

การศึกษาปัจจัยที่ส่งผลต่อความน่าเชื่อถือของเว็บไซต์ข่าวภาษาไทย ด้วยเทคนิคการเรียนรู้ของเครื่อง

องอาจ อุ่นอนันต์^{1,*}, พยุง มีสัจ²

^{1,2} คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

Received: 3 May 2019

Revised: 18 October 2019

Accepted: 18 October 2019

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ในการศึกษาปัจจัยที่ส่งผลต่อความน่าเชื่อถือของเว็บไซต์ข่าวภาษาไทย และนำไปพัฒนาเครื่องมือที่ใช้ในการตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าว โดยทบทวนวรรณกรรมเพื่อรวบรวมปัจจัยที่เกี่ยวข้อง และประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องในการจัดกลุ่มข้อมูลด้วยเทคนิคการจัดกลุ่มแบบเค-มีนส์ (K-Mean Clustering) และเทคนิคการจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering) แล้วนำมาเปรียบเทียบค่าประสิทธิภาพของการจัดกลุ่มข้อมูลด้วยค่าสหสัมพันธ์ซิลลูเอท (Silhouette Coefficient) เมื่อได้การจัดกลุ่มที่ดีที่สุดแล้วจึงนำข้อมูลการจัดกลุ่มดังกล่าวมาทำการอธิบายความหมายของกลุ่ม โดยพบว่าการจัดกลุ่มด้วยเทคนิคเค-มีนส์ โดยค่า $K = 5$ (Silhouette Coefficient = 0.199182) มีประสิทธิภาพดีที่สุด แล้วนำการจัดกลุ่มที่ได้มาพัฒนาเป็นเครื่องมือที่ใช้ในการตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าว

คำสำคัญ : ความน่าเชื่อถือ การเรียนรู้ของเครื่อง การจัดกลุ่ม เว็บไซต์ข่าว

* ผู้ประสานงานหลัก; อีเมล: aongart.a@email.kmutnb.ac.th

The Study of Factors Influencing to Credibility of Thai News Website by Using Machine Learning Techniques

Aongart Aun-a-nan^{1,*}, Phayung Meesad²

^{1,2} The Faculty of Information Technology,
King Mongkut's University of Technology North Bangkok

ABSTRACT

This research aims to study the factors affecting the credibility of Thai news websites and to apply the results to develop the tool for assessing credibility of news websites. The researcher reviewed related literature in order to identify the related factors, and then the K-Mean Clustering and Hierarchical Clustering techniques were applied to be a learning machine for clustering data. After that, the efficiencies of data clustering were compared with the use of Silhouette Coefficient values. When the best data clustering had been identified, the researcher took the information from the data clustering to explain the meaning of the clustering. The results showed that the clustering by K-Means Clustering technique with $K = 5$ (Silhouette Coefficient = 0.199182) gave the optimal efficiency. Finally, results from the data clustering were used to develop the tool for assessing credibility of news websites.

Keywords: credibility; learning machine; data clustering; news website

* Corresponding Author; Email: aongart.a@email.kmutnb.ac.th

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันเทคโนโลยีอินเทอร์เน็ต (Internet) มีบทบาทสำคัญในชีวิตประจำวันของมนุษย์มากขึ้น ทำให้พฤติกรรมการใช้บริการข้อมูลข่าวสารเปลี่ยนไปจากเดิม โดยในอดีตนั้นรับข้อมูลข่าวสารจากวิทยุ หนังสือพิมพ์ วารสาร และโทรทัศน์เป็นช่องทางหลัก แต่ในปัจจุบันเมื่อเทคโนโลยีเข้ามามีบทบาททำให้มีช่องทางในการรับข้อมูลข่าวสารจากช่องทางอื่นเพิ่มขึ้น เช่น เว็บไซต์ (Website) และสื่อสังคม (Social Media) (National Statistical Office, 2018) ที่สามารถส่งต่อ (Share) ข้อมูลข่าวสารได้อย่างสะดวก รวดเร็ว และเป็นปัจจุบันมากยิ่งขึ้น แต่เว็บไซต์และสื่อสังคมนั้นสามารถถูกสร้างขึ้นได้ง่าย จากบุคคลทั่วไปทั้งที่มีส่วนเกี่ยวข้องและไม่มีส่วนเกี่ยวข้องกับการเป็นสื่อสารมวลชนหรือเป็นสำนักข่าว ทำให้มีการฉวยโอกาสสร้างเว็บไซต์หรือสื่อสังคมปลอมที่เลียนแบบสำนักข่าวจริง เพื่อสร้างความเข้าใจผิดและเผยแพร่ข่าวสารอันเป็นเท็จ ประกอบกับพฤติกรรมการใช้งานสื่อสังคมที่มีการส่งต่อข้อมูลข่าวสารระหว่างกัน ทำให้ข้อมูลข่าวสารต่าง ๆ มีการแพร่กระจายได้อย่างรวดเร็วเป็นวงกว้าง ดังนั้นถ้ามีการเผยแพร่หรือส่งต่อข้อมูลข่าวสารอันเป็นเท็จจากแหล่งข่าวที่ไม่น่าเชื่อถือโดยที่ผู้ใช้งานไม่รู้ข้อเท็จจริงของข่าวนั้นก็จะทำให้เกิดผลเสียแก่ผู้รับและผู้ส่งข่าวสาร (Shah et al., 2015) อีกทั้งยังผิดพระราชบัญญัติว่าด้วยการกระทำความผิดเกี่ยวกับคอมพิวเตอร์ (ฉบับที่ 2) พ.ศ. 2560 อีกด้วย

ซึ่งการตรวจสอบข้อมูลข่าวสารว่ามีความน่าเชื่อถือหรือไม่นั้นเป็นเรื่องยาก ต้องอาศัยผู้เชี่ยวชาญในสาขาที่เกี่ยวข้องกับข้อมูลข่าวสารที่ต้องการตรวจสอบ ซึ่งปัจจุบันมีบุคคล หน่วยงาน และสื่อช่องทางต่างๆ ที่พยายามช่วยกันตรวจสอบข้อมูลข่าวสารที่ไม่มีแหล่งที่มาของข่าวชัดเจนหรือมาจากแหล่งที่มาปลอม ให้ได้ข้อเท็จจริงก่อนแล้วจึงนำเสนอข้อเท็จจริงของข่าวสารที่เป็นข้อสงสัย เช่น "รายการข่าวก่อนแชร์" (Polkett & Salathong, 2018) ที่เป็นการตรวจสอบข้อมูลข่าวสารที่ไม่มีแหล่งที่มาของข่าวที่ชัดเจนแล้วถูกส่งต่อในอินเทอร์เน็ตหรือสื่อสังคม โดยนำประเด็นข้อสงสัยไปสอบถามข้อมูลจากผู้เชี่ยวชาญแล้วสรุปข้อเท็จจริง ว่าเรื่องใดจริงเรื่องใดไม่จริงแล้วจึงนำมาเผยแพร่ แต่อย่างไรก็ตามเมื่อเปรียบเทียบกับปริมาณข้อมูลข่าวสารในแต่ละวันกับข้อมูลข่าวสารที่ผ่านการตรวจสอบโดยผู้เชี่ยวชาญนั้นเป็นสัดส่วนที่น้อยมาก ทำให้การตรวจสอบโดยผู้เชี่ยวชาญไม่ครอบคลุมข่าวสารต่างๆ ที่มีอยู่เป็นจำนวนมาก นอกจากผู้เชี่ยวชาญแล้วผู้สื่อข่าว สำนักข่าว หน่วยงาน หรือองค์กรที่ทำหน้าที่เป็นสื่อสารมวลชนสามารถมีส่วนช่วยในการตรวจสอบความน่าเชื่อถือของข้อมูลข่าวสารในเบื้องต้นได้เช่นกัน ซึ่งถ้าข้อมูลข่าวสารนั้นมีแหล่งที่มาของข่าวที่ชัดเจนเป็นแหล่งข่าวที่มีความน่าเชื่อถือก็มีความถูกต้องของเนื้อหามากกว่าข้อมูลข่าวสารที่มาจากแหล่งข่าวที่ไม่ชัดเจนไม่น่าเชื่อถือ (Patricia & Melancon, 2018) หากผู้รับข่าวสารสามารถตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าวที่ได้รับการส่งต่อมาได้ด้วยตนเอง ก็จะสามารถตัดสินใจที่จะเชื่อหรือไม่เชื่อข้อมูลที่ได้รับมาได้อย่างถูกต้องมากยิ่งขึ้น และเป็นเรื่องที่ทำหายที่จะทำการตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าวออนไลน์ เนื่องจากมีการสร้างเว็บไซต์และเครือข่ายสังคมออนไลน์ขึ้นมาใหม่อยู่เสมอ

จากเหตุผลดังกล่าว เนื่องจากแหล่งข่าวที่มีความน่าเชื่อถือจะมีเนื้อหาข่าวที่มีการตรวจสอบมาก่อนเบื้องต้นในระดับหนึ่ง งานวิจัยนี้จึงมีวัตถุประสงค์ในการศึกษาปัจจัยที่มีผลต่อความน่าเชื่อถือของเว็บไซต์ข่าว เพื่อวิเคราะห์จัดกลุ่มข้อมูลของเว็บไซต์ข่าว และพัฒนาเครื่องมือที่ใช้ในการตรวจสอบเว็บไซต์ข่าว โดยมีการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) มาใช้ในการจัดกลุ่มข้อมูลด้วยเทคนิคการจัดกลุ่มแบบเค-มีนส์ (K-Mean Clustering) และเทคนิคการจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering) โดยวิเคราะห์การเชื่อมต่อแบบ Single Linkage, Average Linkage และ Complete Linkage แล้วนำมาเปรียบเทียบประสิทธิภาพของการจัดกลุ่มข้อมูล

ด้วยค่า Silhouette Coefficient เมื่อได้การจัดกลุ่มที่ดีที่สุดแล้วจึงนำข้อมูลการจัดกลุ่มดังกล่าวมาทำการอธิบายความหมายของกลุ่ม แล้วนำการจัดกลุ่มที่ได้มาพัฒนาเป็นเครื่องมือที่ใช้ในการตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าวภาษาไทย

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อความน่าเชื่อถือของเว็บไซต์ข่าวภาษาไทย ด้วยเทคนิคการเรียนรู้ของเครื่อง
2. เพื่อเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูล และเลือกเทคนิควิธีที่เหมาะสมนำไปพัฒนาเครื่องมือที่ใช้ในการตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าว
3. เพื่อพัฒนาเครื่องมือที่ใช้ในการตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าวภาษาไทย

เอกสารและงานวิจัยที่เกี่ยวข้อง

ความน่าเชื่อถือของเว็บไซต์ประกอบด้วยส่วนที่เป็นประสบการณ์ของผู้ใช้งานและส่วนที่เป็นความเชื่อมั่นของเนื้อหาซึ่งอยู่ในรูปของหนังสือ หนังสือพิมพ์ เอกสารอิเล็กทรอนิกส์หรือเอกสารประเภทอื่นๆ ที่มีการเก็บข้อมูลอยู่ใน (Flanagin & Metzger, 2008) ปัจจัยที่เกี่ยวข้องกับความน่าเชื่อถือของเว็บไซต์ ที่นิยมนำมาใช้ในการประเมินความน่าเชื่อถือได้ถูกจัดเป็นหมวดหมู่ (Shah & Ravana, 2014) โดยสามารถนำมาสรุปเป็นหมวดหมู่ได้ 6 ปัจจัยหลักแสดงดังตารางที่ 1

ตารางที่ 1 ปัจจัยที่เกี่ยวข้องกับความน่าเชื่อถือของเว็บไซต์

ปัจจัย	คำอธิบาย
ความถูกต้อง (Accuracy)	ความถูกต้องของข้อมูลที่ได้รับจากผู้เขียน
ผู้มีอำนาจหน้าที่ (Authority)	ประสบการณ์และความนิยมของผู้เขียน ซึ่งรวมถึงคุณสมบัติและข้อมูลประจำตัวของผู้เขียน
สุนทรียภาพ (Aesthetics)	ภาพรวมของเว็บไซต์ พิจารณาจากสี รูปแบบ รูปภาพ วิดีโอ ตัวอักษร สัญลักษณ์ ตารางข้อมูล ที่ใช้บนเว็บเพจ
ความเป็นมืออาชีพ (Professionalism)	นโยบายความเป็นส่วนตัวและคุณสมบัติที่มีอยู่ในเว็บไซต์
ความนิยม (Popularity)	ชื่อเสียงของเว็บไซต์ในหมู่ผู้ใช้เว็บและผู้เชี่ยวชาญ
ความเป็นปัจจุบัน (Currency)	ความถี่ของการอัปเดตเนื้อหาและการแสดงวันที่ในข่าว

ซึ่งแหล่งข้อมูลของข่าวมีจากหลายช่องทาง โดย Noosom & Suttisima (2019) ได้แบ่งกลุ่มของแหล่งข้อมูลออกเป็น 3 กลุ่ม ดังนี้ 1) เว็บไซต์ผู้ผลิตข่าวปลอม โดยคัดมาจากรายชื่อที่อยู่ในบัญชีเว็บไซต์ข่าวปลอม 2) เว็บไซต์หรือแฟนเพจที่ทำการแก้ไขหรือนำเสนอความจริงเกี่ยวกับข่าวปลอมที่ปรากฏบนสื่อสังคมออนไลน์ และ 3) เว็บไซต์หรือแฟนเพจของสำนักข่าวจริงหรือสื่อมวลชนที่เชื่อถือได้ แล้วยังพบว่าข่าวปลอมมีแนวโน้มที่เพิ่มขึ้น ทำให้ผู้ใช้งานต้องใช้วิจารณญาณในการรับข่าวสารที่เพิ่มขึ้น เช่นเดียวกับงานของ Aun-a-nan et al. (2016) ผู้ใช้งานบริโภคข่าวจาก

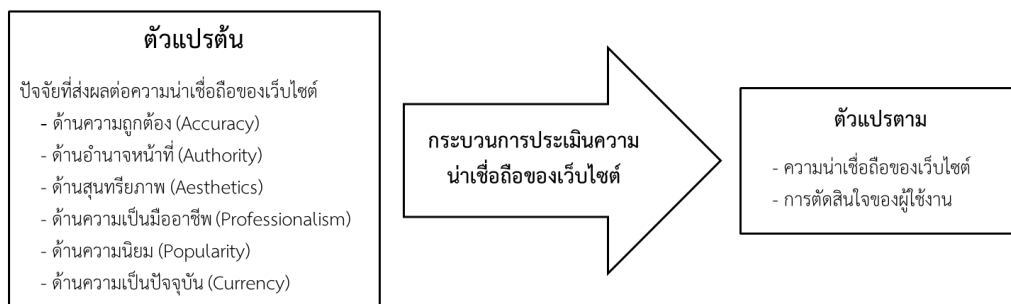
เว็บไซต์ที่เป็นเว็บท่า (Web Portal) และเว็บไซต์สำนักข่าวจริงมีการพบข่าวเท็จบ้างแต่จะพบข่าวเท็จในเว็บไซด์สำนักข่าวจริงน้อยกว่าในเว็บท่า และยังไม่พบว่ามีเครื่องมือที่ใช้ในการตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าวภาษาไทย

Krishnamoorthy et al. (2018) เสนอวิธีการจัดหมวดหมู่บทความข่าวจากแหล่งข่าวออนไลน์และติดป้ายกำกับโดยอัตโนมัติตามโดเมนของข่าวงานนี้ใช้แนวคิดเรื่องการประมวลผลภาษาธรรมชาติและการเรียนรู้ด้วยเครื่อง โดยการใช้เทคนิค K-Means และ Incremental Clustering ความถูกต้องของผลลัพธ์ของรูปแบบที่นำเสนอเป็นที่น่าพึงพอใจ ผลลัพธ์เหล่านี้สามารถใช้เพื่อเปรียบเทียบความสนใจของผู้ชมที่แตกต่างกันในแต่ละแหล่งข่าวได้

Nanayakkara & Ranathunga (2018) ตัวรวบรวมข่าวสารช่วยให้ผู้อ่านสามารถจัดการข่าวสารจำนวนมากได้ โดยการรวบรวมข่าวสารที่เหมือนกันไว้ในที่เดียวกัน ตัวรวบรวมข่าวนี้นี้มีให้สำหรับภาษาอังกฤษและภาษายอดินิยมอื่นๆ อย่างไรก็ตามยังไม่มีเครื่องมือดังกล่าวสำหรับภาษา Sinhala เมื่อต้องการแก้ปัญหาที่บทความนี้จะนำเสนอ การจัดกลุ่มข่าวสาร แล้วนำมาใช้มาตรการคล้ายคลึงกันของคลังข้อมูล แม้จะมีความเรียบง่ายของเทคนิค แต่ก็ประสบความสำเร็จเป็นอย่างมาก

กรอบแนวคิดในการวิจัย

จากการทบทวนวรรณกรรมผู้วิจัยได้พบว่ามีปัจจัยที่ส่งผลต่อความน่าเชื่อถือของเว็บไซต์ข่าวจำนวน 6 ปัจจัย จึงได้กำหนดให้เป็นตัวแปรต้น โดยทางผู้วิจัยได้ออกแบบกระบวนการประเมินความน่าเชื่อถือของเว็บไซต์ข่าว เพื่อให้ผู้ใช้งานสามารถใช้เป็นเครื่องในการประเมินความน่าเชื่อถือของข้อมูลหรือแหล่งข่าวได้เบื้องต้นและนำข้อมูลไปประกอบการตัดสินใจในการเชื่อข่าวสารนั้น โดยกำหนดให้ความน่าเชื่อถือของเว็บไซต์และการตัดสินใจที่จะเชื่อหรือไม่เชื่อข่าวสารนั้นเป็นตัวแปรตาม แสดงดังภาพที่ 1

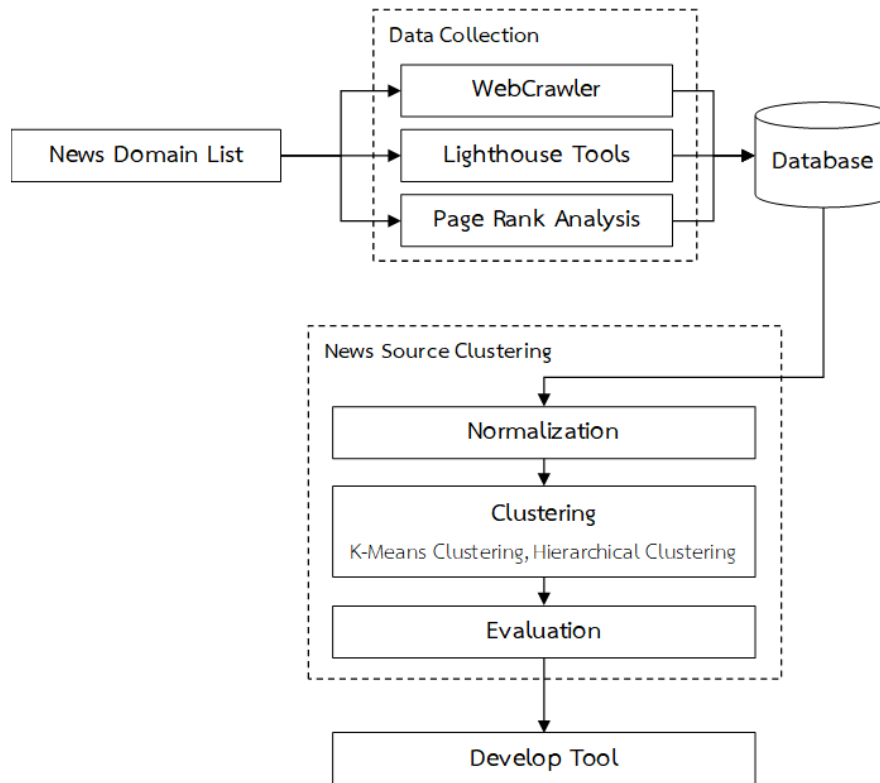


ภาพที่ 1 กรอบแนวคิดในการวิจัย

วิธีดำเนินการวิจัย

ขั้นตอนการวิจัย

ในงานวิจัยนี้ได้กำหนดขั้นตอนการวิจัยแบ่งออกเป็น 3 ขั้นตอนหลัก แสดงดังภาพที่ 2



ภาพที่ 2 ขั้นตอนการวิจัย

การเก็บรวบรวมข้อมูล

หลังจากทำการทบทวนวรรณกรรมและแบ่งกลุ่มของปัจจัยที่ส่งผลต่อความน่าเชื่อถือของเว็บไซต์ข่าวได้ 6 กลุ่มปัจจัยหลัก และทำการวิเคราะห์หาปัจจัยย่อยที่คอมพิวเตอร์สามารถดึงข้อมูลปัจจัยได้อย่างอัตโนมัติ จำนวน 35 ปัจจัยย่อย แล้วดำเนินการเลือกวิธีในการจัดเก็บข้อมูลที่เหมาะสมในแต่ละปัจจัยย่อยโดยใช้เครื่องมือในการเก็บข้อมูล ดังนี้ 1) Web Crawler ที่ผู้วิจัยพัฒนาขึ้นมาโดยเฉพาะ (Aun-a-nan & Meesad, 2018) ใช้ในการเก็บข้อมูลที่แสดงอยู่บนเนื้อหาของเว็บไซต์ 2) Lighthouse Tools เป็นเครื่องมือที่ถูกพัฒนาโดย Google ใช้สำหรับเก็บข้อมูลประสิทธิภาพของ Progressive Web Ap (PWA) และ 3) Page Rank Analysis Tools ใช้ในการเก็บข้อมูลที่เกี่ยวข้องกับความนิยมของเว็บไซต์ โดยดึงข้อมูลจากเว็บไซต์ข่าวภาษาไทยจำนวน 152 เว็บไซต์ ซึ่งประกอบด้วย เว็บไซต์สำนักข่าวหลักในประเทศ เว็บไซต์สำนักข่าวออนไลน์และเว็บไซต์ข่าวที่นิยมถูกส่งต่อกันในสื่อสังคม แล้วนำข้อมูลที่ดึงได้ไปจัดเก็บในฐานข้อมูล

การวิเคราะห์ข้อมูล

ในขั้นตอนนี้ดึงข้อมูลปัจจัยย่อย ที่ถูกจัดเก็บไว้ในฐานข้อมูลแล้วนำมาผ่านกระบวนการทำให้ค่าอยู่ในช่วงปกติ (Normalization) ให้อยู่ในค่าช่วง 0-1 จากนั้นนำมาวิเคราะห์การจัดกลุ่มด้วยเทคนิค K-Mean Clustering และ Hierarchical Clustering โดยใช้ Scikit-learn ในภาษาไพธอน (Python) เป็นเครื่องมือที่ใช้ในการวิเคราะห์ แล้วทำการเปรียบเทียบว่าการจัดกลุ่มแบบใดมีประสิทธิภาพการจัดกลุ่มได้ดีที่สุด โดยการวัดประสิทธิภาพของข้อมูลที่ได้ทำการจัดกลุ่มพิจารณาจากค่า Silhouette Coefficient เป็นการหาค่าความใกล้เคียงกันในข้อมูลระดับอัตราส่วน (Ratio

Scale) และมีการหากลุ่มที่มีความหนาแน่นมากและแยกออกมาจากกลุ่มอื่นอย่างชัดเจน โดยใช้ค่าเฉลี่ยของระยะห่างจุด S_i กับจุดต่าง ๆ ภายในกลุ่มเดียวกัน ส่วนด้วยระยะห่างน้อยที่สุดของจุด S_i กับจุดต่าง ๆ ในแต่ละกลุ่ม โดยค่าที่ได้จะอยู่ในช่วง $[-1, 1]$ ถ้ามีค่าเข้าใกล้ 1 แสดงว่าข้อมูลที่ถูกนำมาจัดกลุ่มมีความเหมาะสม แต่ถ้ามีค่าเข้าใกล้ -1 แสดงว่าข้อมูลที่ถูกนำมาจัดกลุ่มไม่เหมาะสมกับกลุ่มที่จัดไว้ สำหรับฟังก์ชันการคำนวณค่า Silhouette Coefficient (Rousseeuw, 1987) แสดงดังสมการ (1)

$$S_i = \frac{b-a}{\max(a,b)} \quad (1)$$

โดยให้ a คือ ระยะทางเฉลี่ยระหว่างจุดตัวอย่างและจุดอื่นทั้งหมดในกลุ่มเดียวกัน และให้ b คือระยะทางเฉลี่ยระหว่างจุดตัวอย่างและจุดทั้งหมด

หลังจากที่ได้การจัดกลุ่มที่ดีที่สุดแล้วทำการเปรียบเทียบปัจจัยของแต่ละกลุ่มแล้วอธิบายความหมายของแต่ละกลุ่มเพื่อจัดอันดับความน่าเชื่อถือของเว็บไซต์ของแต่ละกลุ่ม

การพัฒนาเครื่องมือตรวจสอบความน่าเชื่อถือ หลังจากการวิเคราะห์ข้อมูลแล้วนำข้อมูลที่ได้จากการจัดกลุ่มมาทำการสร้างเครื่องมือเพื่อให้ข้อมูลความน่าเชื่อถือของเว็บไซต์ข่าว โดยเครื่องมือสามารถรับลิงค์ของข่าวจากผู้ใช้งาน และทำการประมวลผลแสดงกลุ่มความน่าเชื่อถือของเว็บไซต์ข่าวจากลิงค์ที่ผู้ใช้งานต้องการตรวจสอบ

ผลการวิจัย

ผลการดำเนินการวิจัยศึกษาปัจจัยที่ส่งผลต่อความน่าเชื่อถือของเว็บไซต์ข่าวภาษาไทย ด้วยเทคนิคการเรียนรู้ของเครื่อง แบ่งออกเป็น 3 ส่วนตามขั้นตอนในการดำเนินการวิจัยโดยมีรายละเอียด ดังนี้

1. ผลการวิเคราะห์ปัจจัยและการเก็บข้อมูล (Data Collection)

จากการทบทวนวรรณกรรมที่เกี่ยวข้อง พบว่ามีปัจจัยที่ส่งผลต่อความน่าเชื่อถือ 6 ปัจจัยหลัก มีปัจจัยย่อยที่คอมพิวเตอร์สามารถดึงข้อมูลได้ 35 ปัจจัยย่อย ที่จะนำมาใช้ในการวิเคราะห์ข้อมูลด้วยการเรียนรู้ของเครื่อง โดยกำหนดข้อมูลที่จะถูกจัดเก็บในแต่ละปัจจัยรวมถึงวิธีการในการจัดเก็บข้อมูล โดยรายละเอียดของข้อมูลที่จะถูกจัดเก็บในแต่ละปัจจัย แสดงดังตารางที่ 2

ตารางที่ 2 รายละเอียดของปัจจัยย่อยที่ส่งผลต่อความน่าเชื่อถือเว็บไซต์

ที่	ปัจจัยหลัก	ปัจจัยย่อย	เครื่องมือในการจัดเก็บ
1	ความถูกต้อง (Accuracy)	1. ข้อมูลการอ้างอิงแหล่งข่าว	Web Crawler
		2. ข้อมูลลายน้ำในรูปภาพประกอบข่าว	Web Crawler
2	ผู้มีอำนาจหน้าที่ (Authority)	1. ข้อมูลผู้รายงานข่าว/ผู้เขียนข่าว	Web Crawler
		2. ข้อมูลการติดต่อกับแหล่งข่าว (ที่อยู่)	Web Crawler
		3. ข้อมูลการติดต่อกับแหล่งข่าว (เบอร์โทร)	Web Crawler
		4. ข้อมูลการติดต่อกับแหล่งข่าว (อีเมลล์)	Web Crawler

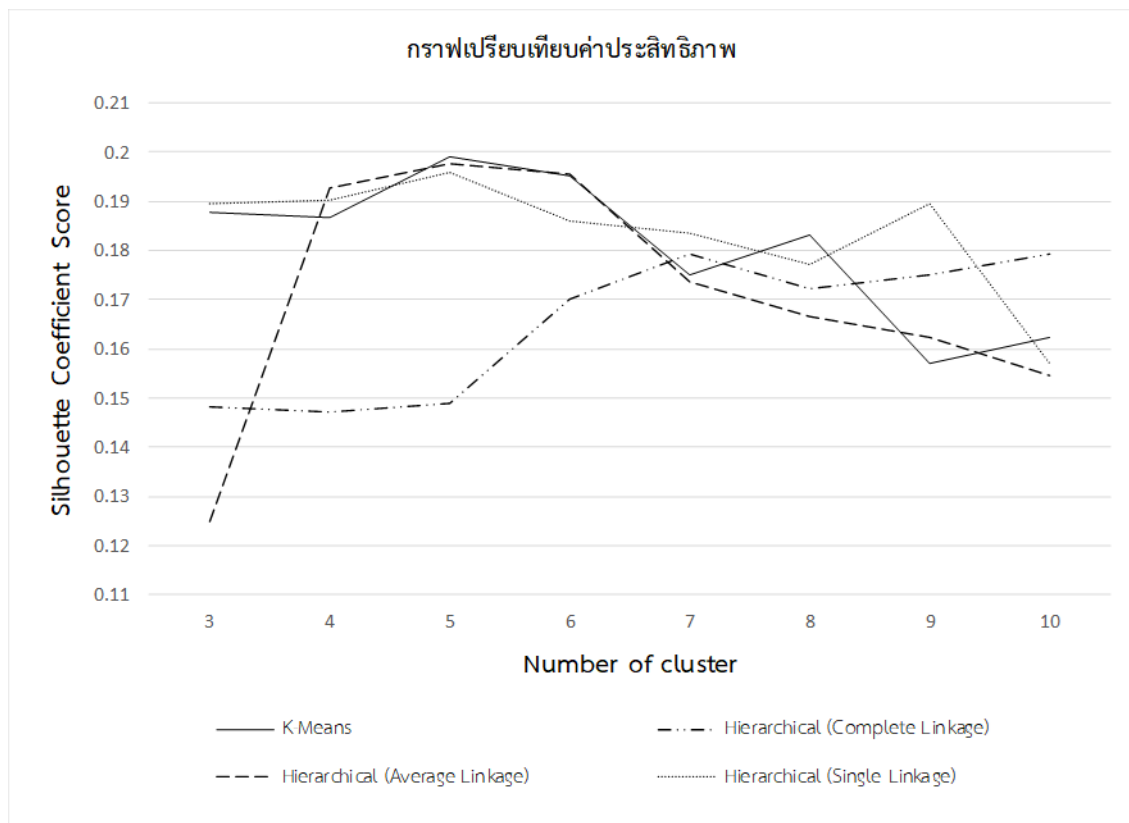
ที่	ปัจจัยหลัก	ปัจจัยย่อย	เครื่องมือในการจัดเก็บ
3	สุนทรียภาพ (Aesthetics)	1. คะแนนการเข้าถึง (Accessibility Score)	Lighthouse Tools
		2. คะแนนการทำตาม Best Practice ของ Modern Web Development (Best Practice Score)	Lighthouse Tools
		3. คะแนนการทำ Search Engine Optimization (SEO Score)	Lighthouse Tools
		4. คะแนน Progressive Web App (PWA Score)	Lighthouse Tools
4	ความเป็นมืออาชีพ (Professionalism)	1. เป็นสมาชิกของชมรมผู้ผลิตข่าวออนไลน์ (Society of Online News Providers: SONP)	Database List
		2. เป็นสมาชิกของสมาคมนักข่าวหนังสือพิมพ์แห่งประเทศไทย (Thai Journalist Association: TJA)	Database List
		3. ข้อมูลปี พ.ศ. เริ่มจด Domain Name	Web Crawler
		4. ข้อมูลปี พ.ศ. หมดอายุ Domain Name	Web Crawler
		5. ข้อมูลนโยบายความเป็นส่วนตัว	Web Crawler
5	ความนิยม (Popularity)	1. ข้อมูล Alexa Traffic Rank	Page Rank Analysis Tools
		2. ข้อมูล Google Pages Indexed	Page Rank Analysis Tools
		3. ข้อมูล Bing Pages Indexed	Page Rank Analysis Tools
		4. ข้อมูล Global Rank	Page Rank Analysis Tools
		5. ข้อมูล Country Rank	Page Rank Analysis Tools
		6. ข้อมูล Total Visits	Page Rank Analysis Tools
		7. ข้อมูล Referring Sites	Page Rank Analysis Tools
		8. ข้อมูล Destination Sites	Page Rank Analysis Tools
		9. ข้อมูล Internal Links	Page Rank Analysis Tools
		10. ข้อมูล External Links	Page Rank Analysis Tools
		11. ข้อมูลการลิงค์ไปยัง Facebook	Page Rank Analysis Tools
		12. ข้อมูลจำนวนผู้ติดตาม Facebook	Page Rank Analysis Tools
		13. ข้อมูลการลิงค์ไปยัง Twitter	Page Rank Analysis Tools
		14. ข้อมูลจำนวนผู้ติดตาม Twitter	Page Rank Analysis Tools
		15. ข้อมูลการลิงค์ไปยัง Instagram	Page Rank Analysis Tools
		16. ข้อมูลจำนวนผู้ติดตาม Instagram	Page Rank Analysis Tools
		17. ข้อมูลการลิงค์ไปยัง YouTube	Page Rank Analysis Tools
		18. ข้อมูลจำนวนผู้ติดตาม YouTube	Page Rank Analysis Tools
6	ความเป็นปัจจุบัน (Currency)	1. ข้อมูลการอัปเดตข่าวสาร	Web Crawler
		2. ข้อมูลการแสดงวันที่ลงข่าวในเนื้อหาข่าว	Web Crawler

2. ผลการจัดกลุ่มข้อมูล (News Source Clustering)

นำข้อมูลปัจจัยที่ส่งผลต่อความน่าเชื่อถือมาทำการจัดกลุ่ม โดยเปรียบเทียบเทคนิคการจัดกลุ่ม 2 เทคนิค ได้แก่ K-Means Clustering และ Hierarchical Clustering โดยกำหนดจำนวนการแบ่งกลุ่มไว้ตั้งแต่ 3 ถึง 10 กลุ่ม สำหรับการประเมินผลการจัดกลุ่มพิจารณาจากค่า Silhouette Coefficient เพื่อดูการกระจายตัวของกลุ่ม ซึ่งผลการประเมินการกระจายตัวของข้อมูลในการจัดกลุ่ม แสดงดังตารางที่ 3 และเปรียบเทียบกับกราฟ แสดงดังภาพที่ 3

ตารางที่ 3 ผลค่าประสิทธิภาพการจัดกลุ่มเว็บไซต์ข่าว

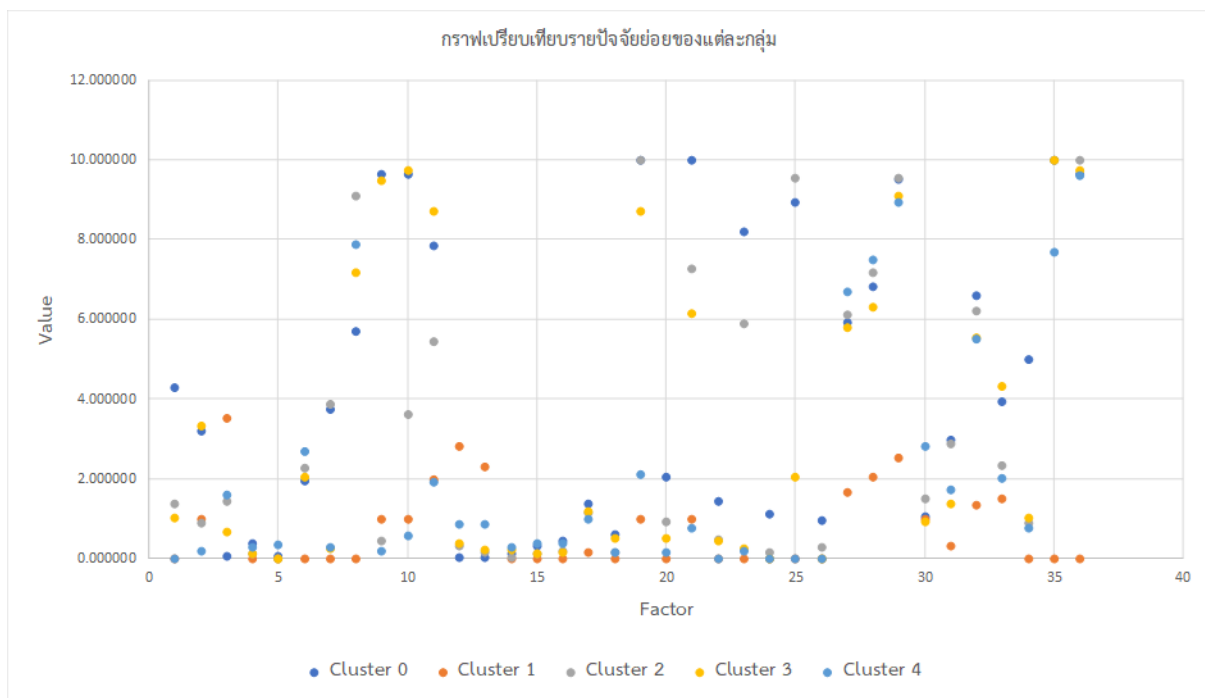
Cluster	Silhouette Coefficient Score			
	K-Means	Hierarchical (Complete Linkage)	Hierarchical (Average Linkage)	Hierarchical (Single Linkage)
3	0.187689	0.148150	0.124906	0.189458
4	0.186799	0.147150	0.192621	0.190231
5	0.199182	0.148919	0.197670	0.195757
6	0.195061	0.169985	0.195653	0.186135
7	0.175228	0.179152	0.173568	0.183528
8	0.183300	0.172215	0.166718	0.177365
9	0.157084	0.175089	0.162353	0.189446
10	0.162346	0.179263	0.154489	0.157237



ภาพที่ 3 กราฟเปรียบเทียบค่าประสิทธิภาพการจัดกลุ่มเว็บไซต์ข่าว

จากตารางและภาพกราฟเปรียบเทียบค่าประสิทธิภาพการจัดกลุ่มเว็บไซต์ข่าวพบว่า เทคนิค K-Means Clustering มีค่าประสิทธิภาพมากที่สุดเมื่อ $K = 5$ (Silhouette Coefficient = 0.199182) เทคนิค Hierarchical Clustering ใช้การเชื่อมต่อแบบ Complete Linkage มีค่าประสิทธิภาพมากที่สุดเมื่อ $K = 10$ (Silhouette Coefficient = 0.179263) เทคนิค Hierarchical Clustering ใช้การเชื่อมต่อแบบ Average Linkage มีค่าประสิทธิภาพมากที่สุดเมื่อ $K = 5$ (Silhouette Coefficient = 0.197670) และเทคนิค Hierarchical Clustering ใช้การเชื่อมต่อแบบ Single Linkage มีค่าประสิทธิภาพมากที่สุดเมื่อ $K = 5$ (Silhouette Coefficient = 0.195757) จะเห็นได้ว่าการแบ่งกลุ่มที่ $K = 5$ มีประสิทธิภาพดีที่สุดจำนวน 3 เทคนิค ส่วน $K = 10$ มีประสิทธิภาพดีที่สุดเพียง 1 เทคนิค ดังนั้นผู้วิจัยจึงทำการแบ่งกลุ่มของเว็บไซต์ข่าวออกเป็น 5 กลุ่ม โดยเทคนิค K-Means Clustering ที่ $K = 5$ และมีค่า Silhouette Coefficient มากที่สุด

เมื่อนำค่าศูนย์กลาง (Centroid) ของแต่ละกลุ่มที่จัดกลุ่มด้วยเทคนิค K-Means Clustering ($K = 5$) มาวาดกราฟเพื่อเปรียบเทียบปัจจัยย่อยของแต่ละกลุ่มเพื่อหาปัจจัยย่อยที่ส่งผลต่อความน่าเชื่อถือของเว็บไซต์ข่าว โดยพิจารณาจากระยะห่างของข้อมูลค่าศูนย์กลางของแต่ละกลุ่ม ถ้าปัจจัยที่มีค่าระยะห่างของค่าศูนย์กลางน้อยแสดงว่ามีอำนาจจำแนกน้อย แต่ถ้าปัจจัยที่มีค่าระยะห่างของค่าศูนย์กลางมากแสดงว่ามีอำนาจจำแนกมาก แสดงดังภาพที่ 4



ภาพที่ 4 กราฟเปรียบเทียบรายปัจจัยย่อยของแต่ละกลุ่ม

จากภาพกราฟเปรียบเทียบรายปัจจัยย่อยของแต่ละกลุ่มจะพบว่า ปัจจัยที่มีค่าระยะห่างมากที่สุด คือ ปัจจัยที่ 25 ข้อมูลการลิงค์ไปยัง YouTube โดยมีค่าระยะห่างของจุดศูนย์กลางเท่ากับ 56.038961 และปัจจัยที่มีค่าระยะห่างน้อยที่สุด คือ ปัจจัยที่ 5 ข้อมูล Bing Pages Indexed โดยมีค่าระยะห่างของจุดศูนย์กลางเท่ากับ 1.480558 เมื่อนำทุกปัจจัยมาทำการแปลความหมายของทั้ง 5 กลุ่มจะได้รายละเอียดของแต่ละกลุ่ม แสดงดังตารางที่ 4

ตารางที่ 4 ผลการแปลความหมายของแต่ละกลุ่ม

กลุ่มที่ 0	
ความถูกต้อง (Accuracy)	- มีการอ้างอิงแหล่งข่าวน้อย - มีการทำลายน้ำในรูปภาพประกอบข่าวมาก
ผู้มีอำนาจหน้าที่ (Authority)	- มีข้อมูลผู้รายงานข่าว/ผู้เขียนข่าวน้อย - มีข้อมูลติดต่อกับแหล่งข่าวมาก
สุนทรียภาพ (Aesthetics)	- เว็บไซต์มีโครงสร้างที่เหมาะสมกับมือถือและ SEO
ความเป็นมืออาชีพ (Professionalism)	- เป็นสมาชิกของสมาคมหรือองค์กรใดที่เกี่ยวข้องสื่อมากที่สุด - มีระยะเวลาในการจด Domain มาก - มีนโยบายรักษาข้อมูลส่วนตัวมากที่สุด
ความนิยม (Popularity)	- เว็บไซต์อยู่ในระดับความนิยมของ Alexa, Google และ Bing มากที่สุด - มีอัตรา Referring Sites มาก - มีการเชื่อมโยงกับ Social Media มาก
ความเป็นปัจจุบัน (Currency)	- มีการอัปเดตข่าวสารตลอดเวลา - มีการแสดงวันที่ลงข่าวในเนื้อหาข่าว
กลุ่มที่ 1	
ความถูกต้อง (Accuracy)	- ไม่มีการอ้างอิงแหล่งข่าว - ไม่มีการทำลายน้ำในรูปภาพประกอบข่าว
ผู้มีอำนาจหน้าที่ (Authority)	- ไม่มีข้อมูลผู้รายงานข่าว/ผู้เขียนข่าว - มีข้อมูลติดต่อกับแหล่งข่าวน้อย
สุนทรียภาพ (Aesthetics)	- เว็บไซต์ไม่มีโครงสร้างที่เหมาะสมหรืออาจถูกปิดไปแล้ว
ความเป็นมืออาชีพ (Professionalism)	- ไม่เป็นสมาชิกของสมาคมหรือองค์กรใดที่เกี่ยวข้องสื่อ - มีระยะเวลาในการจด Domain น้อยที่สุด - ไม่มีนโยบายรักษาข้อมูลส่วนตัว
ความนิยม (Popularity)	- เว็บไซต์อยู่ในระดับความนิยมของ Alexa, Google และ Bing น้อยที่สุด - มีอัตรา Referring Sites น้อยที่สุด - มีการเชื่อมโยงกับ Social Media น้อยที่สุด
ความเป็นปัจจุบัน (Currency)	- ไม่มีการอัปเดตข่าวสาร - ไม่มีการแสดงวันที่ลงข่าวในเนื้อหาข่าว
กลุ่มที่ 2	
ความถูกต้อง (Accuracy)	- มีการอ้างอิงแหล่งข่าวมาก - มีการทำลายน้ำในรูปภาพประกอบข่าวมากที่สุด
ผู้มีอำนาจหน้าที่ (Authority)	- มีข้อมูลผู้รายงานข่าว/ผู้เขียนข่าวมากที่สุด - มีข้อมูลติดต่อกับแหล่งข่าวปานกลาง
สุนทรียภาพ (Aesthetics)	- เว็บไซต์มีโครงสร้างที่เหมาะสมกับมือถือและ SEO

ความเป็นมืออาชีพ (Professionalism)	- เป็นสมาชิกของสมาคมหรือองค์กรใดที่เกี่ยวข้องกับสื่อมาก - มีระยะเวลาในการจด Domain ปานกลาง - มีนโยบายรักษาข้อมูลส่วนตัวน้อย
ความนิยม (Popularity)	- เว็บไซต์อยู่ในระดับความนิยมของ Alexa, Google และ Bing มาก - มีอัตรา Referring Sites ปานกลาง - มีการเชื่อมโยงกับ Social Media มาก
ความเป็นปัจจุบัน (Currency)	- มีการอัปเดตข่าวสารตลอดเวลา - มีการแสดงวันที่ลงข่าวในเนื้อหาข่าว
กลุ่มที่ 3	
ความถูกต้อง (Accuracy)	- มีการอ้างอิงแหล่งข่าวปานกลาง - มีการทำลายน้ำในรูปภาพประกอบข่าวน้อย
ผู้มีอำนาจหน้าที่ (Authority)	- มีข้อมูลผู้รายงานข่าว/ผู้เขียนข่าวปานกลาง - มีข้อมูลติดต่อกับแหล่งข่าวมากที่สุด
สุนทรียภาพ (Aesthetics)	- เว็บไซต์มีโครงสร้างที่เหมาะสมกับมือถือและ SEO
ความเป็นมืออาชีพ (Professionalism)	- เป็นสมาชิกของสมาคมหรือองค์กรใดที่เกี่ยวข้องกับสื่อปานกลาง - มีระยะเวลาในการจด Domain มากที่สุด - มีนโยบายรักษาข้อมูลส่วนตัวน้อย
ความนิยม (Popularity)	- เว็บไซต์อยู่ในระดับความนิยมของ Alexa, Google และ Bing ปานกลาง - มีอัตรา Referring Sites ปานกลาง - มีการเชื่อมโยงกับ Social Media มาก
ความเป็นปัจจุบัน (Currency)	- มีการอัปเดตข่าวสารตลอดเวลา - มีการแสดงวันที่ลงข่าวในเนื้อหาข่าว
กลุ่มที่ 4	
ความถูกต้อง (Accuracy)	- มีการอ้างอิงแหล่งข่าวมากที่สุด - มีการทำลายน้ำในรูปภาพประกอบข่าวปานกลาง
ผู้มีอำนาจหน้าที่ (Authority)	- มีข้อมูลผู้รายงานข่าว/ผู้เขียนข่าวมาก - มีข้อมูลติดต่อกับแหล่งข่าวน้อย
สุนทรียภาพ (Aesthetics)	- เว็บไซต์มีโครงสร้างที่เหมาะสมกับมือถือและ SEO
ความเป็นมืออาชีพ (Professionalism)	- ไม่เป็นสมาชิกของสมาคมหรือองค์กรใดที่เกี่ยวข้องกับสื่อ - มีระยะเวลาในการจด Domain น้อย - มีนโยบายรักษาข้อมูลส่วนตัวน้อย
ความนิยม (Popularity)	- เว็บไซต์อยู่ในระดับความนิยมของ Alexa, Google และ Bing น้อย - มีอัตรา Referring Sites มาก - มีการเชื่อมโยงกับ Social Media น้อย
ความเป็นปัจจุบัน (Currency)	- มีการอัปเดตข่าวสารปานกลาง - มีการแสดงวันที่ลงข่าวในเนื้อหาข่าว

3. ผลการพัฒนาเครื่องมือตรวจสอบความน่าเชื่อถือ

การสร้างเครื่องมือสำหรับตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าวถูกพัฒนาในรูปแบบเว็บแอปพลิเคชัน (Web Application) ซึ่งจะนำข้อมูลที่ทำกรจัดกลุ่มแล้วมาประยุกต์ใช้ในการตรวจสอบจากลิงค์ (URL) ของข่าวสารที่ผู้ใช้งานต้องการตรวจสอบ แล้วแจ้งว่าเว็บไซต์ข่าวถูกจัดกลุ่มความน่าเชื่อถืออยู่ในกลุ่มใด โดยเครื่องมือตรวจสอบจากลิงค์ที่ผู้ใช้นำมาตรวจสอบแล้วทำการสกัดเอา Domain Name ของเว็บไซต์แหล่งข่าว จากนั้นนำมาจัดกลุ่มว่าอยู่ในกลุ่มเว็บไซต์ข่าวที่มีลักษณะอย่างไร แล้วแสดงผลลัพธ์ให้ผู้ใช้งานทราบ ถ้าลิงค์ที่ผู้ใช้งานนำมาตรวจสอบไม่พบอยู่ในกลุ่มที่เคยจัดกลุ่มเอาไว้ ระบบจะทำการบันทึกข้อมูลลิงค์นั้นไว้เพื่อนำมาแจ้งผู้ดูแลระบบ แล้วนำเว็บไซต์ดังกล่าวมาวิเคราะห์ด้วยการจัดกลุ่มในรอบถัดไป โดยสามารถเข้าไปใช้งานเครื่องมือได้ที่ www.aongart.com/news ซึ่งตัวอย่างผลการสร้างเครื่องมือดังกล่าว แสดงดังภาพที่ 5

ปัจจัย	รายละเอียด
ความถูกต้อง (Accuracy)	- มีการอ้างอิงแหล่งข่าวน้อย - มีการทำลายน้ในรูปภาพประกอบข่าวมาก
ผู้มีอำนาจ (Authority)	- มีข้อมูลผู้รายงานข่าว/ผู้เขียนข่าวน้อย - มีข้อมูลติดต่อกับแหล่งข่าวมาก
สุนทรียภาพ (Aesthetics)	- เว็บไซต์มีโครงสร้างที่เหมาะสมกับมือถือและ SEO
ความเป็นมืออาชีพ (Professionalism)	- เป็นสมาชิกของสมาคมหรือองค์กรที่เกี่ยวข้องมากที่สุด - มีระยะเวลาในการจด Domain มาก - มีนโยบายรักษาข้อมูลส่วนตัวมากที่สุด
ความนิยม (Popularity)	- เว็บไซต์อยู่ในระดับความนิยมของ Alexa, Google และ Bing มากที่สุด - มีอัตรา Referring Sites มาก - มีการเชื่อมโยงกับ Social Media มาก
ความเป็นปัจจุบัน (Currency)	- มีการอัปเดตข่าวสารตลอดเวลา - มีการแสดงวันที่ลงข่าวในเนื้อหาข่าว

ภาพที่ 5 เครื่องมือสำหรับตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าว

อภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์ในการศึกษาปัจจัยที่ส่งผลต่อความน่าเชื่อถือของเว็บไซต์ข่าวภาษาไทย และนำไปพัฒนาเครื่องมือที่ใช้ในการตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าว โดยใช้เทคนิคการเรียนรู้ของเครื่องด้วยวิธีการจัดกลุ่ม เนื่องจากเว็บไซต์ข่าวไม่มีการติดป้ายกำกับ (Label) มาก่อนว่าแต่ละเว็บไซต์มีความน่าเชื่อถือหรือไม่ จึงต้องมีการจัดกลุ่มข้อมูล (Data Clustering) ของเว็บไซต์ข่าว ซึ่งเปรียบเทียบเทคนิคในการจัดกลุ่ม 4 เทคนิค ได้แก่ K-Means Clustering, Hierarchical Clustering (Complete Linkage, Average Linkage และ Single Linkage)

โดยดึงข้อมูลจากเว็บไซต์ข่าวจำนวน 152 เว็บไซต์ ข้อมูลประกอบไปด้วยปัจจัยหลัก 6 ปัจจัย ได้แก่ ความถูกต้อง (Accuracy) ผู้มีอำนาจหน้าที่ (Authority) สุนทรียภาพ (Aesthetics) ความเป็นมืออาชีพ (Professionalism) ความนิยม (Popularity) และความเป็นปัจจุบัน (Currency) โดยมีปัจจัยย่อยทั้งหมด 35 ปัจจัย ซึ่งผลลัพธ์ที่ได้จากการจัดกลุ่มจะถูกวัดค่าประสิทธิภาพในการจัดกลุ่มด้วยค่า Silhouette Coefficient เพื่อใช้เป็นตัวแบบในการนำไปพัฒนาเครื่องมือตรวจสอบความน่าเชื่อถือของเว็บไซต์ข่าว ซึ่งใช้การจัดกลุ่มเพื่อติดป้ายกำกับแหล่งข่าวเหมือนงานวิจัยของ Krishnamoorthy, et al. (2018) ทำให้การวิเคราะห์สามารถทำได้มากขึ้น

เมื่อเปรียบเทียบประสิทธิภาพการจัดกลุ่มโดยในขั้นตอนของการจัดกลุ่มข้อมูลเว็บไซต์ข่าวพบว่า เทคนิค K-Means Clustering มีค่าประสิทธิภาพมากที่สุดเมื่อ $K = 5$ (Silhouette Score = 0.199182) ซึ่งสอดคล้องกับงานของ Krishnamoorthy, et al. (2018) และ Nanayakkara & Ranathunga (2018) จึงทำการติดป้ายกำกับแหล่งข่าวด้วยเทคนิค K-Means โดยปัจจัยที่มีค่าอำนาจจำแนกมากที่สุดคือ ปัจจัยที่ 25 ข้อมูลการลิงค์ไปยัง YouTube โดยมีค่าระยะห่างของจุดศูนย์กลางเท่ากับ 56.038961 หลังจากผ่านกระบวนการทำการจัดกลุ่มแล้ว นำมาสร้างเป็นเครื่องมือตรวจสอบในรูปแบบเว็บแอปพลิเคชัน ผู้ใช้งานสามารถคัดลอก (Copy) ลิงค์ของข่าวที่ต้องการตรวจสอบมาตรวจสอบได้อย่างสะดวก และช่วยในการตัดสินใจที่จะเชื่อหรือไม่เชื่อข่าวที่นำมาตรวจสอบได้ โดยเมื่อได้นำข่าวปลอมมาทดสอบแล้วพบว่าข่าวปลอมจะถูกจัดอยู่ในกลุ่มแหล่งข่าวที่มีความน่าเชื่อถือน้อยถึงน้อยที่สุด

ข้อเสนอแนะ

1. ควรมีการพัฒนาเนื่องจากในบางปัจจัยที่ส่งผลต่อความน่าเชื่อถือยังไม่สามารถดึงข้อมูลออกมาได้อย่างอัตโนมัติทำให้การทำงานยังไม่สามารถเป็นอัตโนมัติทั้งระบบได้ เช่น ข้อมูลลายน้ำในรูปภาพประกอบข่าว ข้อมูลการเป็นสมาชิกของชมรมผู้ผลิตข่าวออนไลน์ ข้อมูลนโยบายความเป็นส่วนตัว เป็นต้น
2. ควรมีการพัฒนาเครื่องมือที่จัดเก็บข้อมูลปัจจัยบางอย่างที่ดึงข้อมูลจาก API ซึ่งเป็นผู้ให้บริการจากบุคคลที่สาม (Third-Party) ทำให้มีโอกาสที่ผู้ให้บริการอาจจะรับ ยกเลิก หรือเก็บค่าบริการได้ เช่น คะแนนการเข้าถึง (Accessibility Score) คะแนนการทำตาม Best Practice ของ Modern Web Development (Best Practice Score) คะแนนการทำ Search Engine Optimization (SEO Score) ที่ดึงจาก Lighthouse Tools
3. ควรเพิ่มปัจจัยอื่นที่เกี่ยวข้องเข้ามาช่วยในการจัดกลุ่มเว็บไซต์ข่าวได้ เช่น ปัจจัยด้านความเป็นกลางของเนื้อหา ซึ่งจะต้องทำการวิเคราะห์อารมณ์ของเนื้อหาว่ามีความเป็นกลาง หรือเอนเอียงไปทางใดทางหนึ่งหรือไม่ หรือปัจจัยด้านคุณภาพ ต้องทำการวิเคราะห์ความถูกต้องในการพิมพ์เนื้อหาข่าว หรือเนื้อหาข่าวใกล้เคียงกับแหล่งข่าวอื่นหรือไม่

เอกสารอ้างอิง

- Aun-a-nan, A., Manthung, W., Kaenkaew, S. and Wannapiroon, N. (2016). The Study and Systematic Literature Review: The Analysis of Mobile Usage Behavior and Reliability in Content of News on Online Social Network. *The Proceedings of National Graduate Research Conference (NGRC)*, Pathumthani, Thailand, 884-897. (in Thai)

- Aun-a-nan, A. and Meesad, P. (2018). Web Crawler Optimization for Extracting Thai Language News Online. *The Proceedings of National Conference on Information Technology (NCIT)*, Khonkan, Thailand, 121-127. (in Thai)
- Flanagin, A. J. and Metzger, M. J. (2008). *Digital media and youth: unparalleled opportunity and unprecedented responsibility*. Cambridge. MA: MIT Press.
- Krishnamoorthy, A., Patil, A. K., Vasudevan, N. and Pathari, V. (2018). News Article Classification with Clustering using Semi-Supervised Learning. *The Proceedings of International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, India, 86-91.
- Nanayakkara, P. and Ranathunga, S. (2018). Clustering Sinhala News Articles Using Corpus-Based Similarity Measures. *The Proceedings of Moratuwa Engineering Research Conference (MERCon)*, Moratuwa, 437-442.
- National Statistical Office. (2018). *The 2018 Household Survey on the Use of Information and Communication Technology*. Bangkok: Economic and Social Statistics Bureau. (in Thai)
- Noosom, N. and Suttisima, V. (2019). The Analysis of Fake News and The Level of Media Literacy of Users in Bangkok. *Journal of Communication Arts*, 37 (1), 37-45. (in Thai)
- Patricia, R. T., Melancon, J. (2018). Gender and live-streaming: source credibility and motivation. *Journal of Research in Interactive Marketing*, 12 (1), 79-93.
- Polkett P., Salathong J., (2018). Audience's Perception and Media Literacy Of "Sure and Share" TV Program. *Journal of Communication and Management NIDA*, 4 (3), 47-62. (in Thai)
- Rousseeuw, P. J. (1987). Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics*, 20, 53-65.
- Shah, A. A. and Ravana, S. D. (2014). Evaluating information credibility of digital content using hybrid approach. *International Journal of Information Systems and Engineering*, 2 (1), 92-99.
- Shah, A. A., Ravana, S. D., Hamid, S. and Ismail, M. A. (2015). Web credibility assessment: affecting factors and assessment techniques. *Information Research*, 20 (1), 655-672.