International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Recommendation System Using Deep Learning with Text Mining Algorithm in E-Commerce Framework

[1]Kabilan , [2]Santhosh raj, [3]Selvamanikandan, [4]Prof Kalaivani

[1]E.G.S Pillay Engineering College (Autonomous), Nagapattinam, Tamilnadu, India
E-mail:kabilan150321@gmail.com

[2]E.G.S Pillay Engineering College (Autonomous), Nagapattinam, Tamilnadu, India
E-mail:santhoshegs@gmail.com

[3]E.G.S Pillay Engineering College (Autonomous), Nagapattinam, Tamilnadu, India
E-mail:selvamanikandan003@gmail.com

[4]Associate Professor E.G.S Pillay Engineering College (Autonomous), Nagapattinam, Tamilnadu, India
E-mail:kalaivani@egspec.org

ABSTRACT

E-Commerce has been known as a rapidly growing commercial enterprise, and even though on line purchasing has no longer accompanied those identical boom patterns within the beyond, it's miles now being diagnosed for its capability. Sentiment evaluation is one of the current research subjects in the subject of textual content mining. Opinions and sentiments mining from natural language are very difficult task. Sentiment analysis is the best solution. This gives important information for decision making in various domains. Various sentiment detection methods are available which affect the quality of result. In this paper we are finding the sentiments of people related to the services of E-shopping websites. The sentiments include reviews, ratings and emoticons. The main goal is to identify the fake reviews based on user id and product id to buy the best products and analysing which one is the best. For this we use hybrid learning algorithm which analyse various feedbacks related to the services. Text mining algorithm is used to find scores of each word. Then sentiments are classified as negative, positive and neutral. To find out fake review in the website can be analysed. Finally predict the fake reviews which are posted by user. Only purchased users post the reviews and check the duplicates based on user id and booking id. Genuine reviews are consider for product recommendation.

Keywords: E-Commerce, Review collection, Sentiment Analysis, Recommendation system, Fake reviews monitoring.

1. INTRODUCTION

Social media has emerged due to the usage trends and innovative services offered by the prevalent social networking sites. Social media's impending role is considered an expediter that enhances the learning and analysis of massive data .The development of web-based techniques is unparalleled, wherein social media technologies have to turn into an essential component in everyday lives. The accessibility of social media is no more ostensible than at the universities, wherein the technologies are transforming the ways students interact. The associate editor coordinating the review of this manuscript and approving it for publication was Sabah Mohammed and collaborate. The processor in the telecom system generates precise responses within a particular time interval. Moreover, the trades are not processed in the same order as they occur, but they are processed according to certain priorities. The needs on the quality, usability, reliability, and availability of these models are very rigorous. Generally, the software based on telecom is incessantly upgraded due to the long era of manipulation. To fulfill customer requirements, the software packages are produced, delivered, and specified rapidly in a precise manner. Recently, the ISP containing system downtime is highly important for sales argument. Telecom software development's augmentations should be carefully controlled, predictable, planned, and measurable. However, it is essential to briefly study the classification for detecting the sentiments hidden in the post. Sentiment analysis is extensively categorized into ternary sentiment classification, binary sentiment classification. The purpose of classifying

Twitter sentiment is to inevitably find the divergence of sentiment in tweets as negative or positive. Twitter sentiment classification monitors the machine learning techniques for building a classifier from tweets using polarity labels of manually annotated sentiments. Munjal et al developed a framework for predicting twitter sentiments and also proposed a model for opinion dynamics. The classification of multi-class sentiment determines all the sentiments containned in the provided tweets and attributes with different scores considering the sentiments portraying the weight or to determine how pertinent they are considered in a tweet. The classification of multi-class sentiment is based on the writing patterns considering the set of features to classify ternary sentiment using the tweets. These cores of the sentiment analysis are employed to filter the sentiments using the tweets. A model, namely Glo Ve – Deep Convolution Neural Network (DCNN), is devised to classify the sentiment. Recently, the deep learning methods gained interest, which significantly improved the classification accuracy. The sentiment information is represented as word representation and words embedding that concatenated the feature of prior polarity score. The feature sets are integrated and subjected in the DCNN to predict the sentiment labels using the tweets obtained from the social media that helped in information retrieval methods. Information retrieval (IR) is an extensive and inter disciplinary exploration that contained relevance rankings, search engines, and information indexing. Moreover, the evaluation measures used for sentiment classification involved recall and precision. IR is concerned with determining the critical information based on user queries. The IR arena also covered supporting users in filtering or browsing data or further process the set of retrieved data. The domain of information retrieval has emanated a long way and facilitated the more comfortable and fast discovery of information. The purpose of IR is to counterpart the user's requirements for acquiring the required information.

To compute information retrieval effectiveness in a standardized way, a test collection with three things should be mainly considered. The first is the collection of documents; the second is the generation of queries using test suites, and the third is the set of relevant judgments wherein a binary assessment of either non-relevant or relevant for each pair of query-document. IR integrates the space and dimensions for dealing with classical retrieval issues. Besides, the queries required a collective learning system for determining the matching between the spatial domain and thematic domain. Moreover, it is essential to provide a match between the user's query and data. The relative incompetence of IR models is mainly caused by the inaccuracies using a query . The field of attacking may cause serious issues that users face in the information-hidden world . Due to the exponential growth of information, the IR can play a progressively imperative role in the future. In this research, proposed SMCA-based Deep RNN is employed for classifying the sentiment from the telecom reviews. Here, the pre-processing is adapted to eliminate irrelevant reviews and reduce the seeking time using stemming and removing stop words. The mining of features is carried out with Senti Word Net, wherein positive score, negative score, polarity, and subjectivity features are extracted with other features, like elongated words, punctuation, hash tag, and numerical values. The Deep RNN is employed for classifying the sentiments using the proposed SMCA, which is designed by integrating SMO and CSA. The sorting of sentiments manually is a complex, time-consuming, and expensive task.

2. LITERATURE REVIEW

2.1. Syed Mohammed Anas (2021) et al The focus of this research is to create an environment of online E-commerce industry where consumers build trust in a platform where the products they purchase are genuine and feedbacks posted on these websites/applications are true, are checked regularly by the company where the number of users is increasing day by day, henceforth companies like Twitter, What's app, Facebook use sentiment analysis to check fake news, harmful/derogatory posts and banning such users/organizations form using their platforms. Parallel to that E-commerce (Flipkart, Amazon) industries, hotels booking (Trivago), logistics, tourism (Trip Advisor), job search (LinkedIn, Glass door), food (Twiggy, Zomato), etc. use algorithms to tackle fake reviews, spammers to deceive the consumers in buying below average products/ services. And the users need to be alerted of the spammer like "not verified profile" hence users need not worry about such false users. Manual labeling of the reviews is practically time-consuming and less effective. So supervised learning model is used for labeling the reviews and then predicting the label is not feasible. For example, Mukherjee et al. manually labeled 2431 reviews for over 8 weeks, so automated review labeling should be possible to contain time and energy which is difficult and proposed by Sunil Soumya et.al. Some industries pay money to write fake reviews of their products and services where it is not possible to label the review as spam or not spam. Amazon's "Yelp" dataset has 30% to 40% of spam reviews. Feature selection is an important aspect in selecting and training these models. In this work, the comparison of two models developed to justify the model performance for this "Amazon's yelp" dataset and their relevance to deploy these models in real-time software. The disadvantage is that Random forests model performed well compared to the Naïve Bayes algorithm by a large margin

2.2. Jianrongyao (2021) et al proposed approach can find the best sampling ratio for each classifier effectively and accurately. Next, we conduct an ablation study for feature pruning. Specifically, we calculate the influence of each predictive feature by dropping one of them at a time. The basic idea is that since the performance of a classifier can be evaluated by its F1-score, the usefulness of a feature can be calculated by recording how the F1-score changes when this feature is dropped. Next, the grid search method is applied once again to optimize the parameters for each classifier. To avoid over fitting, 10% of the data are selected as the validation data for the optimization task. After we optimize the base classifiers, we apply the voting and stacking ensemble strategy to integrate them. The ensemble model can reduce the relative weaknesses of single classifiers as they are compensated by the advantages of other classifiers. And this approach has been proven to be effective in improving the performance of classification. And two Yelp datasets are used to evaluate the model. To the best of our knowledge, few studies on fake review detection have combined these processes in a model. The present study contributes to the literature by providing a different way to effectively detect fake reviews. Specifically, we initially look at a very novel approach by combining data resembling with the grid search method to address the data imbalance problem, as this can effectively improve the performance of the model to a large extent on an imbalanced dataset. The results show that the proposed approach can effectively improve the performance for each classifier, especially for the random forest (RF), who's F1- scores are improved by 4.65% and 2.98% on the two Yelp datasets, respectively. Prior studies proposed many features for fake review detection. However, few studies have focused on the usefulness of the selected features, and there is no definition of which features are useless.

2.3. Ahmed M. Elmogy (2021) et al Machine learning techniques can provide a big contribution to detect fake reviews of web contents. Generally, web mining techniques find and extract useful information using several machine learning algorithms. One of the web mining tasks is content mining. A traditional example of content mining is opinion mining which is concerned of finding the sentiment of text (positive or negative) by machine learning

where a classifier is trained to *analyze* the features of the reviews together with the sentiments. Usually, fake reviews detection depends not only on the category of reviews but also on certain features that are not directly connected to the content. Building features of reviews normally involves text and natural language processing NLP. However, fake reviews may require building other features linked to the reviewer himself like for example review time/date or his writing styles. Thus the successful fake reviews detection lies on the construction of meaningful features extraction of the reviewers. To this end, this paper applies several machine learning classifiers to identify fake reviews based on the content of the reviews as well as several extracted features from the reviewers. We apply the classifiers on real corpus of reviews taken from Yelp. Besides the normal natural language processing on the corpus to extract and feed the features of the reviews to the classifiers, the paper also applies several features engineering on the corpus to extract various *behaviors* of the reviewers. The paper compares the impact of extracted features of the reviewers if they are taken into consideration within the classifiers. The papers compare the results in the absence and the presence of the extracted features in two different language models namely TF-IDF with bi-grams and TF-IDF with tri-grams. The results indicate that the engineered features increase the performance of fake reviews detection process. The disadvantage is *that, Developing* Algorithm*s That Can Work Efficiently in t*he Real World.

2.4. Rami mohawesh (2021) et al In this era of the internet, customers can post their reviews or opinions on several websites. These reviews are helpful for the organizations and for future consumers, who get an idea about products or services before making a selection. In recent years, it has been observed that the number of customer reviews has increased significantly. Customer reviews affect the decision of potential buyers. In other words, when customers see reviews on social media, they determine whether to buy the product or reverse their purchasing decisions. Therefore, consumer reviews offer an invaluable service for individuals. Positive reviews bring big financial gains, while negative reviews often exert a negative financial effect. Consequently, with customers becoming increasingly influential to the marketplace, there is a growing trend towards relying on customers' opinions to reshape businesses by enhancing products, services, and marketing. The way consumers openly express and use their feedback has contributed to issues with websites containing customer reviews. Social media (Twitter, Facebook, etc.) allows anyone to freely post feedback or critiques of any company at any time with no obligations or limits. The lack of restrictions, in turn, leads certain companies to use social media to unfairly promote their goods, brands or shops, or to unfairly *criticize* those of their rivals. For example, suppose a few consumers who bought a specific digital camera posted negative reviews on image quality. These reviews portray the digital camera *unfavorably* to the public. Thus, the camera manufacturer might employ an individual or team to post fake positive reviews about the camera. Similarly, in order to promote the company, the producer might ask the hired persons to post negative comments about competitors' products. Reviews published by people who have not personally encountered the items being reviewed are considered fake reviews. This paper presented an extensive survey of the most notable works to date on machine learning-based fake review detection. Firstly, we have reviewed the feature extraction approaches used by many researchers. The disadvantage is *that, Developing* Algorit*hms That Can Work Efficiently i*n *the r*eal *w*orld.

2.5. Rakibul Hassan (2020) et al Thus, online reviews play a very crucial role in decision making while purchasing products online. This also creates opportunity for some groups of bad people to deceive people with fake comments and reviews. They can post fake reviews for the promotion of their goods or to demote the products of the competitor. That is why, detecting fake online reviews is very important for both the users to get benefited from reviews as well as the companies to maintain their goodwill to the consumers. Detecting fake online reviews is basically a binary classification problem. Many researchers are working on it to auto detect the fake reviews using various machine learning techniques. Most of the researches are based on supervised learning. Some semi-supervised and clustering approaches are also taken by some of the researchers. All of these studies, mainly focus on two aspects: reviewers and reviews. Detection system based on only reviews is call content-based study. Classification tasks are done using the features extracted from the content of the reviews that includes what is written in the text. The reviewer-based study is also called user *behavior* based classification. How often a reviewer reviews, how varieties of products are reviewed by one reviewer and also the timing and group connections of the reviewers are *analyzed* here. One of the fundamental problems of all these studies is data *labeling*. It is very difficult to label the dataset for human by watching the content. This can create garbage out of garbage situation if reliable *labeling* is not available. In such cases, semi-supervised or clustering approaches are suggested by many researchers. But, with reliable labeling, supervised classification approaches always perform better. In this research paper, we have introduced some supervised machine learning classification techniques to detect fake online reviews with a good accuracy. Classifications are done using content-based features. The disadvantage is that, Development of a feature set that can perform better classification.

3. ARCHITECTURE DIAGRAM

System architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.
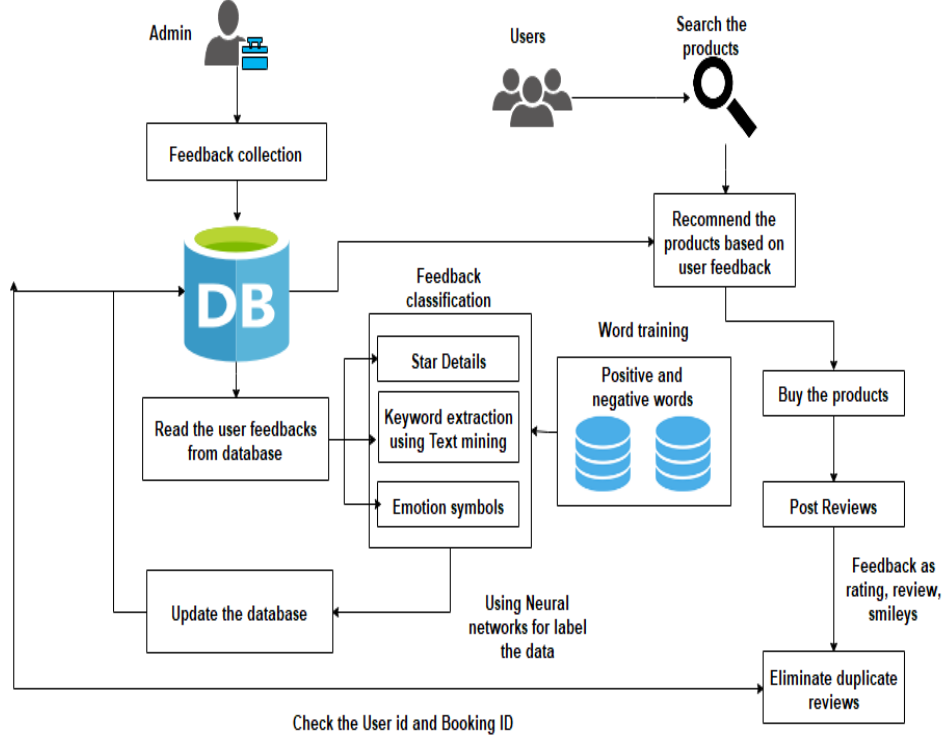
DESCRIPTION:

A recommendation system has been implemented based on hybrid approach of stochastic learning and context based engine. We have tried to combine the existing algorithms for recommendation to come up with a hybrid one. It improves the performance by overcoming the drawbacks of traditional recommendation systems. Recommender systems being a part of information filtering system are used to forecast the bias or ratings the user tends to give for an item. Among different kinds of recommendation approaches, collaborative filtering technique has a very high popularity because of their effectiveness. These traditional collaborative filtering systems can even work very effectively and can produce standard recommendations, even for wide ranging problems. For item based on their neighbour's preferences entropy based technique creates better suggestions than others. Whereas other techniques like content based suffers from poor accuracy, scalability, data sparsity and big-error prediction. To find these possibilities we have used user-based collaborative filtering approach. A product recommendation system is a software tool designed to generate and provide suggestions for items or content a specific user would like to purchase or engage with. Product recommendation will analyse the existing things more specifically, we focus on fashion products and develop a method that only requires a single input image to return a ranked list of similar-style recommendations. In this project we can implement hybrid filtering algorithm which contains the input such as text, rating and reviews. Text can be analysed using text mining algorithm and back propagation neural network algorithm. Comments are read by Natural language processing system to remove the stop words, stemming words and extract the key terms. These key terms are matched with database and classified using back propagation neural network algorithm. Finally recommend the products based on user feedbacks

3.   MODELING AND ANALYSIS

E-commerce framework is used to buy the products in online to easy retrieval the mobile products. This module is used to create android and web site for recommending best mobiles in specific area. Admin is the responsibility for maintaining the all details in server and server can be design in server. There are two accounts such as admin and user account. Admin can login to the system and post item details with expiry dates. User can login to the mobile to choose the language and area. Then view the products with specified filter. This module is used to create web site buy or post products for users. Admin can login to the system and post products with features. User can login to the system to view product details.

REVIEWS COLLECTION:

Admin collect reviews and have various types of reviews. Reviews may be rating reviews, text reviews and smiley's reviews. All reviews are stored in database for future evaluation. Ratings, reviews and emoticons are stored in database. Rating, Reviews and Emoticons are the evaluation or assessment of something, in terms of quality (as with a critic rating a novel), quantity or some combination of both.

SENTIMENT ANALYSIS:

Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and ratings for applications that range from marketing to customer service to buy the products efficiently. Admin can analyze whether the product is positive or negative. In star rating, we can calculate star count values. In text reviews, extract keywords and matched with database. Then smileys reviews are calculated based positive and negative symbols.

RECOMMENDATION SYSTEM:

Recommender systems are a subclass of information filtering system that seek to predict the "rating" or "preference" that a user would give to an item. User can search the product in search bar. And view the list of products based on price and review details. Implement the stochastic learning algorithm to classify the products such as positive or negative. Positive products are display in recommendation panel based on ratings and reviews. If the product has negative review means, automatically the positive products in recommendation panel.

FAKE REVIEWS MONITORING:

In this module, fake reviews are analyzed by admin. A media access control address (MAC address) of a computer is a unique identifier assigned to network interfaces for communications at the data link layer of a network segment. Admin can get user account details, Mobile address and Order id details. So user can post one reviews that will be genuine reviews.

4.   RESULTS AND DISCUSSION

1.   It is difficult to find the product based on customer need in offline market.

2.   In offline market there are sometime short of product or have not idea what product to by with a similar use product .

3.   Various limitations are lack of consumer satisfaction, no personalized recommendation, unable to solve cold start problem, limited resource situation not properly handled, data valid time not handled properly and less efficient

 I. Proposed System

Proposed system is a better E-Commerce recommendation system that can give out effective recommendations for users which are satisfying to them to a great extent. Customers can get many benefits and also the trading volume can be increased and the above mentioned three problems are also overcome. Proposed system aims at implementing the recommendation system for customers to get the items they want. Proposed recommendation system mainly consists of 3 models namely –the user model, the recommended model and the recommendation algorithm as shown in fig-1. Proposed system satisfies the consumers to a better extent. Proposed system makes use hybrid algorithm to overcome the three problems.
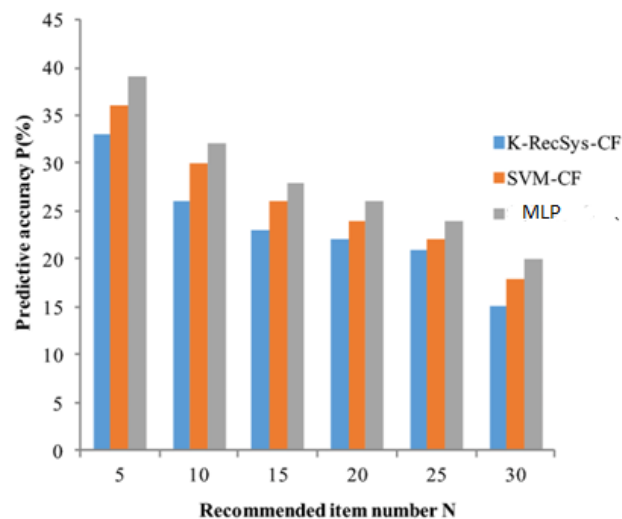
5. CONCLUSION

In this project, we have proposed a novel implementation of a product recommendation system based on deep learning with text mining algorithm. The main advantages of our method are a visual organization of the data based on the underlying structure, and a significant reduction in the size of the search space per result output. And user can easily search the products anywhere and anytime. Ratings, reviews and emoticons are analyzed and categorized as positive and negative sentiments. Search the products based on price based filtering and reviews based filtering. MAC based filtering approach can be used to avoid fake reviews. Supermarket can benefits because easy buying, easy transactions and to get more customers. Our method was evaluated against real user data collected through an online website, by using a subset of the movies liked by each user as input to the system. The current results are notably better than random approach. However, we feel that with a better dataset and a number of improvements to our method, we may achieve better results. Hybrid Recommendations is one of the main modules of the system which helps overcome the drawbacks of the traditional Collaborative and Content Based Recommendations. We have obtained promising results using our current model.

REFERENCES

[1] Anas, Syed Mohammed, and SantoshiKumari. "Opinion Mining based Fake Product review Monitoring and Removal System." 2021 6th International Conference on Inventive Computation Technologies (ICICT). IEEE , 2021.

[2] Yao, Jianrong, Yuan Zheng, and Hui Jiang. "An Ensemble Model for Fake Online Review Detection Based on Data Resampling, Feature Pruning, and Parameter Optimization." IEEE Access 9 (2021): 16914-16927.

[3] Elmogy, Ahmed M., et al. "Fake Reviews Detection using Supervised Machine Learning." Int. J. Adv. Comput. Sci. Appl 12 (2021).

[4] Mohawesh, Rami, et al. "Fake Reviews Detection: A Survey." IEEE Access 9 (2021): 65771-65802.

[5] Hassan, Rakibul, and MdRabiul Islam. "A Supervised Machine Learning Approach to Detect Fake Online Reviews." 2020 23rd International Conference on Computer and Information Technology (ICCIT). IEEE, 2020.

[6] Wang, Jingdong, et al. "Fake review detection based on multiple feature fusion and rolling collaborative training." IEEE Access 8 (2020): 182625-182639.

[7] Liu, Meiling, et al. "Detecting Fake Reviews Using Multidimensional Representations With Fine-Grained Aspects Plan." IEEE Access 9 (2020): 3765-3773.

[8] Wang, Xinyue, et al. "Identification of fake reviews using semantic and behavioral features." 2018 4th International Conference on Information Management (ICIM). IEEE, 2018.

[9] Liu, Wenqian, et al. "A method for the detection of fake reviews based on temporal features of reviews and comments." IEEE Engineering Management Review 47.4 (2019): 67-79.

[10] Hassan, Rakibul, and MdRabiul Islam. "Detection of fake online reviews using semi-supervised and supervised learning." 2019 International conference on electrical, computer and communication engineering (ECCE). IEEE, 2019.