Journal of Engineering Science and Technology Review 17 (1) (2024) 16 - 23

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Sentiment Based Product Recommendation System Using Machine Learning Techniques

N. Vaishnavi and B. Kalpana*

Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore-641043, Tamil Nadu, India.

Received 7 June 2023; Accepted 8 November 2023

Abstract

A sentiment-based product recommendation system is a system that utilizes natural language processing to extract sentiment features from reviews and the machine learning algorithms namely logistic regression, random forest classifier, XGBoost classifier and CatBoost classifier are applied to classify sentiments and these algorithms performance is evaluated based on the performance metrics such as accuracy, recall, precision and F1 score. Memory based collaborative filtering models such as user based and item-based recommendation models are developed and assessed using benchmark data, its performance is evaluated based on the evaluation metric Root mean square error (RMSE) and then the recommendations are fine tuned. While real-time data is obtained through web scraping from the Mamaearth website. Sentiment-based recommendations are implemented on real-time data, and their performance is evaluated using the same key performance metrics and the recommendations are fine-tuned. Finally, a Streamlit web app is created and deployed using Ngrok to enhance the accessibility and utility of the recommendation system.

Keywords: Machine Learning, Collaborative Filtering, User-Based Filtering, Web Scraping, Streamlit and Ngrok.

1. Introduction

Product recommendation systems are commonly used in various industries, including e-commerce, streaming services, and online content platforms. A product recommendation system is designed to suggest relevant items or products to users based on their preferences, behavior, or other relevant factors. It leverages algorithms and data analysis techniques to analyse user data and make personalized recommendations. A sentiment-based recommendation system considers user preferences along with sentiment or opinion expressed by users towards products or services. It utilizes natural language processing (NLP) techniques to extract sentiment-related features from user reviews, ratings, or feedback. Analyses of sentiment, such as positive, negative, or neutral, the system can generate recommendations that align with users' sentiment towards products. This approach helps to personalize recommendations based on users' emotional responses, leading to a more tailored and satisfying user experience. Current ecommerce websites can fine tune their strategies to improve customer satisfaction using sentiment-based recommendations. These innovations help to create a more customized and fulfilling user experience by tailoring recommendations based on users' emotional reactions. This method goes into the psychological and emotional aspects of user behavior and choice, going beyond simple algorithmic recommendations. This research holds significance in e-commerce and other industries as it can enable organizations to improve customer satisfaction. E-commerce websites can improve user engagement and loyalty by adjusting their methods to better match customers' emotional responses by utilizing the power of sentiment-based suggestions.

The primary objective of this research is to create, assess, and improve a sentiment-based product recommendation system that uses natural language processing (NLP) techniques to extract sentiment-related data and provide tailored suggestions. By conducting a thorough assessment, this research seeks to determine how well the system works to enhance user experience and customer happiness in the context of e-commerce, offering useful information and workable solutions to companies operating in this industry.

Sentiment analysis plays a vital role in numerous domains by providing valuable insights into the emotions, opinions, and attitudes of individuals expressed in textual data. It helps businesses understand customer sentiment, brand reputation, market trends, and social media sentiment. Machine learning models are widely employed in sentiment analysis tasks due to their ability to automatically learn patterns and classify sentiment from large datasets [1]. Star ratings in customer reviews serve as a commonly used metric to assess product satisfaction, and sentiment can be effectively incorporated through these ratings. In this study, the researchers classified star ratings into five distinct categories, representing a range of polarity values from strongly negative to strongly positive. By categorizing ratings based on sentiment, valuable insights into product satisfaction and customer opinions can be obtained [2]. By leveraging the power of recommendation systems, users can receive tailored suggestions that align with their tastes and preferences, facilitating a more personalized and satisfying shopping experience. It highlights the significance of recommendation systems in improving user satisfaction and streamlining the product selection process in the domain of apparel [3]. The Root Mean Square Error values of various algorithms are analysed and compared. Through this analysis, the most effective approach for generating

*E-mail address: nvaishnavi206@gmail.com
ISSN: 1791-2377
doi:10.25103/jestr.171.03

accurate recommendations is identified [4 - 13].Given the vast amount of research in this field , it was justified to conduct an in-depth analysis of memory based and model based collaborative filtering . The authors have also highlighted the importance of user satisfaction through ratings [14, 15]. However , the recommendation systems could still benefit from combining user sentiments, hence the motivation arose to include sentiment data to ratings .

While the existing literature studies provide insightful information about user satisfaction through ratings, recommendation systems, and sentiment analysis, this research combines the two to create a shopping experience that is more emotionally impactful. Theory and practice are brought together and e-commerce personalization is emphasized by identifying the most efficient algorithms for accurate recommendations through the analysis of RMSE values.

2. Methodology

The primary activities carried out during the development phase of a sentiment-based product recommendation system can be divided into different modules as shown in Fig 1.

2.1 Load The Dataset

The dataset used for sentiment-based product recommendation system is "sample30.csv" which is obtained from Kaggle website. It comprises of 30,000 records and 15 attributes. These attributes include id, brand, categories, manufacturer, name, reviews date, reviews did purchase, reviews do recommend, reviews rating, reviews_text, reviews_title, reviews_userCity, reviews_userProvince, reviews_username, and user_sentiment.

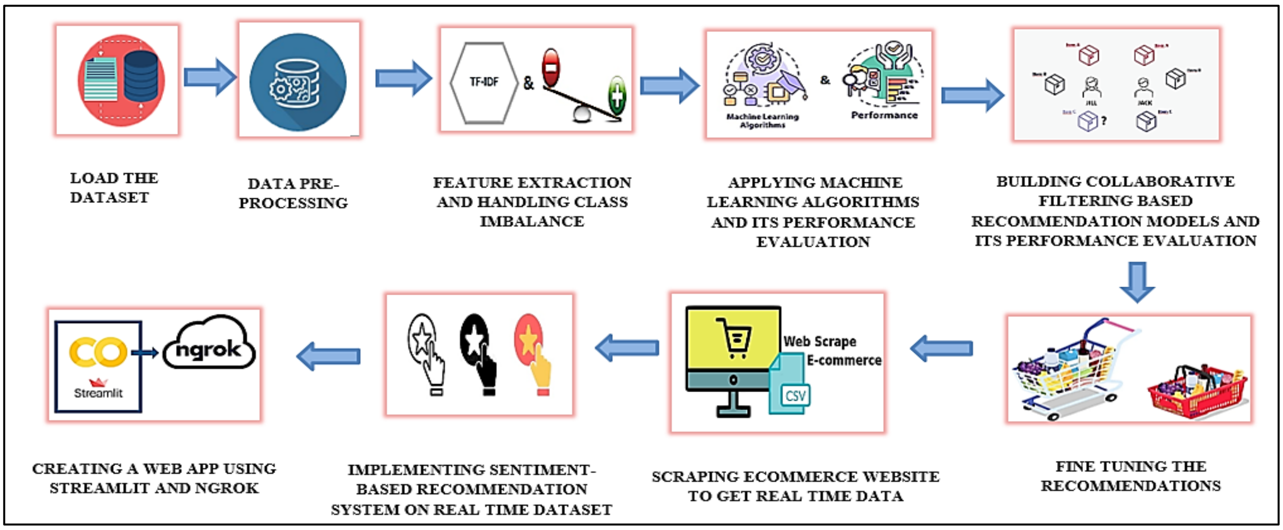


Fig. 1. Steps in recommendation system

2.2 Data Pre-Processing

Data pre-processing involves cleaning and transforming the data into a usable format by handling null or missing values. Text pre-processing focuses on preparing raw text data for natural language processing tasks, such as sentiment analysis. This includes conversion of uppercase letters to lowercase letters, removing punctuation, stop words, and special characters from the text to simplify it and eliminate noise.

Conversion of uppercase letters to lowercase letters

In text data preparation, every text is converted to lowercase. This removes case sensitivity and promotes consistency in activities involving natural language processing, such sentiment analysis. More accurate and consistent results in NLP applications are made possible by this method, which guarantees that uppercase and lowercase variations of the same word are treated as equivalent.

Removing punctuations

The goal of removing question marks, commas, and periods from text data is to make it simpler and rid it of symbols that are not necessary. In order to keep the focus on the words themselves, which convey sentiment and meaning, this phase is essential for reducing noise and maintaining text consistency for downstream analysis.

Removing Stop word

Stop words like “and”,”then”,”in” and so on must be eliminated in order to improve the quality of text data by getting rid of keywords that aren't important. By using a more condensed vocabulary, this step streamlines computational resources and improves the accuracy of natural language processing tasks while reducing noise dramatically.

Lemmatization

Lemmatization is the process of changing words to their most basic or dictionary form (lemma), such as "running" to "run," in order to guarantee that related meanings are consistently represented. The significance of standardization resides in its ability to improve accuracy and facilitate sentiment analysis by identifying common sentiments between words such as "run" and "running." Lemmatization reduces the vocabulary, producing base forms that are more accurate and significant for better sentiment classification.

Exploratory data analysis looks at user sentiment, ratings, and product reviews using visualization techniques like word clouds, bar charts, count plots, and histograms. Histograms show text character length, word clouds show frequently used words, and bar charts show sentiment distribution.

2.3 Feature Extraction and Handling Class Imbalance

The reviews text is vectorized using the Term Frequency and Inverse Document Frequency (TF-IDF) method, converting the textual data into numerical form. This vectorized representation is then utilized for classification tasks with machine learning algorithms, enabling effective analysis and prediction.

Synthetic Minority Over-sampling Technique (SMOTE) is used to handle class imbalance by generating synthetic data points for the minority class. This helped in balancing the data distribution and improves the performance of machine learning models.

2.4 Applying Machine Learning Algorithms and Its Performance Evaluation

The machine learning algorithms are applied to classify the user sentiments into positive and negative. The particular machine learning algorithms—Random Forest, XGBoost, CatBoost, and Logistic Regression—were selected for sentiment categorization due to their unique qualities and benefits that make them appropriate for this task:

Logistic regression

The linear, straightforward model of logistic regression is simple to comprehend and apply. This makes it a wise option if one has to elucidate the sentiment categorization procedure to stakeholders who aren't technically inclined. Sentiment analysis frequently involves binary classification tasks (positive or negative sentiment). Because it represents the likelihood of falling into either one of the two classes, logistic regression is particularly good at binary classification. It works well with big text datasets, which are typical in sentiment analysis, and is computationally efficient.

Random Forest classifier

Random Forest is an ensemble learning technique that generates predictions by combining several decision trees. It works well for sentiment analysis since it can handle high-dimensional data and excels at capturing intricate patterns. When working with real-world text data, Random Forest's resistance to outliers and noisy data is a benefit.

XGBoost classifier

XGBoost has a reputation for having a high degree of predicted accuracy, making it especially useful in situations where exact sentiment classification is essential. Its gradient boosting approach facilitates the discovery of intricate links within the data.

XGBoost offers insights into feature importance, so one can determine which features or words have the most effects on sentiment classification.

CatBoost classifier

CatBoost is created especially to effectively handle categorical features. Text data used in sentiment analysis frequently contains categorical features like brand, category, and user data. The way CatBoost handles these kinds of attributes can increase the accuracy of sentiment classification. CatBoost makes use of gradient boosting techniques to capture sentiment patterns in text data that are subtle but effective.

Combining these algorithms provides a thorough method of classifying sentiments by utilizing each one's unique advantages in terms of ease of use, interpretability, precision, and adaptability when processing various kinds of data. In the context of research, this choice strengthens and improves the sentiment analysis framework.

The performance of logistic regression, random forest classifier, XGBoost classifier and CatBoost classifier is evaluated using the metrics namely, accuracy, precision, recall and F1 score. Based on the evaluation metrics. CatBoost classifier was found to be the best model to classify user sentiment into positive and negative.

2.5 Building Collaborative Filtering Based Recommendation Model and Its Performance Evaluation

The dataset has been split into a training set comprising 80% of the data and a testing set comprising the remaining 20%.

User based recommendation model

User based recommendation model works by the following steps as shown in fig 2

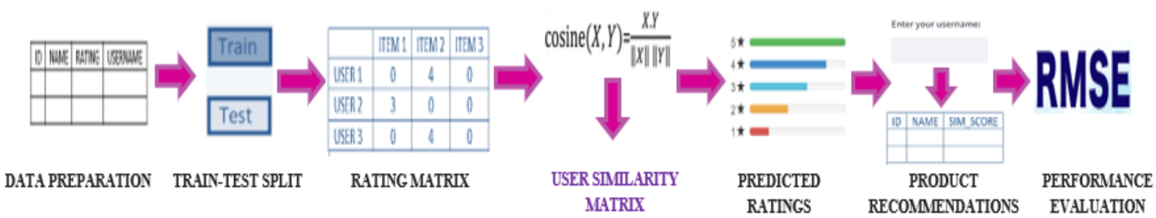


Fig. 2. Steps involved in User based recommendation model

Rating Matrix: User interactions with products are represented by the rating matrix. User ratings are represented by values, product names by columns, and users by rows. The recommendations are based on this matrix.

Cosine Similarity: The similarity between users is computed using cosine similarity. The cosine of the angle formed by two vectors that represent user preferences is measured. Users with similar preferences are indicated by higher cosine similarity.

User Similarity Matrix: Cosine similarity and a pivot table are used to create a user similarity matrix. The similarity between users is captured by this matrix, which is based on their product ratings and interactions.

Predicted Ratings: The pivot table from the training set is multiplied by the User Similarity Matrix to determine the predicted ratings for unrated products. In this step, products that users have not yet rated are given predicted ratings.

Recommendations: Predicted ratings for unrated products are sorted to produce recommendations. The top 20 products are suggested for each and every specific user based on user interactions and ratings, the predicted ratings for the testing set are modified to match user preferences.the top 20 products are suggested for each and every specific user.

Item Based Recommendation model

Item based recommendation model works by the following steps as shown in Fig 3.

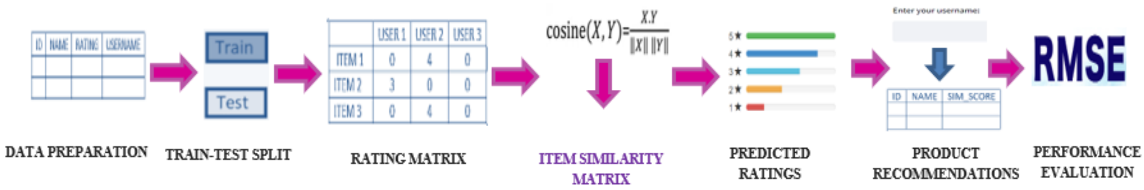


Fig. 3. Steps involved in Item based recommendation model

Rating matrix: User-item interactions are represented by the rating matrix, in which user ratings are the values, rows stand for items, and columns for users. It acts as the starting point for making recommendations.

Cosine similarity: Based on user ratings, cosine similarity is used to calculate how similar two items are to one another. It measures the degree of similarity between two item vectors in terms of user preferences by computing the cosine of the angle between them.

Item similarity matrix: Cosine similarity is used to build the item similarity matrix while accounting for user ratings. It is an essential part of creating recommendations as it measures the similarity between items.

Predicted Ratings: Using the item similarity matrix, predicted ratings are calculated for user-item pairs. In this step, the user's preferences expressed in the matrix are used to calculate how they might rate an item they haven't rated yet.

Product recommendations: Users are provided with recommendations based on the predicted ratings arranged in descending order. The user is presented with a list of the top 20 products with the highest predicted ratings.

The performance of the user based and item-based recommendation models are evaluated using Root Mean Square Error (RMSE). The model with lowest Root Mean Square Error is considered to be the best model. Based on this User based recommendation model was found to be the best model to recommend the products.

2.6 Fine Tuning the Recommendations Using CatBoost Classifier

Using the user-based collaborative filtering model, products were recommended along with their cosine similarity scores, taking the user's name as input. To further enhance recommendations, a CatBoost classifier was employed to predict the sentiment of user reviews for the recommended products. This classifier was trained on a product review dataset, and its predictions were utilized to sort the recommended products based on their positive sentiment percentage. The positive sentiment percentage was calculated as the proportion of positive reviews among the total number of reviews for each product.

2.7 Scraping Ecommerce Website to Get Real Time Data

Fig 4 shows the process involved in web scraping techniques, which refer to the automated extraction of data from websites. It begins by making HTTP requests to the Mamaearth website's API, retrieving JSON data in return. This JSON data is then parsed to extract specific information according to the desired criteria, such as ID, product names, price, SKU, slug, URL, value, unit, pack size, date of creation, product category, custom attributes, review count, review date/time, rating, review content. The extracted data is organized and stored in a structured format, typically a data frame, which allows for easy analysis and manipulation. Finally, the scraped data can be exported as a CSV file, enabling further exploration and utilization.

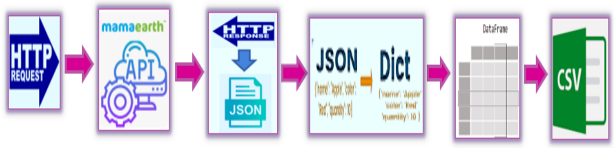


Fig. 4. Steps to get real time data

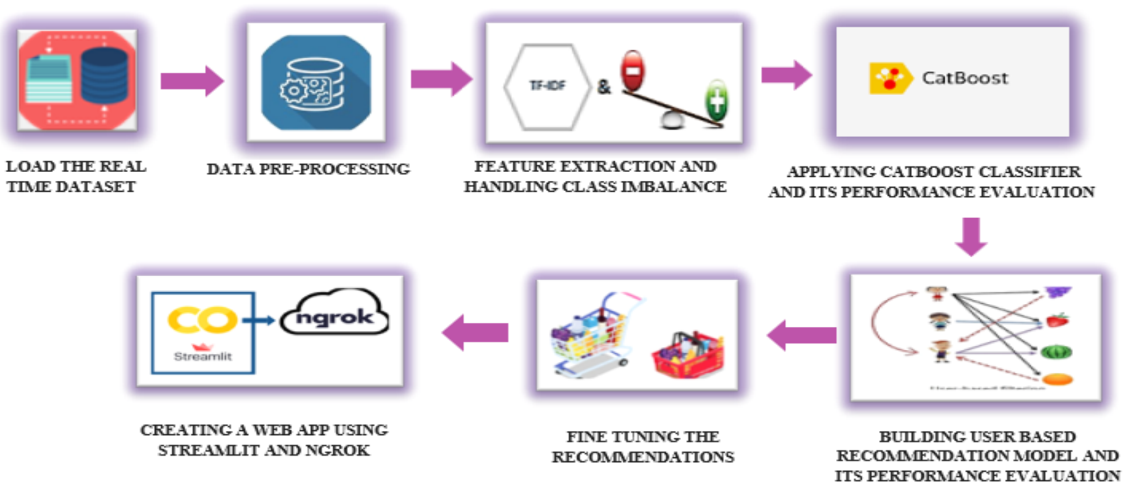


Fig. 5. Sentiment Based Product Recommendation Implementing steps

2.8 Implementing Sentiment Based Product Recommendation System on Real Time Data

Fig 5 shows the steps involved in implementing sentiment-based product recommendation system. Initially, the real time dataset is imported. For the target class, the sentiment polarity such as strongly negative, negative, neutral, positive, strongly positive are incorporated into the dataset through rating. Then the text pre-processing techniques namely converting the upper-case letters to lowercase letters, removing stop words and punctuations, and lemmatization are performed. Exploratory data analysis is performed to gain insights about the data. Feature extraction is performed using the TF-IDF (Term frequency and inverse document frequency) vectorization method. Class imbalance is handled using SMOTE (synthetic minority over-sampling technique). A CatBoost classifier is used to classify, and a user-based recommendation model was built to suggest the top 20 products. The recommendations were further refined using the CatBoost classifier, which is trained on product reviews from "MAMAEARTH REALTIME DATA". The trained model was utilized to predict the sentiment of reviews for the products recommended to the user. The predictions are then used to sort the recommended products based on their positive sentiment percentage.

2.9 Creating A Web App Using Streamlit and Ngrok
Streamlit is a Python library for creating web applications with a simple interface, while Ngrok is a tool for exposing local web servers to the internet. Ngrok generates URL to make the web application accessible. The web application has multiple pages, such as "Get Recommendations' ' where users enter their name for product suggestions, and an "About" page displaying recommended products with positive sentiment percentage.

3. Results and Discussion

The experimental results carried out in this work are the evaluation of machine learning algorithms, including logistic regression, random forest classifier, XGBoost classifier, CatBoost classifier, as well as the evaluation of recommendation models, such as user-based and item-based recommendation models. Additionally, the sentiment-based recommendations are evaluated using real time data, and the output of the web app using Streamlit is analysed.

Performance Evaluation of Machine Learning Algorithms
Model comparison is an essential aspect of machine learning that involves comparing the performance of Logistic Regression, Random Forest classifier, XGBoost classifier and CatBoost classifier on Ecommerce product reviews dataset, namely “sample30.csv”. This comparison is done by evaluating the models using various metrics such as accuracy, precision, recall, and F1 score.

Table 1. Performance evaluation of machine learning algorithms

Algorithm	Evaluation metrics (%)			
	Accuracy	Recall	Precision	F1 score
Logistic regression	88	97	90	93
Random Forest	92	90	96	96
XGBoost	84	96	85	90
CatBoost	94	94	99	97

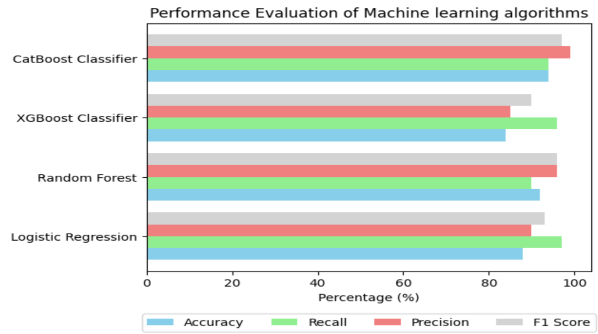


Fig. 6. Machine learning models accuracy comparison

The results in the table 1 and the Fig 6 shows that, CatBoost classifier has the highest accuracy of 94%, followed by Random Forest with 92%, XGBoost classifier with 84%, and Logistic regression with 88%.

Fig. 7. Performance evaluation of machine learning algorithms

Fig 7 shows the performance evaluation of machine learning through the evaluation metrics such as accuracy, precision, recall, F1 score. Because of its outstanding categorical data handling, implicit feature scaling, gradient boosting mechanism for intricate pattern recognition, and strong handling of missing values, CatBoost performed better than other algorithms. Because of its good generalization over both benchmark and real-time data, it has the potential to improve customer retention, Ecommerce recommendations, and product feedback analysis for the mutual benefit of users and businesses.

The practical implications of CatBoost's strong performance include better product recommendations, personalized shopping, increased customer retention, easier product feedback analysis, and faster user preference adaptation.

The exceptional outcomes of CatBoost indicate that it has the ability to offer companies and e-commerce platforms useful information and resources for enhancing customer happiness and engagement.

Performance Evaluation of Collaborative Filtering Based Recommendation Models
The collaborative filtering types such as user based and item-based recommendation models are built and these models are evaluated using the Root mean square error.

The results in table 2 and Fig 8 shows that the user-based recommendation model has a lower Root Mean Square Error value of 2.09 compared to the item-based recommendation model with a Root Mean Square Error value of 3.55. The user-based recommendation model is found to be the best model for product recommendations. Then the products are recommended by the user-based recommendation model, the recommendations are fine-tuned using CatBoost classifier to predict the sentiment of reviews and the percentage of positive reviews of each recommended product was calculated as shown in Fig 9.

Table 2. Performance evaluation of collaborative based recommendation models

Algorithm	Evaluation Metric
	Root Mean Square Error (%)
User based Recommendation model	2.09
Item based Recommendation model	3.55

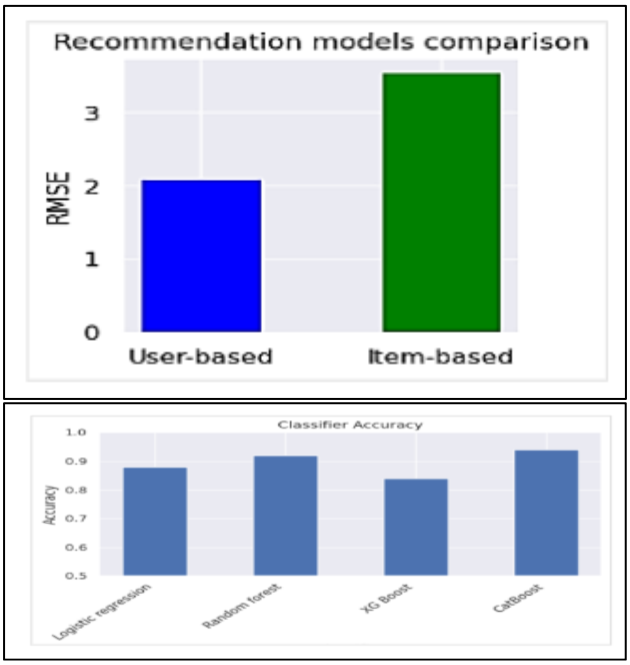
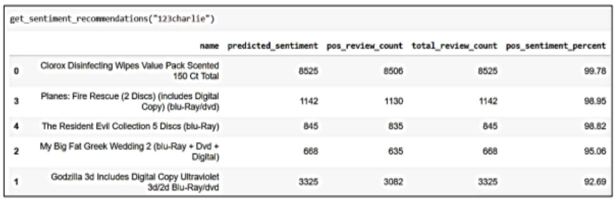


Fig. 8. Recommendation models comparison



```
get_sentiment_recommendations("123charlie")
```

	name	predicted_sentiment	pos_review_count	total_review_count	pos_sentiment_percent
0	Clorox Disinfecting Wipes Value Pack Scented 150 Ct Total	8525	8506	8525	99.78
3	Planes: Fire Rescue (2 Discs) (includes Digital Copy) (blu-Ray/dvd)	1142	1130	1142	98.95
4	The Resident Evil Collection 5 Discs (blu-Ray)	845	835	845	98.82
2	My Big Fat Greek Wedding 2 (blu-Ray + Dvd + Digital)	668	635	668	95.06
1	Godzilla 3d Includes Digital Copy Ultraviolet 3d/2d Blu-Ray/dvd	3325	3082	3325	92.69

Fig. 9. Top 5 products recommended for the user "123charlie"

Performance Evaluation of CatBoost Classifier on Realtime Data

The performance of a sentiment-based product recommendation system is being evaluated on real-time data. As earlier mentioned, the CatBoost classifier algorithm was identified as the most effective in classifying user sentiments as either positive or negative. However, the algorithm's efficiency is now being evaluated with a modification to classify sentiments into five categories, namely, strongly negative, negative, neutral, positive and strongly positive on real time data. Based on table 3, it can be concluded that the CatBoost Classifier performed relatively well in classifying user sentiment into five classes and it performs well on both real-time data and benchmark data.

Because of its ability to adapt effectively to changing data streams and its resilience when it comes to categorical features, the CatBoost classifier is an excellent option for dynamic sentiment analysis and performs exceptionally well on real-time data.

Table 3. Performance evaluation of CatBoost Classifier on real time data

Algorithm	Evaluation Metrics (%)			
	Accuracy	Recall	Precision	F1 score
CatBoot classifier	83.5	79	83.5	79.1

Evaluating the Performance of User Based Recommendation Model on Real Time Data

A sentiment-based product recommendation system is currently being assessed using real-time data. As mentioned above, the user-based recommendation model was found to be more effective in predicting user preferences or ratings for products. The user-based recommendation model is being utilized on real-time data to evaluate its performance.

Table 4. Performance evaluation of user-based recommendation model on real time data

Algorithm	Evaluation metric (%)
	Root Mean Square Error
User Based Recommendation Model	2.62

The results in the table 4 shows that, the user based recommendation model is showing promising results in the evaluation of a sentiment-based product recommendation system using real-time data. The products are recommended by user-based recommendation model, the recommendations are fine-tuned using CatBoost classifier to predict the sentiment of reviews and the percentage of positive reviews of each recommended product was calculated on real time data as shown in the Fig 10.

A sentiment-based recommendation system can help businesses and e-commerce platforms by improving customer satisfaction, driving sales, and raising user engagement. This system lowers churn rates and facilitates effective inventory management by offering insightful data on user sentiments and product feedback.

Web application using streamlit

The web app contains two pages namely Get Recommendation page and About page. In "Get Recommendations" page users can enter their name for product suggestions as shown in Fig 11 and Fig 12, and an "About" page displays the recommended products with positive sentiment percentage as shown in Fig 13 and Fig 14. Providing information about the sentiment analysis process can be very useful for users. It can help them understand how the recommended products and sentiment scores are generated.

get_sentiment_recommendations("Aathira")

	PRODUCT_NAME	predicted_sentiment	strong_pos_review_count	total_review_count	strong_pos_sentiment_percent
0	CoCo Body Wash With Coffee and Cocoa For Skin Awakening - 300 ml	46	46	46	100.00
4	Vitamin C Foaming Face Wash with Vitamin C and Turmeric for Skin Illumination - 150ml	1234	1231	1234	99.76
3	Tea Tree Foaming Face Wash with Tea Tree and Salicylic Acid for Acne and Pimples - 150ml	609	607	609	99.67
2	Green Tea Face Wash With Green Tea & Collagen For Open Pores - 100 ml	63	62	63	98.41
1	Flowers of Youth Essence Serum with Hyaluronic Acid & Hibiscus for Youthful Skin – 30 ml	68	61	68	89.71

Fig. 10. Top 5 products recommended for the user "Aathira" on real time data

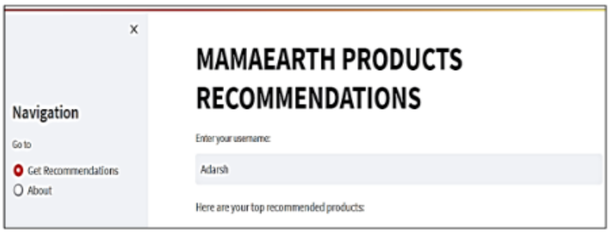


Fig. 11. Entering the username as "Adarsh" in the web app on Get Recommendations page

	PRODUCT_NAME
0	Ubtan Nourishing Bathing Soap With Turmeric and Saffron - 375g
1	Vitamin C Foaming Face Wash with Vitamin C and Turmeric for Skin Illumination - 150ml
2	Tea Tree Foaming Face Wash with Tea Tree and Salicylic Acid for Acne and Pimples - 150ml
3	Aqua Glow Gel Face Moisturizer With Himalayan Thermal Water and Hyaluronic Acid for 72 Hours Hydration - 100ml
4	Tea Tree Face Serum for Acne and Pimples - 30 ml

Fig. 12. The Top 5 products Recommended for the user "Adarsh"

Fig. 13. The product recommendations in About page

	PRODUCT_NAME	predicted_	strong_pos_	total_review	strong_pos_sen
0	Ubtan Nourishing Bathing Soap With Turmeric and Saffron - 375g	98	98	98	100.00
1	Vitamin C Foaming Face Wash with Vitamin C and Turmeric for Skin Illumination - 150ml	1234	1233	1234	99.92
2	Tea Tree Foaming Face Wash with Tea Tree and Salicylic Acid for Acne and Pimples - 150ml	609	608	609	99.84
3	Aqua Glow Gel Face Moisturizer With Himalayan Thermal Water and Hyaluronic Acid for 72 Hours Hydration - 100ml	162	161	162	99.38
4	Tea Tree Face Serum for Acne and Pimples - 30 ml	162	160	162	98.77

Fig. 14. The products recommended in the about page along with sentiment percentage

4. Conclusion and Future Scope

In this work, the CatBoost classifier demonstrated superior performance for sentiment classification, while the user-based recommendation model proved to be the best for product recommendations. Fine-tuning recommendations using the CatBoost classifier improved sentiment prediction for each recommended product. The sentiment-based product recommendation system was evaluated using real-time data and successfully deployed as a web application using Streamlit and Ngrok to enhance product recommendations and drive customer satisfaction and engagement. In summary, this research emphasizes how well CatBoost performs in sentiment classification and how useful the user-based recommendation model is for making product recommendations. Sentiment predictions for suggested products were further enhanced by fine-tuning using the CatBoost classifier. When used as a web application, the sentiment-based recommendation system improves user engagement and satisfaction.

The future scope of this project includes integrating deep learning algorithms for more accurate sentiment analysis, incorporating SEO techniques to optimize website content and attract more traffic, and integrating chatbots for real-time support and personalized recommendations based on customer emotions and preferences. These advancements will further enhance the recommendation system, improving customer satisfaction and loyalty. Chatbots and SEO strategies will optimize content and offer real-time support, while deep learning algorithms will improve sentiment analysis accuracy. All of these developments will eventually increase customer satisfaction and loyalty.

References

[1] A. Naresh, P. Venkata Krishna,"Recommender System for Sentiment Analysis using Machine Learning Models". *Turk. J. Comput. Math. Educ*, vol.12, pp.583-588. Apr, 2021.

[2] S. Kausar, X. Huahu, W. Ahmad, M. Y. Shabir, and W. Ahmad, "A Sentiment Polarity Categorization Technique for Online Product Reviews," *IEEE Acc.*, vol. 8, pp. 3594–3605, Dec. 2020, doi: 10.1109/ACCESS.2019.2963020.

[3]M. Tahir, R. N. Enam, and S. M. Nabeel Mustafa, "E-commerce platform based on Machine Learning Recommendation System," in *2021 6th Int. Multi-Topic ICT Conf. (IMTIC)*, Jamshoro & Karachi, Pakistan: IEEE, Nov. 2021, pp. 1–4. doi: 10.1109/IMTIC53841.2021.9719822.

[4] Sakthivel, K. Janaki, A. Joseph, R. K. Nandeshvar, S. Anandharajan, "Product recommendation system using machine learning". *Int. Res. J. Modern. Eng. Technol. Sci.,* Vol. 4, pp. 343-349, Dec. 2022, doi: 10.56726/IRJMETS32024.

[5] Bobadilla, J. Ortega, F. Hernando, A. Gutierrez, "Recommender systems survey". *Knowl. Based Syst.,* vol.46, pp.109-132, Jul. 2013, doi: 10.1016/j.knosys.2013.03.012.

[6] K. Haruna, M. A. Ismail, Suhendroyono , Damiasih, A. C. Pierewan, H. chiroma, "Context-aware recommender systems: a Review of Recent Developmental Process and Future Research Direction". *ACM Trans. Intell. Syst. Technol.,* vol.7, pp.2-25, Dec. 2017, doi: 10.3390/app7121211.

[7] Dimah, Z. Xiao-Jun. "Improving Recommendation Using Trust and Sentiment Inference from OSNs", *Int. J. Knowl. Eng.,* vol.12, pp.9-17. Dec. 2017, doi: 10.7763/ijke.2015.v1.2

[8] Ekstrand, M. D. Ludwig, J. A. Konstan, "Rethinking the recommender research ecosystem: Reproducibility, openness, and

lensing". in *Proc. 2012 ACM Conf. Recommender Syst.,* Aug. 2012, pp. 133-140, doi: 10.1145/2043932.2043958.

[9] A. Chinnalagu and A. K. Durairaj "A survey of automatic text-based sentiment analysis in product reviews", *PeerJ Comput. Sci.,* Vol. 29, pp.2-25, Dec. 2021, doi: 10.7717/peerj-cs.813.

[10] N. Afriliana, N. M. S. Iswari, Suryasari, "Sentiment analysis of user-generated content: a bibliometric Analysis", *J. Syst. Manag. Sci.,* vol. 12, no. 6, pp.583-598, Dec. 2022, doi: 10.33168/JSMS.2022.0634.

[11] X. Su, T. M. Khoshgoftaar, "A survey of collaborative filtering techniques", *Adv. Artif. Intell.,* vol.09, pp.421-427, Oct. 2009, doi: 10.1155/2009/421425.

[12] J. A Konstan, J. Riedl, "Recommender systems: From algorithms to user experience". *User Model. User-Adapt. Interact.,* vol. 22, pp.101-123, Mar. 2012, doi: 10.1007/s11257-011-9112-x.

[13] G. N. Kottage, D. K. Jayathilake, K. C. Chankuma, G. U. Ganegoda, and T. Sandanayake, "Preference based recommendation system for apparel e-commerce sites," in *Proc. 17th Int. Conf. Comput. Inf. Sci.,* 2018, pp. 122–127, doi: 10.1109/ICIS.2018.8466382.

[14] A. Roshini, B. Kalpana, "Recommendation System for Music using machine learning techniques", in *Proc. 20th Int. Conf. Artif. Intell. Soft Comput.,* Aug. 2021, pp. 19-35

[15] A. Roshini, B. Kalpana, "IOT based smart home automation and personalization system", in *Proc. Int. Conf. Next Gener. Comput. Sys. (ICNGCS - 2021),* Mar. 2021, pp. 92 – 96.