การจำแนกข้อความหลอกลวงและข้อความจริงจาก SMS

Classification of fraudulent and real message from SMS

พีท อ่อนทอง[1], นายอภิรักษ์ ดุลยเกษม[1]

อาจารย์ที่ปรึกษา : ดร. กรรณิกาณ์ หิรัญกสิ[2], ผศ.ดร. สิรักข์ แก้วจำนงค์[3], ผศ. วรรณภา พนิตสุภากมล[4]

[1]สาขาวิชาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร, [2]ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร, [3]ภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร, [4]ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ จำแนกข้อความหลอกลวงและข้อความจริงจากบริการข้อความสั้น (SMS) ที่ส่งมายังมือถือ โดยทำการเก็บรวบรวมข้อมูลจากโทรศัพท์มือถือที่มีการส่งข้อความสั้นเข้ามาจริง ๆ ซึ่งทำการเก็บรวบรวมทั้งหมด 400 ข้อความ โดยแบ่งเป็นข้อความจริงจำนวน 200 ข้อความ และข้อความหลอกลวงจำนวน 200 ข้อความ ทำการวิเคราะห์ข้อมูลด้วยกันทั้งหมด 4 วิธีการคือ Naïve Bayes, Random Forest, Long short-term memory และ Support vector machine เพื่อเปรียบเทียบว่าวิธีการใดที่สามารถจำแนกได้แม่นยำที่สุด โดยจากผลลัพธ์วิธีการของ Naïve Bayes ได้ค่าความถูกต้องสูงที่สุด ซึ่งมีค่าเท่ากับ 97%

คำสำคัญ : บริการข้อความสั้น, หลอกลวง, มิจฉาชีพ, ข้อความจริง

Abstract

This research purposes to classify fraudulent messages and real messages from short message service (SMS) sent to mobile phones. By collecting data from mobile phones that send short messages. A total of 400 messages are collected, divided into 200 real messages and 200 fake messages. Data are analyzed by 4 methods: Naïve Bayes, Random Forest, Long short-term memory and Support vector machine. Comparing each

method is shown to be classified the most accurately. We have found Naïve Bayes method is the highest accuracy (97%).

Keywords : Short message service, Fraudulent, Criminal, Real Message

1. บทนำ

การสื่อสาร คือ การแลกเปลี่ยนสารระหว่างบุคคลหนึ่งไปยังอีกบุคคลหนึ่ง ซึ่งเป็นก้าวแรกและก้าวสำคัญที่ทำให้เกิดการสื่อสารกันระหว่างมนุษย์คือการพูด ตลอดจนทักษะในการประดิษฐ์เครื่องมือที่ซับซ้อนต่อมาได้การพัฒนาเป็นการใช้สัญลักษณ์เพื่อสื่อความหมายโดยเห็นได้จากจารึกต่าง ๆ หรือแม้กระทั่งการอาศัยสัตว์เป็นผู้ส่งสาร หลังจากนั้นมนุษย์เริ่มเรียนรู้ประโยชน์จากการใช้ไฟฟ้าและเริ่มมีการเติบโตทางนวัตกรรมจึงเกิดเป็นที่มาของการสื่อสารของโทรเลข โทรสาร โทรศัพท์ วิทยุ เปรียบเสมือนเป็นพัฒนาการครั้งสำคัญจนกลายมาเป็นปัจจุบัน ที่คอมพิวเตอร์และเครือข่ายต่าง ๆ เข้ามามีบทบาทในการดำรงชีวิตมากขึ้น โดยมีระบบดิจิทัล เข้ามาแทนที่ ทำให้เกิดการสื่อสารในรูปแบบ คอมพิวเตอร์ อินเทอร์เน็ต จดหมายอิเล็กทรอนิกส์และพัฒนาต่อมาเป็นโทรศัพท์เคลื่อนที่ ตลอดจนเทคโนโลยีต่าง ๆ เข้ามามีบทบาทในชีวิตมากขึ้น ส่งผลให้มีพัฒนาการของการสื่อสารอยู่ในรูปแบบของ สมาร์ทโฟน SMS Chat Apps Social Network ต่าง ๆ ซึ่งวิวัฒนาการการสื่อสารของมนุษย์ถูกพัฒนาเพื่อทำให้เรียนรู้และเข้าใจกันมากขึ้น เมื่อเทคโนโลยีได้พัฒนาขึ้นสามารถช่วยให้ติดต่อสื่อสารระหว่างกันสามารถทำได้อย่างได้กว้างขวางขึ้น สอดคล้องกับขนาดสังคมของมนุษย์ที่มีขนาดใหญ่มากขึ้นเช่นกัน [1]

ในปัจจุบันเมื่อเทคโนโลยีมีความทันสมัยมากขึ้น ทำให้ง่ายต่อการเข้าถึงแหล่งข้อมูล ต่าง ๆ แต่อาจก่อให้เกิดภัยร้ายที่คาดไม่ถึงอีกมากมายเช่นกัน จากการจัดอันดับของหน่วยนวัตกรรมและเทคโนโลยีความมั่นคงปลอดภัยของมหาวิทยาลัยเชียงใหม่ ในปีพ.ศ. 2565 ได้ระบุถึง 3 อันดับภัยไซเบอร์ใกล้ตัวที่คนไทยถูกหลอกมากที่สุด ได้แก่ 1. มิจฉาชีพบน Social Media 2. อีเมลหลอกลวง (Phishing) 3. การขโมยข้อมูลส่วนบุคคล (Data Theft) [8] จากทั้ง 3 ข้อซึ่งล้วนเกี่ยวของกับการหลอกลวงภายใต้ความต้องการต่าง ๆ ของผู้กระทำความผิด นอกจากนี้ผู้วิจัยยังพบว่า ภัยจาก SMS ที่เป็นผลจากการพัฒนาเทคโนโลยีการสื่อสารของมนุษย์ โดยสถิติปีพ.ศ. 2564 ที่ผ่านมาจากผู้พัฒนาแอปพลิเคชันฮูส์คอล (Whoscall) พบว่ามีการใช้โทรศัพท์เพื่อหลอกลวงในประเทศไทยมากกว่า 6.4 ล้านครั้ง เพิ่มขึ้นถึง 270% จากปีก่อนหน้า มีข้อความ SMS หลอกลวง เพิ่มขึ้นถึง 57% การหลอกลวงด้วยวิธีการส่งข้อความมีมากขึ้นสาเหตุอาจเป็นเพราะมีต้นทุนที่ต่ำประกอบกับการเข้าถึงกลุ่มเป้าหมายหรือเหยื่ออยู่ในอัตราสูง ทำให้จำนวนข้อความหลอกลวงเพิ่มขึ้นทุก ๆ เดือน ตั้งแต่ปีพ.ศ. 2563 และเพิ่มขึ้นสูงสุดในปีพ.ศ. 2564 ซึ่งข้อความเหล่านี้อาจนำไปสู่การเข้าถึงข้อมูลส่วนบุคคล หรือการสูญเสียทรัพย์สินได้ [2]

ผู้วิจัยจึงมีความสนใจที่จะทำการศึกษาเรื่อง การจำแนกข้อความหลอกลวงและข้อความจริง เพื่อที่สามารถจำแนกข้อความหลอกลวงต่าง ๆ ที่ส่งผ่านเข้ามายังโทรศัพท์เคลื่อนที่ เพื่อป้องกันการตกเป็นเหยื่อของมิจฉาชีพโดยไม่ได้ตั้งตัว

1.1 วัตถุประสงค์

เพื่อศึกษาการจำแนกข้อความหลอกลวงและข้อความจริงจาก SMS จากการเรียนรู้ของเครื่อง ทั้ง 4 โมเดล ได้แก่ Naive Bayes, Random Forest, Support Vector Machine และ Long short-term memory

1.2 ขอบเขตงานวิจัย

ในการศึกษาการจำแนกข้อความหลอกลวงและข้อความจริง จะทำการศึกษาจำแนกข้อมูลจากคำนิยามของข้อความหลอกลวงและข้อความจริงในโทรศัพท์เคลื่อนที่ ด้วยการเรียนรู้ของเครื่อง เพื่อจำแนกข้อความ SMS ที่ส่งผ่านเข้ามายังโทรศัพท์เคลื่อนที่

ข้อมูลที่ใช้ภายในงานวิจัยนี้ได้รวบรวมมาจากการรวบรวมข้อความ SMS ผ่านโทรศัพท์เคลื่อนที่ของส่วนที่เป็นข้อความจริง 200 ข้อความ และข้อความหลอกลวง 200 ข้อความ รวมทั้งสิ้น 400 ข้อความที่จะนำมาใช้ในงานวิจัยนี้

โมเดลที่ใช้ในการวิเคราะห์ความสามารถในการจำแนกข้อความจริงและข้อความหลอกลวง มีทั้งหมดโมเดลทั้งหมด 4 โมเดล ได้แก่ Naive Bayes, Random Forest, Support Vector Machine และ Long short-term memory

2. ความรู้พื้นฐาน

2.1 นิยามศัพท์

ข้อความหลอกลวง (Defraud) คือ เป็นสแปมในรูปแบบที่ใช้บริการส่งข้อความ SMS เป็นสื่อกลาง [6] ข้อความหลอกลวงส่วนใหญ่ทำเพื่อการโฆษณาเชิงพาณิชย์ มักจะเป็นที่น่าสงสัย หรือเป็นบริการที่ก้ำกึ่งผิดกฎหมาย [7] โดยผู้ส่งอาจจะเสียค่าใช้จ่ายในการส่งไม่มากนัก แต่ค่าใช้จ่ายส่วนใหญ่จะตกอยู่กับผู้รับ SMS ถึงแม้บางครั้งจะเหมือนไม่อันตราย แต่ก็สร้างความน่ารำคาญใจแก่ผู้รับ

ข้อความจริง (Real) คือ ข้อความ SMS ที่ส่งมาโดยยึดถือความต้องการของผู้รับเป็นหลัก ข้อความประเภทนี้มักจะเป็นข้อความสำคัญที่เกี่ยวข้องกับผู้รับ โดยผู้รับอาจเสียค่าใช้จ่ายหรือไม่เสีย ทั้งนี้ขึ้นอยู่กับความต้องการของผู้รับเอง

2.2 พื้นฐานศัพท์

อักขระพิเศษ หมายถึงอักขระที่เป็นเครื่องหมายต่าง ๆ เช่น + - * / $ = , " " ฯลฯ รวมทั้งสระ วรรณยุกต์ ในภาษาไทย บางทีเรียกรวม ๆ ว่า เครื่องหมายวรรคตอน นอกจากนั้นยังหมายรวมถึงตัวอักขระที่ไม่สามารถกดได้จากแป้นพิมพ์ แต่อาจใช้เทคนิคบางอย่างทำให้พิมพ์ออกมาได้

การตัดคำ (Tokenize) ในปัจจุบันข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) มีจำนวนมากขึ้น เช่น รูปภาพใน Facebook หรือ Instagram (IG) หรือ ข้อความต่าง ๆ ไม่ว่าจะเป็นในเว็บไซต์ต่าง ๆ ตลอดจนแม้กระทั่ง บริการข้อความ SMS แต่ข้อมูลเหล่านี้ยังไม่สามารถนำไปใช้วิเคราะห์ได้โดยตรงเนื่องจากไม่ได้อยู่ในรูปแบบของตาราง ดังนั้นจึงต้องมีขั้นตอนในการแปลงข้อความเหล่านี้ให้เป็นตารางเสียก่อน ซึ่งในที่นี้การตัดคำเป็นขั้นตอนการตัดข้อความหรือประโยคต่าง ๆ ออกเป็นคำ (word) ในงานวิจัยนี้ทำการวิจัยด้วยภาษา Python โดยใช้ ไลบรารีในการตัดคำ 2 แบบคือ ไลบรารีที่ช่วยในการตัดคำภาษาไทย คือThai Natural Language Processing in Python (PyThaiNLP) และไลบรารีที่ช่วยในการตัดคำภาษาอังกฤษ คือ Natural Language Toolkit (Nltk)

2.3 งานวิจัยที่วิเคราะห์เกี่ยวกับบริการข้อความ SMS

H. Jain และ R. K. Maurya [3] ได้ทำการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอลิทึมต่าง ๆ จากการตรวจจับ SMS Spam โดยได้เก็บรวมรวมข้อมูลจาก Kaggle จำนวน 5,574 Record แล้วจึงทำความสะอาดข้อมูล และลบอักขระจำพวกวรรคตอนออก และแยกอักขระที่เป็นตัวอักษร หลังจากนั้นทำการแปลงอักขระทั้งหมดเป็นตัวพิมพ์เล็ก และใช้การตัดคำในการตัดแบ่งคำ หลังจากนั้นใช้กระบวนการ lemmatization โดยแปลงจากข้อความ ไปเป็นรูปแบบ Root form ก่อนนำข้อมูลเข้าโมเดลเพื่อทำการวิเคราะห์ผล โมเดลทั้ง 4 วิธีได้แก่ Naïve Bayes, Random Forest, K-Neighbors และ Support Vector Machine จากผลการศึกษาวิธีการ Random Forest สามารถจำแนกข้อความ Ham และ Spam ได้ค่าความถูกต้องมากถึง 0.974537

A. Ordonez, R. E. Paje และ R. Naz [4] ได้ทำการศึกษาการจัดประเภทข้อความ SMS ที่ส่งผ่านอุปกรณ์เคลื่อนที่ ในงานวิจัยนี้ได้มีข้อมูล SMS 5 ประเภท ได้แก่ Spam จำนวน 578 ข้อความ Invalid จำนวน 629 ข้อความ Alert 1 จำนวน 372 ข้อความ Alert 2 จำนวน 295 ข้อความ และ Alert 3 จำนวน 406 ข้อความ รวมทั้งสิ้น 2280 ข้อความ พบว่าวิธีการ Naive Bayes สามารถจำแนกประเภทของ SMS ได้ถูกต้อง 89% สามารถจำแนกข้อความ SMS ที่แตกต่างกัน 5 ประเภทโดยข้อความที่คัดแยกไว้จะต้องทำการทำความสะอาด ข้อมูลก่อน เช่น Stop Words, Noise

2.4 งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อความ

รวิสุดา เทศเมือง และ นิเวศ จิระวิชิตชัย [5] การวิเคราะห์ความคิดเห็นของการรีวิวสินค้าออนไลน์โดยใช้ขั้นตอนวิธีซัพพอร์ตเวกเตอร์แมทชีน การบริการห้องพัก รีสอร์ท โรงแรม ซึ่งได้มีผู้ที่เข้ามาใช้บริการจองห้องพัก รีสอร์ท โรงแรม และเข้ามาติชมหรือแสดงความคิดเห็นต่อการให้บริการและการใช้บริการ หรือรีวิวสินค้าผ่านทางเว็บไซต์ จำนวนมาก มีการนำข้อมูลเหล่านั้นมาวิเคราะห์ความคิดเห็นของภาษาไทยเกี่ยวกับการรีวิวออนไลน์ทำให้เพิ่มความ สะดวกสบายแก่ผู้ที่เข้ามาใช้บริการหรือกำลังตัดสินใจที่จะใช้บริการ โดยผู้ที่จะเข้ารับบริการสามารถอ่านรีวิวที่มีผู้ใช้บริการก่อนหน้านี้เพื่อประกอบการตัดสินใจในการใช้บริการของตนเองโดยทำการเก็บข้อมูลของ Agoda Thailand และ Twitter Thailand รวม 2,890 ข้อความ ใช้วิธีการวิเคราะห์ข้อมูล 4 แบบ แล้วนำมาเปรียบเทียบกันได้แก่ Support Vector Machine, Decision Tree, Naïve-Bayes, และ K-Nearest Neighbor จากการวิเคราะห์ข้อมูลพบว่า คุณลักษณะที่ดีสุดคือ Support Vector Machine ระดับรองลงมาเป็น Naïve-Bayes, Decision Tree และ K-Nearest Neighbor ตามลำดับ [5]

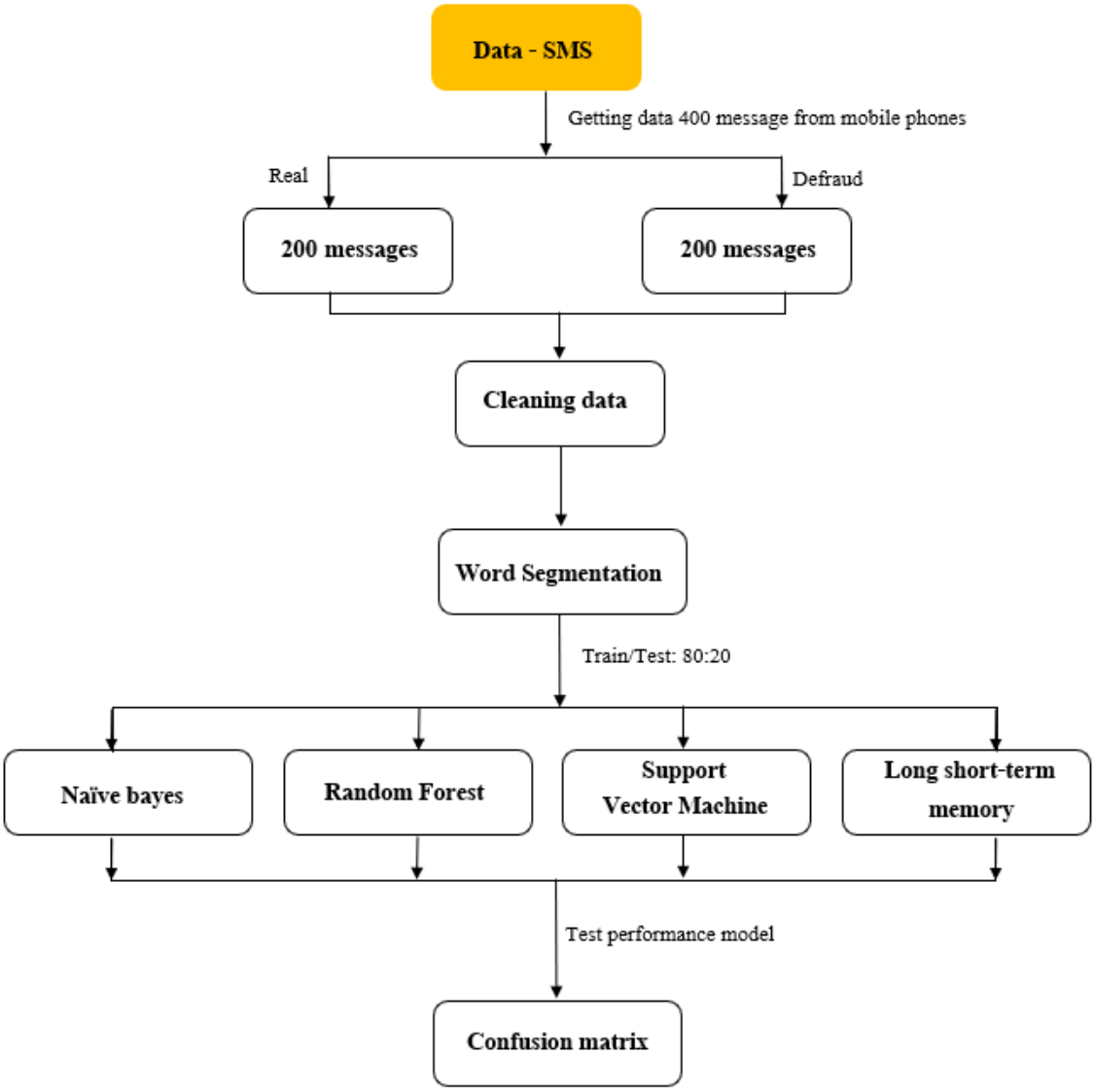
2.5 กระบวนการตัดคำ

TF-IDF (Term Frequency and Inverse Document Frequency) โดยปกติคอมพิวเตอร์ไม่มีความสามารถในการเรียนรู้ภาษาธรรมชาติของมนุษย์จากเอกสารต่าง ๆ ได้โดยตรง ดังนั้นจึงต้องแปลงเอกสารให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถใช้ในการเรียนรู้ได้ ในขั้นตอนในการแปลงเอกสาร เรียกว่า การทำดัชนี (Indexing) เพื่อสร้างตัวแทนเนื้อหาของเอกสาร (Document Representation) เพื่อสร้างตัวแทนที่ใช้ประมวลผลสำหรับคำนั้น ๆ ทั้งนี้ TF-IDF จะเป็นการจำแนกหมวดหมู่คำที่เกี่ยวข้องในเอกสารด้วยการเรียนรู้ด้วยคอมพิวเตอร์ ซึ่งนิยมใช้ลักษณะของความสำคัญคำในเอกสาร โดยไม่สนใจตำแหน่ง และดดับของคำ [9] กล่าวอีกนัยหนึ่งคือ TF-IDF เป็นเทคนิคการคัดแยกคำตามความสำคัญของคำนั้น ๆ ซึ่งจะถูกใช้ในการสร้างเวกเตอร์ โดยเทคนิคนี้จะใช้ในการประเมินความสำคัญของคำในข้อความทั้งหมด โดยความสำคัญของคำจะมีสัดส่วนเพิ่มตามจำนวณครั้งของคำนั้น ๆ ที่เกิดขึ้นในข้อความทั้งหมด [10] ยิ่งถ้ามีคำนั้น ๆ ปรากฏบนเอกสารมากหมายความว่าคำนั้น ๆ มีความสำคัญมาก

3. ขั้นตอนการวิจัย

สำหรับขั้นตอนของงานวิจัยชิ้นนี้ แสดงกระบวนการทำงานของงานวิจัยมีดังนี้ ในขั้นแรกทางผู้วิจัยได้ทำการรวบรวมข้อมูล (Getting data) จาก SMS จริงที่ส่งมายังโทรศัพท์เคลื่อนที่ประกอบไปด้วยจำนวน 400 ข้อความ แบ่งเป็นข้อความจริงจำนวน 200 ข้อความ และข้อความหลอกลวงจำนวน 200 ข้อความ และขั้นตอนถัดไปคือการการทำความสะอาดข้อมูล และกระบวนการทำ Word Segmentation เพื่อเตรียม

ข้อมูลสำหรับเข้าโมเดลทั้ง 4 โมเดลคือ Naïve bayes, Random forest, Support vector machine และ Long short term memory เมื่อทำการเทรนตามโมเดลต่าง ๆ แล้ว ในขั้นต่อนถัดไปคือการวัดประสิทธิภาพของโมเดลผ่าน Confusion matrix เพื่อหาโมเดลที่ดีที่สุดที่สามารถจำแนกข้อความจริงและข้อความหลอกลวงจาก SMS ดังแสดงในรูปที่ 1



รูปที่ 1 กระบวนการทำงาน

ข้อมูลทั้ง 400 ข้อมูลที่เก็บรวบรวมจะถูกแบ่งเก็บเป็น 2 ส่วน

1. Class หมายถึง ผลเฉลยของบริการข้อความ SMS ซึ่งประกอบด้วยข้อความจริง จำนวณ 200 ข้อความ และข้อความหลอกลวง จำนวณ 200 ข้อความ โดยก่อนการนำไปใช้ประมวล เพื่อจำแนกข้อความจริงหรือ

ข้อความหลอกลวง ต้องทำการแปลงข้อความจริง เป็น 0 และข้อความหลอกลวง เป็น 1 เพื่อใช้ตัวเลขเหล่านี้แทนผลเฉลยของข้อความที่จะนำไปใช้ประมวลผลต่อไป

2. Text หมายถึง ข้อความ SMS ที่ส่งมายังโทรศัพท์เคลื่อนที่ โดยทำการรวมรวบข้อมูลจากข้อความจริง จำนวณ 200 ข้อความ และข้อความหลอกลวง 200 ข้อความ หลังจากนั้นจะนำมาทำความสะอาดข้อมูล โดยในที่นี้ผู้วิจัย จะทำการตัดข้อความเฉพาะส่วนที่เป็นช่องว่างระหว่างคำ หรือระหว่างประโยคออกเท่านั้น เนื่องจากข้อความ SMS มีจำนวณมากที่มีส่วนประกอบของ อักขระพิเศษต่าง ๆ ดังแสดงในตารางที่ 1

ลำดับที่	ตัวอย่างข้อความ
1	“ชำระ 55.00บ บัตร x-2636@Foodpanda Thailand 22:02น”
2	“OTP = 290291 [รหัสอ้างอิง:48NEAD] เพื่อทำรายการผ่าน 'แอปไทยชนะ' ภายใน 5 นาที”

ตารางที่ 1: ตัวอย่างข้อความ

จากตารางที่ 1 จะเห็นได้ว่าข้อความที่ 1 จะมีอักขระพิเศษ ได้แก่ - , @ , : และข้อความจะมีอักขระพิเศษ ได้แก่ = , [,] , : , ‘ , ’ จะเห็นได้ว่า องค์ประกอบของ ข้อความ SMS จะมีอักขระพิเศษต่าง ๆ ประกอบอยู่ด้วย ดังนั้นแนวคิดของผู้วิจัย ภายในงานวิจัยนี้จึงถือว่า อักขระพิเศษต่าง ๆ เหล่านี้เป็นส่วนหนึ่งของ ข้อความ SMS จึงไม่ทำการตัดอักขระพิเศษต่าง ๆ เหล่านี้ออก

จากตัวอย่างข้อความข้างต้นก่อนหน้านี้ จะสังเกตได้ว่า ข้อความที่เก็บรวบรวมมาจะมีทั้งข้อความที่เป็นภาษาไทยเพียงอย่างเดียว หรือภาษาอังกฤษเพียงอย่างเดียว หรือรูปแบบผสมทั้งภาษาไทยและภาษาอังกฤษในข้อความเดียว เนื่องจากข้อมูลของข้อความสั้นที่รับเข้ามาอยู่ในรูปแบบของประโยค ที่มีความยาวแตกต่างกัน ดังนั้นข้อความทั้งหมดจะถูกตัดให้เป็น คำ 1 คำเพื่อให้มีขนาดของการประมวลผลที่เท่ากัน ดังนั้นในการตัดคำของงานวิจัยนี้ สำหรับภาษาทั้งสองภาษาโดยทั้งนี้ เราจะใช้ ไลบรารีใน Python มาช่วยในการตัดคำ ได้แก่ Pythainlp ที่ใช้ตัดคำภาษาไทย และ Nltk ที่ใช้ตัดคำภาษาอังกฤษ ดังนั้นข้อความทั้งหมดจะถูกตัดให้เป็น คำ 1 คำเพื่อใช้ในการประมวลผล

ในการตัดคำตามปกติ ไลบรารีใน Python จะไม่เข้าใจตัวเลขต่าง ๆ เพราะตัวเลขเป็นค่าเฉพาะของแต่ละข้อความ และไม่ถือเป็นคำในภาษานั้น ๆ ดังนั้น ในงานวิจัยนี้จะมีการเปลี่ยนตัวเลขเหล่านี้ให้กลายเป็น “num” เพื่อให้ตัวเลขทุกตัวที่มีขนาดความยาวแตกต่างกัน มีค่าเฉพาะที่แตกต่างกัน ถูกเปลี่ยนเป็นรูปแบบเดียวกันเพื่อให้ง่ายต่อการประมวลผล

จากกระบวนการตัดคำ และแปลงข้อมูลตัวเลข จะนำคำที่ถูกตัดคำและแปลงเรียบร้อยแล้ว เปลี่ยนให้อยู่ในรูปแบบอาร์เรย์ เนื่องจากคอมพิวเตอร์ไม่มีความสามารถในการเรียนรู้ภาษาของมนุษย์เป็นคำ แต่มีความสามารถในการเรียนรู้คำต่าง ๆ ที่อยู่ในรูปแบบอาร์เรย์ โดยไลบรารี Python คือ TfidfVectorizer เข้ามาช่วยในการแปลงคำที่ถูกตัดและแปลงข้อมูลตัวเลขให้อยู่ในรูปแบบเป็นอาร์เรย์

ต่อมาผู้วิจัยจะทำการประเมินผลโมเดลโดยใช้ค่า Accuracy, Precision, Recall, และ F1-Score ซึ่งมีความหมายและสูตรการคำนวณดังนี้

Accuracy คือการวัดความถูกต้องของโมเดล โดยพิจารณารวมทุกคลาส มีสมการดังนี้

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision คือการวัดความแม่นยำของข้อมูล โดยพิจารณาแยกทีละคลาส มีสมการดังนี้

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall คือการวัดความถูกต้องของโมเดล โดยพิจารณาแยกทีละคลาส มีสมการดังนี้

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1 Score คือค่าประสิทธิภาพโดยรวม นำค่า Precision และ Recall มาพิจารณาร่วมกัน มีสมการดังนี้

$$F1 - Score \quad = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

เมื่อสร้างและฝึกฝนโมเดลเสร็จเรียบร้อย จึงนำชุดข้อมูลสำหรับการทดสอบมาทดสอบโมเดลเพื่อดูความถูกต้องและแม่นยำเบื้องต้นในรูปแบบของ confusion matrix ซึ่งเป็นตารางวัดความสามารถของ โมเดลที่นำมาแก้ปัญหาจำแนกข้อความหลอกลวงและข้อความจริงจาก SMS โดยใช้การเปรียบเทียบค่า True Positive

(TP), False Positive (FP), False Negative (FN) และ True Negative (TN) ดังตารางที่ 2 ประกอบด้วยค่าสำคัญดังนี้ ค่า Accuracy สมการที่ (1) , ค่า Precision สมการที่ (2) , ค่า Recall สมการที่ (3) , ค่า F1 Score สมการที่ (4)

		Actual	
		0 (Real)	1 (Defraud)
Predict	0 (Real)	TP	FP
	1 (Defraud)	FN	TN

ตารางที่ 2: ตารางแสดงรูปแบบ Confusion matrix

จากตารางที่ 2 สามารถอธิบายค่าต่าง ๆ ที่เกิดขึ้นได้ว่า

True Positive (TP) คือ ผลลัพธ์ที่โมเดลทำนายว่าเป็น ‘ข้อความจริง’ และมีค่าเป็น ‘ข้อความจริง’

True Negative (TN) คือ ผลลัพธ์ที่โมเดลทำนายว่าเป็น ‘ข้อความหลอกลวง’ และมีค่าเป็น ‘ข้อความหลอกลวง’

False Positive (FP) คือ ผลลัพธ์ที่โมเดลทำนายว่าเป็น ‘ข้อความจริง’ แต่มีค่าเป็น ‘ข้อความหลอกลวง’

False Negative (FN) คือ ผลลัพธ์ที่โมเดลทำนายว่าเป็น ‘ข้อความหลอกลวง’ แต่มีค่าเป็น ‘ข้อความจริง’

4. ผลการวิจัย

ในงานวิจัยนี้ ผู้วิจัยเลือกใช้โมเดลในการวิเคราะห์ความสามารถในการจำแนกข้อความจริงและข้อความหลอกลวง โดยจะใช้โมเดลทั้งหมด 4 โมเดล ได้แก่ Naive Bayes, Random Forest, Support Vector Machine และ Long short-term memory ด้วยภาษา Python ซึ่งได้ผลลัพธ์ดังนี้

4.1 Naive Bayes

		Actual	
		0 (Real)	1 (Defraud)
Predict	0 (Real)	36	0
	1 (Defraud)	2	42

ตารางที่ 3: ตาราง Confusion matrix ของโมเดล Naive Bayes

จากตารางที่ 3 แสดงตาราง Confusion matrix ของโมเดล Naïve Bayes จะได้ค่า True Positive เท่ากับ 36, True negative เท่ากับ 42, False Positive เท่ากับ 0 และ False negative เท่ากับ 2 สรุปได้ว่า โมเดล Naïve Bayes สามารถทำนายได้ถูกจำนวน 78 ข้อความ และทำนายผิดจำนวน 2 ข้อความ จากชุดข้อมูลทดสอบ 20%

	precision	recall	f1-score	support
0 (Real)	0.95	1	0.97	36
1 (Defraud)	1	0.95	0.98	44
Accuracy	0.97			

ตารางที่ 4: ผลลัพธ์ของโมเดล Naive Bayes

จากตารางที่ 4 สรุปได้ว่าค่าความถูกต้องของวิธีการ Naive Bayes มีค่าเท่ากับ 0.97 หรือ โมเดล Naive Bayes มีความสามารถในการจำแนกข้อความจริงและข้อความหลอกลวง ที่ 97%

4.2 Random Forest

		Actual	
		0 (Real)	1 (Defraud)
Predict	0 (Real)	34	2
	1 (Defraud)	2	42

ตารางที่ 5: ตาราง Confusion matrix ของโมเดล Random Forest

จากตารางที่ 5 แสดงตาราง Confusion matrix ของโมเดล Random Forest จะได้ค่า True Positive เท่ากับ 34, True negative เท่ากับ 42, False Positive เท่ากับ 2 และ False negative เท่ากับ 2 สรุปได้ว่าโมเดล Naïve Bayes สามารถทำนายได้ถูกจำนวน 76 ข้อความ และทำนายผิดจำนวน 4 ข้อความ จากชุดข้อมูลทดสอบ 20%

	precision	recall	f1-score	support
0 (Real)	0.94	0.94	0.94	36
1 (Defraud)	0.95	0.95	0.95	44
Accuracy	0.95			

ตารางที่ 6: ผลลัพธ์ของโมเดล Random Forest

จากตารางที่ 6 สรุปได้ว่าค่าความถูกต้องของวิธีการ Random Forest มีค่าเท่ากับ 0.93 หรือ โมเดล Random Forest มีความสามารถในการจำแนกข้อความจริงและข้อความหลอกลวง ที่ 93 %

4.3 Support Vector Machine

		Actual	
		0 (Real)	1 (Defraud)
Predict	0 (Real)	35	1
	1 (Defraud)	2	42

ตารางที่ 7: ตาราง Confusion matrix ของโมเดล Support Vector Machine

จากตารางที่ 7 แสดงตาราง Confusion matrix ของโมเดล Support Vector Machine จะได้ค่า True Positive เท่ากับ 35, True negative เท่ากับ 42, False Positive เท่ากับ 1 และ False negative เท่ากับ 2 สรุปได้ว่าโมเดล Naïve Bayes สามารถทำนายได้ถูกจำนวน 77 ข้อความ และทำนายผิดจำนวน 3 ข้อความ จากชุดข้อมูลทดสอบ 20%

	precision	recall	f1-score	support
0 (Real)	0.95	0.97	0.96	36
1 (Defraud)	0.98	0.95	0.97	44
Accuracy	0.96			

ตารางที่ 8: ผลลัพธ์ของโมเดล Support Vector Machine

จากตารางที่ 8 สรุปได้ว่าค่าความถูกต้องของ Support Vector Machine มีค่าเท่ากับ 0.96 หรือ โมเดล Support Vector Machine มีความสามารถในการจำแนกข้อความจริงและข้อความหลอกลวง ที่ 96%

4.4 Long short-term memory

		Actual	
		0 (Real)	1 (Defraud)
Predict	0 (Real)	35	1
	1 (Defraud)	3	41

ตารางที่ 9: ตาราง Confusion matrix ของโมเดล Long shot-term memory

จากตารางที่ 9 แสดงตาราง Confusion matrix ของโมเดล Long short-term memory จะได้ค่า True Positive เท่ากับ 35, True negative เท่ากับ 41, False Positive เท่ากับ 1 และ False negative เท่ากับ 3 สรุปได้ว่าโมเดล Naïve Bayes สามารถทำนายได้ถูกจำนวน 76 ข้อความ และทำนายผิดจำนวน 4 ข้อความ จากชุดข้อมูลทดสอบ 20%

	precision	Recall	f1-score	support
0 (Real)	0.92	0.97	0.95	36
1 (Defraud)	0.98	0.93	0.95	44
Accuracy	0.95			

ตารางที่ 10: ผลลัพธ์ของโมเดล Long short-term memory

จากตารางที่ 10 สรุปได้ว่าค่าความถูกต้องของ Long short-term memory มีค่าเท่ากับ 0.95 หรือโมเดล Long short-term memory มีความสามารถในการจำแนกข้อความจริงและข้อความหลอกลวง ที่ 95%

5. สรุปผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อ จำแนกข้อความหลอกลวงและข้อความจริงจากบริการข้อความ SMS ที่ส่งมายังโทรศัพท์เคลื่อนที่ โดยทำการเก็บรวบรวมข้อมูลจากโทรศัพท์เคลื่อนที่ ที่มีการส่งข้อความสั้นเข้ามาจริง ๆ ซึ่งทำการเก็บรวบรวมข้อมูลทั้งหมด 400 ข้อความ โดยแบ่งเป็นข้อความจริงจำนวน 200 ข้อความ และข้อความหลอกลวงจำนวน 200 ข้อความ ทั้งนี้จะแบ่งการเก็บข้อมูลออกเป็น 2 ส่วนคือส่วนของ Class (ผลเฉลยของข้อความสั้น) และส่วนของ Text (ข้อความสั้นที่ส่งมายังโทรศัพท์เคลื่อนที่) และภายในงานวิจัยจะทำการทำความสะอาดข้อความที่รับเข้ามาด้วยการตัดช่องว่างออกเพียงอย่างเดียว โดยจะไม่ตัดอักขระพิเศษของข้อความออก พร้อมทั้งทำการแปลงข้อความที่ผ่านการทำความสะอาดข้อความแล้ว ให้อยู่ในรูปแบบอาร์เรย์เพื่อใช้ในการวิเคราะห์ข้อมูล ทั้งนี้เพื่อที่จะให้โมเดลเรียนรู้การจำแนกข้อความหลอกลวงและข้อความจริงจากบริการข้อความ SMS ซึ่งในงานวิจัยมีวิธีวิเคราะห์ข้อมูลทั้งสิ้นรวม 4 วิธีการคือ 1. Naïve Bayes 2. Random Forest 3. Long short-term memory 4. Support vector machine โดยมีจุดประสงค์ของการใช้ทั้ง 4 วิธีการคือเพื่อเปรียบเทียบว่าวิธีการใด มีความสามารถในจำแนกข้อความหลอกลวงและข้อความจริงได้อย่างแม่นยำที่สุด โดยได้จากผลลัพธ์วิธีการทั้ง 4 ดังตารางที่ 11

Model	Accuracy
1.Naive Bayes	0.97
2.Random Forest	0.95
3.Long Short-term memory	0.95
4.Support Vector Machine	0.96

ตารางที่ 11: ตารางแสดงค่า Accuracy ของแต่ละโมเดล

จากผลลัพธ์โดยรวมพบว่า วิธีการที่ได้ผลลัพธ์ดีที่สุด คือ Naive Bayes ซึ่งมีความสามารถในการจำแนกข้อความจริงและข้อความหลอกลวง ที่ 97% ซึ่งสูงที่สุดเมื่อเปรียบเทียบจากทั้ง 4 วิธีการ

จากการศึกษาในครั้งนี้ผู้วิจัยมีความเห็นว่าควรเพิ่มข้อมูลให้มากขึ้น เพื่อให้โมเดลเรียนรู้รูปแบบของข้อความสั้นที่หลากหลาย และเพิ่มจำนวนการเก็บข้อมูลที่หลากหลายมากขึ้นจากหลาย ๆ เครื่องโทรศัพท์มือถือเนื่องจากข้อความหลอกลวงที่ส่งมาจากมิจฉาชีพในบางเครื่องจะมีรูปแบบข้อความที่แตกต่างกันออกไป แต่จะมีข้อจำกัดเรื่องความเป็นส่วนตัวของการเก็บข้อมูล

6. เอกสารอางอิง

[1] Wanwisa Thuanyod. “วิวัฒนาการการสื่อสารของแต่ละยุคแบบเข้าใจง่าย The Evolution of Communication.” thinknet.co.th. https://shorturl.asia/8f6WG

[2] ไทยพีบีเอส (Thai PBS). “ฮูส์คอล" เปิดสถิติโทรศัพท์หลอกลวงพุ่ง 6.4 ล้านครั้ง.” www.thaipbs.or.th. https://shorturl.asia/FRuOm

[3] H. Jain and R. K. Maurya, "A Review of SMS Spam Detection Using Features Selection," *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, Sonepat, India, 2022, pp. 101-106, doi:10.1109/CCiCT56684.2022.00030.

[4] A. Ordonez, R. E. Paje and R. Naz, "SMS Classification Method for Disaster Response Using Naïve Bayes Algorithm," *2018 International Symposium on Computer, Consumer and Control (IS3C)*, Taichung, Taiwan, 2018, pp. 233-236, doi: 10.1109/IS3C.2018.00066.

[5] รวิสุดา เทศเมือง และ นิเวศ จิระวิชิตชัย, "การวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการ รีวิวสินค้าออนไลน์โดยใช้ขั้นตอนวิธีซัพพอร์ตเวกเตอร์แมทชีน," 2560 Engineering Journal of Siam University, Volume 18, Issye 1,มกราคม-มิถุนายน 2560,https://e-library.siam.edu/e-journal/wp-content/uploads/2018/11/EJSU_No.34_pp_1-12.pdf

[6] Manajit Chakraborty, Sukomal Pal, Rahul Pramanik, C. Ravindranath Chowdary, "Recent developments in social spam detection and combating techniques: A survey" Information Processing & Management, Volume 52, Issue 6, November 2016, pp. 1053-1073, doi.org/10.1016/j.ipm.2016.04.009

[7] Thaiware. “Spam คืออะไร ? มีที่มาจากไหน ? Spam มีกี่ประเภท ? ทำไมเราถึงตกเป็นเป้าหมาย?.” tips.thaiware.com. https://shorturl.asia/OMdf3

[8] หน่วยเทคโนโลยีสารสนเทศ. “3 อันดับภัยไซเบอร์ใกล้ตัวที่คนไทยถูกหลอกมากที่สุด” . it.edu.cmu.ac.th. https://shorturl.asia/eokLV

[9] T. Kongmanee, S. Vanichayobon and W. Wettayaprasit, "The TF-IDF and Neural Networks Approach for Translation Initiation Site Prediction," *2009 2nd IEEE International Conference on Computer Science and Information Technology*, Beijing, China, 2009, pp. 318-322, doi: 10.1109/ICCSIT.2009.5234582.

[10] บุษบงก์ คชินทรโรจน์, เดือนเพ็ญ ธีรวรรณวิวัฒน์, และ พาชิตชนัต ศิริพานิช, “การสร้างตัวแบบหัวข้อและตัวแบบจัดประเภทการเกลียดกลัวคนต่างชาติบนทวิตเตอร์ในช่วงการแพร่ระบาดของ COVID-19,” *Thai Journal of Operations Research* : TJOR, vol. 9, no. 1, pp. 31–44, 2021, Accessed: Mar. 17, 2023. [Online]. https://shorturl.asia/wHgqN