

Text Classification Cyberbullying Comments on Online Social Media.

การจำแนกประเภทข้อความความคิดเห็นที่เป็นการกระหนทางไซเบอร์ในสื่อสังคมออนไลน์

Ekarat Rattagan and Paradorn Boonpoor*

เอกรัตน์ รัฐกาญจน์ และ ภารดร บุญพอ*

Graduate School of Applied Statistics, National Institute of Development Administration

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ กรุงเทพมหานคร

*Corresponding Author, Tel. -, E-mail: paradorn.boon@stu.nida.ac.th

*ผู้แทนประสานงาน โทรศัพท์ - อีเมล: paradorn.boon@stu.nida.ac.th

Abstract

This research presents the development of a model for classifying Thai texts as cyberbullying using LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units), Bidirectional LSTM and Bidirectional GRU by creating Word2Vec generated from the dataset and using the Pretrained Model from WangchanBerta. and Naïve Bayes algorithm used Tf-Idf as word representation. Compare the algorithms to assess the performance of the text classification using Accuracy, Precision, Recall. And F1-Score which found that the algorithm using Word Representation from Word2Vec developed from the dataset. more efficient than Naïve Bayes algorithm It uses Tf-Idf Word Representation and WangchanBerta Pretrained Model.

Keywords: Text Classification, Cyberbullying, Social Media

บทคัดย่อ

งานวิจัยนี้ได้นำเสนอการพัฒนาตัวแบบในการจำแนกประเภทข้อความภาษาไทยที่เป็นการกระหนทางไซเบอร์ โดยใช้อัลกอริทึม LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units), Bidirectional LSTM, และ Bidirectional GRU โดยการนำ Word Representation แบบ Word2Vec ที่สร้างขึ้นจากชุดข้อมูล และการใช้ Pretrained Model จาก WangchanBerta นอกจากนี้ยังใช้อัลกอริทึม Naïve Bayes โดยนำ Word Representation แบบ Tf-Idf มาเปรียบเทียบกับอัลกอริทึมเพื่อประเมินประสิทธิภาพในการจำแนกข้อความทั้งในด้านความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ความระลึก (Recall) และค่าความถ่วงดุล (F1-Score) ซึ่งพบว่า อัลกอริทึมที่ใช้ Word Representation จาก Word2Vec ที่พัฒนาจากชุดข้อมูล มีประสิทธิภาพมากกว่า การใช้อัลกอริทึม Naïve Bayes ที่ใช้ Word Representation แบบ Tf-Idf และให้ประสิทธิภาพดีกว่าการใช้ Pretrained Model แบบ WangchanBerta

คำสำคัญ: การจำแนกประเภทข้อความ การกระหนทางไซเบอร์ สื่อสังคมออนไลน์