไว้ 5 ค่า สุ่มค่าจากที่เลือก 1 ใน 5 หาค่าอยู่ระหว่างค่าที่เลือกตอนแรกและค่าที่สุ่มมาตอนหลัง เพื่อนำค่าที่ได้มาเพิ่มจำนวนข้อมูล ดังสมการที่ 2

$$X_{new} = X_i + (\hat{X_i} - X_i) \times \delta \quad (2)$$

X_{new} คือ ข้อมูลใหม่

X_i คือ ข้อมูลที่สุ่มในตอนแรก

X_i คือ ข้อมูลที่สุ่มมาเพิ่ม

δ คือ ค่าสุ่มตั้งแต่ 0-1

วิธีดำเนินการวิจัย

ในงานวิจัยนี้มีกระบวนการในการทำเหมืองข้อความ 4 ขั้นตอนดังนี้

1. การเก็บรวบรวมข้อมูล

เก็บรวบรวมข้อมูล ตั้งแต่วันที่ 1 กุมภาพันธ์ 2562 ถึงวันที่ 15 กุมภาพันธ์ 2562 บน Google Play จำนวน 3,798 ความคิดเห็น โดยใช้โปรแกรม WebHarvy ในการดึงข้อมูลความคิดเห็นออกมา ดัง Table 1

Table 1 Examples of reviews

No	User	Date	Review
1	JAMEMY CHA	1/2/2019	ชอบมากเลยครับ พัฒนาต่อไปๆ
2	Song Kung	1/2/2019	ข้าวมันไก่เจ๊แหม่มอร่อยมากครับ
3	ผู้ใช้ Google	1/2/2019	เกมดีมากแต่เวลาโหลดโหลดยากมาก
4	ยา สาม	1/2/2019	สวดยอดเลยล้วกเพ้ เป็นอิกหนึ่งเกมที่เล่นสนุกมาก
5	ผู้ใช้ Google	1/2/2019	ชอบอยุนะแต่กระตุกไปหน่อย
6	ผู้ใช้ Google	1/2/2019	สนุกเพื่นดีคะบ่างทีแถบลืมกินเข่าเลยคะยุดไม่ได้
7	ผู้ใช้ Google	1/2/2019	เกมpingมันมากมาก
8	ผู้ใช้ Google	1/2/2019	ดีมากๆๆชอบเกมนี้เป็นอันดับ1ของใจ
9	ผู้ใช้ Google	1/2/2019	ดีมากๆเลยครับ ขอบคุณสำหรับเกมดีๆครับ
10	ผู้ใช้ Google	1/2/2019	เกมนี้ดีมากคับเออฟเฟคสวยคัยไม่ได้เร้นพลาดมาก

2. การเตรียมข้อมูล

1) ตรวจสอบความถูกต้องของข้อมูล (Data Cleaning) เป็นการนำเอาข้อความความคิดเห็นที่ได้เก็บรวบรวมมาแก้ไขคำที่ผิดให้ถูกต้องตามพจนานุกรม จากนั้นกำจัดตัวเลข ตัวอักษรภาษาอังกฤษ สัญลักษณ์ และช่องว่าง จากนั้นทำการคัดแยกข้อคิดเห็นที่เกี่ยวข้องกับหัวข้อที่กำหนดและจะต้องเป็นข้อความภาษาไทยเท่านั้น โดยหลังจากผ่านกระบวนการนี้จะเหลือข้อความความคิดเห็นจำนวน 3,123 ความคิดเห็น

2) การตัดคำและกำจัดคำหยุด หลังจากผ่านตรวจสอบความถูกต้องของข้อมูลแล้วจะนำข้อมูลที่ได้มาทำการตัดคำและกำจัดคำหยุดออกไปดังแสดงใน Table 2 และ Table 3

Table 2 Examples of words segmentation

No.	Review
1	ชอบ, มาก, เลย, ครับ, พัฒนา, ต่อไป
2	เกม, ดีมาก, แต่, ดาวน์โหลด, ยาก, มาก
3	สุดยอด, เลย, ลูกพี่, เป็น, อีก, หนึ่ง, เกม, ที่, สนุก
4	ชอบ, อยู่, นะ, แต่, กระตุก, ไป, หน่อย
5	สนุก, เพลิน, ดี, ค่ะ, บางที, แทบ, ลืม, กินข้าว

Table 3 Examples of removing stop-words

No.	Review
1	ชอบ พัฒนา ต่อไป
2	เกม ดีมาก เวลา ดาวน์โหลด
3	สุดยอด เกม เล่น สนุก
4	ชอบ กระตุก
5	สนุก เพลิน ดี แทบ ลืม กินข้าว ไม่ได้

3) การเลือกคำบ่งชี้ หลังจากผ่านการตัดคำและกำจัดคำหยุดแล้วจะได้คำทั้งหมดจำนวน 1,212 คำ จากนั้นจะนำคำแต่ละคำมาให้ประเภทของคำตามพจนานุกรม โดยใช้พจนานุกรมอิเล็กทรอนิกส์ฉบับราชบัณฑิตยสถาน ซึ่งพัฒนาโดย NECTEC ดังแสดงใน Table 4

Table 4 Examples of word types

Rows	word	type	notice
1	เกม	n	คำนาม
2	สนุก	adv	คำวิเศษณ์
3	เล่น	v	คำกริยา
4	ดี	adv	คำวิเศษณ์
5	ไม่ได้	adv	คำวิเศษณ์
6	ชอบ	v	คำกริยา
7	กระตุก	v	คำกริยา
8	ผม	pron	คำสรรพนาม
9	ดีมาก	adv	คำวิเศษณ์
10	มันส์	adv	สแลง

จากนั้นเลือกคำวิเศษณ์ และเลือกคำสแลงที่มีความหมายของคำเป็นคำวิเศษณ์ จะได้ทั้งหมด 208 คำ แล้วทำการแบ่งคำบ่งชี้คุณลักษณะออกมาโดยผู้เชี่ยวชาญ จะได้คำบ่งชี้คุณลักษณะเชิงบวกจำนวน 70 คำ คำบ่งชี้คุณลักษณะเชิงลบจำนวน 57 คำ และคำที่ไม่บ่งชี้คุณลักษณะใดเลยจำนวน 81 คำ

4) วิธีการสร้างตัวแทนเอกสาร จะใช้วิธีการนำคำบ่งชี้คุณลักษณะในชุดข้อมูลมาเรียงกันเพื่อทำการนับความถี่ของการเกิดขึ้นของคำนั้นๆ จากนั้นจึงนำค่าจำนวนความถี่ของคำมาสร้างเวกเตอร์ตัวแทนเอกสาร และคำบ่งชี้ที่ไม่ปรากฏบนเอกสารจะมีค่าเป็น 0 จากนั้นนับจำนวนความถี่ของค่าคุณลักษณะในแต่ละคุณลักษณะว่ามีจำนวนเท่าใด เพื่อนำค่าความถี่ในการใช้คำของแต่ละคุณลักษณะมาทำการเปรียบเทียบกันโดย

1. จำนวนความถี่ของคุณลักษณะความคิดเห็นเชิงบวกมากกว่าความถี่ของคุณลักษณะเชิงลบให้ตัวแปรตามเป็นความคิดเห็นเชิงบวกแทนด้วย P

2. จำนวนความถี่ของคุณลักษณะความคิดเห็นเชิงลบมากกว่าความถี่ของคุณลักษณะเชิงบวกให้ตัวแปรตามเป็น ความคิดเห็นเชิงลบ แทนด้วย N

3. จำนวนความถี่ของคุณลักษณะความคิดเห็นเชิงบวกและความถี่ของคุณลักษณะเชิงลบเท่ากันให้ตัวแปรตามเป็น ความคิดเห็นเป็นกลาง แทนด้วย B

Table 5 Examples of defined a class

Rows	ฉิบหาย	ดี	ดีมาก	บ่า	มันส์	...	ไม่ทัน	P	N	Class
1	0	1	0	0	1	...	0	3	1	P
2	0	0	0	0	0	...	0	1	3	N
3	0	0	0	0	0	...	0	1	1	B
4	0	1	0	0	0	...	0	2	1	P
5	0	1	0	1	1	...	0	2	1	P
6	1	2	0	0	0	...	0	4	2	P
7	0	0	0	0	0	...	0	1	1	B
8	0	1	0	0	0	...	0	1	1	B
9	0	0	1	0	0	...	1	2	4	N
10	0	0	2	0	0	...	0	2	1	P

โดยจะเลือกเพียงชุดข้อมูลที่เป็นคลาส P และ N ซึ่งมีจำนวนทั้งหมด 1,875 ข้อมูล แบ่งเป็นคลาส P จำนวน 1,576 ข้อมูล และคลาส N จำนวน 299 ข้อมูล

5) ปรับความไม่สมดุลของข้อมูล ปรับสมดุลของข้อมูลด้วยวิธีสังเคราะห์ข้อมูลใหม่ (SMOTE) ซึ่งเป็นวิธีการเพิ่มคลาสที่มีกลุ่มน้อยให้เพิ่มมากขึ้นเพื่อให้ใกล้เคียงหรือเท่ากันกับอีกคลาส หลังจากการปรับความสมดุลแล้วมีจำนวนข้อมูลเพิ่มขึ้น 1,277 ชุดข้อมูล ดังนั้นจะมีจำนวนข้อมูลทั้งสิ้นก่อนการสร้างแบบจำลองจำนวน 3,152 ข้อมูล โดยแบ่งเป็นคลาส P จำนวน 1,576 ชุดข้อมูลและคลาส N จำนวน 1,576 ชุดข้อมูล

3. การสร้างแบบจำลอง

งานวิจัยนี้ได้ใช้เครื่องมือคือโปรแกรม RapidMiner Studio เวอร์ชั่น 9.6.0 และสร้างแบบจำลองด้วยเทคนิค Naïve Bayes เทคนิค Support Vector Machine เทคนิค K-Nearest Neighbor เทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิค Random Forest ซึ่งใช้ค่าพารามิเตอร์ของแต่ละแบบจำลอง โดย Naïve Bayes ไม่มีพารามิเตอร์ Support Vector Machine มีพารามิเตอร์ kernel=polynomial, degree=1.0, C=0.0 K-Nearest Neighbor มีพารามิเตอร์ K=3 ต้นไม้ตัดสินใจ C4.5 มีพารามิเตอร์ maximal depth=40 และ Random Forest มีพารามิเตอร์ number of trees=100, maximal depth=40

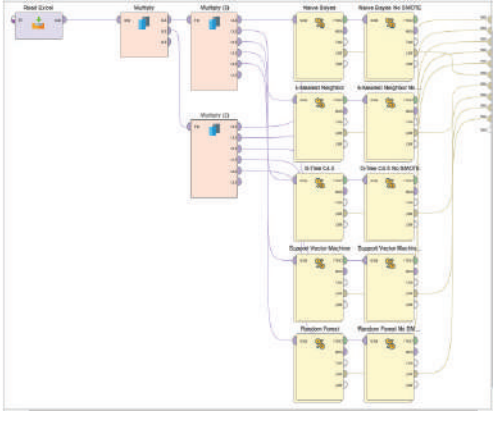

Figure 1 Modeling on RapidMiner Studio

4. การวัดประสิทธิภาพของแบบจำลอง

ในการประเมินผลของแบบจำลอง ได้มีการใช้เทคนิคการวัดประสิทธิภาพแบบ 10-Fold cross validation โดยแบ่งออกเป็นทั้งหมด 10 กลุ่ม ทั้งนี้จะแบ่งกลุ่มข้อมูลเพื่อใช้เป็นข้อมูลทดสอบ (Test data) 1 ชุด และที่เหลือจะเป็นข้อมูลฝึก (Training data) ซึ่งคิดเป็นอัตราข้อมูลทดสอบต่อปริมาณข้อมูลฝึก คิดเป็นอัตราร้อยละ 10:90 ซึ่งค่าที่ใช้ในการวัดประสิทธิภาพของแบบจำลองได้แก่ ค่าความแม่นยำ (Precision) คือ การวัดความสามารถในการที่จะขจัดเอกสารที่ไม่เกี่ยวข้องออกไปโดยที่ค่าความแม่นยำนั้นจะเป็นอัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและได้มีการถูกดึงออกมา เพื่อเทียบกับจำนวนเอกสารที่ถูกดึงออกมาทั้งหมด ค่าความระลึก (Recall) คือ การวัดความสามารถของระบบในการดึงเอกสารที่เกี่ยวข้องออกมา โดยค่าความระลึกนั้นจะใช้อัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและได้มีการถูกดึงออกมา เทียบกับจำนวนเอกสารที่เกี่ยวข้องทั้งหมด และค่าความถูกต้อง (Accuracy) คือ การวัดผลจากผลลัพธ์ของการเรียนรู้ในการทำนายกลุ่มตัวอย่างชุดใหม่ได้อย่างถูกต้อง โดยมี Confusion Matrix ดัง Table 6

Table 6 Confusion Matrix

	Class	Actual Class	
		Negative	Positive
Prediction Class	Negative	TN	FP
	Positive	FN	TP

และสามารถคำนวนได้ดังสมการที่ 3, 4 และ 5 ตามลำดับ

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

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

โดย

TP คือ จำนวนข้อมูลที่ถูกดึงออกมาอย่างถูกต้อง

TN คือ จำนวนข้อมูลที่ถูกต้องแต่ไม่ถูกดึงออกมา

FP คือ จำนวนข้อมูลที่ผิดพลาดที่ถูกดึงออกมา

FN คือ จำนวนข้อมูลที่ผิดพลาดแต่ไม่ถูกดึงออกมา

ผลการวิจัย

งานวิจัยนี้ได้ทำการวิเคราะห์ความคิดเห็นของผู้เล่นที่มีต่อเกมมือถือผับจี โดยนำเทคนิคของการทำเหมืองข้อความมาใช้ในการวิเคราะห์ทั้งหมด 5 เทคนิค ได้แก่ เทคนิค Naïve Bayes เทคนิค Support Vector Machine เทคนิค K-Nearest Neighbor และเทคนิคต้นไม้ตัดสินใจ C4.5 ผ่านการใช้งานโปรแกรม RapidMiner Studio เวอร์ชั่น 9.6.0 โดยข้อมูลที่ใช้ในงานวิจัยนี้เก็บรวบรวมจาก Google Play ตั้งแต่วันที่ 1-15 กุมภาพันธ์ 2562 จำนวน 3,798 ชุดข้อมูล ซึ่งหลังจากผ่านกระบวนการต่าง ๆ แล้วจะเหลือชุดข้อมูลที่เตรียมเข้าสู่โมเดลทั้ง 5 จำนวน 1,875 ชุดข้อมูล แบ่งเป็นคลาส P จำนวน 1,576 ข้อมูล และคลาส N จำนวน 299 ข้อมูล ซึ่งทั้งสองคลาสมีความไม่สมดุลกันของข้อมูล จากปัญหาดังกล่าว ผู้วิจัยจึงได้นำวิธีการ SMOTE[10] ซึ่งเป็นวิธีการในการแก้ปัญหาข้อมูลไม่สมดุล (Imbalanced data) วิธีการนี้เป็นการเพิ่มคลาสที่มีกลุ่มน้อยให้เพิ่มมากขึ้นเพื่อให้ใกล้เคียงกันกับอีกคลาส โดยวิธีนี้จะสุ่มเพิ่มชุดข้อมูลของคลาสจำนวนน้อย และสร้างแบบจำลองด้วยเทคนิค Naïve Bayes เทคนิค Support Vector Machine เทคนิค K-Nearest Neighbor เทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิค Random Forest โดยใช้ 10-Fold Cross Validation ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ และวัดประสิทธิภาพของแบบจำลองด้วยค่าความแม่นยำ ค่าความระลึก และค่าความถูกต้อง

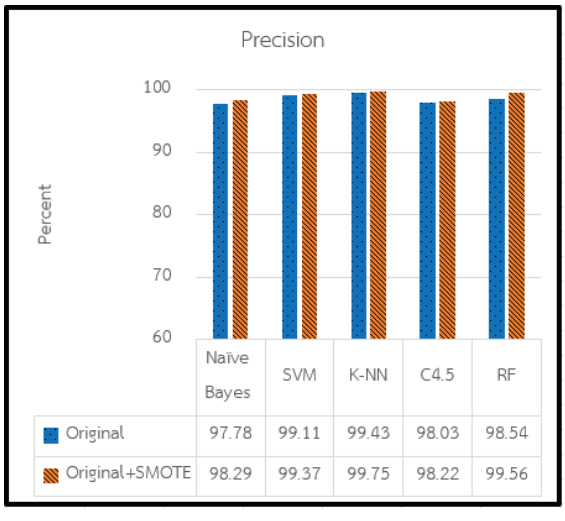

Figure 2 Comparison of precision

จาก Figure 2 จะเห็นได้ว่าหลังจากปรับความสมดุลของข้อมูลด้วย SMOTE แล้ว ทุกเทคนิคมีค่าความแม่นยำเพิ่มขึ้น โดยเฉพาะเทคนิค K-Nearest Neighbor มีค่าความแม่นยำสูงสุด 99.75% รองลงมาคือเทคนิค Random Forest 99.56% ตามมาด้วยเทคนิค Support Vector Machine 99.37% ตามด้วยเทคนิค Naïve Bayes 98.29% และเทคนิคต้นไม้ตัดสินใจ C4.5 99.22% ตามลำดับ

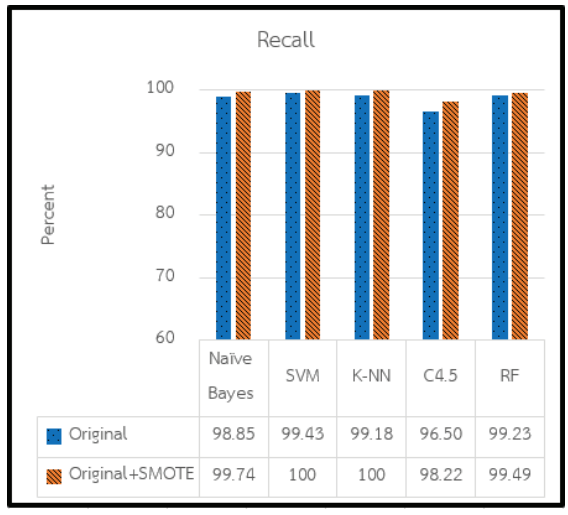

Figure 3 Comparison of recall

จาก Figure 3 จะเห็นได้ว่าหลังจากปรับความสมดุลของข้อมูลด้วย SMOTE แล้ว ทุกเทคนิคมีค่าความระลึกเพิ่มขึ้น โดยเฉพาะเทคนิค Support Vector Machine และเทคนิค K-Nearest Neighbor มีค่าความระลึกสูงสุดเท่ากัน 100% รองลงมาคือเทคนิค Naïve Bayes 99.74% ตามด้วยเทคนิค Random Forest 99.49% และเทคนิคต้นไม้ตัดสินใจ C4.5 98.22% ตามลำดับ

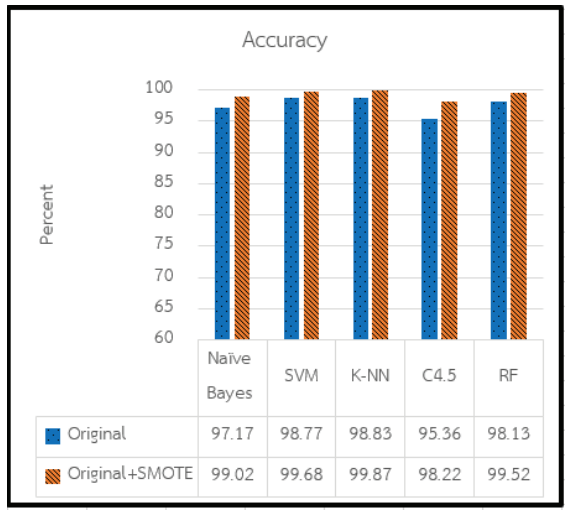

Figure 4 Comparison of accuracy

จาก Figure 4 แสดงการเปรียบเทียบค่าความถูกต้องของ 5 เทคนิคที่นำมาใช้ในการสร้างแบบจำลอง ประกอบด้วยเทคนิค Naïve Bayes เทคนิค Support Vector Machine เทคนิค K-Nearest Neighbor เทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิค Random Forest จากผลการทดลองพบว่าหลังจากปรับความสมดุลของข้อมูลด้วย SMOTE แล้ว ทุกเทคนิคมีค่าความถูกต้องของแบบจำลองเพิ่มขึ้น โดยเทคนิค K-Nearest

Neighbor ค่าความถูกต้องสูงสุด 99.87% รองลงมาคือเทคนิค Support Vector Machine 99.68% ตามมาด้วยเทคนิค Random Forest 99.52% ตามด้วยเทคนิค Naïve Bayes 99.02% และเทคนิคต้นไม้ตัดสินใจ C4.5 98.22% ตามลำดับ

Table 7 Confusion Matrix for Non-SMOTE

Model	TP	FP	TN	FN
Naïve Bayes	1541	35	281	18
SVM	1562	14	290	299
K-NN	1567	9	286	13
C4.5	1545	31	243	56
Random Forest	1553	23	287	12

Table 8 Confusion Matrix for SMOTE

Model	TP	FP	TN	FN
Naïve Bayes	1549	27	1572	4
SVM	1566	10	1576	0
K-NN	1572	4	1576	0
C4.5	1548	28	1548	28
Random Forest	1569	7	1568	8

วิจารณ์และสรุป

งานวิจัยฉบับนี้ มีวัตถุประสงค์เพื่อค้นหาเทคนิคการทำเหมืองข้อความที่มีประสิทธิภาพในการวิเคราะห์ความคิดเห็นของผู้เล่นที่มีต่อเกมมือถือผับจีซึ่งถูกเขียนขึ้นด้วยภาษาไทย ซึ่งเก็บรวบรวมข้อความความคิดเห็นจาก Google Play ตั้งแต่วันที่ 1 กุมภาพันธ์-15 กุมภาพันธ์ 2562 จำนวนทั้งหมด 3,798 ข้อความ โดยงานวิจัยนี้ได้ทำการแบ่งคุณลักษณะออกเป็น 2 กลุ่ม คือ คุณลักษณะเชิงบวกและคุณลักษณะเชิงลบ โดยนำเอาคำวิเศษณ์และคำสแลงจากความคิดเห็นของผู้เล่นที่มีต่อเกมมือถือผับจี ซึ่งคำวิเศษณ์นี้สามารถแสดงถึงอารมณ์เชิงบวกและเชิงลบได้ดี[18] จากนั้นได้นำเอาเทคนิควิธีการวิเคราะห์เหมืองข้อความมาทำการวิเคราะห์ข้อความความคิดเห็นจำนวน 5 เทคนิค คือ เทคนิค Naïve Bayes เทคนิคSupport Vector Machine เทคนิค K-Nearest Neighbor เทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิค Random Forest

เนื่องจากชุดข้อมูลที่เตรียมเข้าสู่การสร้างแบบจำลองมีความไม่สมดุลกันของข้อมูล ผู้วิจัยจึงได้แก้ปัญหาโดยการนำวิธีการ SMOTE มาใช้เพื่อปรับให้ข้อมูลมีความสมดุลกัน และใช้หลักการ 10-Fold cross validation ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ และวัดประสิทธิภาพของแบบจำลองด้วยค่าความแม่นยำ ค่าความระลึก และค่าความถูกต้อง จากผลการวิจัยพบว่าวิธีการ SMOTE สามารถทำให้ประสิทธิภาพโดยรวมของแบบจำลองเพิ่มขึ้น โดยค่าความแม่นยำเพิ่มขึ้นเฉลี่ย 0.46% ค่าความระลึกเพิ่มขึ้นเฉลี่ย 0.85% และค่าความถูกต้องเพิ่มขึ้นเฉลี่ย 1.61% โดยเทคนิคที่มีประสิทธิภาพมากที่สุดคือเทคนิค K-Nearest Neighbor ที่ให้ค่าความแม่นยำ 99.75% ค่าความระลึก 100% และค่าความถูกต้อง 99.87%

กิตติกรรมประกาศ

ขอขอบคุณคณาจารย์คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ที่ให้คำปรึกษาและคำแนะนำต่างๆ ในการดำเนินงานวิจัยในครั้งนี้

เอกสารอ้างอิง

1. Mthai. PUBG Mobile: Bluehole ; 2019 [Available from: https://game.mthai.com/mobile-games/112880.html.
2. Zagal J, Ladd A, Johnson T. Characterizing and understanding game reviews 2009. 215-22 p.
3. ยืน ภู่วรวรรณ. การประมวลผลภาษาธรรมชาติ. กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี ; 1992. 241 p.
4. กานดา แผ่วัฒนากุล, ปราโมทย์ ลือนาม. การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์. บทความ วิชาการวารสารการจัดการสมัยใหม่ปีที่ 11. 2013 ; 11(2).
5. ประพัฒน์ พรมน้ำอ่าง, วสุวรรธน์ พงศ์ขจร, นิเวศ จิระวิชิตชัย. การจำแนกกลุ่มข้อความรีวิวโดยใช้เทคนิคเหมืองข้อมูล. วารสารวิทยาศาสตร์และเทคโนโลยี มหรธัญบุรี ปีที่: 6 2016 ; 1: 94-102.
6. พัชรนิกานต์ พงษ์ธนู. วิเคราะห์ความพึงพอใจของลูกค้าจากข้อความคำแนะนำโดยการทำเหมืองความคิดเห็น. โครงการประชุมวิชาการนานาชาติ Knowledge and Smart Technologies. 2012(1): 53-60
7. Reddy KG, Gopal J. Twitter sentiment analysis of game reviews using machine learning techniques. Journal of Chemical and Pharmaceutical Sciences. 2017 ; 10(1).
8. Kang H-N, Yong H-R, Hwang H-S. A Study of Analyzing on Online Game Reviews using a Data Mining Approach: STEAM Community Data. International Journal of Innovation, Management and Technology. 2017 ; 8(2): 90-4.

9. Bais R, Odek P, Ou S. Sentiment Classification on Steam Reviews. 2017.
10. Chawla N, Bowyer K, Hall L, Kegelmeyer W. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res (JAIR). 2002 ; 16: 321-57.
11. เชาวนันท์ โสโท. แบบจำลองการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกด้วยโครงข่ายประสาทเทียม. KKU Res J (GS). 2013 ; 13: 39-49.
12. สิริวรรณ ปัญญากาศ. ความสัมพันธ์ระหว่างความรุนแรงในสื่อเกมออนไลน์กับพฤติกรรมก้าวร้าว ของเด็กวัยรุ่นตอนปลายในจังหวัดเชียงใหม่. มหาวิทยาลัยเชียงใหม่ ; 2008.
13. Carter C. Understanding Playerunknown's Battlegrounds 2019 [Available from: https://www.polygon.com/playerunknowns-battlegrounds-guide/2017/6/9/15721366/pubg-how-to-play-blue-wall-white-red-circle-map-weapon-vehicle-inventory-air-drop.
14. Backfried G, Göllner J, Qirchmayr G, Rainer K, Kienast G, Thallinger G, *et al.*, editors. Integration of Media Sources for Situation Analysis in the Different Phases of Disaster Management: The QuOIMA Project. 2013 European Intelligence and Security Informatics Conference ; 2013 12-14 Aug. 2013.
15. Berzal F, Matín N. Data mining: concepts and techniques by Jiawei Han and Micheline Kamber. ACM SIGMOD Record. 2002 ; 31: 66-8.
16. Joachims T, editor Text categorization with Support Vector Machines: Learning with many relevant features. Machine Learning: ECML-98 ; 1998 1998// ; Berlin, Heidelberg: Springer Berlin Heidelberg.
17. Cutler A, Cutler D, Stevens J. Random Forests. 452011. p. 157-76.
18. Sun Y, Quan C, Kang X, Zhang Z, Ren F. Customer emotion detection by emotion expression analysis on adverbs. Information Technology and Management. 2015 ; 16(4): 303-11.