

# E-Commerce Based Chat Bot System Using Text Mining Algorithm

Asst. Prof.Mr.M.Anand, Asst. Prof.Mr.M.Saravanan, Atla Sasikanth Reddy,  
Bathini Sai Bharath Kumar, Chigicherla Dhanunjaya  
Department Of Artificial Intelligence and Data Science,  
Muthayammal Engineering College,Rasipuram.

**Abstract-**Internet purchasing is booming in the current e-commerce landscape. As a result, there is room for advancement in product recommendation systems. Because users need a connection to the system. The user experiences personalised attraction as a relationship progresses. The technology encourages customers to return and spend more money in addition to monitoring and analysing their purchasing behaviour. The tiresome task of people looking through endless categories for what they want is eliminated by the suggestion system. Instead, they use the conversation to weed out superfluous information and provide the consumer what they want. Online shopping provides many benefits, but there are also restrictions and disadvantages that need to be taken into account. The consumer can be upset if the requested product and the one actually received do not always match. Enhancing the current functioning of these systems has become essential since customer requirements change on a regular basis. The history of internet shopping indicates that there will soon be a big need for recommendation systems. A conversational bot that recommends things to customers based on their requirements is being introduced by research. With little user input, the chat bot effectively processes orders and suggests the best item. The product database is utilised in this instance, however this may be done on a much larger scale. The consumer communicates details about the scent to the chatbot. Based on the user's description, it will also suggest relevant items.

**Index Terms-**E-Commerce, Chat bot, Natural Language processing, Machine learning, Recommendation system.

## I. INTRODUCTION

An electrical source supplies energy to a transmitter, which in E-commerce websites, smartphone applications, and messaging services like Facebook Messenger and WhatsApp may all incorporate an e-commerce chatbot for product recommendations. It can run round-the-clock, offering clients round-the-clock support, which may help organisations boost customer happiness and enhance customer service. Different machine learning techniques, including collaborative filtering, content-based filtering, and hybrid methods, can fuel the chatbot's recommendation engine. A method called collaborative filtering suggests items based on previous interactions between users and the actions of other users with similar tastes. Contrarily, content-based filtering suggests goods based on their attributes, such as their specs and descriptions, and matches them to the user's preferences. To provide suggestions that are more precise, a hybrid strategy combines the two methodologies. By using information from social media

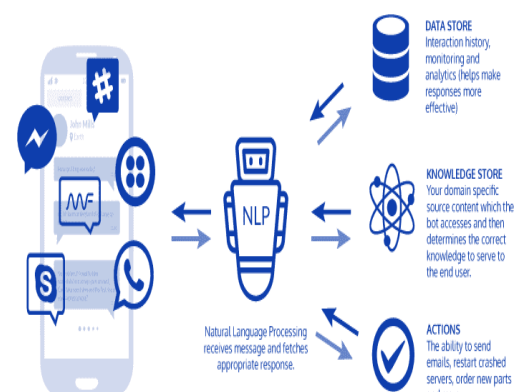


Fig 1: CHATBOT Architecture

accounts and past purchases, the chatbot may also offer tailored suggestions. For instance, the chatbot might propose complementary products or items from the same brand if a user has discussed a certain brand or product on social media. An e-commerce chatbot for product suggestion may help businesses increase sales by raising the likelihood that a user will make a purchase, which is one of its main advantages. The chatbot may direct customers towards things they are

more likely to be interested in and buy by making them personalised suggestions, which will enhance sales and income. In conclusion, an e-commerce chatbot for product recommendation is a useful tool for companies trying to enhance customer service, boost revenue, and provide customers a better overall shopping experience. The Basic chatbot flow is shown in fig 1.

## II. RELATED WORK

**Syed Mohammed Anas, et.al,[1]** To demonstrate the model performance for this "Amazon's yelp" dataset and their applicability for deployment in real-time applications, a comparison of two models was created. As a result, the Random forests model outperformed the Naive Bayes method by a wide margin. The goal is to choose an algorithm to complete the work of detecting fake reviews and eliminating them. The fake review detection problem is thoroughly handled and provides a fair look into its legality and need. Future research can test hybrid models and brand-new models for the bogus review detection model. The research can accelerate the execution process by utilising Google Co-Lab and NVIDIA graphics GPU. The approaches of Naive Bayes and Random Forest are used to create this model. One can rapidly determine the quantity of spam reviews on a website or application by using these models. A sophisticated algorithm that has been trained on millions of reviews is needed to combat these scammers. The "Amazon Yelp dataset" is utilised in this study to train the models. Its very tiny dataset is used for training on a very small scale and may be expanded to get great accuracy and flexibility.

**Jianrongyao, et.al,[2]** a methodology for detecting false reviews using an ensemble approach is suggested. Data resampling, feature reduction, parameter optimisation, and classifier ensembling are the four phases that make up the model. The initial three stages are suggested for improving the fundamental classifiers. On two example Yelp datasets with unbalanced distributions, the performance of the ensemble model is greatly enhanced by optimising the basic classifiers; this performance is noticeably superior to that of several conventional models and methods. The study has a number of theoretical and practical consequences. Since it outperforms other meta-classifiers in the proposed model and the parameters are also optimised using the grid search method, we choose the Gaussian kernel support vector machine (SVM) classifier as the meta-classifier in the current study (because it performs best on the proposed datasets). To further optimise the model, we additionally use the meta-classifier with the suggested data resampling technique.

**Ahmed M. Elmogy, et.al,[3]** suggests using machine learning to detect false reviews. This research uses numerous features engineering techniques in addition to the reviews' feature extraction procedure to extract different reviewer behaviour.

The performance of several tests conducted on a genuine Yelp dataset of restaurant reviews with and without characteristics derived from user behaviours is compared in the study. We evaluate the performance of various classifiers in both scenarios, including KNN, Naive Bayes (NB), SVM, Logistic Regression, and Random Forest. The assessments also take into account several n-gram language models, particularly bi-gram and trigram. The research utilises numerous features engineering techniques to the corpus in addition to the standard natural language processing to extract and feed the features of the reviews to the classifiers. This allows for the extraction of different reviewer behaviour. The research examines the effects of reviewers' extracted characteristics when they are taken into account by the classifiers. The studies examine the outcomes of two distinct language models, TF-IDF with bi-grams and TF-IDF with tri-grams, in the presence and absence of the retrieved characteristics. The findings show that the built features improve the effectiveness of the method for identifying bogus reviews.

**Rami mohawesh, et.al,[4]** 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. Then, we detailed the existing datasets with their construction methods. Then, we outlined some traditional machine learning models and neural network models applied for fake review detection with summary tables. Traditional statistical machine learning enhances text classification model performance by improving the feature extraction and classifier design. In contrast, deep learning improves performance by enhancing the presentation learning method, algorithm's structure, and additional knowledge. We also provided a comparative analysis of some neural network model-based deep learning and transformers that have not been used in fake review detection. The outcomes showed that Robert achieved the highest accuracy on both datasets. Further, recall, precision, and F1 score proved the efficacy of using Robert in detecting fake reviews. Finally, summarized the current gaps in this research area and the possible future direction to get robust outcomes in this domain.

**Rakibul Hassan, et.al,[5]** developed an effective supervised machine learning approach to classify fake online reviews using a dataset that contains hotel reviews from online websites. And detecting fake online reviews is basically a binary classification problem. Purchasing online products is one of our daily activities. Now-a-days, we can get almost everything from various online marketplaces. When we think about purchasing something, almost every one of us first check the product in websites like

**Amazon, AliExpress, eBay etc.** In case of traveling: hotel booking, purchasing air tickets and all forms of other tasks also can be done with the help of online service providers.

But, as we cannot know physically, what products or services we are purchasing, we check what other people talk about the services or products. 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. Many researchers are working on it to auto detect the fake reviews using various machine learning techniques. Most of the research are based on supervised learning. Some semi-supervised and clustering approaches are also taken by some of the researchers.

**Jindong Wang, Haitaokan, et.al,[6]** proposed a fake review detection model based on the combination of multi-feature fusion and rolling collaborative training. Experimental results show that this method is more effective than traditional algorithms. It uses unlabelled data to improve the performance of the classification system and has better classification accuracy. At the same time, the consistency of sentiment and score is analysed, and the feature extraction of the review is carried out through the text representation model, and the feature fusion is combined with the external features of the text, which can effectively improve the classification effect of the classification model. The main reason is that the writing methods of professional fake writers will be changed according to the update of the detection mechanism of the e-commerce platform, and we will try our best to bypass. Through the detection mechanism, it is more difficult to be found, and corresponding countermeasures can be made based on consumer psychology and times like shopping festivals. And hope that the next researchers can continue their research from the direction of dynamic update detection strategy.

**Meilingliu, et.al,[7]** propose an attention-based multilevel interactive neural network model with aspect constraints that mines the multilevel implicit expression mode of reviews and integrates four dimensions, namely, users, review texts, products, and fine-grained aspects, into review representations. We model the relationships between users and products and use these relationships as a regularization term to redefine the model's objective function. The experimental results from three public datasets show that the model that we propose is superior to the state-of-the-art methods: thus, showing the effectiveness and portability of our model. To verify the effectiveness of the MIANA, we conducted extensive experiments on three public datasets. The experiments showed that the classification effect has been significantly improved, that the MIANA outperforms the state-of-the-art methods for fake review detection tasks and proved the effectiveness and feasibility of proposed scheme.

**Xinyue Wang, et.al,[8]** provide two types of features, one is behavioural feature set, the other is related to semantic. We propose three feature sets by performing semantic analysis. Our study shows that although behavioural features reported very high detection accuracy with supervised classifiers, semantic features are also work well on real-life fake reviews in the commercial setting of Yelp.com. Among the semantic features, readability features, topic features and gram features provide approximately the same results in performance. For n-gram features, the feature vectors were constructed using Count and Tf-Idf values of the review content and classifiers of SVM and NB are used for classification process. Overall, the classifier of LR performed the best, which achieves up to 97.2% accuracy with all features. Nowadays with the advent of the e-commerce, an increasing number of people are taking pleasure of shopping online, and then sharing their opinions on the electronic business website. These online opinions may be used by customers and merchants when they make purchase and other decisions. For the online reviews, positive reviews play a simulative role in recapping economic benefits and well-deserved reputation for merchants' businesses. Thus, it makes merchants have strong intentions to manoeuvre their fame and employ specialized imposters posting higher opinions on the shopping sites.

**Wenqian Liu, et.al,[9]** studied the review records of online shopping sites and proposed a novel approach for the detection of fake reviews of products. This review outlier detection method detects the outlier products by analysing the temporal trends of reviews and comments. Such perspective makes our method more advantageous than some existing methods. Also, detecting review records is efficient and easy. To accomplish the research, we first analysed the characteristics of online reviews based on an Amazon China dataset. We then proposed an isolation forest-based method for fake review detection. We also compared method with several temporal outlier detection methods to prove the effectiveness and the efficiency of our method. And proposed method can better detect outlier products according to the abnormal changing trends, the experiment also showed that the selection of window size for the review records plays an important role. A proper window size will yield better performance in terms of the accuracy of detection. We observed that outlier reviews are generally bursting, generally happening in only a few days or weeks.

**Rakibul Hassan, et.al,[10]** shown several semi-supervised and supervised text mining techniques for detecting fake online reviews in this research. We have combined features from several research works to create a better feature set. Also, we have tried some other classifier that were not used on the previous work. Technologies are changing rapidly. Old technologies are continuously being replaced by new and sophisticated ones. These new technologies are enabling

people to have their work done efficiently. Such an evolution of technology is online marketplace. We can shop and make reservation using online websites. Almost, every one of us checks out reviews before purchasing some products or services. Hence, online reviews have become a great source of reputation for the companies. Also, they have large impact on advertisement and promotion of products and services. With the spread of online marketplace, fake online reviews are becoming great matter of concern. People can make false reviews for promotion of their own products that harms the actual users. Also, competitive companies can try to damage each other's reputation by providing fake negative reviews.

### III. BACKGROUND OF THE WORK

Consumers use the evaluation or opinion of other people as an important information source. People like to get recommendations when they perceive a risk in making a purchase decision or when they want to simplify their buying decision. Recommendation becomes even more important on the Internet-based shopping environment where consumers do not make physical contact with products and face higher cognitive risk. In addition, e-commerce sites offer a very large number of alternatives since they do not have any physical constraint on inventory or shelf space. Hence, consumers may be confused by the number of choices. If the consumer is not familiar with the Internet, the problem becomes even more serious to solve these problems, several e-commerce sites are employing recommender systems to help their customers make their purchase decisions more efficiently. A recommender system is an electronic agent that helps customers to find the most valuable products/services based on their historical preferences or tastes. In existing system implement the collaborative filtering algorithm to check user details and text-based query system applied to recommend the products.

### IV. PROPOSED WORK

Chatbots have the potential to revolutionise customer service and internet support. Fashion firms have started using chatbots to offer individualised customer experiences as e-commerce has grown. Technology improvements and customer behaviour have been the focus of research on chatbots for fashion e-commerce, but little has been done to analyse chatbot characteristics from a holistic standpoint. This project's goal is to provide an interdisciplinary evaluation by thoroughly classifying current research on the subject and to guide future studies in the field. With the help of machine learning algorithms and natural language processing, we can use the framework for product recommendations in this project, which has a higher accuracy rate. NLP integrates cognitive algorithms including statistical, machine, and deep

learning algorithms with computational linguistics, which is the rule-based modelling of spoken human language. The intelligent voice assistants and chatbots that you may use in daily life are made possible by the combination of these technologies. Artificially intelligent chatbots are designed to replicate human characteristics and behaviours, as the name indicates. Such chatbots are greatly made possible by NLP, or natural language processing, which helps computers grasp the nuances and undertones of human speech. A genuinely intelligent chatbot is produced when NLP and artificial intelligence are coupled. This chatbot can answer to complex inquiries and learn from every contact to produce more appropriate responses in the future. The AI chatbots were created to help human consumers with various goods. The proposed work's general architecture schematic is shown in Fig 2.

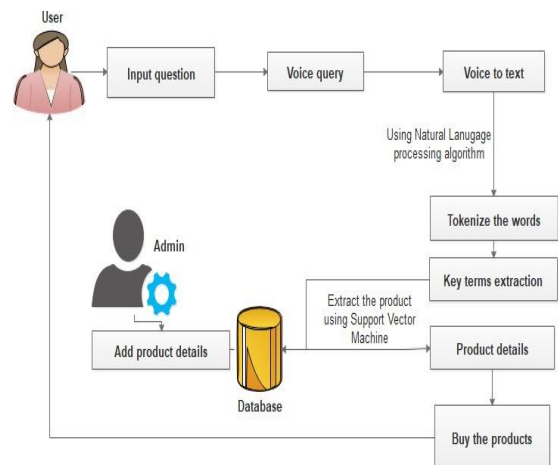


Fig. 2 Proposed Architecture

#### 1. Interface Creation

E-commerce platforms are framed using recommendation system. This system tries to recognise customer's behaviour and then recommend the products according to their interest. 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 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. User can login to the system to view the interface. Then view the products with specified filter. This module is used to create web site buy or post products for users.

#### 2. Post Questions

A Chabot or chatter bot is a software application used to conduct an on-line chat conversation via text or text-to-speech, in lieu of providing direct contact with a live human



agent. A Chabot is a type of software that can help customers by automating conversations and interact with them through messaging platforms. A chat bot will help the user to get the input in the form of search query and then provide the output as a recommended multiple products what user is looking for. Initial work for research is collecting a data. Data required for this research contained the details in the form of name, brand, text descriptions, reviews, and a list of notes.

### 3. Keyword Extraction

As we are using natural language processing, the text data must be pre- processed. It covers some tasks like making text data to lower case, removing stop words, tokenization, stemming, etc. Lowercasing – Lowercasing is the first step in data pre- processing. The step is simple but important. Making the entire text to lowercase is essential step to get the consistent output. Tokenization – Once lowercasing is done, tokenization will take place. It means the sentences are divided into substrings known as Tokens. These tokens can be used to find the words in sentences. Tokenization basically refers to splitting up a larger body of text into smaller lines, words. The various tokenization functions in-built into the nltk module itself. A simple regular expression based tokenize RegexpTokenizer provided by NLTK was used which splits the text into punctuations and whitespaces words get removed we can focus on important, meaningful words. A Natural language toolkit (NLTK) was used to load the stop words and remove them.

### 4. Top- K Results

In this module implement Support vector machine algorithm to provide the products. In machine learning, support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyse data and recognize patterns, used for classification and regression analysis. SVMs work based on the principle of Structural Risk Minimization Principle. SVMs work very well with text data and Chabot's because of the high dimensional input space due to large number of text features, linearly separable data, and the prominence of sparse matrix. It is one of the most popularly used algorithms for text classification and intent identification. This allows us to assess how likely an input is to being in one category or another. Cross validation is the most common way of testing this algorithm, assessing how accurate the output produced from this procedure are based on training and test sets. Precision and recall metrics are also used to assess the performance of this model.

### 5. Buy the Products

Recommendation systems have come a long way since collaborative filtering and content-based systems. Considering comments as an important piece of date, which needs to be

processed to extract information out of it and possibly combine its use cases with other recommendation systems. Since user comments are a direct form of data from the user it contains important keywords based on which further recommendations can be made. This algorithm can be implemented not only to an Ecommerce application but also various applications of the same kind where users and a service or product is involved. After the results, user buy the products and updated into databases.

## V. EXPERIMENTAL RESULTS

In this research, the framework for an E-Commerce framework login page that includes a chat bot interface may be implemented. The findings are displayed in the figures below.

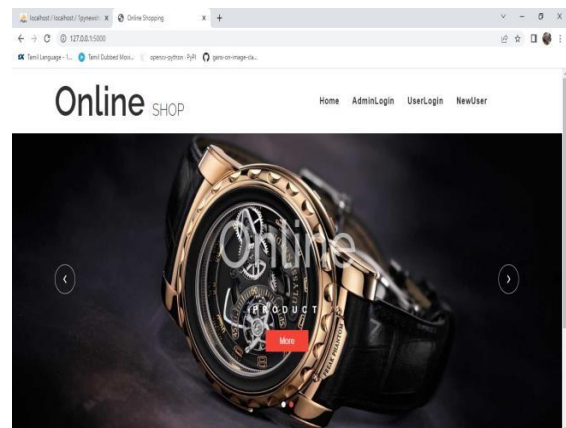
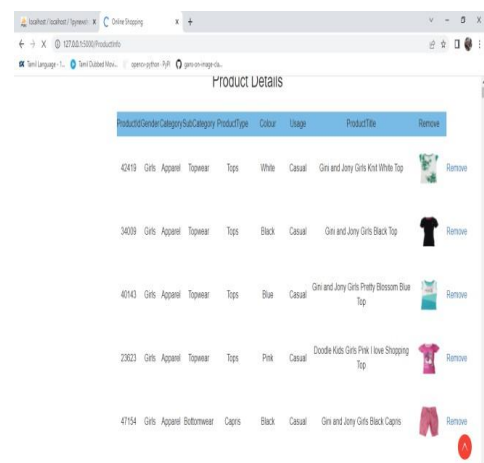


Fig 3: Home page



ProductID	Sender	Category	SubCategory	ProductType	Colour	Usage	ProductTitle	Remove
42418	Girls	Apparel	Topwear	Tops	White	Casual	Girls and Jony Girls Kid White Top	Remove
34039	Girls	Apparel	Topwear	Tops	Black	Casual	Girls and Jony Girls Black Top	Remove
40143	Girls	Apparel	Topwear	Tops	Blue	Casual	Girls and Jony Girls Pretty Blossom Blue Top	Remove
23823	Girls	Apparel	Topwear	Tops	Pink	Casual	Doode Kids Girls Pink Love Shopping Top	Remove
47154	Girls	Apparel	Bottomwear	Caprs	Black	Casual	Girls and Jony Girls Black Caprs	Remove

Fig 4: Product datasets



Fig 5: Chat bot interface

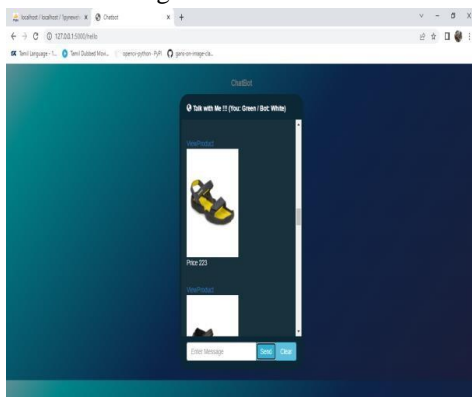


Fig 6: Product recommendation

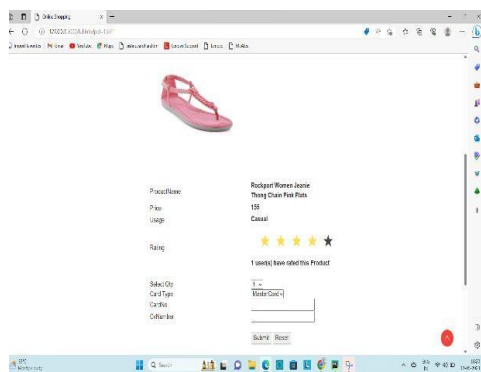


Fig 7: Buy the products

## IV. CONCLUSION

Parallel to the development of artificial intelligence (AI), chatbots have been on the increase in recent years. The basic components of a modern chatbot design may be broken down into four categories: 1) knowledge (open or closed domain), 2) answer production (retrieval or generative), 3) text processing (vector embedding or Latin alphabet), and 4) machine learning (ML) model (typically utilising neural

network). In this project, we've demonstrated a cutting-edge use of a chatbot-based product recommendation system. Building such a modular system was done in order to make the system accessible on more platforms. An architectural design and implementation review of contemporary chatbot systems is presented in this research. This review's objective is to offer a general understanding of design practise and implementation strategy in contemporary chat-bot systems. The administrators' categorised training data is used by the current NLP engine to train its classifier. Additionally, it is SVM-based.

## REFERENCES

1. Anas, Syed Mohammed, and Santoshi Kumari. "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 DataResampling, 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. Compute. 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 (ICIT). IEEE, 2020.
6. Wang, Jindong, 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 behavioural 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.