Hybrid Feature Selection-Based Sentiment Analysis Concept: Case-Study of E-Commerce Opinion Mining Model

Eddie Bouy B. Palad[1], Khin Sandar Kyaw[2*], Tharnpas Sattayaraksa[3], Praman Tepsongkroh[4] and Farida Sasha[5]

[1] Assist. Prof., MSU-Iligan Institute of Technology, Iligan City, Philippines.

[2] Dr., Lecturer, Department of International Business Management, Didyasarin International College, Hatyai University, Thailand.

[3] Dr., Lecturer, Department of International Business Management, Didyasarin International College, Hatyai University, Thailand.

[4] Assoc. Prof. Dr., Dean, Department of Airline Business, Didyasarin International College, Hatyai University, Thailand.

[5] Dr., Lecturer, Department of Management, Hatyai Business School, Hatyai University, Thailand.

[*]**Corresponding author, E-mail:** khinsandarkyaw@hu.ac.th

Abstract

As technology continuously advances, trading is progressively transforming from a manual offline platform to an online web-based platform known as E-commerce. In the E-commerce system, selling and buying, transaction, and customer reviews are performed easily, and business owners can obtain customers information including personal data, customer behavior information including shopping record, web usage, and opinion. This information can be used to improve customer relationship management in the future. Customer opinion has become the most influential parameter that expresses a product image to business owner. This paper presents a case study of E-commerce based on customer opinion mining using sentiment analysis. In addition, the concept of automatic sentiment analysis-based opinion mining model using hybrid feature selection with supervised machine learning model is proposed in this paper for business knowledge discovery from the E-commerce system.

Keywords: Opinion Mining, Classification, Supervised Learning, Web Mining, Business Intelligent, Natural Language Processing

Introduction

The E-commerce system has become a mainstream channel for products purchase and sell due to the rapid development of Internet Technology. On the E-commerce platform, customers can submit their reviews and comments based on personal experience regarding products quality and service. Costumers' perception and opinion holds greater potential for knowledge discovery and decision making by presenting valuable insight to business owner for the development of customer relationship management, product quality control, market prediction system, and supply change management system.

Therefore, opinion mining has become a hot trend of research for optimizing decisions and business intelligence given the growing rate of online resources and rich resources of opinion shared via online review sites and personal blogs. Opinion mining for E-commerce includes web mining and text mining scheme with sentiment analysis on numerous user reviews on the online web-based platforms (Jabbar, 2019). Opinion mining is the automatic extraction of knowledge from the individual opinions on a particular interest or problem. In addition, classification of customer reviews using sentiment analysis helps businesses owners monitor brand and product sentiment in customer feedback and understand customer needs. Sentiment analysis is a natural language processing technique used for determining polarity (positive, negative, neutral), emotions (happy, sad, etc), urgency (urgent, not urgent), and even intentions (interested, not interested) (Zuheros, Martínez-Cámara, Herrera-Viedma, & Herrera, 2021).

There are several types of sentiment analysis including fine-grained sentiment analysis, emotion detection, aspect-based sentiment analysis, and multilingual sentiment analysis (Mowlaei, Abadeh, & Keshavarz, 2020). Sentiment analysis works together with natural language processing (NLP) and machine learning algorithms, to determine the opinion tone of online text data from E-commerce, social media, blog, etc. Sentiment algorithms can be classified into three types including rule-based, automatic, and hybrid. In the rule-based sentiment algorithm, sentiment analysis is performed based on a set of manually crafted rules whereas the automatic systems rely on the machine learning techniques to adapt with data, and the hybrid systems incorporates both rule-based and automatic approaches. This paper conceptualizes automatic customer opinion mining using hybrid feature selection-based sentiment analysis with supervised learning model as

a case study for the E-commerce system. This article reviews trends in techniques and methods that provides opinion-oriented information from various customer opinion mining-based business intelligence models. In addition, the research presents challenges related to sentiment analysis and opinion mining.

Objectives

This research intends to achieve the following outcomes:

1. To enhance customer relationship management (CRM) in the E-commerce system through customer opinion mining

2. To establish a link between natural language processing (NLP) and E-commerce trading using sentiment analysis

Literature Review

A novel semi-supervised fuzzy product ontology mining algorithm was proposed to extract semantic knowledge from online customer reviews with positive or negative labels. They performed sentiment analysis at fine-grained level to explore eWOM of products. They applied a feature-based and context-sensitive sentiment analysis mechanism that can leverage the sheer volume of customer reviews on social media sites. The proposed system achieved remarkable performance improvement over baseline methods (Sun, Niu, Yao, & Yan, 2019). However, this work suggested to continuous enhancement of positive and negative opinion words extraction and polarity computation and to explore the new sentiment analysis system. Jabbar (2019) implemented a model on E-commerce application in which sentiment analysis and Support Vector Machine (SVM) was used to classify the polarity of product reviews from Amazon.com. The validity of the proposed algorithm was proved through comparison with well-known sentiment information extraction algorithms, general word counting and SentiStrength. Applying the labelled customer feedbacks on the Amazon dataset, the proposed algorithm extracted sentiments more correctly than the general word counting and SentiStrength algorithms, especially in the negative cases. The authors suggested to improve the sentiment analysis scheme using advanced machine learning and deep learning approaches to handle complex structure of sentence. In (Park, 2020), an online review-based process for evaluation of customer satisfaction process was represented to sort out issue of inefficient

manual monitoring customer satisfaction and service quality. The authors developed a systematic approach for the evaluation of relative customer satisfaction via sentiment analysis and statistical data analysis, and interpreted determinants of positive and negative opinions via TF-IDF analysis. To illustrate the efficacy and applicability of the proposed system, an empirical case study was conducted on the global top 26 cosmetics brands, and the average positive of 35%, negative of 10%, and neutral opinions of 55 %were 35, 10, and 55% were obtained. The authors found that the positive opinion ratio was higher than the negative opinion ratio for most brands. As a future work, the authors pointed out to develop a more highly accurate sentiment for more sophisticated sentiment analysis.

Kothalawala & Thelijjagoda (2020) proposed Aspect-based Sentiment Analysis model for getting insights into consumer opinions and consumers purchase decisions. It consisted of several stages including data gathering, pre-processing, aspect extraction and polarity detection and followed a sequential approach to achieve the intended goal. The implemented system demonstrated an accuracy of 85% from the test data for overall aspects, enabling consumers to get an immediate idea about public opinion for manufacturers to identify their strong and weak points. The authors provided the further improvement of dynamic detecting the aspects and the accuracy of the system in terms of both aspect and polarity should be enhanced with the aid of more training data.

Xiao, et al. (2020), proposed a new sentiment analysis model (SLCABG) based on sentiment lexicon that combines Convolutional Neural Network (CNN) and attention-based Bidirectional Gated Recurrent Unit (BiGRU). The proposed SLCABG model combines the advantages of sentiment lexicon and deep learning technology and overcomes the shortcomings of existing sentiment analysis model for product reviews. Crawled of the real book evaluation from dangdang.com was used for training and testing model. The experimental results showed that the proposed model can effectively improve the performance of text sentiment analysis. Although the proposed system performed well predicting product reviews, further improvement and optimization will be needed to achieve satisfying effects in the prediction of some short texts or unseen data set.

Research Methodology

Background Knowledge

This section describes the background knowledge related to this research. It includes web mining, text mining and feature engineering, opinion mining, and supervised learning model-based sentiment analysis.

Web Mining

Web mining is one of the data mining application employed to discover patterns, structures, and knowledge from the Web using machine learning methods based on the various mathematical models. Web mining includes learning information distribution on the WWW such as categorization of web page, summarization of the web content, association and other relationships among web pages, online users, communities, and web-based activities.

According to the analysis targets, web mining approach can be classified into three categories including web content mining, web structure mining and web usage mining. However, web content mining is concerned with opinion mining mainly because it is the process of extracting useful information from the contents of web documents. The most common application of web content mining includes topic discovery, extracting association patterns, clustering of web documents, web page content summarization, and other information retrieval process. Research activities in this field also involve using techniques from other disciplines such as information retrieval (IR) and natural language processing (NLP).

Text Mining and Feature Engineering

Text mining is a concept of finding patterns in text. In this regard, text is very amorphous, and difficult to handle with than numeric data in the process of data mining. In addition, text mining is applied on several applications such as information extraction, computational linguistics, categorization, clustering, topic tracking spam filtering, sentiment analysis, etc. However, the application domains of text mining can be broadly organized into two groups, namely, document exploration and analysis tools.

Meanwhile, feature engineering process plays an important role to extract and select relevant feature from high-dimensional text feature for specific problem domain such as customer opinion mining. Feature engineering process constructs feature set using feature extraction and feature selection schemes to leverage performance of machine

learning algorithms by reducing feature dimension. Feature extraction performs combinations of original features and projects them into a lower dimensional space to form new features. There are various text feature extraction schemes including information gain, entropy (Xie, Ge, Hu, Xie, & Jiang, 2019), mutual information (Gao, Hu, & Zhang, 2020), gini index (Manek, Shenoy, Mohan, & Venugopal, 2017), term frequency- inverse document frequency (Kadhim, 2019), etc.

Feature selection is the selection of a subset of features that are highly relevant using a criterion measure. Feature selection techniques can be broadly categorized into three including filter (Bommert, Sun, Bischl, Rahnenführer, & Lang, 2020), wrapper (Gokalp, Tasci, & Ugur, 2020) and hybrid (Zarisfi Kermani, Eslami, & Sadeghi, 2019). Filter and Wrapper methods constitutes the group of supervised feature selection techniques. Filter methods utilizes a statistical measure such as correlation (Karegowda, Manjunath, Ratio, & Evaluation, 2010), Chi-Square (Alshaer, Otair, Abualigah, Alshinwan, & Khasawneh, 2021) and Information Gain (Novelty Octaviani Faomasi Daeli, 2020). In addition, filter methods prevent feature bias from interacting with training algorithm since they are not dependent on training algorithm. On the other hand, wrapper models heuristically select features depending on the machine learning algorithm and hence they are computationally expensive and infeasible if the number of features is huge.

Opinion Mining

Opinion mining deals with opinion extraction and sentiment analysis from unstructured text documents. Opinion mining is performed through opinion extraction and structurization for aggregation and analysis of opinions concerned with the pre-defined subjects. Basically, opinion extraction is processed through several stages including the identification of opinion holder, the subject being reviewed, the part or the feature of the subject that is being evaluated and classification of opinion as either positive or negative or some other pre-defined opinion category. Meanwhile, structurization transforms the extracted opinion expressions into structures suitable for assimilation and analysis.

The fundamental opinion mining framework is integrated with two core modules including data acquisition and pre-processing, and opinion extraction and mining. The data acquired using site-specific crawlers are stored along with associated information available. Pre-processing is the process of removing noise from the acquired data which

are customizable to adapt to different domains. On the other hand, opinion extraction performs identification of fragments within a sentence that express opinions about a relevant subject and stores them in pre-defined templates which are subjected to different analytics for generating collective opinions.

Supervised Learning Model-based Sentiment Analysis

Learning model approaches employed to manipulate sentiment analysis can be categorized into two groups including machine learning, and lexicon-based approach (Pitogo & Ramos, 2020). Word reference based and corpus-based methodology can be utilized to consolidates AI with the vocabulary-based methodology. Lexicon-based approach utilize a lexicon to present opinions through counting and measuring words related to sentiments. Lexicon-based approach involves several schemes such as dictionary-based (Kumar & Babu, 2020), corpus-based (Darwich, Mohd Noah, Omar, & Osman, 2019), and hybrid scheme. Machine learning approach includes supervised, semi-supervised, and unsupervised. Unsupervised learning is the process of learning by observation such as clustering and association approach. The common unsupervised learning algorithms used in sentiment analysis are k-means for clustering algorithms, mixture model, and hierarchical clustering. Semi-supervised learning is suitable for unlabeled data classification problem with a small subset of observations because it can hold a lot of class information on joint distribution over classification features.

However, supervised learning is based on the labeled dataset experience during decision-making to produce appropriate outputs. Supervised learning is divided into four categories including linear classification, rule-based and case-based classification (Berka, 2020), probabilistic classification, and decision tree concepts. Basically, supervised machine learning comprises of the training and testing pats. In other words, training part includes corpora utilized to gain proficiency with a “classifier” model and testing part utilizes the training data set to characterize inconspicuous information. Therefore, supervised machine learning model has an over-fitting issue which fits the training data too well but generally poor to testing data. Some common supervised machine learning classifiers are Support Vector Machine (SVM) (Dey et al., 2020), Naïve Bayes (NB) (Bayhaqy, Sfenrianto, Nainggolan, & Kaburuan, 2018), Artificial Neural Network (ANN) (Huang et al., 2020), and Decision Tree (DT) classifiers (Adnan, Sarno, & Sungkono, 2019). This article focuses on the supervised machine learning in sentiment classification.

Background Methodology

The proposed system concept includes several operation phases including crawling data from e-commerce; extracting main content from corresponding domain; cleaning the collected data; feature engineering including feature extraction, feature selection, and feature reduction; and opinion discovery process based on sentiment analysis with supervised machine learning model. The detailed functionalities are described in the following sub-sections.

Proposed Sentiment Analysis-based Opinion Mining Model

The proposed sentiment analysis-based opinion mining model is illustrated in Figure 1 and it comprises several phases to classify the customer opinion from E-commerce platform. In E-commerce data crawling phase, the E-commerce website of interest is plugged with crawler tool to collect web content. The collected web content is passed into main content extraction phase to rid it of irrelevant content for the corresponding problem domain by manipulating the text density with text block concept. The extracted text document is given into the data cleaning phase as an input for opinion mining model. Data cleaning phase utilize the text mining concept to eliminate noisy data using tokenization, article removing, stop-word handling, stemming, lemmatizing, spelling correction, and normalization.

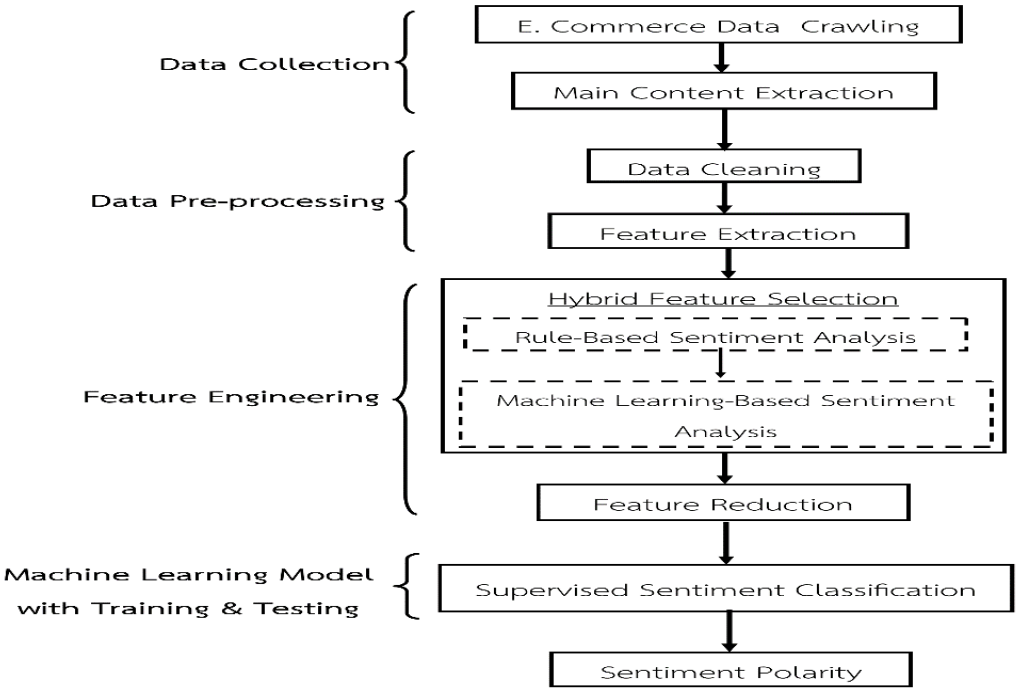

Figure 1 Sentiment Analysis-based Opinion Mining Model

Feature extraction phase accepts the cleaned data and manipulates important feature based on text frequency-inverse document frequency (TF-IDF) to build the feature vector. The proposed system uses the hybrid feature selection scheme which is integrated with rule-based sentiment analysis and Machine-Learning based sentiment analysis to select relevant feature for opinion mining model. The selected feature has the characteristic of multi-dimensional and therefore feature reduction scheme such as principal component analysis is applied to compress the feature by mapping a new feature onto the original feature. The compressed features (opinion) are passed into the supervised Machine Learning model to classify the sentiment polarity based on the training model.

E-Commerce Data Crawling and Main Content Extraction

Web crawlers are the programs for data acquisition from the Web by following hyperlinks. These are also called as spiders or robots or wanderers. A basic Web crawler considers a set of seed URLs as input and adds URLs found on this webpage to the list of URLs to be visited further and generates a set of crawled webpages as output. Basically, every Web crawler follows a set of policies including politeness policy, parallelization policy, revisit policy, and robustness policy. Politeness policy defines only allowed webpages can grant crawling access while parallelization policy includes the rule for assigning new URLs discovered during crawling process to different threads running in parallel. Revisit policy presents either a uniform revisit policy or a proportional revisit policy can be used for deciding when to revisit a webpage. In robustness policy, a crawler must be immune to malicious behavior of any Web server.

Main content extraction is applied on web crawling process to obtain relevant web content for specific problem domain. Basically, main content extraction integrates two aspects namely, noise removal (filtering) and selection of main content (extraction of text content from the given HTML) document. Although several schemes for main content extraction have been developed, the proposed system applies HTML document object model (DOM) tree-based web crawling with TF-IDF based text-block concept to extract text documents (e-commerce reviews / comments) from the corresponding tags.

Data Cleaning and Feature Extraction

Data cleaning is the process of removing irrelevant and redundant data from the accumulative data set. The fundamental pre-processing of text data for opinion mining includes tokenization, stop-words handling, stemming, lemmatizing, spelling checking and normalization. The functional explanation for the data cleaning steps is described in Figure 2. In addition, TF-IDF based feature extraction scheme is used to implement feature vector for future feature selection process. TF-IDF is a statistical weighting method for retrieving the importance of a term in a document for text mining with consideration of the number of occurrences of term in a document and the importance of a term in a whole collection.

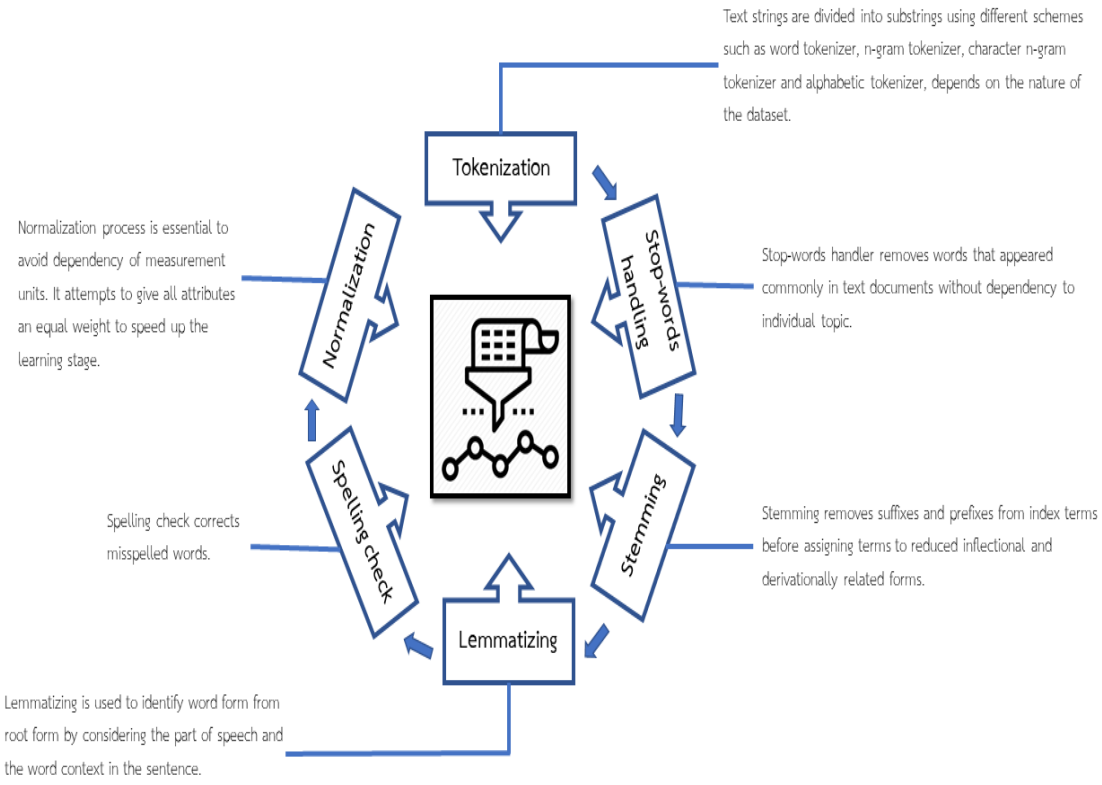

Figure 2 Process Diagram of Data Cleaning

Hybrid Feature Selection-based Sentiment Analysis

Feature selection process performs relevant feature selection for the specific problem domain. This research proposes the hybrid feature selection model to select relevant feature for opinion mining in E-commerce. Rule-based feature selection is integrated with machine learning-based feature selection in the proposed model. Rule-based feature selection is applied on the extracted feature vector to select opinion

expression using part of speech (POS), dictionary-based, corpus and lexicon-based, and so on. The selected opinion features are passed into machine-learning based to discover the relationship between features and their corresponding class (positive or negative sentiment).

Machine Learning-based feature selection has two main schemes such as filter and wrapper. The proposed model uses filter approach to select relevant feature which is independent of the learning model. In specific, correlation-based feature selection method is proposed for this research to overcome the problem of bias in selecting features. Its hypothesis includes heuristic approach for considering features that are highly correlated with predictive label, but uncorrelated with other labels. In addition, it calculates the correlation between features to search for feature subset.

Feature Reduction

Feature reduction is the process of reducing the dimension of text feature in feature vector. Although there are several schemes for the feature reduction process, the proposed system uses Principal Component Analysis (PCA) to create new features by using linear combination of original features. The processing module of PCA includes normalization, covariance matrix, Eigen value and Eigen vector, transformation of column vector, and forming principal component using the multiplication of transformed feature vector and transformed scaled feature.

Supervised Machine Learning-based Opinion Classification

Supervised sentiment or opinion classification aims to categorize the comments/ reviews of customers for certain products or services from E-commerce system into positive or negative opinions based on customers' experience training data set.

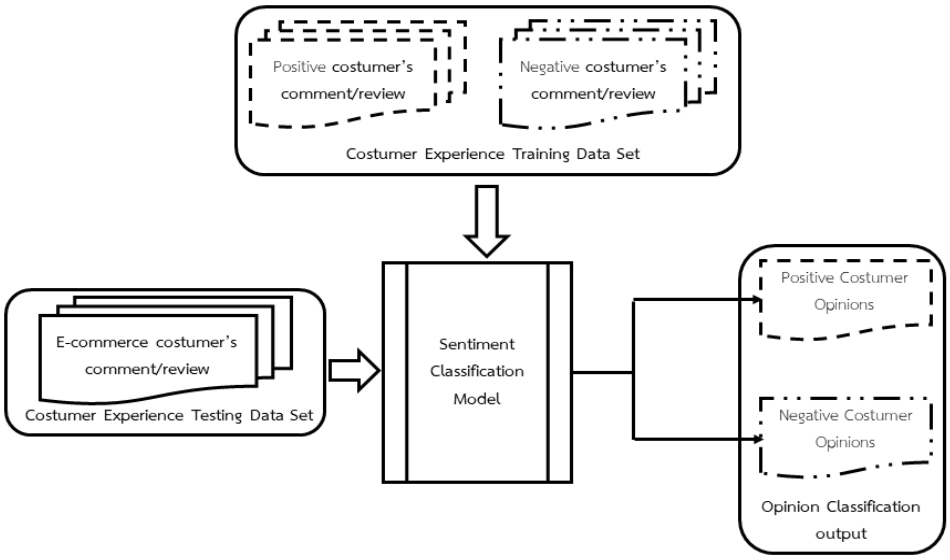

Figure 3 Opinion Classification Model using Supervised Machine Learning Approach

Figure 3 demonstrates the functional diagram of supervised opinion classification model. The most well-known supervised machine learning methods such as the support vector machine (SVM), Naïve Bayes, and the N-gram model can be used to build the sentiment classification model, and the proposed system will use Naïve Bayes, and support vector machine to classify the polarity of opinion.

Performance Evaluation Scheme

Performance evaluation is the process that measure the correctness of the proposed model. Since this proposed system is concerned with opinion classification, the evaluation will be considered by classification performance such as accuracy, precision, recall and F-score. The equation and definition for each measurement is summarized in Table 1.

Table 1 Performance Evaluation Scheme for Sentiment Classification

Equation	Definition
$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$	▪ Correctness of true positive and true negative
$Precision = \frac{TP}{TP + FP}$	▪ Ratio of predicted positive observations to the total number of positive observations
$Recall = \frac{TP}{TP + FN}$	▪ Ratio of correctly predicted positive observations to all observations in actual class
$F_{score} = \frac{2 * Precision * Recall}{Precision + Recall}$	▪ Weighted average of recall and precision

Conclusion and Future Work

The proposed system is intended to enhence customer relationship management (CRM) in the E-commerce system through the classification of the polarity of the customer using hybrid feature selection-based sentiment analysis and supervised machine learning model. In other words, the proposed system is excepted to establish a link between natural language processing and E-commerce trading using sentiment analysis scheme after implementing the system in future. The implementation of the proposed system with experimental set up and performance testing of the proposed system with other systems including rule-based only, machine learning based only on both benchmark data set and real-time crawled data set from E-commerce will be presented as the future works for this research.

Acknowledgement

The authors wish to acknowledge the Department of International Business Management, Didyasarin International College, Hat Yai University and Dr Ozioma F. Nwabor for his assistance throughout the manuscript preparation phase.

References

Adnan, M., Sarno, R., & Sungkono, K. R. (2019). Sentiment Analysis of Restaurant Review with Classification Approach in the Decision Tree-J48 Algorithm. Proceedings - 2019 International Seminar on Application for Technology of Information and Communication: Industry 4.0: Retrospect, Prospect, and Challenges, ISemantic 2019, 121–126. https://doi.org/10.1109/ISEMANTIC.2019.8884282

Alshaer, H. N., Otair, M. A., Abualigah, L., Alshinwan, M., & Khasawneh, A. M. (2021). Feature selection method using improved CHI Square on Arabic text classifiers: analysis and application. Multimedia Tools and Applications, 80(7), 10373–10390. https://doi.org/10.1007/s11042-020-10074-6

Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes. 2018 International Conference on Orange Technologies, ICOT 2018, 1–6. https://doi.org/10.1109/ICOT.2018.8705796

Berka, P. (2020). Sentiment analysis using rule-based and case-based reasoning. 51–66.

Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., & Lang, M. (2020). Benchmark for filter methods for feature selection in high-dimensional classification data. Computational Statistics and Data Analysis, 143, 106839. https://doi.org/10.1016/j.csda.2019.106839

Darwich, M., Mohd Noah, S. A., Omar, N., & Osman, N. A. (2019). Corpus-Based Techniques for Sentiment Lexicon Generation: A Review. Journal of Digital Information Management, 17(5), 296. https://doi.org/10.6025/jdim/2019/17/5/296-305

Dey, S., Wasif, S., Tonmoy, D. S., Sultana, S., Sarkar, J., & Dey, M. (2020). A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews. 2020 International Conference on Contemporary Computing and Applications, IC3A 2020, 217–220. https://doi.org/10.1109/IC3A48958.2020.233300

Gao, W., Hu, L., & Zhang, P. (2020). Feature redundancy term variation for mutual information-based feature selection. Applied Intelligence, 50(4), 1272–1288. https://doi.org/10.1007/s10489-019-01597-z

Gokalp, O., Tasci, E., & Ugur, A. (2020). A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. Expert Systems with Applications, 146, 113176. https://doi.org/10.1016/j.eswa.2020.113176

Huang, M., Xie, H., Rao, Y., Liu, Y., Poon, L. K. M., & Wang, F. L. (2020). Lexicon-Based Sentiment Convolutional Neural Networks for Online Review Analysis. IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2020.2997769

Jabbar, J. (2019). Real-time Sentiment Analysis On E-Commerce Application. 391–396.

Kadhim, A. I. (2019). Term Weighting for Feature Extraction on Twitter: A Comparison between BM25 and TF-IDF. 2019 International Conference on Advanced Science and Engineering, ICOASE 2019, 124–128. https://doi.org/10.1109/ICOASE.2019.8723825

Karegowda, A. G., Manjunath, A. S., Ratio, G., & Evaluation, C. F. (2010). Comparative study of Attribute Selection Using Gain Ratio. International Journal of Information Technology and Knowledge and Knowledge Management, 2(2), 271–277. Retrieved from https://pdfs.semanticscholar.org/3555/1bc9ec8b6ee3c97c524f9c9ceee798c2026e.pdf%0Ahttp://csjournals.com/IJITKM/PDF 3-1/19.pdf

Kothalawala, M., & Thelijjagoda, S. (2020). Aspect-based sentiment analysis on hair care product reviews. Proceedings - International Research Conference on Smart Computing and Systems Engineering, SCSE 2020, 228–233. https://doi.org/10.1109/SCSE49731.2020.9313040

Manek, A. S., Shenoy, P. D., Mohan, M. C., & Venugopal, K. R. (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World Wide Web, 20(2), 135–154. https://doi.org/10.1007/s11280-015-0381-x

Mowlaei, M. E., Abadeh, M. S., & Keshavarz, H. (2020). Aspect-based sentiment analysis using adaptive aspect-based lexicons. Expert Systems With Applications, 148, 113234. https://doi.org/10.1016/j.eswa.2020.113234

Novelty Octaviani Faomasi Daeli, A. (2020). Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor. Journal of Data Science and Its Applications, 3(1), 1–007. https://doi.org/10.34818/JDSA.2020.3.22

Park, J. (2020). Framework for sentiment-driven evaluation of customer satisfaction with cosmetics brands. IEEE Access, 8, 98526–98538. https://doi.org/10.1109/ACCESS.2020.2997522

Pitogo, V. A., & Ramos, C. D. L. (2020). Social media enabled e-Participation: A lexicon-based sentiment analysis using unsupervised machine learning. ACM International Conference Proceeding Series, 518–528. https://doi.org/10.1145/3428502.3428581

Roobaert, D., Karakoulas, G., & Chawla, N. V. (2006). Information gain, correlation and support vector machines. Studies in Fuzziness and Soft Computing, 207(November 2008), 463–470. https://doi.org/10.1007/978-3-540-35488-8_23

Kumar, C. S. P., & Babu, L. D. D. (2020). Evolving dictionary based sentiment scoring framework for patient authored text. Evolutionary Intelligence, (september 2011). https://doi.org/10.1007/s12065-020-00366-z

Sun, Q., Niu, J., Yao, Z., & Yan, H. (2019). Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. Engineering Applications of Artificial Intelligence, 81(December 2018), 68–78. https://doi.org/10.1016/j.engappai.2019.02.004

Xiao, S., Wang, H., Ling, Z., Wang, L., & Tang, Z. (2020). Sentiment Analysis for Product Reviews Based on Deep Learning. Journal of Physics: Conference Series, 1651(1). https://doi.org/10.1088/1742-6596/1651/1/012103

Xie, X., Ge, S., Hu, F., Xie, M., & Jiang, N. (2019). An improved algorithm for sentiment analysis based on maximum entropy. Soft Computing, 23(2), 599–611. https://doi.org/10.1007/s00500-017-2904-0

Zarisfi Kermani, F., Eslami, E., & Sadeghi, F. (2019). Global Filter–Wrapper method based on class-dependent correlation for text classification. Engineering Applications of Artificial Intelligence, 85(May), 619–633. https://doi.org/10.1016/j.engappai.2019.07.003

Zuheros, C., Martínez-Cámara, E., Herrera-Viedma, E., & Herrera, F. (2021). Sentiment Analysis based Multi-Person Multi-criteria Decision Making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using TripAdvisor reviews. Information Fusion, 68(July 2020), 22–36. https://doi.org/10.1016/j.inffus.2020.10.019