Business Intelligent Framework Using Sentiment Analysis for Smart Digital Marketing in the E-Commerce Era

Khin Sandar Kyaw[1*], Praman Tepsongkroh[2], Chanwut Thongkamkaew[3] and Farida Sasha[4]

[1,3]Department of International Business Management, Didyasarin International College, Hatyai University, Hatyai 90110, Thailand
[2]Department of Tourism Industry, Didyasarin International College, Hatyai University, Hatyai 90110, Thailand
[4]Department of Management, Hatyai Business School, Hatyai University, Hatyai, 90110 Thailand

[*]Corresponding author's e-mail: khinsandarkyaw@hu.ac.th

Received: *November 30, 2021* ***Revised:*** *March 28, 2022* ***Accepted:*** *September 2, 2022*

Abstract

Since trading has been transformed into online platforms, marketing strategies have adapted to digital systems in order to enhance the Customer Relationship Management (CRM) in the E-commerce era. E-commerce systems are the most widely used digital platforms where customer information including personal, and behavioral information, flows as a big data stream. Conducting business intelligent observation on digital big data assists to improve digital marketing policy through the customer intention prediction, decision-making to advertise based on the target group clustering, and customer assist recommendation. To discover the business intelligent, sentiment analysis technology can assist as a solution to understand the customer behavior through the opinion mining where the natural language processing, text analysis, computational linguistics, and biometrics are conducted to analysis the customer information and feedbacks, for smart digital marketing applications. This research observes the applications of sentiment analysis in E-commerce systems as a comprehensive study, and the critical role of discovering business intelligent for smart digital marketing in E-commerce platforms is pointed out according to the technical perspective. Furthermore, the concept of a business intelligent framework integrated with the modelling of decision-making, prediction, and recommendation systems using the contribution of hybrid feature selection which is based on rule-based and machine learning-based sentiment analysis, is proposed for the future innovative smart digital marketing trend.

Keywords: Business knowledge, Market basket analysis, Polarity classification, Web business knowledge, Product recommendation, Web mining, Artificial neural network

Introduction

Since understanding customer behavior assists to build up a successful customer Relationship Management System (CRM) for the business owner (Gil-Gomez et al., 2020), customer behavior analysis is the influencing factor for digital marketing development in E-commerce era. In the E-commerce system, customer data including personal information, feedbacks such as review, and rating, and website visiting rate, are becoming the fundamental and critical resources to finding insight for the customer behavior analysis to enhance the digital marketing policy. Since most of the customer information is presented as the text data type on the E-commerce platforms, sentiment analysis can be applied to understand the customers behavior or opinion like positive or negative or neutral, toward products and service.

Sentiment analysis (Bueno et al., 2022) helps business to classify customer sentiment toward products, brands, or services in online conversations and feedback. Sentiment analysis is the task of extracting and analyzing opinions, sentiments, attitudes, and perceptions for the fast evolution of Internet-based applications like websites, social networks, and blogs. Sentiment analysis is intended as a powerful tool for businesses, governments, and researchers to extract and analyze the mood and views, gain business insight, and make better decisions. Sentiment analysis scheme (Birjali et al., 2021) is classified into three levels such as aspect-level sentiment analysis (Mai & Le, 2021), sentence-level sentiment analysis (Li et al., 2021), and document-level sentiment analysis. Since sentiment analysis is the interpretation and classification of emotions (Subhashini et al., 2021) including positive, negative, and neutral from the text data, text analysis is needed to consider as a step of extracting keywords for the identification of polarity. In text analytic, several schemes can be used such as part of speech extraction, bag of word (Zhang et al., 2010), dictionary detection (Kumar & Babu, 2020), tokenization, n-gram calculation (Dey et al., 2018), term frequency-inverse document frequency (Qaiser & Ali, 2018), vectorization (Badugu, 2022) , normalization (Sohail et al., 2018), and so on.

Since customer feedbacks including reviews, ratings, satisfaction surveys and comments are the sources of discovering business intelligent for digital marketing policy enhancement (Behera et al., 2020), and competition for E-commerce age, the interpretation of sentiment can be explored through natural language processing and machine learning scheme. Customer sentiment can be collected via various E-commerce platforms (Xu et al., 2020) such as product websites which are application-based and search engine-based, social media business platforms, blogs, and so on. Although customer sentiments are presented using different languages and data types, this paper proposes a flexible framework that can adapt for the various language settings. Business intelligent discovery is a critical task in online digital marketing for the individual business owners by optimizing the search engine capability with intelligent-based features like chatbots (Nursetyo et al., 2018) for dynamic interaction to customer. In studying business intelligent cases, application-based and browser-based customer feedbacks are gathered as a primary step and the raw data is passed through into the analysis stage where information extraction, visualization, identification, classification, and prediction processes are performed to discovery the business intelligent for smart goals-based digital marketing applications (Wankhade et al., 2022) including target group marketing analysis, promotion marketing analysis, quality control, and customer demand analysis, customer interest analysis, market basket analysis, and customer service analysis, and so forth.

This paper explores the applications of sentiment analysis in E-commerce systems as a comprehensive study and proposes a business intelligent framework using sentiment analysis for the purpose of transforming into smart digital marketing. The critical role of business intelligent discovery for smart digital marketing in E-commerce era will be pointed out according to the technical perspective. Furthermore, business intelligent-based frameworks including decision-making, recommendation, and prediction models based on hybrid features selection concept, and their corresponding purposes for the digital marketing advantages for future innovative smart digital marketing trends will be presented in this research.

Literature review

This section presents a comprehensive literature review for the different applications of sentiment analysis in E-commerce systems. Table 1 summarizes sentiment analysis-based E-commerce applications with their corresponding proposed system methodologies, analysis outcomes, and limitation and future work. According to the summary of related works regards with sentiment analysis-based E-commerce applications, customer opinions and preferences derived from different datasets upon the various categories and criteria or features of products, are the keys to discovering the business intelligent which is useful for the development smart digital marketing policy implementation through prediction, recommendation, and decision-making systems. On the other hand, research gap is still existed as the exploration of flexible and automatic business intelligent system for E-commerce with the purpose of digital marketing enhancement in term of customer relationship management intention. To overcome the issue, the business intelligent framework is integrated with the concept of decision-making system to defining the marketing policy for target group, prediction system for customer intention, and recommendation system for customer assist and production demand manipulation. Moreover, the contribution of hybrid feature selection approach is embedded to manipulate the sentiment score as the feature to feed into the training model for smart digital marketing purpose.

Table 1 Summary of related works for sentiment analysis-based E-commerce applications

Reference	Proposed System	Analysis Outcome	Limitation and Future Work
Ray et al. (2021)	Ensemble-based hotel recommender system using sentiment analysis and aspect categorization	Achieved a Macro F1-score of 84% and test accuracy of 92.36% in the classification of sentiment polarities.	Suggest examining how customers' approval of reviews change over time since recent reviews are more regularly browsed. Recommend investigating the methods for solving class imbalance such as Random Over-Sampling or Under-Sampling, Bootstrap Aggregating (Bagging) based techniques, or Boosting-based techniques.

Reference	Proposed System	Analysis Outcome	Limitation and Future Work
			Suggest conducting experiments using Fuzzy ensemble method or Dempster–Shafer method, etc., to form an ensemble of the methods. Exist human-level error prone since the categorization results could be checked manually. Need changing the vector encoding mode if proper multilingual datasets are found.
Abbasi-Moud et al. (2021)	Tourism recommendation system based on semantic clustering and sentiment analysis	Provided the improved f-measure criterion.	Exploration of the other contextual information such as the traffic of routes, various conditions of individuals, as well as group users and their environment, have been neglected. Suggest extending the proposed recommendation system for group scenarios as users usually travel and visit attractions in groups.
Nawaz et al. (2021)	Sentiment analysis-based eWOM of a women's clothing company using Artificial Neural Networks techniques	Long short-term memory (LSTM) method outperformed Convolutional Neural Network (CNN) and achieved classification accuracy (91.69%), specificity (92.81%), sensitivity (76.95%), and F1-score (56.67%).	Suggest considering textual analysis techniques such as word frequencies analysis, thematic analysis, word correlations, and networks, to attain an in-depth understanding of emotions. Recommend integration data from various platforms including Facebook, Twitter, web portals, and blogs, for sentiment analysis. Consider bi-LSTM for sentiment analysis and compare the findings with this study.
Pugsee and Niyomvanich (2015)	Sentiment analysis of food recipe comments	The proposed system demonstrated that the accuracy of neutral and positive comment classification is about 90%. In addition, the accuracy of negative comment identification is more than 70%.	Suggest collection the personal profiles of people who comment on the recipes, such as nationality and age, to analyze recipe comments by the groups of people.
Zikang et al. (2020)	Sentiment analysis of agricultural product on E-commerce using deep learning	Compared with the Bi-LSTM and BERT fine-tuning models, the Text-CNN model's accuracy, precision, recall, and F1 value are all about 3 to 8 percentage points higher. The accuracy of the Text-CNN convolutional neural network reached 99.92%.	Performs the excellent emotional sentiment classification using Text-CNN in short texts such as agricultural Chinese comment data, but still have challenge regard with language and long text setting as a future work.
Aulawi et al. (2021)	Naive Bayes classifier (NBC), text association, and	Achieved average value of sentiment accuracy prediction	Develops sentiment analysis for social media using other machine learning methods.

Reference	Proposed System	Analysis Outcome	Limitation and Future Work
	focus group discussion (FGD) based consumer sentiment analysis	(99.66%), specificity (99.22%), and sensitivity (100%).	
Gharzouli et al. (2021)	Topic-based sentiment analysis of hotel reviews	Decreased the percentage of neutral opinions. Provided the English language translation using the proposed system.	Since only Textblob was used as API for processing textual data, another API should implement for other kind of classifiers. Uses larger datasets to test the developed system.
Zhou et al. (2019)	User sentiment analysis based on social network information in consumer reconstruction intention	Through the calculation of satisfaction, the degree of consumer repurchase intention for the products was determined based on trust and promotion effort of five sportswear brands in Taobao. Provided the analysis results for the relationship between the initial purchase intention and the repurchase intention of consumers, to assist the marketing strategy development and brand segmentation of online stores.	As only five sport wear data sets were used to evaluate the proposed model, another popular sport wear customer review should be used to analysis the proposed model performance. Moreover, the other machine learning model should try to implement to make the decision for the marketing strategies based on the analysis result on the relationship between the initial purchase intention and the repurchase intention of customers.
AL-Sharuee et al. (2021)	Automatic contextual analysis and ensemble clustering (ACAEC) algorithm-based sentiment analysis for dynamic and temporal clustering of product reviews	Provided average accuracy rates of segregated window clustering (SWC) and window sequential clustering (WSC) with the values of 87.54% and 83.87%, respectively.	Utilizes the unified feature set of WSC to enhance ensemble learning or develops a new algorithm using window correlation.
Jiang et al. (2021)	Observation on the relationship between reviews' information and reputation changes of the products using the evaluation of product reviews based on text sentiment analysis	There was a linear relation between star-rating and sentiment score. Specific quality descriptors of text-based reviews are associated with rating levels to some degree as they are not obvious in neutral reviews. In accordance with the word clouds of the products' reviews, size, warranty, quality, and power are mentioned often for microwave hair dryer, and size, color, delicacy for pacifier, which are worth the market directors' consideration.	Explores sentiment analysis on the customer review text from three datasets using logistic regression training model to classify the emotions including positive, neutral, and negative which used to find the relation between star rating and review text. Develops the other machine learning models and do more testing on the other product review datasets either benchmark or real time dataset.
Liu et al. (2020)	Latent Dirichlet Allocation (LDA)	Categorization of the 12 factors identified by LDA showed that 5	Regarding customer service, focusing on improving employees' service

Reference	Proposed System	Analysis Outcome	Limitation and Future Work
	topic model-based sentiment analysis of B2C online pharmacy reviews	factors were related to logistics including 38.5% of the drug reviews. The number of factors related to drug prices is 3 factors, and it includes 25.5% of the reviews. Customer service and drug effects each had two related factors, and a smaller percentage of these reviews (13.95%) were related to drug effects.	attitudes is necessary to consider as future work.
Wan et al. (2021)	A sentiment analysis-based expert weight determination method for large-scale group decision making driven by social media data	Provided superior accuracy and stability. Improved the quality and satisfaction of the decision results.	Develops more advanced big data analysis tools or methods like the adoption of a distributed system and special big data frameworks for processing and storing data. Considers several rounds of expert decision-making since the accuracy of the data may be relatively insufficient due to the occurrence of lower-probability events. Chooses different sentiment analysis algorithms for testing new and huge social media data. Explores the types of behavior of big data to support decision-making.
Zhang et al. (2020)	Sentiment analysis of E-commerce text reviews based on sentiment dictionary	Attained the effective sentiment analysis results. The results of part-of-speech tagging were biased.	Need optimization in terms of part-of-speech tagging and improves the algorithm to reduce the number of manual proofreading.
Liu et al. (2021)	Multilingual review-aware deep recommender system via aspect-based sentiment analysis for rating prediction tasks	Multilingual aspect-based sentiment analysis module (MABSA) was used to extract aligned aspects and their associated sentiments in different languages simultaneously, with only requiring overall ratings. Multilingual recommendation module explored aspect importance of both the user and item by considering different contributions of multiple languages and estimated aspect utility via a dual interactive attention mechanism integrated with aspect-specific sentiments from MABSA. Overall ratings can be inferred by a prediction layer adopting the aspect utility value and aspect importance as inputs. Extensive experimental results on nine real-world datasets demonstrated the superior	Recommend detecting special sentiments in sentences and thereby integrating them into recommendation tasks.

Reference	Proposed System	Analysis Outcome	Limitation and Future Work
		performance and interpretability of the proposed model.	
Osman et al. (2019)	Contextual sentiment-based recommender system for electronic products domain	Illustrated better performance in terms of root-mean-square error (RMSE) and mean absolute error (MAE) measurements than the conventional collaborative filtering approach in electronic product recommendation.	Considers the implementation and evaluation for the other domains such as medicine.
Karthik and Ganapathy (2021)	Product recommendation system using a new fuzzy logic with sentiment analysis and ontology in E-commerce	Achieved better performance than the existing product recommendation systems in terms of prediction accuracy of the relevant products for target users and the time taken.	Considers an intelligent neuro-fuzzy approach for making effective decisions and new pre-processing techniques.
Bineet Kumar Jha (2021)	Sentiment analysis for E-commerce products using PySpark, and resilient distributed dataset (RDD) based sentiment analysis	Utilized FLASK-based Restful APIs and Scrapy for web scrapping and Spark NLP to categorize the sentiments as positive, or neutral or negative from the context of the Amazon e-commerce site for an electronics product. Achieved most of the positive reviews are given by customers of age between 30-50 years in frequency distribution analysis. Provided accuracy level of the proposed algorithm with 90% to 93% whereas the precision is 85% to 90%.	Suggests a better methodology for sentiment analysis with an improved deep learning approach in Spark NLP to relationship analysis.
Singh et al. (2013)	Computing sentiment polarity of texts at document and aspect levels on online review data	The proposed approach achieved a more useful sentiment profile for movies and has accuracy levels equivalent to the document-level approach. Moreover, the algorithmic formulation used for aspect-level sentiment profile generation is very simple, quick to implement, fast in producing results and does not require any previous training. Different aspects of interest can also be used as an add-on step in movie recommendation systems for content-filtering, collaborative-filtering, or hybrid approaches.	Suggests developing other supervised machine learning classifiers and hybrid sentiment analysis model with the different types of datasets.
Zhai et al. (2021)	A social media opinion leader identification model based on online	The crawl online comments were analyzed to establish emotional indicators of different attributes, calculate the sentiment value of the	Since there were the crawler limitations of the Weibo platform, the experiment cannot crawl all the fan groups and friend groups of opinion

Reference	Proposed System	Analysis Outcome	Limitation and Future Work
	comment sentiment analysis	text data, and finally use artificial neural network technology to train an opinion leader recognition model. The experimental results showed that emotional communication is a very important factor in opinion leader identification, and the proposed model can identify opinion leaders more accurately.	leaders, which will have a certain impact on the calculation results of close centrality The domain sentiment dictionary was used in this experiment, but there are new Internet buzzwords in the current Internet environment that are not included, which will also bring certain errors to the calculation of sentiment values.
Lian (2021)	Personalized recommendation algorithm based on online comment sentiment analysis	Improved the traditional model and verified the effectiveness of the algorithm on a real data set.	Considers a new artificial intelligence technique to improve the proposed model.

Background theory

In this section, fundamental theory and knowledge concerned with digital marketing framework, E-commerce system, sentiment analysis, business intelligent model, text data preprocessing, text feature manipulation, and supervised machine learning-based text classification are described according to the technical perspective.

Digital marketing framework

In the implementation of digital marketing framework (Santos & Martinho, 2021), SMART goals-based, a forward-thinking plan is built up with the technical provision of digital agency services including website design, Search Engine Organization (SEO), paid ads, content marketing and social media marketing. According to a business perspective, digital marketing is composed of SWOT (Strength, Weaknesses, Opportunities, Threats) analysis, SMART business goals, market segmentation, and building buyer persona. User interface layers, including web-based or application-based, is implemented using web development (Necula et al., 2018) and software development technology, to bridge the connection between consumers and business owners. In backend operation, search engines perform customer behavior analysis to enhance the customer experience with the action of content mining, opinion mining, and usage mining on web platforms (Bertea, 2019). Furthermore, natural language processing-based sentiment analysis (Solangi et al., 2019) and machine learning-based artificial intelligent technology assist the search engine optimization for the digital marketing sector. To acquire customer information and behavior for digital marketing systems, information retrieval process is embedded on the E-commerce website, and social media-based trading platforms. The extracted E-commerce information is connected to an analysis module in the digital marketing framework (Naim, 2021) to discover patterns of digital marketing policy for CRM systems (Zhang et al., 2022). Several goals including finding target customer groups, customer intention and satisfaction, product quality control and market competition, clustering customer type, notification of product launching, partnership product investment, can be set up in a smart-digital marketing framework (Manoharan & Narayanan, 2021).

E-commerce system

Since the E-commerce system is a web-based online trading platform, the architectural framework consists of several layers of services including application services, brokerage and data management, interface layer, secure messaging, middleware services, and network infrastructure. According to technical descriptions, a variety of technologies such as web server, server software, web tools, database system, networking, browser compatibility, ports, domain name, are needed to implement the E-commerce platform. On the other hand, social media business suits can be utilized for trading as a low-cost and easy customer-engage platform. Therefore, demand for the type of E-commerce platform is dependent on the type of business such as enterprise business, small and medium size business (SME), Business to Business (B2B), Business to Consumer (B2C), and so on. However, the customized E-commerce platform can provide benefits such as unique service, low security risk, good privacy for customers, control right, etc. Furthermore, intelligent-based technology is added as a new feature in E-commerce such as chatbot, customer history record, tracking transaction, to attain customer trust with the provision of flexibility.

Sentiment analysis

The interpretation of consumers' polarity related with the product, brand and new product perception, reputation management and service available on the E-commerce (Jabbar, 2019) can be manipulated using sentiment analysis. In the sentiment process, finding relevant documents, finding relevant sections, finding the overall sentiment, quantifying the sentiment, and aggregating all sentiments are involved to classify customer opinions such as positive, negative, and neutral. There are various types of opinion mining tasks such as document level (Choi et al., 2020), sentence level and feature level (Ahuja et al., 2019), and therefore the process of performance evaluation of sentence analysis is also becoming a challenge due to the lack of ground truth. Sentiment analysis assists businesses to attain insights of social sentiment of the brand, product or service while monitoring online conversations. The two main important usages of sentiment analysis (Lee et al., 2018) are to achieve key aspects of a brand's product and service, and consumer's underlying intentions and reactions concerning those aspects.

Different sentiment score calculation schemes can be applied based on the interested customer data type such as numerical, and text. However, most sentiment analysis focuses on the customer reviews and comments. Therefore, the calculation of sentiment score using text analysis tools (Zucco et al., 2020) can be employed to find customer polarity. Text analysis involves cleaning, preprocessing, extraction, and selection of sentiment from the collected customer reviews and comments. However, the rule for all these processes might be different based on the characteristics of the collected data and smart goal of the E-commerce system.

Business intelligent model

Business Intelligent (BI) is the insight of business data such as E-commerce data, organization data, production data, etc. Business intelligent can be achieved by conducting analysis process using either qualitative or quantitative method. Moreover, data mining (Khoa Dam, 2019), and machine learning based artificial intelligent can be employed to discover business knowledge for the development of intelligent-based model to monitoring of business statistics, adjusting the business policy including quality control, and marketing policy, and improving the customer experience and satisfaction. Business intelligence model is diversified according to the business smart goal. It consists of business recommendation system (Zhang et al., 2019), business decision-making system, business prediction model (Lye & Teh, 2022) using classification and clustering techniques (Bandyopadhyay et al., 2021).

In the implementation of BI model, business data analysis is carried out as the most fundamental and critical step to understand the business data characteristic. Data analytic is conducted using a variety of mathematical calculations based on the data types, and the analysis result is presented as a visual description. The analytic result is plugged with a modelling system where execution of information extraction, scoring, training, and testing processes are performed. In the development of recommendation models for E-commerce, consumer information including activity, opinion, and behavior, is extracted as primary data to analyze the score for understanding customer demand. Based on the individual customer score, the business intelligent rules are defined to categorize the customer group for the specific product and train the intelligent model using machine-learning algorithms (Tudoran, 2022) based on the historical customer data to provide suitable product recommendation in future. On the other hand, recommendation model results can be related in decision-making systems for implementing effective digital marketing policy. Furthermore, sentiment analysis is conducted on customer reviews to classify polarity, and clustering of customer groups is performed based on the classified polarity results to predict supplier and customer demands in future. Moreover, market competition for new product launching can be addressed using the prediction model results.

Text data preprocessing

Since text preprocessing is the critical step in the pipeline of the text mining process, data cleaning is conducted as the first step of preprocessing to filter the unnecessary information. Normally, tokenization, Part-of-Speech (POS) tagging, stemming and lemmatization, and stop words removal are the common preprocessing steps (Uysal & Gunal, 2014). From the perspective of tools, several open-source tools for different language settings such as Jieba for Chinese (Gao et al., 2004), and NLTK (Bird & Loper, 2004) for English language, etc., are available. Moreover, ontology-based preprocessing for domain-specific text-based research such as WordNet, and domain-specific lexicon are also used for preprocessing.

Text feature manipulation

Feature manipulation transforms the raw textual data into the feature-based description by passing through feature extraction and selection phases. Feature extraction forms new features using the combinations of original features and projects them into a lower dimensional space. The popular text feature extraction schemes include term frequency- inverse document frequency (Kadhim, 2019), Gini Index (Manek et al., 2017), information gain, entropy (Xie et al., 2019), mutual information (Gao et.al, 2020), etc. Moreover, feature selection is the selection of a subset of relevant features that are highly related to the criterion measure. Feature selection techniques can be classified as filter (Bommert et al., 2020), wrapper (Gokalp et al., 2020) and hybrid (Zarisfi Kermani et al., 2019). Filter methods measure the correlation (Karegowda et al., 2010), Chi-Square (Alshaer et al., 2021) and Information Gain. Since filter methods are not dependent on training algorithm, they can avoid feature bias case. However, wrapper models are computationally expensive and infeasible if the number of features is huge since they heuristically select features depending on the machine learning algorithm.

Supervised machine learning-based text classification

Supervised machine learning (Sarker, 2021) is one of the well-known techniques for intelligent system development in several domains including E-commerce, medical, engineering, industry, education, etc. The core concept of supervised machine learning involves training and testing phases for the labeled data. Supervised learning (Caruana & Niculescu-Mizil, 2006) is divided into four categories including decision tree concepts (Tanha et al., 2017), rule-based and case-based classification (Berka, 2020), linear classification, and probabilistic classification. In other words, the training part includes corpora utilized to gain proficiency with a “classifier” model and testing part utilizes the training data set to characterize inconspicuous information. However, supervised learning is based on the labeled dataset experience during decision-making. Therefore, supervised machine learning models have an overfitting issue which fits the training data too well but generally poorly to testing data. The common popular classification algorithms are linear classifier, Artificial Neural Network (Huang et al., 2020), support vector machine (Dey et al., 2020), decision tree, Naïve Bayes (Bayhaqy et al., 2018), etc. Text classification is an important task for information management applications by automatically allocating a specified document to one or more predefined classes. Automatic text classification is treated as a supervised machine learning technique (Miklosik et al., 2019). The goal of this technique is to determine whether a given text belongs to the given category or not by looking at the words or terms of that category.

Background methodology

In this section, the methods concerned with the proposed framework implementation will be presented. Since the proposed framework practices the sentiment analysis-based business intelligent framework to implement the conceptual models of decision making, prediction and recommendation, data collation, noise elimination, sentiment score manipulation, feature vector implementation using hybrid feature selection, intelligent model training and model evaluation scheme, are needed as critical steps according to the knowledge from the literature review and background theory section. Furthermore, the detailed proposed system workflow and functionalities will be presented in the following subsection.

Data collation

Since the proposed framework is intended for customer opinion mining in E-commerce platforms, web crawlers (Yang & Thiengburanathum, 2020) or robots can be used to extract the E-commerce reviews as the real time data from several online platforms including E-commerce websites, social media, and blogs. Web crawler is a program used to acquire data from the Web by following hyperlinks. Web crawler is composed with a set of policies including politeness policy, parallelization policy, revisit policy, and robustness policy. Politeness policy defines only allowed web pages that 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 that 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.

To acquire relevant web content for customer opinions' sentiment analysis framework, content extraction process (Zhang et al., 2013) should be applied on web crawling processes. Fundamentally, the content extraction process integrates two phases namely, noise removal (filtering) and selection of content (extraction of text content from the given Hypertext Markup Language, HTML) document. Although several schemes for main content extraction have been developed, the HTML Document Object Model (DOM) tree feature (Utiu & Ionescu, 2018) based web crawling is the most effective approach to extract text documents (e-commerce reviews / comments) from the corresponding tags. However, this proposed framework intends to practice on the standard and benchmark different E-commerce datasets which are derived from the UCI machine learning repository, Kaggle data repository, and Amazon big data repository.

Noise elimination for opinion mining

Since customer reviews include noisy text, several steps are conducted for the purpose of noise elimination for opinion mining process to enhance the execution cost by reducing the irrelevant text from the collected data. It involves pre-processing for cleaning noisy data, opinion expression identification and extraction, and opinion assimilation and aggregation. In the first step, tokenization, stop words removing, normalization including negation handling, stemming, lemmatization, substitution, removal of special characters, standalone punctuation, the identification of correct improper casing, sentence boundaries, and correct spelling error are applied using the domain knowledge. Part of speech (POS) tagging, phrase structure extraction and dependency analysis are conducted in the step of opinion expression identification and extraction. Domain ontology is employed to assimilate extracted opinions and aggregate the information at multiple levels of specificity.

Sentiment score manipulation

Opinion lexicon-based bags of words can be used to extract the opinion word from the noise eliminated text or customer review. As a basic and simple score manipulation scheme, absolute score calculation can be applied to count the polarity by subtracting the total number of counts for the different types of polarity. The N-gram-based TF-IDF method can be used to calculate the sentiment score for complicated text data. According to the characteristics of the collected dataset from the noise elimination step, different n-gram words are extracted to identify the polarity of the text data and identify the most frequent word based on the term frequency inverse document frequency (TF-IDF) scheme.

Feature vector implementation using hybrid-feature selection

Normalize the extracted sentiment score and select the relevant sentiment using hybrid feature selection scheme which is illustrated in Figure 1. It integrates with rule-based sentiment analysis and machine learning based sentiment analysis. Rule-based feature selection (Kang & Zhou, 2017) is applied on the extracted feature vector to select opinion expression using part of speech (POS), dictionary-based, corpus and lexicon-based (Taj et al., 2019; Zhang et al., 2018), etc. The selected opinion features are passed through the machine-learning model to discover the relationship between features and their corresponding class such as positive, negative, and neutral sentiment. Machine learning-based feature selection has two main schemes such as filter and wrapper. The proposed model uses a filter approach to select relevant features which is independent of the learning model. In specific, correlation-based feature selection method (Michalak & Kwasnicka, 2010) is proposed for this research to overcome the problem of bias in selecting features. Its hypothesis includes a heuristic approach for considering features that are highly correlated with predictive labels, but uncorrelated with other labels. In addition, it calculates the correlation between features to search for feature subset.

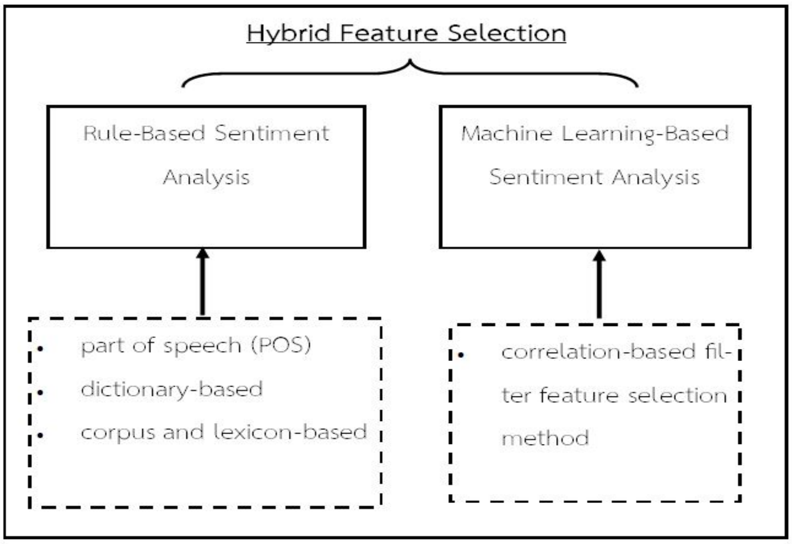

Figure 1 Hybrid feature selection for sentiment analysis model

Intelligent model training

Learning-based Intelligent models are developed using training and testing parts. While historical data is used to train the learning model in the training phase, a testing phase is performed using a new testing data set. Artificial Neural Network (ANN) (Abiodun et al., 2018) will be used to train the model. It is a supervised procedure where input, hidden, and output are involved as the operation layers. Furthermore, it consists of an activation function that defines the output of a neuron in terms of a local induced field. Activation functions are a single line of code that gives the neural nets non-linearity and expressiveness. There are many activation functions such as identity function, binary step function, and sigmoid function.

Evaluation scheme for opinion mining model

Evaluation scheme is the testing for the performance of a learning model. Ten-folds cross validation, accuracy in Eq. (1) and precision in Eq. (2) are used for validating the proposed intelligent model. Ten-fold cross validation is a technique to evaluate the models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. Accuracy is the measurement of a ratio of correctly predicted observations to the total observations, whereas precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

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

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

where TP means true positive, FP presents false positive, TN refers true negative, and FN represents false negative.

Proposed Sentiment Analysis-Based Business Intelligent Framework

Figure 2 illustrates the workflow diagram of sentiment-analysis for E-commerce data with its corresponding functions. Data from different E-commerce channels are collated and kept in the consumer data warehouse. The collated raw data includes the noise and therefore connected to the filter process to eliminate the noise. A sentiment score manipulation step is conducted to calculate the score of each customer sentiment data before passing it through the feature selection process. Feature vector implementation is performed by hybrid feature selection to achieve relevant features. The selected feature set is fed to the intelligent learning model to train the data for future polarity classification.

In Figure 3, the polarity analysis outcome from the intelligent model is applied for different applications in the digital marketing sector. Upon the three-polarity analysis outcome, the clustering technique can be applied to define the marketing policy for grouping the target customers that is the customer who give positive polarity for the product or service will be defined as our target group for marketing. Similarly, the polarity analysis outcome can be used to train the prediction model where the customers' opinions are recorded as the training dataset to predict the possibility of customer positive or negative or neural opinion to find the customer intention for the testing dataset. The feature of customer recommendation system for customer assist such as auto product or service recommendation including product selection, new product launching with promotion sale, and production demand for supply chain management can be implemented using the customers' polarity analysis outcomes.

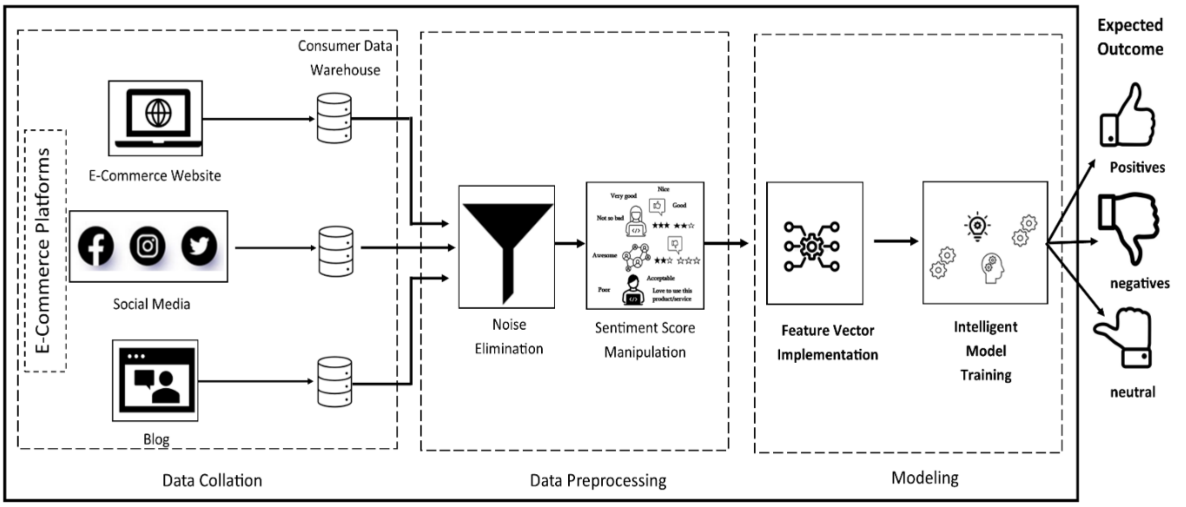

Figure 2 Functional diagram of sentiment analysis-based business intelligent system

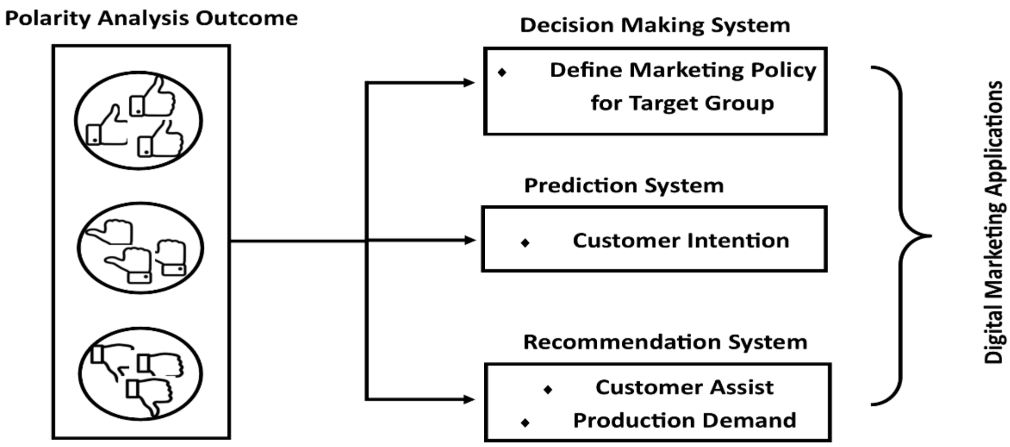

Figure 3 Application of polarity analysis outcome in smart digital marketing

Experimental set up and expected outcome

As a case study, the proposed business intelligent framework for smart digital marketing, will collect the E-commerce product and service review datasets from three common data repositories to conduct the sentiment analysis including UCI machine learning repository for sentiment dataset, Kaggle data repository for Twitter product review dataset, and Amazon big data repository (Hawlader et al., 2021). High-level general purpose Python programming language and its NLTK (Natural Language Toolkit) will be applied for sentiment analysis process as the analysis tool. The attained sentiment polarity will be used to train customer recommendation system, decision making system, and prediction system as the simulation models. As an expected outcome, a simulated smart business intelligent model using sentiment analysis will provide the domain intelligent model for E-commerce and the other researchers can use the proposed framework as a knowledge base in future. Moreover, the research related with search engine optimization can use the outcome of the proposed framework as a domain knowledge to improve their customer intention service for enhancing the web-based commerce system in future.

Theoretical and practical implications

The expected findings of this research present theoretical implications and provide several useful practical suggestions for Business Intelligent (BI) users including BI managers and BI developer in different business industries to improve the digital marketing experience in E-commerce era. Since this research proposed the business intelligent framework using sentiment analysis for smart digital marketing, understanding the customer behavior and characteristic by collecting, analysis and modelling customer data using a technology-oriented process like sentiment analysis is the main concept to derive the business intelligent or the insights for assisting organizations inaccurate decision-making. Meanwhile, social media, E-commerce website, and blog are the common platform for digital marketing strategies since the huge amount of customer data is generated drastically that support for business intelligent deriving process for smart decision-making system. As the additional practical perspective, the outcome of the proposed conceptual framework would support more perception towards intelligent-based smart digital marketing platform for the search engine-oriented E-commerce business platform by knowing the customer through the effective real-time sentiment analytics.

The principal usage of proposed business intelligent framework in digital marketing application intends for online marketer who need to identify the real-time customer data for keeping a check on return on investment from campaigns to justify their existence. Moreover, smart decision-making systems for the making digital policy can utilize the proposed framework since the proposed framework includes the concept of retrieving the deep insights of the client’s preferences and attitudes by passing through the several knowledge discovery process steps including data collection from E-commerce platform, data preprocessing with noise elimination and sentiment score manipulation, modelling with feature vector implementation and intelligent model training for smart digital marketing systems including decision-making, prediction and recommendation services.

Furthermore, this research can provide the suggestion to incorporate business intelligent framework into the smart digital marketing area by integrating with sentiment analysis-based customer behavior analysis technology. In fact, the inference of the customer opinion analysis which can be used as the knowledge-based management system for business data analytic or search engine optimization for intelligent E-commerce system with dynamic user interaction experience in future.

The proposed system can provide the advantage for enhancing the Customer Relationship Management (CRM) by optimizing the E-commerce system with the smart intelligent-based search engine capability which utilizes the training model derived from the customer behavior analysis (polarity classification) using hybrid feature selection-based sentiment analysis and supervised machine learning model. In other words, the proposed system is expected to establish a link between natural language processing and E-commerce trading using sentiment analysis scheme-based on rule-based and machine learning-based sentiment score calculation scheme in future.

Moreover, the hybrid feature selection for sentiment analysis model of the proposed framework result will be generalized by conducting the comparison with other systems such as rule-based only, machine learning based only on both benchmark data set and real-time crawled data set from E-commerce, and benchmark sentiment analysis models, by using the standard evaluation scheme for opinion mining model such as accuracy and precision.

Conclusions and future work

The proposed framework aims to enhance smart digital marketing policy using sentiment analysis in E-commerce era. Three business intelligent models using sentiment analysis based on hybrid feature selection and supervised machine learning including decision making system for defining marketing policy for the target customer group; prediction system for customer intention; and recommendation system for the customer assist and production demand, are suggested as the conceptual models. The experimental implementation for the proposed business intelligent framework with three business intelligent models for smart digital marketing will be presented as a future work.

Acknowledgements

This work was supported by the Department of International Business Management, Didyasarin International College, Hatyai University. The authors wish to thank Dr Ozioma F. Nwabor for his assistance throughout the manuscript preparation phase.

References

Abbasi-Moud, Z., Vahdat-Nejad, H., & Sadri, J. (2021). Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Systems with Applications*, *167*, 114324.

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A. E., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11), e00938.

Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, *152*, 341-348.

AL-Sharuee, M. T., Liu, F., & Pratama, M. (2021). Sentiment analysis: dynamic and temporal clustering of product reviews. *Applied Intelligence, 51*(1), 51-70.

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.

Aulawi, H., Karundeng, E., Kurniawan, W. A., Septiana, Y., & Latifah, A. (2021). *Consumer sentiment analysis to E-commerce in the Covid-19 pandemic era* (pp. 1-5). In Proceedings of the 2021 International Conference on ICT for Smart Society (ICISS). Bandung, Indonesia: IEEE.

Badugu, S. (2022). A comparative study on classification algorithms using different feature extraction and vectorization techniques for text. *Turkish Online Journal of Qualitative Inquiry, 12*(7), 8216-8225.

Bandyopadhyay, S., Thakur, S. S., & Mandal, J. K. (2021). Product recommendation for E-commerce business by applying principal component analysis (PCA) and K-means clustering: Benefit for the society. *Innovations in Systems and Software Engineering, 17*(1), 45-52.

Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018, October). *Sentiment analysis about E-Commerce from Tweets using decision tree, K-Nearest Neighbor, and Naïve Bayes* (pp. 1-6). In Proceedings of the 2018 International Conference on Orange Technologies (ICOT). Bandung, Indonesia: IEEE.

Behera, R. K., Gunasekaran, A., Gupta, S., Kamboj, S., & Bala, P. K. (2020). Personalized digital marketing recommender engine. *Journal of Retailing and Consumer Services, 53*, 101799.

Berka, P. (2020). Sentiment analysis using rule-based and case-based reasoning. *Journal of Intelligent Information Systems, 55*(1), 51-66.

Bertea, A. F. (2019). *Data web mining in E-commerce: Progress and perspectives* (pp. 85-90). In Proceedings of the 18th International Conference on Informatic in Economy Education. Bucharest, Romania: Research and Business Technologies.

Bineet Kumar Jha, S. G. G. V. K. R. (2021). Sentiment analysis for E-commerce products using natural language processing. *Annals of the Romanian Society for Cell Biology, 25*(5), 166-175.

Bird, S., & Loper, E. (2004). *NLTK: The Natural Language Toolkit* (pp. 69-72). In Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions. Sydney: Association for Computational Linguistics

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems, 226*, 107134.

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.

Bueno, I., Carrasco, R. A., Ureña, R., & Herrera-Viedma, E. (2022). A business context aware decision-making approach for selecting the most appropriate sentiment analysis technique in e-marketing situations. *Information Sciences*, *589*, 300-320.

Caruana, R., & Niculescu-Mizil, A. (2006). *An empirical comparison of supervised learning algorithms* (pp.161-168). In Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh, PA: ACM.

Choi, G., Oh, S., & Kim, H. (2020). Improving document-level sentiment classification using importance of sentences. *Entropy*, *22*(12), 1336.

Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-N-Gram: An n-gram lexicon for sentiment analysis. *Expert Systems with Applications*, *103*, 92-105.

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* (pp. 217-220). In Proceedings of the 2020 International Conference on Contemporary Computing and Applications (IC3A). Lucknow, India: IEEE.

Gao, J., Wu, A., Li, M., Huang, C.-N., Li, H., Xia, X., & Qin, H. (2004). *Adaptive Chinese word segmentation*. Retrieved from https://doi.org/10.3115/1218955.1219014

Gao, W., Hu, L., & Zhang, P. (2020). Feature redundancy term variation for mutual information-based feature selection. *Applied Intelligence*, *50*(4), 1272-1288.

Gharzouli, M., Hamama, A. K., & Khattabi, Z. (2021). Topic-based sentiment analysis of hotel reviews. *Current Issues in Tourism*, 1-8.

Gil-Gomez, H., Guerola-Navarro, V., Oltra-Badenes, R., & Lozano-Quilis, J. A. (2020). Customer relationship management: digital transformation and sustainable business model innovation. *Economic Research-Ekonomska Istrazivanja* , *33*(1), 2733-2750.

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.

Han, H., Zhang, J., Yang, J., Shen, Y., & Zhang, Y. (2018). Generate domain-specific sentiment lexicon for review sentiment analysis. *Multimedia Tools and Applications*, *77*(16), 21265-21280.

Hawlader, M., Ghosh, A., Raad, Z. K., Chowdhury, W. A., Shehan, M. S. H., & Ashraf, F. Bin. (2021). *Amazon product reviews: Sentiment analysis using supervised learning algorithms* (pp. 1-6). In Proceeding of the 2021 International Conference on Electronics, Communications, and Information Technology (ICECIT). Khulna, Bangladesh: IEEE.

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.14*(8), pp 1-12.

Jabbar, J. (2019). *Real-time sentiment analysis on E-commerce application* (pp. 391-396). In Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC). Banff, AB, Canada: IEEE.

Jiang, Y., Wang, H., & Yi, T. (2021). *Evaluation of product reviews based on text sentiment* *analysis* (pp. 1-8). In Proceedings of the 2021 2nd International Conference on Artificial Intelligence and Information Systems (ICAIIS'21). Chongqing, China: ACM.

Kadhim, A. I. (2019). *Term weighting for feature extraction on Twitter: A comparison between BM25 and TF-IDF* (pp. 124-128). In Proceedings of the 2019 International Conference on Advanced Science and Engineering (ICOASE 2019). Zakho - Duhok, Iraq: IEEE.

Kang, Y., & Zhou, L. (2017). RubE: Rule-based methods for extracting product features from online consumer reviews. *Information and Management*, *54*(2), 166-176.

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 Management*, *2*(2), 271-277.

Karthik, R. V., & Ganapathy, S. (2021). A fuzzy recommendation system for predicting the customers interests using sentiment analysis and ontology in e-commerce. *Applied Soft Computing*, *108*, 107396.

Khoa Dam, N. A. (2019). Marketing intelligence from data mining perspective - A literature review. *International Journal of Innovation, Management and Technology*, *10*(5), 184-190.

Kumar, C. S. P., & Babu, L. D. D. (2020). Evolving dictionary based sentiment scoring framework for patient authored text. *Evolutionary Intelligence*, *14*(2), 657-667.

Lee, S. Y., Qiu, L., & Whinston, A. (2018). Sentiment manipulation in online platforms: An analysis of movie tweets. *Production and Operations Management*, *27*(3), 393-416.

Li, X., Sun, X., Xu, Z., & Zhou, Y. (2021). *Explainable sentence-level sentiment analysis for amazon product reviews* (pp. 88-94). In Proceedings of the 2021 5th International Conference on Imaging, Signal Processing and Communications (ICISPC). Kumamoto, Japan: IEEE.

Lian, Q. (2021). Personalized recommendation algorithm based on online comment sentiment analysis. *Journal of Physics: Conference Series*, *1873*(1), 012086.

liu, jingfang, zhou, yingyi, jiang, xiaoyan, & zhang, wei. (2020). Consumers' satisfaction factors mining and sentiment analysis of B2C online pharmacy reviews. *BMC Medical Informatics and Decision Making*, *20*(1), 1-13.

Liu, P., Zhang, L., & Gulla, J. A. (2021). Multilingual review-aware deep recommender system via aspect-based sentiment analysis. *ACM Transactions on Information Systems, 39*(2), 1-33.

Lye, S. H., & Teh, P. L. (2022). *Customer intent prediction using sentiment analysis techniques* (pp. 185-190). In Proceedings of the 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS*).* Cracow, Poland: IEEE.

Mai, L., & Le, B. (2021). Joint sentence and aspect-level sentiment analysis of product comments. *Annals of Operations Research*, *300*(2), 493-513.

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.

Manoharan, G., & Narayanan, S. (2021). A research study to investigate the feasibility of digital marketing strategies in advertising. *PalArch's Journal of Archaeology of Egyptology*, *18*(9), 450-456.

Michalak, K., & Kwasnicka, H. (2010). Correlation based feature selection method. *International Journal of Bio-Inspired Computation*, *2*(5), 319-332.

Miklosik, A., Kuchta, M., Evans, N., & Zak, S. (2019). Towards the adoption of machine learning-based analytical tools in digital marketing. *IEEE Access*, *7*, 85705-85718.

Naim, A. (2021). Applications of marketing framework in business practices. *International Journal of Innovative Analyses and Emerging Technology*, *1*(6), 171-186.

Nawaz, Z., Zhao, C., Nawaz, F., Safeer, A. A., & Irshad, W. (2021). Role of artificial neural networks techniques in development of market intelligence: A study of sentiment analysis of eWOM of a women's clothing company. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(5), 1862-1876.

Necula, S. C., Pavaloaia, V. D., Strîmbei, C., & Dospinescu, O. (2018). Enhancement of E-commerce websites with semantic web technologies. *Sustainability*, *10*(6), 1-15.

Nursetyo, A., Setiadi, D. R. I. M., & Subhiyakto, E. R. (2018). *Smart chatbot system for E-commerce assitance based on AIML* (pp. 641-645). In Proceedings of the 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI 2018). Yogyakarta, Indonesia: IEEE.

Osman, N. A., Noah, S. A. M., & Darwich, M. (2019). Contextual sentiment based recommender system to provide recommendation in the electronic products domain. *International Journal of Machine Learning and Computing*, *9*(4), 425-431.

Pugsee, P., & Niyomvanich, M. (1970). Sentiment analysis of food recipe comments. *ECTI Transactions on Computer and Information Technology*, *9*(2), 182-193.

Qaiser, S., & Ali, R. (2018). Text mining: Use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, *181*(1), 25-29.

Ray, B., Garain, A., & Sarkar, R. (2021). An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing*, *98*, 106935.

Santos, F., & Martinho, R. (2021). Architectural challenges on the integration of E-commerce and ERP systems: A case study. *1*, 313-319.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), 1-21.

Singh, V. K., Piryani, R., Waila, P., & Devaraj, M. (1970). Computing sentiment polarity of texts at document and aspect evels. *ECTI Transactions on Computer and Information Technology*, *8*(1), 67-79.

Sohail, O., Elahi, I., Ijaz, A., Karim, A., & Kamiran, F. (2018). *Text classification in an under-resourced language via lexical normalization and feature pooling*. In Proceedings of the 22nd Pacific Asia Conference on Information Systems - Opportunities and Challenges for the Digitized Society (PACIS 2018). Japan: Association for Information

Systems.

Solangi, Y. A., Solangi, Z. A., Aarain, S., Abro, A., Mallah, G. A., & Shah, A. (2019). *Review on Natural Language Processing (NLP) and its toolkits for opinion mining and sentiment analysis* (pp. 22-23). In Proceedings of the 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS 2018). Bangkok, Thailand: IEEE.

Subhashini, L. D. C. S., Li, Y., Zhang, J., Atukorale, A. S., & Wu, Y. (2021). Mining and classifying customer reviews: A survey. In *Artificial Intelligence Review*, *54*, 6343-6389.

Taj, S., Shaikh, B. B., & Fatemah Meghji, A. (2019). *Sentiment analysis of news articles: A lexicon based approach* (pp. 1-5). In Proceedings of the 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (ICoMET 2019). Sukkur, Pakistan: IEEE.

Tanha, J., van Someren, M., & Afsarmanesh, H. (2017). Semi-supervised self-training for decision tree classifiers. *International Journal of Machine Learning and Cybernetics, 8*(1), 355-370.

Tudoran, A. A. (2022). A machine learning approach to identifying decision-making styles for managing customer relationships. *Electronic Markets*, 1-24.

Utiu, N., & Ionescu, V. S. (2018). *Learning web content extraction with DOM features* (pp. 5–11). In Proceedings of the 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP 2018). Cluj-Napoca, Romania: IEEE.

Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing and Management*, *50*(1), 104-112.

Wan, Q., Xu, X., Zhuang, J., & Pan, B. (2021). A sentiment analysis-based expert weight determination method for large-scale group decision-making driven by social media data. *Expert Systems with Applications*, *185*(932), 115629.

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. In *Artificial Intelligence Review*, 1-50.

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.

Xu, F., Pan, Z., & Xia, R. (2020). E-commerce product review sentiment classification based on a Naïve Bayes continuous learning framework. *Information Processing and Management*, *57*(5), 102221.

Yang, D., & Thiengburanathum, P. (2020). *A comparison of open source web crawlers for E-commerce websites* (pp. 200-205). In Proceedings of the 2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT and NCON 2020). Pattaya, Thailand: IEEE.

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,* 619-633.

Zhai, Y., Wang, Z., Zeng, H., & Hu, Z. (2021). *Social media opinion leader identification based on sentiment analysis* (pp. 436-440). In Proceedings of the 2021 International Conference on Bioinformatics and Intelligent Computing (BIC 2021). Harbin, China: Association for Computing Machinery.

Zhang, H., Zang, Z., Zhu, H., Uddin, M. I., & Amin, M. A. (2022). Big data-assisted social media analytics for business model for business decision making system competitive analysis. *Information Processing and Management*, *59*(1), 102762.

Zhang, Yanrong, Sun, J., Meng, L., & Liu, Y. (2020). S*entiment analysis of E-commerce text reviews based on sentiment dictionary* (pp. 1346-1350). In Proceedings of 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). Dalian, China: IEEE.

Zhang, Yin, Abbas, H., & Sun, Y. (2019). Smart E-commerce integration with recommender systems. *Electronic Markets*, *29*(2), 219-220.

Zhang, Yin, Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*, *1*(1-4), 43-52.

Zhang, Y., Mukherjee, R., & Soetarman, B. (2013). Concept extraction and E-commerce applications. *Electronic Commerce Research and Applications*, *12*(4), 289-296.

Zhou, Q., Xu, Z., & Yen, N. Y. (2019). User sentiment analysis based on social network information and its application in consumer reconstruction intention. *Computers in Human Behavior*, *100*, 177-183.

Zikang, H., Yong, Y., Guofeng, Y., & Xinyu, Z. (2020). *Sentiment analysis of agricultural product E-commerce review data based on deep learning* (pp. 1-7). In Proceedings of the 2020 International Conference on Internet of Things and Intelligent Applications (ITIA 2020). Zhenjiang, China: IEEE.

Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H., & Cannataro, M. (2020). Sentiment analysis for mining texts and social networks data: Methods and tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(1), 1-32.