Recommendation System Using Deep Learning in E-Commerce

Karthik Jayanthi [1] Karthik Raja [2] Sai Chaitanya[3] Shanmukha Rahul[4] Sai Prabhas [5]
1*School of Computer Science Engineering, BML Munjal University,
67th Milestone, NH 48, Kapriwas, 122413, Haryana, India.
*Corresponding author(s). E-mail(s): dmvsvkarthik.jayanthi.21cse@bmu.edu.in;
nichenametla.karthikraja.21cse@bmu.edu.in;
sivapuramshanmukha.rahul.21cse@bmu.edu.in;
burugupallisai.chaitanya.21cse@bmu.edu.in
mallidisai.prabhas.21cse@bmu.edu.in
*These authors contributed equally to this work

Abstract:

With their ability to forecast preferences based on vast user histories, recommendation systems are essential to current services. Using the Amazon reviews dataset, this study examines the efficacy of classical machine learning (ML) models, long short-term memory (LSTM) networks, and recurrent neural networks (RNNs) in e-commerce suggestions. Sophisticated methods like collaborative filtering are investigated to improve suggestion precision. We evaluate these models' performance and compare their accuracies through extensive testing to determine which model is best for forecasting user behavior and enhancing the quality of recommendations. Comprehensive research and significant insights into the efficacy of recommendation systems are made possible by utilizing the extensive user input and preferences included in the Amazon reviews dataset. Our results aid in the e-commerce industry's recommendation system optimization by providing helpful advice on which.

Keywords:- Recommendation systems, E-commerce, Amazon reviews dataset, RNN, LSTM, Machine learning.

1 Introduction

In today's big online stores, there are so many products that it's hard for users to find what they want. The need for efficient and personalized recommendation systems has become paramount. These recommendation systems help users discover things the users might like, and help stores sell more. These systems look at what users have done in the past and suggest things the users might like based on that. These systems use fancy techniques like deep learning and understanding language to figure out what users might want.

Recommendation systems are helpful not just for online stores, but also in other areas like education and research [13]. They deal with lots of information and help make sense of it all. The good thing about recommendation systems is that they make shopping easier for users [6]. They suggest things that match what users like, which makes them happy. And they help stores by suggesting products to people who are likely to buy them.

Online shopping can be overwhelming with so many choices. But recommendation systems help by suggesting items that match what users like. This not only makes shopping easier but also helps stores

sell more. These systems are super important in today's online shopping world because they make users happy and boost sales for stores [6].
It's challenging to create effective recommendation systems. Given the abundance of products, the recommendations must be appropriate for every consumer. Furthermore, while using user data for recommendations, caution about privacy needs to be considered. Maintaining these systems' performance as their user base increases is equally challenging. Accurate suggestions are sometimes difficult to make when there is little information available about a new user or item. This kind of issue is called the "cold start" problem [1]. In addition, the "data sparsity" problem refers to the difficulty the system has in making useful suggestions when there is a lack of information on specific individuals or goods [6]. However, with hard work and innovation, ways to overcome these challenges can be found, making online shopping even better for everyone.
Global retail website traffic peaked at 14.3 billion visits in March 2020 [1]. At the same time, online sales in Turkey soared by 98.3 percent across the country [1]. Big companies like Amazon have been using recommendation systems for a long time. But now, even smaller stores can use them to make their customers happy and sell more stuff. Deep learning is a big deal in making recommendation systems better. It's a smart way of teaching computers to understand things better, which helps them make better recommendations. There are different kinds of recommendation systems. Some look at what other users like (collaborative filtering), while others focus on what the product is like (content based) [7][10]. Some combine both approaches to give even better recommendations [7].
Platforms like Netflix and YouTube exemplify the fusion of collaborative and content-based filtering, wherein factors like genre play a pivotal role in recommendation algorithms [6]. Collaborative filtering and content-based algorithms, renowned for their versatility and effectiveness, continue to underpin the fabric of recommendation systems across diverse domains.
In this project, our aim is to utilize deep learning techniques to develop a recommendation system tailored for e-commerce. This system will streamline the shopping experience for users while simultaneously boosting sales for online retailers.

2 Literature review

This section will highlight the previously published works on the subject of "Recommendation system using deep learning in ecommerce," as suggested by a few writers.

The paper [1] presents a recommendation system that uses Deep Learning and Natural Language Processing (NLP) techniques to deliver individualized recommendations based on user ratings. The system integrates LSTM, RNN, encoder, decoder, and attention mechanisms using TF-IDF on pre-processed reviews. Remarkably, it obtains a Root Mean Squared Error (RMSE) of 0.9357 and a Mean Absolute Error (MAE) of 0.6586, demonstrating better recommendation accuracy than previous methods.

To extract product qualities from reviews, the authors [2] introduced Latent Dirichlet Allocation (LDA) to incorporate ratings and review texts into an aspect-based latent component model. A user's attributes matrix is then created using these attributes. To get over sparsity problems, a deep neural network converts this sparse matrix into dense features, which makes Matrix Factorization (MF) an effective method for generating recommendations and ratings. The model is exceptionally successful in a variety of domains as seen by its noteworthy Mean Square Error (MSE) values of 1.806, 1.249, and 1.064 in the Electronics, Movies, and Music datasets, respectively.

The article [3] presents a plug-and-play cloud service designed to ease the difficulties small businesses encounter while creating recommendation systems. With the help of an LSTM deep learning model that was trained over 20 epochs and had its batch size tuned at 512, the system can provide product suggestions with an astounding accuracy rate of 64%. This gives small businesses a workable way to include efficient recommendation systems into their workflows.

In their detailed research the authors [4] are paying special attention to deep learning approaches. To assess the value of reviews, they experimented with a variety of models, including feedforward networks and LSTM. Remarkably, they discovered that "matrix factorization" was the most successful strategy, outperforming even tried-and-true techniques like collaborative filtering. Notably, in their study, LSTM performed better than the feedforward network. Root Mean Square Error (RMSE) comparisons of model correctness were performed, clarifying the efficacy of these methods in evaluating review helpfulness.

Two different recommender systems were investigated by the authors in the study described in paper [5]. The first solution used a simple NumPy-based strategy and made predictions based on popularity. The second approach used the SVD technique in addition to user-based closest neighbour collaborative filtering to integrate matrix factorization. The research also promoted deep learning methods integration for improved performance. The model's impressive RMSE and MAE values of 0.00275 and 0.0019, respectively, highlight how effective it is at making recommendations.

The research [6] does a systematic literature review (SLR) with a focus on deep learning-based recommender systems (RS) to solve problems such as data sparsity and cold start. It mentions that Movie Lens is a well-liked choice for movie recommendation systems, while Amazon review datasets are popular for multi-domain recommendation systems. Autoencoder (AE) models are the most widely utilized deep learning architectures for reinforcement learning (RS), with convolutional neural networks (CNNs) and recurrent neural networks (RNNs) following closely behind. Precision and root mean square error are the two key metrics used to assess the effectiveness of deep learning-based recommendation systems.

The paper [7] suggests a hybrid movie recommendation system that combines sentiment analysis of tweets about movies, content-based filtering (CBF), and collaborative filtering (CF). It displays better precision values at the top-5 and top-10 recommendations when compared to pure hybrid and sentiment similarity models. The paper takes into account other techniques such as the FFM algorithm and deep learning-based models that make use of pretrained word embeddings and attention processes but concludes that their effectiveness is limited by issues with context understanding and data availability. All things considered, the suggested hybrid system successfully uses movie tweets to improve recommendation accuracy, showing encouraging outcomes in comparison to current methods.

This research paper [8] explores the integration of Word2vec and Node2vec techniques for movie recommendation on the Movie Lens dataset. Word2vec is utilized for word embedding, while Node2vec is employed for graph embeddings. The study differentiates these methods based on their influence on user interactions and movie features to compute low-dimensional embeddings. Additionally, it demonstrates the injection of movie features into the model to enhance movie embeddings, leading to improved recommendation performance. The paper emphasizes the training and sampling strategies of both neural models and illustrates their application in recommending movies based on various aspects such as genre and user preferences.

The authors in [9] offer the T-RECSYS approach, which combines collaborative and content-based filtering algorithms in an innovative way for music recommendation. The system evaluates songs based on user preferences and historical data to identify key metadata elements such as genre, tempo, and mood. The system achieves an impressive 88% accuracy score, demonstrating its effectiveness in delivering accurate music recommendations.

The article [10] presents a hybrid collaborative filtering system based on deep learning. Neural networks are used to produce recommendations once inputs are mapped via Stacked DAE to higher-level representations. Offline experiments show that compared to conventional User-based CF algorithms, the EIN network performs noticeably better in terms of coverage metrics and accuracy. It's interesting to note that the EIN network performs around 20% more precisely than the User-based CF algorithm.

The paper [11] provides a description of the study performed on the Bangladesh e-commerce review dataset. It provides valuable insights on the preferences and actions of consumers in Bangladesh's e-commerce industry. The results of these studies demonstrate the importance of annotated datasets, like the one that was examined in relation to Bangladeshi e-commerce, in advancing the field of sentiment analysis research and improving e-commerce processes by providing valuable insights into customer behavior and sentiment dynamics.

A method for using Natural Language Processing (NLP) to sentiment analysis of customer feedback in Bangla is given in [12]. The study evaluates several classification methods in addition to Support Vector Machine (SVM), such as KNN, Decision Tree, Random Forest, and Logistic Regression. Notably, the SVM model performs remarkably well in the sentiment analysis of Bangla customer comments, with an incredible accuracy of 88.81%.

While there are numerous deep learning methods being employed for recommendation system using deep learning in ecommerce, we have divided into 2 categories which were mentioned in the table-1.

ML/DL techniques	NLP and recommendation techniques
1. LSTM 2. RNN (Recurrent Neural Network) 3. Encoder and Decoder architectures 4. Autoencoder (AE) 5. Word2vec 6. Support Vector Machine (SVM) 7. KNN (K-Nearest Neighbour's) 8. Decision Tree 9. Random Forest	1. TF-IDF (Term Frequency-Inverse Document Frequency) 2. Latent Dirichlet Allocation (LDA) 3. Matrix Factorization (MF) 4. Collaborative Filtering(CF)

Table-1: Deep-learning methods for recommendation system

Methodology Overview: -

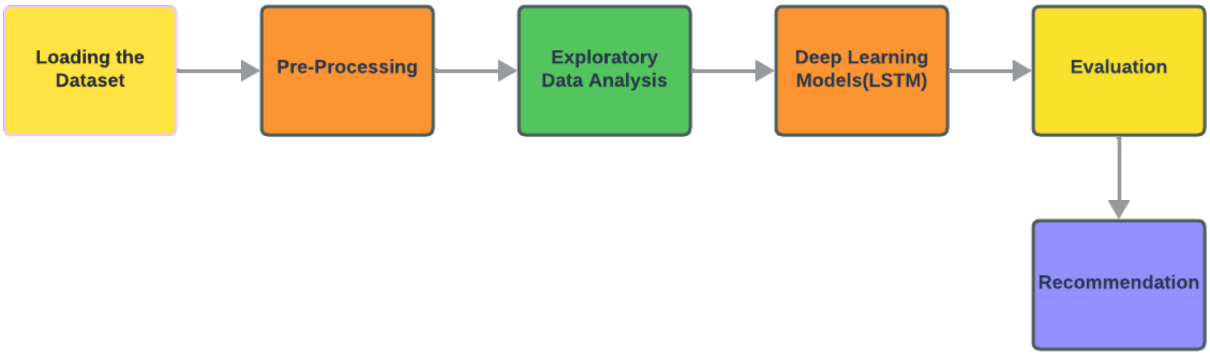

Fig 1: Displaying the modular workflow for the whole process.

3. Methodology

3.1 Dataset: -

The Amazon user Reviews dataset includes millions of reviews from 1995 to 2015, providing insights into user opinions on a variety of items. Scholarly research in domains like Natural Language Processing (NLP) and Machine Learning (ML) is made possible by information such as marketplace, product ID, star rating, and review date. It offers a wealth of data on helpful votes, confirmed purchases, and review durations, making it an invaluable tool for cross-regional analysis of consumer behavior and product perception. With this original and well-documented dataset, researchers may explore trends and patterns in product evaluations over time with ease since it gives them a thorough grasp of client experiences and preferences. The Amazon user review dataset contains 37 files of products that belong to various categories. The data that belongs to Beauty, Electronics, Grocery, and Health & Personal care are selected for the task of recommendation.

3.2 Pre-processing:

After combining the datasets into a single dataset. The review text is processed to ensure consistency and reduce recurrence of terms with various cases, the text data was standardized by transforming all characters to lowercase. This improved the accuracy of subsequent research.

3.2.1 Stop Word Removal:

Common stop words (such as "the," "is," and "and") were eliminated from the text data to reduce noise and concentrate on more significant words and phrases that provide crucial details about the attitudes and views of users.

3.2.2 Part-of-Speech (POS) Tagging:

By giving contextual information on word usage and syntactic structure, POS tagging enables more accurate lemmatization by assigning grammatical tags (such as noun, verb, and adjective) to every word in the text data.

3.2.3 Lemmatization:

Lemmatization is a text normalization approach that reduces inflected words to their root forms for improved analysis and interpretation by transforming words in the text data into their base or dictionary form (lemmas). When lemmatization and POS tagging are combined, word form variations are efficiently handled, improving the text data's consistency and interpretability. This leads to more precise sentiment analysis and topic modeling when developing recommendation systems.

3.2.4:

The data preprocessing strategy involves removing single-letter terms and meaningless terms from the dataset to improve data quality. This is achieved by filtering out terms with a length less than 3 characters, as they are often not informative. Additionally, terms with characters repeated more than 3 times are eliminated, such as "aaaamazing" or "zzzz," which are unlikely to contribute meaningfully to the analysis. These steps help in cleaning the dataset and focusing on more relevant and informative textual content for further analysis or modeling tasks.

3.3 Exploratory data Analysis (EDA):-

3.3.1 Count of Product Categories:

The graph(fig-2) displays the distribution of reviews across four product categories: beauty, Electronics, Grocery, and Health and Personal Care. Each bar in the graph is a visual representation of a product category, and the height of the bar vividly corresponds to the number of reviews in that category. The color-coded segments within each bar further enhance the visual appeal and represent different review elements, such as marketplace, customer_id, review_id, and more. Beauty has the highest number of reviews, followed by Health and Personal Care. Electronics and Groceries have significantly fewer reviews.

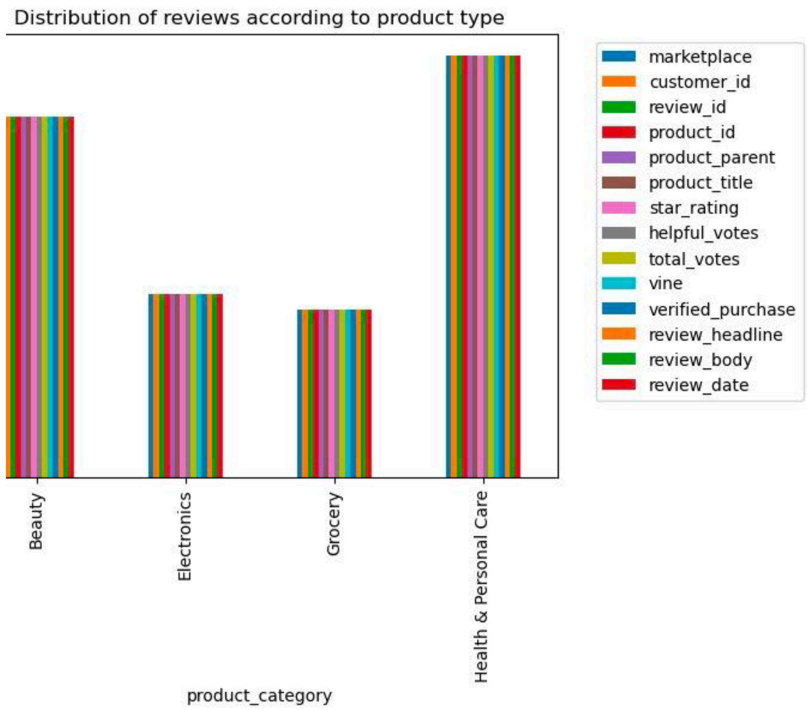

Fig2: Distribution of reviews according to the product type

3.3.2 Count of Star Ratings:

The bar(fig-3) for 5-star ratings is significantly taller, indicating a high count compared to other ratings. The bar for 4-star ratings is shorter, suggesting a lower count. There's almost no count for 2-star ratings as its bar is very short. The bar for 1-star ratings has a moderate height.

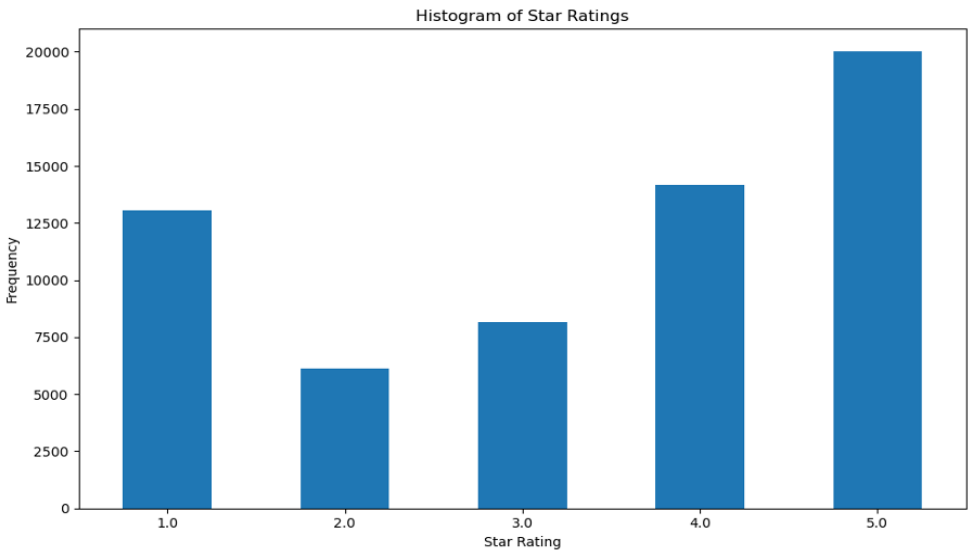

Fig3: histogram representation of star ratings

3.3.4 Details of the vocab size:

The user's dataset comprises 2,207,494 words in total, with a vocabulary size of 64,648 unique words. This indicates a diverse range of terms used within the dataset, reflecting its complexity and richness in textual content.

3.3.3 Most Repeating words:

Here's the data presented in a table-2 says about the top-10 most repeated words that are obtained from dataset after pre-processing, so it will give us the insights about how well our dataset is cleaned.

use	32192
product	29772
Good	22963
Work	22750
Get	21104
Like	20329
One	19754
Great	19617
well	14922

buy	14570

Table-2: top 10 words repeated in the dataset

3.4 Feature Selection:

The data preparation for the RNN model involved transforming reviews into TF-IDF vectors using a vocabulary size of 64,648. Additionally, numerical encoding via label encoding was applied to customer IDs and product IDs. Ratings were then converted into a suitable format for model training through one-hot encoding. These transformed features, comprising customer ID, product ID, and review text vectors, were selected as the primary inputs for constructing the LSTM model aimed at predicting ratings.

3.4.1 How can the user provide recommendations?

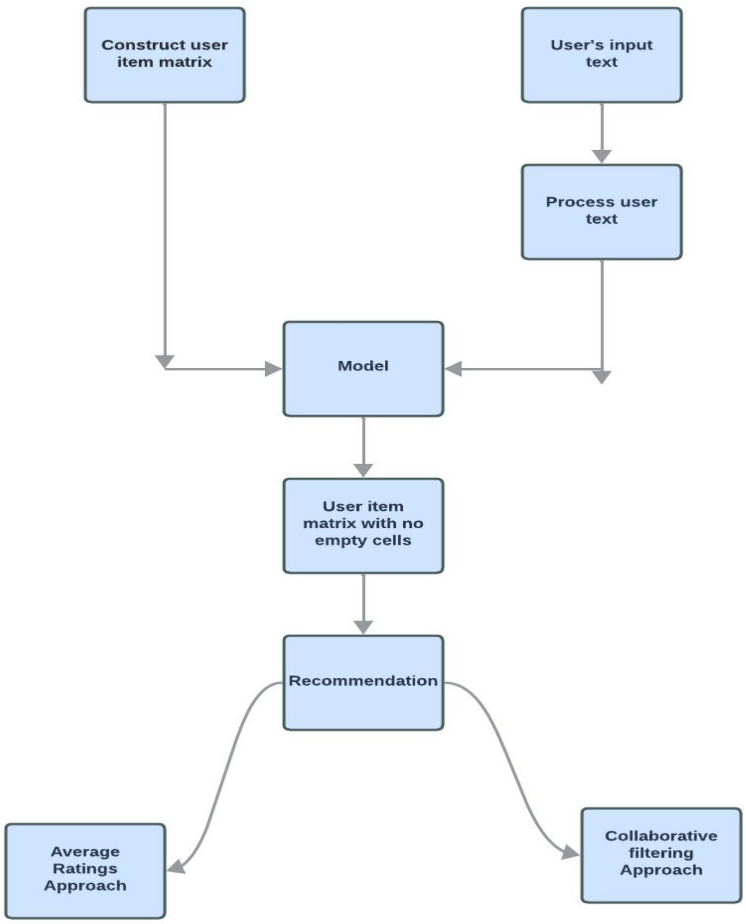

Figure4: Flow showing how a user recommends.

To provide recommendations, the user begins by selecting a sample of 50 users and 50 items. They then construct a user-item matrix by populating cells with ratings given by these users for the products as represented in the (figure-4). Since not every user rate all products, a model is employed to predict ratings for all items, thereby filling in empty values (NaN) in the matrix for improved accuracy.

The recommendation process involves two key approaches. Firstly, the average ratings of all products are calculated using the matrix, and products with the highest average ratings are recommended. Secondly, collaborative filtering is utilized, focusing on a specific user. Cosine similarity is applied to the matrix to identify users with similar preferences to the selected user. Products favored by these similar users are then recommended, leveraging shared interests to enhance the recommendation process.

3.5 Predictive Modelling :

We opted for LSTM (Long Short-Term Memory) in our recommendation system due to its effectiveness in processing textual data and capturing specific features. This approach is specifically chosen for recommending product categories based on its ability to handle text sequences and extract relevant information.

Reasons for choosing LSTM:

1.Sequential Data Processing:
LSTMs are great at understanding the order of user actions or item sequences, which helps in making accurate recommendations.

2. Long-Term Dependency Modeling:
LSTMs can remember past interactions over long sequences, allowing them to predict future preferences based on previous behaviors.

3.Personalized Recommendations:
LSTMs analyze past interactions to create recommendations tailored to each user's unique interests, providing a personalized experience.

The model includes input layers for item data and category labels. Item data is reshaped to fit an LSTM layer, which predicts sequence patterns. Following LSTM, some information is discarded to prevent overfitting, while labels are flattened. These sets are combined and reshaped for category prediction. Training optimizes internal settings using RMSprop, with batches of 64 size and validation data for performance evaluation. Post-training, the model achieved 59% accuracy in categorization for both seen and unseen data.

4. Performance and Result

Classes	Precision	recall	f1-score	support
0	0.64	0.78	0.70	2632
1	0.38	0.15	0.21	1200
2	0.42	0.35	0.38	1604
3	0.51	0.49	0.50	2858
4	0.68	0.77	0.72	4010

Table 3: Evaluation metrics of the model across different classes.

The LSTM classification model exhibits promising opportunity, as shown by the performance metrics obtained on the test data. With an overall 59.2% accuracy rate, the model's prediction performance is impressive. Furthermore, the model's efficiency in capturing the complexity of the data and generating precise predictions can be seen by the precision, recall, and F1-score metrics, which stand at 56.88%, 59.2%, and 57.27%, respectively. These results highlight the model's potential for refinement and optimization, with opportunities to further enhance its performance and predictive prowess.

It is clear from a deeper look at the [Table 3] that the model performs differently in each class. Higher accuracy, recall, and F1-scores are shown for classes 0 (1-star rating) and 4 (5-star rating), suggesting strong abilities to predict for those groups. On the other hand, Classes 1 (2-star rating) and 2 (3-star rating) have lower ratings, indicating that there is room for development in terms of accurately predicting these classes. The results from [Table 3] offer valuable insight into the model's positive and negative aspects, leading future efforts to improve the model's architecture, and eliminate any biases or problems with the quality of the data. The model's predictive reliability and accuracy may be increased via thorough evaluation and repeated improvement, which will increase its usefulness in practical applications.

5 Discussion

Our results, on average 59.2% accurate, demonstrate the success of our recommendation system. This number shows that how good our system is at providing recommendations in comparison to the other models which are already existed. Apart from this achieving high accuracy rates, this model also achieves that by applying sophisticated analysis methods such as machine learning and sentiment analysis too. These methods allow the systems to process the big data sets by going through it deeply to draw out subtle points and understand that what the user really wants or likes.

Therefore, not only does our recommendation model propose the products based on user behavior but also it considers the sentiment analysis while making another perception about suggested items. This approach works towards improving both customer satisfaction levels and business value realized from leveraging large amounts of customer feedback contained within reviews or comments left by them about different products they have used before.
So, where we have also used the user based collaborative filtering technique which mean recommending to the products to a user based on similar user likes and dislikes so that the user experience will be improved which also highlights the model.

Additionally traditional methods for creating suggestions can be stretched beyond their limits if we incorporate machine learning and sentiment analysis in recommendations. The reason why we say so is because with these advanced techniques our model has shown an ability to learn from its mistakes more deeply than any other time before and adjust itself according to new changes brought about by personal taste shifts among consumers over a period influenced greatly by market forces.

6. Conclusion

The overall outcomes of the recommendation system accuracy and precision were occurred of 59.2% and 56.88 which can represent how well the model is able to suggest the products to a user based on other user likes which is nothing but similarity score. Here for calculating the similarity as we mentioned have used the cosine-similarity to calculate the similarity score between every two users which makes model more robust and user-based collaborative filtered.

The results of the model highlighting the robustness of model and by successively recommending the products to the user our model demonstrates its capability to provide valuable insights for fashion systems-commerce websites, and some other applications which need to integrate the recommendation system for the better user-satisfaction.

7. Acknowledgments

We would like to express our sincere thanks to BML Munjal University for all of their help with this initiative. We are especially grateful to our project guide, Dr. Atul Mishra, whose counsel, and direction greatly influenced the development of our project review and summary. It would have been very difficult to do this assignment without his help. We also like to express our gratitude to all of the instructors and support personnel for their unwavering support and altruistic help at every turn throughout our endeavor. We sincerely appreciate the Department of Computer Engineering's assistance and recommendations.

We also express our gratitude to everyone that helped us collect data for our study. Throughout this journey, our families have also been an invaluable source of emotional support and encouragement.

8. Abbreviations

1. LSTM - Long Short-Term Memory
2. RNN - Recurrent Neural Network
3. TF-IDF - Term Frequency-Inverse Document Frequency
4. LDA - Latent Dirichlet Allocation
5. MF - Matrix Factorization
6. POS – Parts of speech tagging
7. RMSE – Root Mean Square Error
8. MAE – Mean Absolute Error

9.References

[1] Almahmood, R.J.K.; Tekerek, A. Issues and Solutions in Deep Learning-Enabled Recommendation Systems within the E-Commerce Field. *Appl. Sci.* **2022**, *12*, 11256. https://doi.org/10.3390/app122111256

[2] B. Maleki Shoja and N. Tabrizi, "Customer Reviews Analysis with Deep Neural Networks for E-Commerce Recommender Systems,"

[3] Dhage, R., Nehete, S., Hon, S., Patankar, T., Kale, L. (2023). Implementation of Recommendation System's Service Model Using Amazon E-commerce Dataset. In: Tuba, M., Akashe, S., Joshi, A. (eds) ICT Systems and Sustainability. ICT4SD 2023. Lecture Notes in Networks and Systems, vol 765. Springer, Singapore. https://doi.org/10.1007/978-981-99-5652-4_37

[4] Zhou, Z., & Xu, L. (2009). Amazon Food review classification using deep learning and recommender system. In Stanford University Stanford.

[5] Dwivedi, R., Anand, A., Johri, P., Banerji, A., & Gaur, N. (2020). Product based recommendation system on amazon data. Int J Creat Res Thoughts–IJCRT.

[6] Da'u, A., & Salim, N. (2020). Recommendation system based on deep learning methods: a systematic review and new directions. Artificial Intelligence Review, 53(4), 2709-2748.

[7] Kumar, S., De, K., & Roy, P. P. (2020). Movie recommendation system using sentiment analysis from microblogging data. IEEE Transactions on Computational Social Systems, 7(4), 915-923.

[8] Khoali, M., Tali, A., & Laaziz, Y. (2021). A Survey of Artificial Intelligence-Based E-Commerce Recommendation System. In Emerging Trends in ICT for Sustainable Development: The Proceedings of NICE2020 International Conference (pp. 99-108). Springer International Publishing.

[9] Fessahaye, F., Perez, L., Zhan, T., Zhang, R., Fossier, C., Markarian, R., ... & Oh, P. (2019, January). T-recsys: A novel music recommendation system using deep learning. In 2019 IEEE international conference on consumer electronics (ICCE) (pp. 1-6). IEEE.

[10] Rehman, E. U. (2021). Recommendation System For Movies Using Knowledge Base. International Research Journal of Modernization in Engineering Technology and Science, 3(02), 10.

[11] Sun, Y., Lv, H., Liu, X., Xu, P., Huang, Y., & Sun, Y. (2018, April). Personalized recommendation for Weibo comic users. In 2018 Wireless Telecommunications Symposium (WTS) (pp. 1-6). IEEE.

[12] Rashid, M. R. A., Hasan, K. F., Hasan, R., Das, A., Sultana, M., & Hasan, M. (2024). A comprehensive dataset for sentiment and emotion classification from Bangladesh e-commerce reviews. Data in Brief, 53, 110052.

[13] Shafin, M. A., Hasan, M. M., Alam, M. R., Mithu, M. A., Nur, A. U., & Faruk, M. O. (2020, December). Product review sentiment analysis by using nlp and machine learning in bangla language. In 2020 23rd International Conference on Computer and Information Technology (ICCIT) (pp. 1-5). IEEE.

[14] Ahmed, Z., Shanto, S. S., & Jony, A. I. (2023). Advancement in Bangla Sentiment Analysis: A Comparative Study of Transformer-Based and Transfer Learning Models for E-commerce Sentiment Classification. Journal of Information Systems Engineering & Business Intelligence, 9(2).