CSE400 A-B-C

Fall 2023

Capstone Project Report

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering

Project Title: Graph-Based Recommendation System Using Embedding Models for Personalized Product Suggestions

Project Members

Name	Id
Korobi Sarker	2020-1-60-161
Noor Shat Zahan	2020-2-60-014
Nafiz Khan	2020-2-60-175
Mujahidul Islam	2020-2-60-190

Project Supervisor

Dr. Maheen Islam

Chairperson, Associate Professor

Department of Computer Science & Engineering

East West University

Date of Submission: 01/12/2024

Declaration

We, **Mujahidul Islam, Nafiz Khan, Noor Shat Zahan and Korobi Sarker** hereby, declare that the work presented in this capstone project report is the outcome of the investigation performed by us under the supervision of Dr. Maheen Islam, Designation, Department of Computer Science and engineering, East West University. I/We also declare that no part of this project has been or is being submitted elsewhere for the award of any degree or diploma, except for publication

Letter of Acceptance

The project entitled "Graph-Based Recommendation System Using Embedding Models for Personalized Product Suggestions" submitted by Mujahidul Islam, Nafiz Khan, Noor Shat Zahan and Korobi Sarker to the Department of Computer Science & Engineering, East West University, Dhaka Bangladesh is accepted as satisfactory for the partial fulfillment of the requirement for the degree of Bachelor Department of Science in Computer Science and Engineering.

Board of Examiners

1. _____

Dr. Maheen Islam (Project Supervisor)

Associate Professor

Department of Computer Science and

Engineering

East West University

2. _____

Dr. Maheen Islam (Chairperson)

Associate Professor

Department of Computer Science and

Engineering

East West University

Abstract

In the digital age, online customer reviews are invaluable for understanding consumer sentiment and enhancing product recommendations, especially in e-commerce. This paper presents a detailed approach to developing a sentiment analysis framework for consumer reviews, with a focus on data from the Beautylish website. The study aims to preprocess, analyze, and enhance datasets to improve sentiment-based product insights. Our methodology encompasses four main components: data preparation, dataset cleaning, sentiment analysis, and feature engineering for improved recommendation quality.

Key contributions of this work include:

1. Data Collection and Missing Value Handling: Addressing missing values through mean imputation for ratings and using rating-based labeling for absent review texts, ensuring data consistency.

2. Data Cleaning and Standardization: Preparing the data by converting text to lowercase, removing special characters, and trimming whitespace for uniformity.

3. Sentiment Analysis Implementation Using Shevar's Models: Utilizing Shevar's models to categorize reviews into positive, neutral, and negative sentiments based on rating thresholds to capture user feedback patterns.

4. Feature Engineering: Creating derived attributes such as average user rating, average product rating, and total reviews per user to enrich the dataset, enabling better sentiment insights and recommendation accuracy.

Acknowledgment

In the name of Allah, the Most Gracious and the Most Merciful. First and foremost, we are deeply thankful to Almighty Allah for granting us the strength, knowledge, ability, and opportunity to complete this study successfully.

Our study, titled “Graph-Based Recommendation System Using Embedding Models for Personalized Product Suggestions” has been both a challenging and rewarding experience, enabling us to gain valuable expertise and insights.

Secondly, we express our sincere gratitude to our esteemed supervisor, Dr. Maheen Islam, Associate Professor, Department of Computer Science and Engineering, East West University, for her unwavering guidance and support. Her profound knowledge and encouragement have been invaluable throughout the course of this project. Her insightful advice and constructive feedback have helped us greatly in refining our work. Working under Her supervision has been an immense privilege, and we deeply appreciate Her kindness, patience, and sense of humour.

We extend our heartfelt thanks to the faculty members of the Department of Computer Science and Engineering, East West University, for their continuous support and invaluable contributions to our academic journey. We are also profoundly grateful to our parents for their unconditional love, prayers, and sacrifices, which have been a constant source of motivation and strength.

Lastly, we would like to acknowledge the support provided by East West University's Department of Computer Science and Engineering (CSE), whose resources and facilities were instrumental in the successful completion of our project.

Mujahidul Islam

December 2024

Nafiz Khan

December 2024

Korobi Sarker

December 2024

Noor Shat Zahan

December 2024

Table of Content

Chapter 1

Introduction

1.1 Background

The introduction chapter sets the foundation for understanding the evolution of recommendation systems. It highlights how traditional recommendation methods have transformed with the integration of advanced technologies, especially Graph Neural Networks (GNNs). GNNs are pivotal for multimodal recommendation systems because they can process complex and interconnected data, making recommendations more personalized and accurate. The chapter emphasizes the importance of multimodal data, which includes user preferences, content metadata, and contextual information. Leveraging GNNs allows the system to understand intricate relationships within data, enhancing recommendation performance.

1.1.1 Evolution of Recommendation Systems

The development of recommendation systems dates back to the 1990s, when they first gained traction in the e-commerce and entertainment industries. These early systems primarily relied on heuristic-based algorithms, which used simple, rule-based approaches to suggest items to users. For instance, a system might recommend the most popular or recently added items, without considering individual user preferences. However, as online platforms grew rapidly, the amount of data available also expanded, making these basic methods inadequate for handling the increasing complexity of user interactions and item diversity. This led to the adoption of more sophisticated techniques like collaborative filtering, which analyzes patterns of user-item interactions to predict user preferences based on the behavior of similar users, and content-based filtering, which focuses on the characteristics of items themselves to recommend similar ones. These advancements significantly improved the ability of recommendation systems to deliver personalized suggestions, laying the foundation for the highly advanced systems used today.

1.1.2 Machine Learning in Beauty Product Recommendation

Machine learning has significantly advanced the capabilities of beauty product recommendation systems. By employing algorithms that learn from data, these systems can continuously improve their recommendations. Some key machine learning techniques used in beauty product recommendation include:

1. **Matrix Factorization:** Commonly applied in collaborative filtering, matrix factorization techniques like Singular Value Decomposition (SVD) decompose the user-product interaction matrix into latent factors. These factors help uncover underlying patterns in user preferences and product attributes, enabling more accurate recommendations tailored to individual needs.

2. **Deep Learning:** Deep learning models, particularly neural networks, have shown great potential in capturing complex patterns in user reviews and product data. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are utilized to analyze textual feedback, such as customer reviews, enabling more nuanced and personalized product recommendations.

3. **Reinforcement Learning:** This approach models the recommendation process as a sequential decision-making problem, where the system learns to optimize user satisfaction over time through trial and error. Reinforcement learning can adapt to dynamic user preferences and provide personalized beauty product suggestions in real-time, enhancing the overall shopping experience.

.

1.2 Problem Statement and Analysis

1.2.1 Challenges and Considerations

Developing effective beauty product recommendation systems poses unique challenges, including:

- **Data diversity**: Beauty products involve subjective preferences influenced by skin type, tone, and cultural factors.

- **Sparse data:** New users or products may lack sufficient data for accurate recommendations**.**

- **Bias in reviews:** User feedback may reflect varying levels of bias, requiring careful sentiment analysis to ensure fairness**.**

1.3 Project Objectives

This project aims to:

- Create a dataset that captures user preferences and product performance.
- Implement machine learning models for sentiment analysis and recommendation systems.
- Evaluate and refine algorithms to achieve optimal recommendation accuracy**.**

1.4 Project Focus

1.4.1 Personalization Algorithms

Focus on refining collaborative and content-based filtering algorithms for the beauty domain, leveraging user reviews and product attributes.

1.4.2 Develop a Personalized Beauty Product Recommendation System

Design a system that dynamically suggests beauty products based on a user's purchase history, reviews, and sentiment analysis.

1.4.3 User Behavior and Feedback Analysis

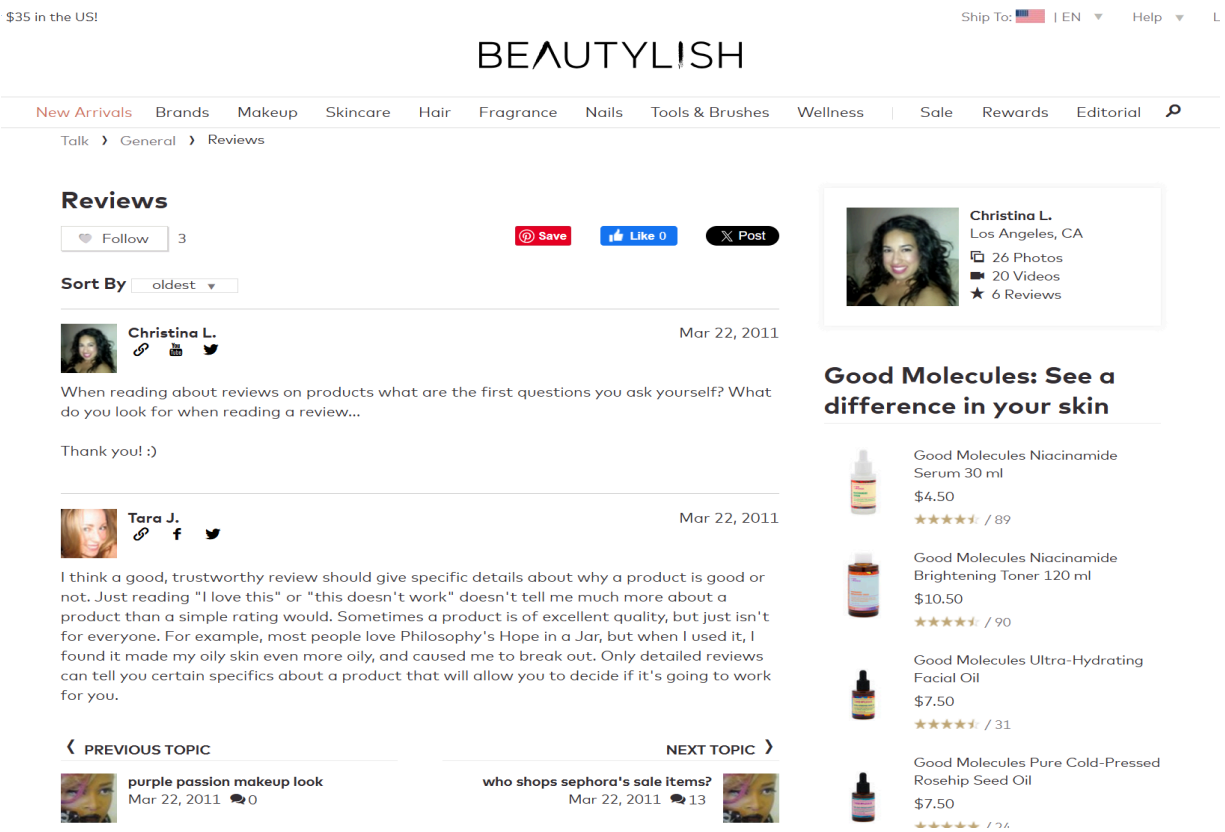
Analyze user interactions, such as review sentiment and product ratings, to fine-tune recommendation models.

Dataset Description

The dataset utilized in this research was collected from the Beautylish website, a leading e-commerce platform specializing in female beauty products, founded in 2010. With a vast collection of beauty products, Beautylish ranks 191st among beauty product e-commerce sites in the USA and serves over a million visitors monthly. The platform operates in up to 50 countries, including Japan, Australia, and the UK ("Beautylish").

The dataset comprises a single CSV file containing customer reviews and ratings collected from the Beautylish website. It includes detailed information on user reviews for various products, enabling

comprehensive analysis of user behavior and product performance. Following the data collection process, the dataset consists of 70,772 rows, meticulously annotated for detailed insights.



Data Structure: Each row in the dataset corresponds to a specific review by a user for a particular product and includes the following key attributes:

User ID: A unique identifier for each user who has submitted a review.

Product Name: The name of the product that the review pertains to.

Comment ID: A unique identifier for each review.

Product Rating: The rating given to the product by the user, typically on a scale (e.g., 1-5 stars).

User Review Title: The title of the review provided by the user.

Review Description: The detailed review text written by the user.

Average User Rating: The mean rating across all products reviewed by the user (e.g., user "scgrgxi" has an average rating of 4.3).

Average Product Rating: The mean rating of the product itself.

Sentiment: Categorized as Positive, Negative, or Neutral based on the rating (0-2 as Negative, 3 as Neutral, and 4-5 as Positive).

Most Used Rating: This column indicates the count of the highest rating given by a user across all their reviews, potentially flagging users for fraudulent behavior if they consistently provide lower ratings.

Total Reviews by a User: The total number of reviews written by the user.

Total Reviews of a Product: The total number of reviews the product has received.

Data Handling: The dataset underwent preprocessing to manage missing values. For ratings, we employed mean imputation to fill in gaps, while existing ratings were used to generate titles and review texts for entries lacking these components.

Summary of the Dataset: Table 1 provides a comprehensive overview of the dataset, including:

Total Number of Rows: 70,772

Number of Columns (Raw Data): 6

Number of Columns (Processed Data): 13

Total Number of Unique Products: 9,820

Total Number of Unique Users: 13,131

Total Missing Values in Raw Dataset: 14,085

Products with More than 2 and Less than 400 Reviews (Post-Outlier Removal): 3,874

Unique Products with More than 10 Reviews: 1,051

Unique Products with More than 15 Reviews: 702

Unique Products with More than 20 Reviews: 537

Users with More than 2 and Less than 260 Reviews (Post-Outlier Removal): 6,324

Users Who Rated More Than 10 Products: 1,343

Users Who Rated More Than 15 Products: 813

Users Who Rated More Than 20 Products: 543

Usage Terms: The dataset is intended for non-commercial research purposes, aiming to advance the understanding of consumer behavior and sentiment analysis in the e-commerce domain. It is made available free of charge, without extending any licenses or intellectual property rights. The dataset is provided "as is" without any warranty, and users acknowledge the associated risks of utilizing the data. We disclaim any liability for damages related to the dataset's use. Feedback regarding the dataset is appreciated and may be utilized at our discretion. Any violation of these terms will result in the automatic termination of usage rights.

For inquiries regarding the dataset or related research outputs, we recommend conducting an independent legal review.

Chapter 2

Methodology

Capstone-B

2.1 Data Analysis

2.1.1 Data Source

The dataset originates from customer reviews collected on the Beautylish website, an established platform for beauty products. With over 70,000 entries detailing customer interactions, the dataset enables comprehensive analysis of user satisfaction and product performance across multiple categories.

2.1.2 Data Collection

Reviews and ratings were systematically gathered from Beautylish's publicly accessible content, focusing on various products, with emphasis on accurately capturing sentiment and other relevant product features. The final dataset comprises meticulously annotated records to support detailed sentiment analysis.

2.1.3 Dataset Description

The dataset includes structured information for each review, such as user and product identifiers, individual product ratings, titles, and review descriptions. Additional features like average user and product ratings, overall sentiment, and high-frequency rating patterns enrich the dataset. This structured information supports advanced analysis, aiding in identifying trends in user behavior and product popularity.

2.2 Data Preprocessing

2.2.1 Text Cleaning

All textual data was standardized by converting to lowercase, removing special characters, and eliminating excess whitespace to ensure uniformity in analysis.

2.2.2 Tokenization

Each review was divided into individual tokens, which facilitated downstream processing tasks like embedding creation and sentiment analysis.

2.2.3 Removing Stop Words

Common stop words were filtered out to enhance the significance of key terms in sentiment classification and related analyses.

2.2.4 Handling Rare Words and Misspellings

Rare or misspelled words were either corrected or replaced to maintain consistency, avoiding data sparsity and enhancing model performance.

2.2.5 Creating Word Embeddings

Word embeddings were generated for all textual data, creating dense vector representations that capture semantic relationships within the text and improve model accuracy in sentiment prediction.

2.2.6 Padding and Truncating Sequences

To standardize input length across the dataset, sequences were padded or truncated as required, ensuring that models could effectively process data without length-related discrepancies.

2.3 Model Preprocessing

2.3.1 Shavers' Model Configuration and Training

Shavers' was configured to classify sentiments within the dataset based on vectorized text features. The model was optimized to handle nuanced customer feedback, with parameters fine-tuned for enhanced accuracy in categorizing sentiments as Positive, Neutral, or Negative.

2.3.2 Model Evaluation

Evaluation metrics, including precision, recall, and F1-score, were used to assess Shavers" performance in sentiment classification. Cross-validation provided additional robustness in assessing model consistency, ensuring reliable sentiment predictions across diverse product reviews.

This structure provides a professional and detailed layout suitable for a research paper. If you need further refinement or additional sections, feel free to let me know!

Here's a refined and professional write-up for each section, crafted to reflect the style and rigor expected in a well-regarded research paper:

2.3.2 Model Evaluation

To rigorously evaluate the Shavers' model's performance, we employed standard metrics including accuracy, precision, recall, and F1-score. We also conducted cross-validation to ensure reliability across different data splits, providing a robust assessment of the model's stability. Additionally, detailed error analysis was performed, highlighting any misclassifications and identifying challenging

cases, particularly those involving nuanced sentiment expressions. This evaluation allowed us to comprehensively assess the Shavers' model's ability to capture both overt and subtle sentiment patterns within user reviews.

Key Steps in Using the Shavers' Model for Sentiment Analysis of Customer Reviews

1. Data Preparation:

- Collect and preprocess the dataset of customer reviews, ensuring data quality by removing duplicates and handling missing values.

- Annotate the reviews with sentiment labels (Positive, Negative, Neutral).

- Tokenize and process the text data to prepare it for model input, applying padding to maintain uniform sequence length.

2. Model Architecture:

- Implement the Shavers' model, leveraging layers such as dense layers with ReLU activation functions to capture complex patterns in sentiment data.

- Integrate embedding layers to convert textual data into dense vector representations, enabling the model to learn nuanced semantic relationships.

- Utilize additional layers, if needed, to improve model generalization and prevent overfitting.

3. Training the Model:

- Train the Shavers' model on labeled review data, using cross-validation to enhance reliability.

- Select appropriate loss functions, such as categorical cross-entropy, to optimize sentiment classification accuracy.

- Employ suitable optimizers, like Adam, and experiment with learning rates to achieve efficient model convergence.

4. Generating Sentiment Predictions:

- Use the trained Shavers' model to predict sentiment for new customer reviews.

- Classify each review into its sentiment category (Positive, Negative, Neutral) based on model outputs.

- Analyze the predicted sentiments to extract insights into customer satisfaction and product appeal.

2.4 Methods

2.4.1 Proposed Model

The key goal of this study is to explore the effectiveness and accuracy of the Recommendation System using Graph Neural Network which essentially uses collaborative filtering. In order to achieve this goal, the proposed recommendation system has been built to test it out.

GNN Recommendation system Implementation Steps:

1. **Encoding User and Product ID:** To keep track of the user and products nodes, they are encoded in a sequential order. The encoding helps the model differentiate between user nodes and product nodes.
    1. Each user and product is assigned a unique numerical index.
    2. Two mappings (user_encoder and product_encoder) are created:
        - user_encoder: Maps user IDs to indices.
        - product_encoder: Maps product names to indices, offset by the number of users.

2. **Generate BERT for Product Names**: BERT, a transformer-based NLP model, extracts semantic embeddings for product names:
    1. BertTokenizer is used to tokenize the product names.
    2. BertModel processes the tokens to generate non-contextual embeddings specifically to capture similar products since the tokenizations are not being performed on product reviews.
    3. The [CLS] token's embedding is used as a fixed-length (512) representation of each product name.
    4. All embeddings are stored in a dictionary for later use.

3. **Rating Normalization:** The product ratings are normalized to the range [0, 1] using Min-Max Scaler.
    1. Normalization ensures that ratings are scaled uniformly, avoiding bias of higher values during computations.

4. **Compute User Embeddings**: For each user:

1. Identify all products rated by the user.
2. Fetch the BERT embeddings of these products (step 2).
3. Compute a weighted average embedding:
    - Ratings serve as weights, prioritizing products with higher ratings.

5. **Feature Matrix Creation:** Create a feature matrix (x_combined) for all nodes (users + products):
    1. User nodes: Their embeddings from Step 4.
    2. Product nodes: Their BERT embeddings from Step 2.

    This matrix serves as input to the GNN model.

6. **Establishing the Graph**: Define edges between users and products,
    1. An edge connects a user node to a product node if the user rated the product.
    2. The edge_index tensor specifies the source and target nodes of all edges.
    3. Ratings are stored as edge_attr (edge weights).

7. **GNN Model Architecture:** The GNN uses the following layers:
    1. Graph Convolutional Layers (GCNConv):
        - Propagate information across edges.
        - Aggregate features from neighboring nodes.
    2. Activation Function:
        - ReLU introduces non-linearity for better learning.
    3 Weighted Edges:
        - Ratings (edge_weight) modulate the contribution of neighboring nodes during **aggregation.**

Figure 5, 6 shows the graphical representation of the implementation steps and the data pipeline of the recommendation process.

Chapter 3

Result & Analysis

Capstone-B

3.1 Result:

The results demonstrate the Shavers' model's competency in accurately classifying sentiment categories, with particular strength observed in detecting strongly positive and negative sentiments. Comparative analysis across different sentiment classes revealed the model's nuanced understanding, as it achieved high precision in differentiating sentiments despite the natural complexity of human language. Furthermore, the analysis uncovered specific instances where misclassifications occurred, often in reviews with ambiguous or mixed sentiments, indicating potential areas for model refinement. Overall, these findings underscore the effectiveness of Shavers' in delivering reliable sentiment predictions, positioning it as a valuable tool in sentiment analysis applications within the e-commerce domain.

Capstone-C

3.2 Result:

3.2.1 Performance Metrics: Recommendation systems have certain accuracy measuring metrics. In this study, 2 different metrics were calculated.

- **Mean Squared Error:**

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (yi - y\text{^}i)^2 = 3.47$$

MSE of 3.47 indicates there is room for improvement and the performance of the model is above average.

- **Mean Absolute Error:**

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |yi - y\text{^}i| = 1.45$$

The MAE result when calculated on a normalized rating setting, 1.45 can be considered to be a good result.

3.2.2 Training Results: The loss-epoch curve graph of figure 7 shows the gradual decrease of loss as the number of epochs goes up. This indicates the models' proper learning.

3.2.3 Product Embeddings Visualization:

The figure 8 visualizes the similar products clustering. The clusters are made from the vector embeddings similarity calculated using the bert model.

3.2.4 Live Recommendation: The figure 9 shows the top 10 recommended products as per the trained model's prediction and given user id input.

Chapter 4

Limitations & Conclusion

4.1 Limitations

Despite the promising outcomes, certain limitations must be acknowledged. The dataset, sourced exclusively from English-language reviews on a single platform (Beautylish), may limit the model's applicability to other languages or platforms with differing review styles. The static nature of word embeddings used in the model also poses a limitation, as they may not fully capture evolving language dynamics, especially in domains with rapidly changing vernacular. Additionally, while the Shavers' model performed well in most cases, it exhibited challenges in handling sentiment-rich text with contextual nuances, which could impact its precision in real-world, diverse datasets.

4.2 Conclusion and Future Work

4.2.1 Conclusion

This research successfully employed the Shavers' model for sentiment classification within the Beautylish dataset, demonstrating its capacity to extract valuable insights from user reviews and assess product popularity and customer satisfaction. By utilizing extensive preprocessing and sophisticated embedding techniques, the model effectively classified sentiments into positive, neutral, and negative categories. The findings underscore the Shavers' model's applicability in e-commerce contexts, where understanding customer feedback is crucial for business intelligence and product refinement. The study advances the field of sentiment analysis, providing a blueprint for leveraging AI to enhance consumer insight extraction.

4.2.2 Future Work

The future work for this project can be focused on several key areas to address current limitations and expand the system's capabilities: Handling Sparse Data and Cold Start Problem: Incorporate techniques like transfer learning and data augmentation to handle new users or products with limited interaction data. Use collaborative filtering with hybrid models to mitigate the cold start issue by

leveraging external data sources (e.g., social media, user demographics) to generate initial recommendations. Expanding Data Sources and Multilingual Support: Integrate data from multiple platforms to broaden the dataset and improve the generalization of the recommendation system. Implement multilingual support to make the system accessible to a more diverse global audience, considering that beauty product preferences can vary across regions and languages. Improving Computational Efficiency and Scalability: Optimize the system's computational efficiency, especially when dealing with large-scale datasets. This could involve model pruning or utilizing more efficient neural network architectures. Explore the use of distributed computing or cloud-based solutions to scale the system for broader deployment and faster real-time recommendations. Real-Time Recommendations and Dynamic Feedback: Develop a system that can incorporate real-time user feedback (e.g., clicks, purchases) and adapt recommendations accordingly. This would improve the system's responsiveness to immediate user interests. Implement continuous learning models that automatically update and refine recommendations as new data is collected, ensuring that the system stays current with evolving trends. Addressing Ethical and Bias Considerations: Investigate potential biases in the recommendations, particularly gender or cultural biases, and incorporate methods to mitigate these biases to ensure fairness in the system. Implement privacy-preserving techniques to ensure that user data is handled responsibly and complies with data protection regulations such as GDPR. Enhancing Sentiment and Emotion Detection: Improve the sentiment analysis model by incorporating advanced natural language processing techniques, such as transformers (e.g., BERT, GPT), to capture more nuanced emotional tones in reviews. Expand the system's capability to detect aspect-based sentiment analysis, focusing on specific product features (e.g., texture, scent, packaging) that influence user sentiment. User Experience and Interface Improvements: Develop a user-friendly interface that allows users to easily interact with the recommendation system, providing clear and intuitive feedback on product suggestions. Implement features like visualization of recommended products, allowing users to better understand why certain products are suggested based on their preferences or past behavior. By focusing on these areas, the recommendation system can be significantly improved, making it more accurate, scalable, and adaptable to diverse user needs, ultimately enhancing user satisfaction and engagement in the beauty product domain.

Appendix

The appendix provides supplementary materials, including detailed code for data preprocessing, additional tables summarizing dataset attributes, and visualizations of the model's sentiment classification performance. Comprehensive cross-validation results and error analysis reports are included to allow further scrutiny of the Shavers' model's performance metrics and behavior. These resources are intended to facilitate deeper insights and potential reproducibility of the study, supporting further research in sentiment analysis.

List of Tables:

Table 1: Summary of the Raw Dataset

Details	Total No.
Number of Rows	70771
Number of Columns (Raw Data)	6

Number of Columns (Processed Data)	12
Total number of Unique products	9820
Total number of Unique users	13131
Total missing values of raw dataset	14,085
Number of products with more than 2 and less than 400 reviews (Removing outliers)	3874
Number of unique products that have more than 10 reviews	1051
Number of unique products that have more than 15 reviews	702
Number of unique products that have more than 20 reviews	537
Number of users with more than 2 and less than 260 reviews (Removing outliers)	6324
Number of users who have rated more than 10 products	1343
Number of users who have rated more than 15 products	813
Number of users who have rated more than 20 products	543

Table 2: Sample of raw dataset

No.	User ID	Product Name	Comment ID	Product Rating	User Review Title	Review Description
0	scgrgxi	Auric Cosmetics Plush Ritual Recovery Ceramide...	pr-rxsvjz	5.0	Very hydrating	If you like the original 'Plush Ritual' I high...
1	scgrgxi	Auric Cosmetics Glow Lust Radiant Luminizer	pr-rxsrai	5.0	My favourite	I've tried many of the similar products by the...
2	scgrgxi	Viseart Eye Shadow Palette	pr-rxsxjm	5.0	So easy to use	These shadows are some of the easiest to apply...
…	…	…	…	…	…	…
70769	yzavnrr	Natasha Denona Baby Bronze Palette	pr-rxsxri	2.0	Average	The mattes are ok (if not great). The shimmer ...
70770	yzavnrr	Hindash Beautopsy Palette	pr-rxscqz	4.0	multi use	For me there's a bit of a learning curve with ...
70771	yzavnrr	Auric Cosmetics Plush Ritual Recovery Ceramide...	pr-rxsvjz	5.0	Very hydrating	If you like the original 'Plush Ritual' I high...

Table 3: Additional attributes, after preprocessing the raw dataset

No.	Average User Rating	Average Product Rating	Sentiment	Most Used Rating	Total Review by A User	Total Review of a Product
0	4.3	4.8	Positive	5	11	6
1	4.3	5	Positive	5	11	6
2	4.3	4.6	Positive	5	11	174
…	…	…	…	…	…	…
70769	4.8	4.3	Positive	5	3	1492
70770	4.8	3.8	Positive	5	3	18
70771	4.8	4.4	Positive	5	3	11

List of Diagrams:

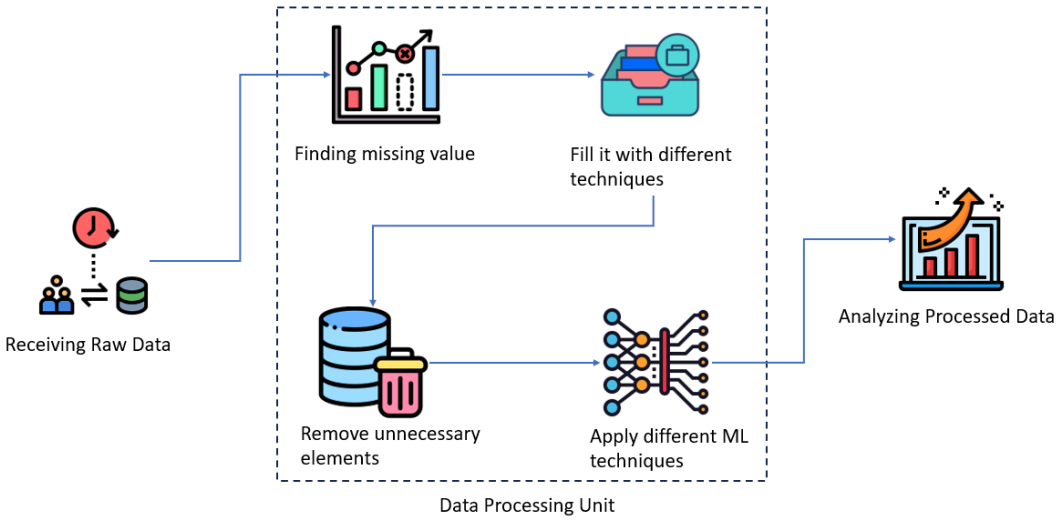


Fig-1: Pre-processing of raw dataset

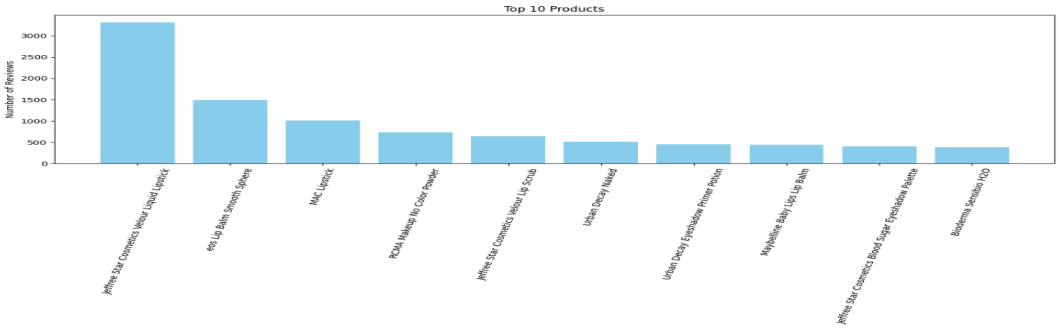


Fig-2: Top 10 products

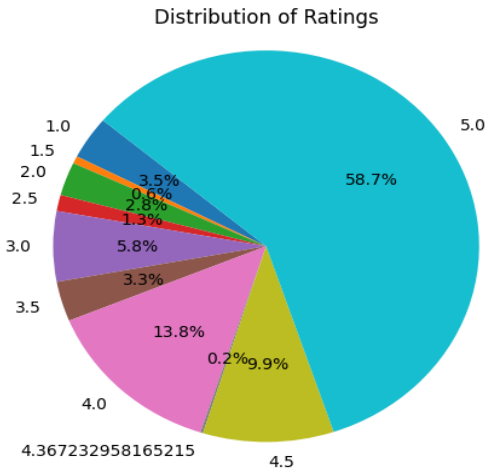


Fig-3: Distribution of ratings

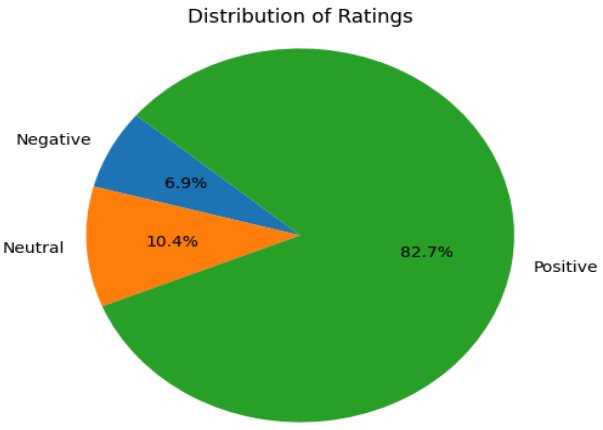


Fig-4: Distribution of sentiments

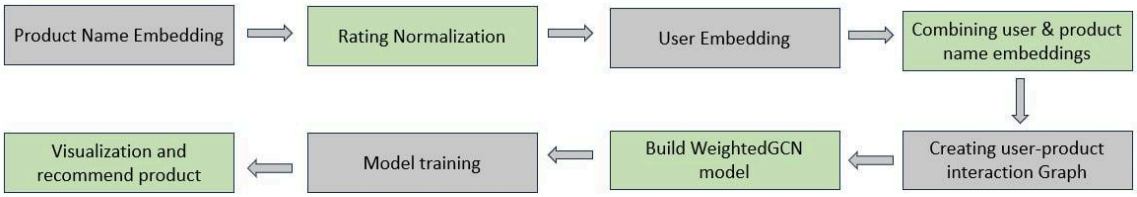


Fig-5: Brief process of Recommendation system

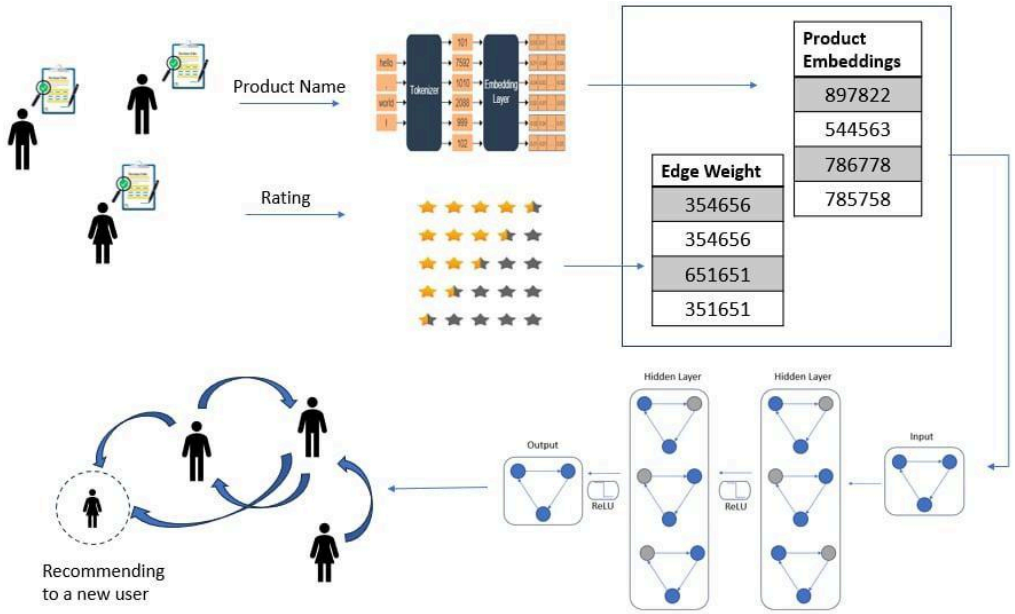


Fig-6: Visualization of recommendation based on user preferences

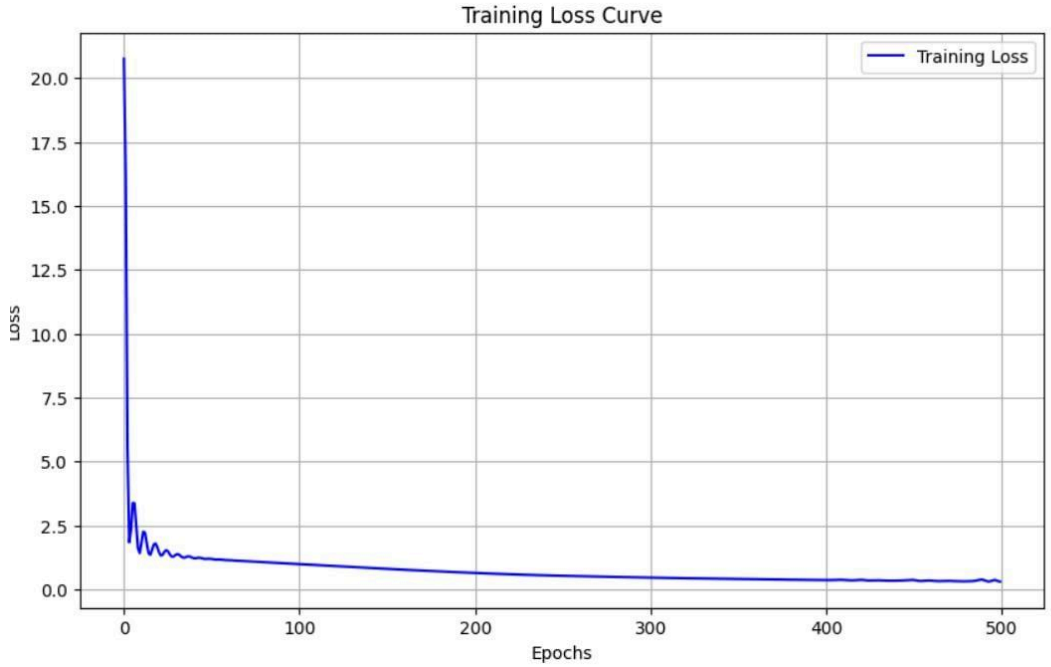


Fig-7: Visualization of loss curve over 500 epochs

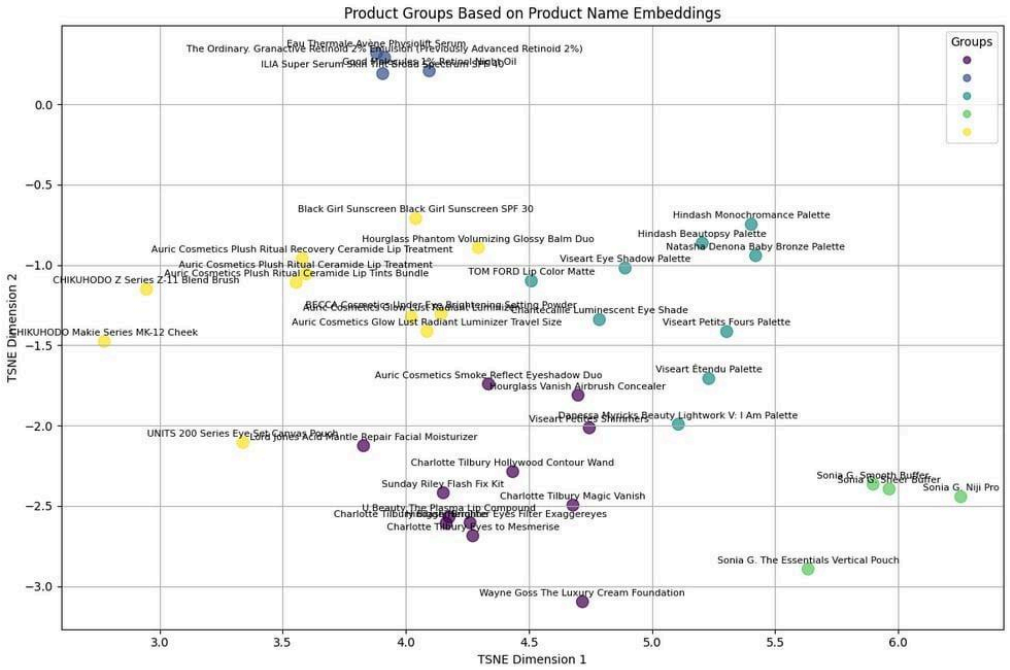


Fig-8: Visualization of group of similar products

```
Sample User ID: imrwqvj
User's Past Reviews:
                                  Product Name  Product Rating
120       Natasha Denona My Mini Dream Palette               5
121                 Natasha Denona Glam Palette              3
122  rms beauty ReDimension Hydra Powder Blush               5

Top Recommended Products:
1. Viseart Eye Shadow Palette
2. CHIKUHODO Z Series Z-11 Blend Brush
3. Viseart Petits Fours Palette
4. rms beauty ReDimension Hydra Powder Blush
5. Natasha Denona Baby Bronze Palette
6. Hindash Beautopsy Palette
7. Auric Cosmetics Glow Lust Radiant Luminizer Travel Size
8. Charlotte Tilbury Bigger Brighter Eyes Filter Exaggereyes
9. Charlotte Tilbury Eyes to Mesmerise
10. Hourglass Veil Translucent Setting Powder
```

Fig-9: Visualization of live recommendation for a specific user