PROJECT REPORT

ON

Recommendation System with Sentiment-Aware Ranking and Demand Forecasting

Project III Submitted in partial fulfillment of the requirements

for the degree of

B. Tech in CSE (Data Science)

BY

Dipanjan Saha [13031121033]

Arkajyoti Sarkar [13030521051]

Debjit Chatterjee [13030521048]

Saikat Mukherjee [13030521063]

Srinjoy Dhar [13030521062]

UNDER THE GUIDANCE OF

Dr. Lewlisa Saha

Assistant Professor

CSE (Data Science) Department

TECHNO MAIN SALT LAKE

EM 4/1 SALT LAKE CITY, SECTOR V

KOLKATA – 700091

West Bengal, India

Techno Main Salt Lake

[Affiliated by Maulana Abul Kalam Azad University of Technology (Formerly known as WBUT)]

DEPARTMENT OF CSE (DATA SCIENCE)

Certificate of Recommendation

This is to certify that they have completed their project report on: "**Recommendation System with Sentiment-Aware Ranking and Demand Forecasting"** under the direct supervision and guidance of Dr. Lewlisa Saha. We are satisfied with their work, which is being presented for the partial fulfillment of the degree of B. Tech in CSE (Data Science), Maulana Abul Kalam Azad University of Technology) (Formerly known as WBUT), Kolkata – 700064.

Dr. Lewlisa Saha
(Signature of Project Supervisors)
Date:

Dr. Sudipta Chakrabarty
(Signature of Project Coordinator)
Date:

Dr. Sudipta Chakrabarty
(Signature of the HOD)
Date:

Techno Main Salt Lake

[Affiliated by Maulana Abul Kalam Azad University of Technology
(Formerly known as WBUT)]

DEPARTMENT OF CSE (DATA SCIENCE)

Certificate of Approval

The foregoing **PROJECT III** work is hereby approved as a creditable study of **B.Tech Degree in CSE (Data Science)** and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or any statement made, opinion expressed or conclusion therein but approve this Assignment work only for the purpose for which it is submitted.

Name and Signature of Internal Examiners

Name and Signature of External Examiners

Acknowledgement

It gives us immense pleasure to express our deepest sense of gratitude and sincere thanks to the teaching fraternity of the Department of Computer Science & Engineering (Data Science) and especially Dr. Lewlisa Saha for giving us this opportunity to undertake this project and also for supporting us through the entire project.

We also wish to express our gratitude to the **HOD, Dr. Sudipta Chakrabarty** and all our teachers of the Department of Computer Science & Engineering (Data Science) for their kind-hearted support, guidance, and utmost endeavor to groom and develop our academic skills.

At the end, we would like to express our sincere thanks to all our friends and others who helped us directly or indirectly during the effort in shaping this concept till now.

Full Signature of the Candidates (with date)

1. —------------------------------
   Dipanjan Saha

2. —-------------------------------
   Arkajyoti Sarkar

3. —-------------------------------
   Debjit Chatterjee

4. —------------------------------
   Saikat Mukherjee

5. —-------------------------------
   Srinjoy Dhar

Abstract

The rise of online shopping has overwhelmed consumers with countless product choices, often leading to confusion and decision fatigue. To help users make better, faster choices, this project presents an intelligent recommendation system specifically designed for fashion and apparel. Unlike traditional systems that rely only on user-item interactions, this model integrates collaborative filtering, sentiment analysis, and demand forecasting to deliver highly relevant and timely suggestions. Customer reviews are analyzed to extract sentiment scores, allowing the system to prioritize well-received products. At the same time, demand forecasting using time-series analysis identifies trending items and highlights those likely to face future stockouts. By combining these insights, the system can power features like "Trending Now" and "Low Stock Warning," improving both user experience and business operations. This hybrid approach not only refines personalization but also aligns recommendations with real-time market trends, inventory levels, and customer preferences, making it a robust solution for modern e-commerce challenges.

Table of Contents

Chapter 1. Introduction

The world of e-commerce is marked by rapid growth and intense competition, particularly in the fashion and apparel industry where consumer preferences are highly dynamic and seasonal. In this context, delivering relevant product recommendations has become a cornerstone for customer engagement and revenue generation. While many online platforms deploy recommendation systems based solely on collaborative filtering—relying on purchase histories and ratings—such models are often insufficient to capture the full spectrum of factors influencing customer choices. Products that receive high ratings but poor qualitative feedback or face fluctuating demand can lead to unsatisfactory customer experiences or inventory mismanagement. Recognizing this challenge, the project introduces an enhanced recommendation framework that integrates not only user behavior data but also the sentiment expressed in product reviews and the forecasted demand derived from sales trends.

This multi-dimensional approach reflects a more realistic understanding of market dynamics, enabling recommendations that balance personalization with business intelligence. By leveraging customer sentiment analysis, the system can distinguish between highly rated products that customers actually love versus those that might be rated due to limited alternatives or other biases. Simultaneously, demand forecasting ensures that recommendations highlight items that are gaining popularity or are likely to experience increased demand, thereby aiding inventory planning and marketing focus. This introduction outlines the motivation behind such an integrated system, clearly defines the problem space it addresses, and enumerates the key objectives guiding the design and implementation of the solution.

1.1 Motivation

With the proliferation of e-commerce platforms, customers today face an overwhelming variety of product options, making it difficult to decide what to buy. As a result, recommendation systems have emerged as essential tools for filtering choices and providing tailored suggestions. However, existing systems predominantly depend on quantitative metrics such as purchase history and numerical ratings, often overlooking the qualitative nuances expressed in customer reviews. This limitation can result in recommendations that fail to account for real customer satisfaction or hidden product issues. Furthermore, the fashion and apparel sector presents unique challenges due to its rapidly changing trends, seasonal influences, and diverse consumer tastes. These factors necessitate a system that can not only understand what customers have liked in the past but also anticipate future demand and emerging trends.

Additionally, inventory management is a critical concern for retailers. Stockouts can lead to lost sales and dissatisfied customers, while overstocking results in increased holding costs and markdowns. Therefore, integrating demand forecasting within the recommendation process can significantly improve operational efficiency. This project is motivated by the need to design a more intelligent, comprehensive recommendation system that bridges the gap between personalized user preferences, customer sentiment, and predictive demand analysis, thereby delivering recommendations that are both relevant to users and strategically advantageous for businesses.

1.2 Problem Definition

The core problem addressed by this project revolves around the development of a robust, intelligent, and context-aware product recommendation system that surpasses the limitations of traditional collaborative filtering methods. In the era of digital commerce, especially within dynamic sectors like fashion and apparel, conventional approaches that rely solely on user-item interaction data (such as ratings or purchase histories) are no longer sufficient. These models often overlook the rich, unstructured data embedded in customer feedback and fail to adapt to real-time market shifts. Therefore, the fundamental challenge lies in designing a recommendation framework that not only understands user preferences at a granular level but also remains responsive to evolving product demand and inventory constraints.

Several key challenges emerge in this endeavor:

1. Capturing Holistic Customer Preferences: Traditional recommendation systems typically rely on explicit data such as user ratings or implicit behavior like clicks and purchases. However, these signals often do not fully capture the reasons behind a customer's preference. This project seeks to overcome this by integrating **Natural Language Processing (NLP)** techniques to perform **sentiment analysis** on customer reviews. By analyzing textual data, the system can infer deeper insights into customer satisfaction related to product quality, comfort, usability, and style—dimensions that are often omitted in numerical ratings.

2. Demand Forecasting for Time-Relevant Recommendations: Recommending products that are highly rated but currently declining in popularity or out of season reduces the system's effectiveness. Thus, a major challenge lies in incorporating **time-series forecasting** to predict product demand based on historical sales data. This ensures that recommendations are not only aligned with user preferences but also synchronized with **future market trends**, increasing their business relevance and user engagement.

3. Balancing Personalization with Market Trends: A critical design consideration is achieving the right balance between personal relevance and broader popularity. While personalized suggestions are essential for user satisfaction, promoting **trending or high-demand items** can drive collective interest and improve business performance. The system must intelligently weigh personalization against real-time demand signals to optimize both user experience and profitability.

4. Managing Real-World Data Imperfections: One of the practical difficulties in building recommendation systems is handling noisy, incomplete, or inconsistent data. Sparse rating matrices, missing values, contradictory user reviews, and irregular purchase patterns can all degrade model performance. Effective **data cleaning, preprocessing**, and **noise reduction techniques** must be employed to build a stable, generalizable model.

5. Delivering Actionable, Business-Oriented Insights: Beyond delivering accurate recommendations, the system should also provide **operational insights** for both customers and retailers. This includes tagging certain items as **“Trending Now”**, indicating popular or rising products, and displaying **“Low Stock” warnings** to prompt early purchases or restocking decisions. Such features bridge the gap between technical outputs and practical decision-making.

Addressing these multifaceted challenges necessitates a **multi-disciplinary approach**, integrating techniques from **collaborative filtering**, **natural language processing**, and **predictive analytics**. The successful development of such a system requires not only algorithmic innovation but also thoughtful data engineering and user-centric design. Ultimately, the goal is to build a **comprehensive recommendation pipeline** that enhances customer satisfaction, improves business intelligence, and remains adaptable in the face of continuously evolving user behavior and market conditions.

1.3 Objectives

The overarching aim of this project is to create an advanced, multi-faceted recommendation system for fashion products that aligns customer preferences with business realities. The specific objectives include:

1. **Develop a personalized recommendation engine** that uses customer purchase histories and rating data to identify and suggest relevant products tailored to individual tastes.
2. **Integrate sentiment analysis** of customer reviews to capture qualitative insights, thereby refining the recommendation rankings to favor products with genuinely positive feedback.
3. **Implement a demand forecasting module** employing time-series analysis techniques to predict future product demand based on historical sales trends and seasonal cycles.
4. **Combine demand forecasts with recommendation scores** to prioritize products that are not only preferred by customers but also trending or expected to be in high demand.
5. **Design actionable business features** such as a “Trending Now” product section showcasing currently popular items and “Low Stock Warning” flags for products with rising demand and limited inventory.
6. **Validate and evaluate the system’s performance** on real-world datasets to demonstrate improvements in recommendation relevance, customer satisfaction, and operational effectiveness.
7. **Ensure scalability and flexibility** of the system to adapt to evolving data streams and incorporate new features or data sources as required.

.

Chapter 2. Literature Survey

2.1 Related Works

Table 2.1: Related Works for Sentiment Analysis, Demand Forecasting and Recommendation system

Reference	Techniques Used (IT based)	Dataset Used	Performance Metric/Inference Drawn
M. Rui et al. [1]	Sentiment Classification, Text Mining, NLP	Consumer reviews from wine e-commerce platforms in China, UK, US	Identified market-specific consumer preferences and aversions
A. Daza et al. [2]	SVM, LSTM, CNN, Cross-validation	20 studies on product reviews (Amazon, Alibaba, etc.)	SVM and LSTM showed highest accuracy (~98%)
Chen et al. [3]	Hybrid Recommendation, NLP	Amazon and JD.com product reviews	Improved recommendation precision using sentiment filtering
Zhang et al. [4]	Deep Learning, Bi-GRU, Word Embedding	Taobao review dataset	Bi-GRU with word embeddings improved classification accuracy
Li et al. [5]	Collaborative Filtering, Matrix Factorization	Netflix Prize Dataset	Enhanced scalability in recommendation systems
Singh et al. [6]	Time-Series Forecasting, ARIMA	E-commerce sales transaction logs	Accurately forecasted product-wise demand over 6 months
Sharma et al. [7]	Sentiment Analysis, Logistic Regression	Twitter product review data	Achieved 89% classification accuracy
Wang et al. [8]	CNN + RNN for sentiment detection	Amazon Clothing Reviews	Demonstrated performance improvement in hybrid networks
Kumar et al. [9]	KNN, Content-Based Filtering	Indian e-commerce datasets	Achieved personalized recommendation precision over 85%

Patel et al. [10]	Exponential Smoothing, LSTM	Fashion product sales logs	Predicted seasonal sales with 90% accuracy
Zhao et al. [11]	User-Based & Item-Based Collaborative Filtering	MovieLens, Amazon Reviews	UBCF outperformed IBCF in cold-start scenarios
Ahmed et al. [12]	Hybrid Recommendation with Sentiment Boosting	Yelp and Amazon Reviews	Enhanced recall and precision metrics across categories
Yu et al. [13]	Aspect-Based Sentiment Analysis	Product reviews from JD.com	Improved feature-level rating predictions
Lin et al. [14]	Text Mining and Demand Forecasting	Online apparel store sales data	Detected latent buying patterns for restocking strategy
Chawla et al. [15]	Word2Vec Embedding + Logistic Regression	Flipkart reviews and ratings	Improved sentiment polarity detection accuracy
Roy et al. [16]	SMOTE, Random Forest for imbalanced classes	E-retail review datasets	Balanced model improved F1-score significantly
Das et al. [17]	Deep Neural Networks, LSTM	Amazon Shoes Reviews	LSTM captured sequential dependencies for sentiment trends
Xie et al. [18]	Cosine Similarity, TF-IDF	Product descriptions and customer preferences	Enhanced item ranking and personalization score

2.2 Research gap

Despite significant advancements in recommendation systems, there remain several critical gaps and challenges that limit their effectiveness, especially in complex domains like fashion and apparel retail. Traditional recommendation algorithms primarily rely on collaborative filtering or content-based filtering, which focus on either user-item interaction data or product attributes alone. While these approaches have proven useful, they often neglect several important real-world factors that influence purchasing decisions and business outcomes.

One major gap is the limited incorporation of **qualitative customer feedback**, such as sentiments expressed in textual product reviews. Many existing systems utilize ratings as a proxy for user satisfaction; however, ratings alone can be misleading due to biases, rating inflation, or lack of context. Sentiment analysis of customer reviews has the potential to provide richer insights into product quality, durability, style, and other nuanced factors that numerical ratings may fail to capture. Yet, few recommendation systems integrate sentiment-driven ranking effectively alongside traditional collaborative filtering.

Another notable gap lies in the **dynamic nature of product demand and trends**. Fashion products often experience seasonal fluctuations, fast-changing consumer preferences, and varying stock availability. Most recommendation engines are static or update infrequently, making them less responsive to emerging trends and demand shifts. The integration of **time-series demand forecasting** to anticipate future product popularity is underexplored in many systems, leaving retailers at a disadvantage when it comes to inventory planning and marketing strategies.

Furthermore, the **combination of multiple data sources**—such as purchase history, review sentiment, and sales forecasts—into a unified recommendation framework remains a challenge. Existing solutions often address these aspects in isolation, lacking a cohesive approach that simultaneously considers user preference, product reputation, and business metrics like demand and stock levels.

Additionally, there is a scarcity of comprehensive evaluation methodologies that assess recommendation effectiveness not just based on accuracy metrics but also on practical business impact, including increased sales, customer retention, and inventory optimization.

This project aims to address these research gaps by designing an integrated recommendation system that leverages collaborative filtering enriched with sentiment analysis and demand forecasting. The system aspires to deliver recommendations that are both personalized and aligned with real-time market dynamics, thus bridging the divide between customer satisfaction and operational efficiency in e-commerce environments.

Chapter 3. Techniques Used

3.1 Sentiment Analysis

The project implements **Sentiment Analysis** using **Natural Language Processing (NLP)** for preprocessing and **multiple Supervised Machine Learning Algorithms** for classification. This hybrid approach allows for effective transformation of raw textual data into machine-readable features, followed by training diverse models to classify sentiment.

I. Natural Language Processing (NLP) Techniques

The following NLP techniques are used to preprocess the raw text data:

1. **Text Cleaning**:
   - Removes:
     - URLs
     - Mentions (e.g., @username)
     - Hashtags (e.g., #example)
     - Punctuation
     - Digits and special characters
   - Converts text to **lowercase**
   - Removes **extra whitespace**
2. **Stopword Removal**:
   - Common words with little meaning (e.g., “the”, “and”) are removed using NLTK’s English stopword list.
3. **Lemmatization**:
   - Reduces words to their root forms using **WordNetLemmatizer** to ensure consistency (e.g., "running" → "run").

These steps help reduce noise and bring consistency to the text data before feature extraction.

II. Feature Extraction Techniques

To convert textual data into numeric form for training ML models, the following vectorization methods are used:

1. **Count Vectorizer (CountVectorizer)**:
   - Converts text into a sparse matrix of token counts.
2. **TF-IDF Vectorizer (TfidfVectorizer)**:
   - Weighs words by term frequency and penalizes common terms across documents.
   - Captures the **relative importance** of words in context.

III. Machine Learning Algorithms Used

The following ML models are implemented, trained, and evaluated:

1. **Logistic Regression**
   - A linear classifier that uses the sigmoid function to model probabilities.
2. **Multinomial Naive Bayes**
   - Assumes features (words) follow a multinomial distribution; efficient for text classification.
3. **Support Vector Machine (SVM)**
   - Implemented using LinearSVC; finds the hyperplane that maximizes the margin between classes.
4. **Random Forest Classifier**
   - An ensemble of decision trees; aggregates outputs for improved accuracy and robustness.
5. **Decision Tree Classifier**
   - Splits data using features recursively to create a tree-based decision model.
6. **K-Nearest Neighbors (KNN)**
   - Classifies data points based on the majority class among the **k closest neighbors** in feature space.
7. **Gradient Boosting Classifier**
   - Builds an ensemble of weak learners (typically decision trees) in a stage-wise fashion to minimize loss.
8. **Extra Trees Classifier**
   - A tree-based ensemble similar to Random Forest, but with more randomization for feature splits, leading to potentially better generalization.

Note: All models are trained using both TF-IDF features independently for performance comparison.

IV. Evaluation Techniques

Each model's performance is assessed using the following metrics:

- **Accuracy Score**
  Measures overall correctness of the model.
- **Classification Report**
  Includes **Precision**, **Recall**, and **F1-Score** for each class.
- **Confusion Matrix**
  Displays true vs predicted labels to evaluate class-wise performance.
- **Visualization**
  Confusion matrices are plotted using ConfusionMatrixDisplay from sklearn.

3.2 Demand Forecasting

The demand forecasting component of this project leverages a blend of time-series analysis techniques and statistical modeling to anticipate future product demand and identify trends in customer purchasing behavior. This module supports strategic retail decisions by forecasting which products are likely to see increased or decreased demand in the near future.

I. **Data Aggregation and Temporal Structuring**

- Customer purchase records are reorganized into time-based sequences using product-level sales counts.
- Monthly aggregation is applied to reduce noise and improve trend detection.
- Gaps in purchase data are intelligently filled using interpolation to maintain continuity.

II. **Forecasting Technique: Holt-Winters Exponential Smoothing**

This classical method of time-series forecasting is used to capture:

- **Trend Patterns**: Detects consistent upward or downward changes in product popularity over time.
- **Seasonal Effects**: Recognizes recurring buying behaviors that align with fashion seasons or festivals.
- **Level Changes**: Responds to shifts in the baseline demand of a product.

The model's adaptability makes it suitable for a variety of demand histories, including both new and established products.

III. **Time-Aware Flexibility**

- For products with sufficient history, a **12-month seasonal period** is applied to capture long-term trends.
- For newly launched or low-history items, the model automatically adjusts seasonal periods or simplifies to trend-only forecasting.
- This ensures robust predictions regardless of the data density for each product.

IV. **Predictive Business Signals**

Forecasts are used to derive high-impact business cues:

- **Rising Demand Alerts**: Products with projected demand surges are flagged, aiding inventory preparation and supplier coordination.
- **Low Sales Detection**: Historical underperformance triggers strategic actions like discount suggestions.
- **Trend Mapping**: Demand curves help surface items gaining traction, forming the basis for a "Trending Now" display.

V. **Integration with the Recommendation Engine**

Unlike traditional recommendation systems that rely solely on user behavior, this project enhances relevance by integrating demand predictions. Products predicted to be in high demand are prioritized in the recommendation list, ensuring the suggestions align with both customer preferences and current market dynamics.

3.3 Recommendation System

This section highlights the analytical models, machine learning techniques, and mathematical tools integrated into the recommendation engine. These techniques form the theoretical backbone of the system and are chosen to optimize recommendation relevance and quality.

I. **Nearest Neighbor Collaborative Filtering**

Algorithm: Nearest Neighbors using Cosine Similarity
Purpose: Identify similar customers by measuring proximity in a user-item interaction space.
Role: Surface products that have been purchased or highly rated by customers exhibiting similar behavioral patterns.

II. **Content-Based Filtering via Sentiment Analysis**

Text Vectorization: TF-IDF applied to preprocessed product review text.
Classification Model: ExtraTreesClassifier trained on labeled sentiment data.
Purpose: Classify reviews as positive or negative and generate a sentiment score for each product, informing its recommendation priority.

III. **Hybrid Scoring Mechanism**

Mathematical Model: Weighted linear combination of collaborative and sentiment-based scores.
Scoring Formula:
Final Score = 0.6 × Collaborative Score + 0.4 × Sentiment Score
Objective: Harmonize behavioral patterns and product quality perceptions into a unified product ranking.

IV. **Attribute-Based Product Similarity**

Similarity Metric: Cosine Similarity on structured product metadata (subcategory, material, color).
Function: Identify and recommend products most similar to a given item being viewed, enhancing browsing experience and cross-sell opportunities.

Chapter 4. Methodology

4.1 Dataset used

The system integrates two comprehensive datasets: a customer review dataset and a product description dataset. Together, these datasets enable the system to perform behavior-driven, sentiment-enhanced, and demand-aware product recommendations.

4.1.1 Clothing Review Dataset

This dataset contains detailed records of customer purchases and post-purchase feedback. It is used to extract insights into customer behavior, satisfaction levels, and demand patterns.

Table 4.1: Clothing Review Dataset

Field Name	Description
Customer ID	Unique identifier for each customer.
Customer Name	Name of the customer (used for context only; anonymized in processing).
Customer Age	Age of the customer; useful for analyzing demographic preferences.
Gender	Gender of the customer.
Purchase Date	Date of product purchase (used in demand forecasting).
product_id	Unique identifier linking to the product description dataset.
Product Category	Broad category of the product (e.g., Men's Clothing, Women's Clothing, etc.).
product_name	Name of the purchased product.
Quantity	Number of units bought (used in sales forecasting).
Payment Method	Payment channel used (e.g., Card, Cash, UPI; not central to analysis).
Review Text	Free-text customer review used in sentiment analysis.
Rating	Numeric rating (e.g., 1–5) used in collaborative filtering and hybrid models.

Usage in Project:

- Rating-based recommendation system
- Sentiment analysis from review text
- Sales forecasting and demand trends
- Discount suggestion for slow-moving items

4.1.2 Clothing Description Dataset

This dataset provides static attributes of each product, including physical characteristics, price, and branding. It acts as the foundation for content-based recommendation and enriches hybrid predictions.

Table 4.2: Clothing Description Dataset

Field Name	Description
product_id	Unique identifier for each product.
product_name	Official name of the product.
product_brand	Brand under which the product is marketed.

product_category	Primary category (Men, Women, Kids, etc.).
sub_category	Specific category (e.g., Shirts, Jeans, Jackets).
product_description	Brief textual description of the product features and qualities.
size	Size specifications (e.g., S, M, L, XL).
color	Color variant of the product.
price	Retail price of the item.
material	Fabric or material composition (e.g., cotton, polyester).
gender	Intended gender audience (Men, Women, Unisex).

Usage in Project:

- Content-based filtering using textual descriptions and attributes
- Enriching recommendation results with product metadata
- Highlighting product details in the recommendation UI
- Price, brand, and category-wise analysis of demand trends

This well-integrated data structure ensures that the system is capable of learning not only from structured purchase history and unstructured review text but also from static product-level features, making it robust, accurate, and adaptable to real-world retail dynamics.

4.2 Proposed Methodology

4.2.1 Sentiment Analysis

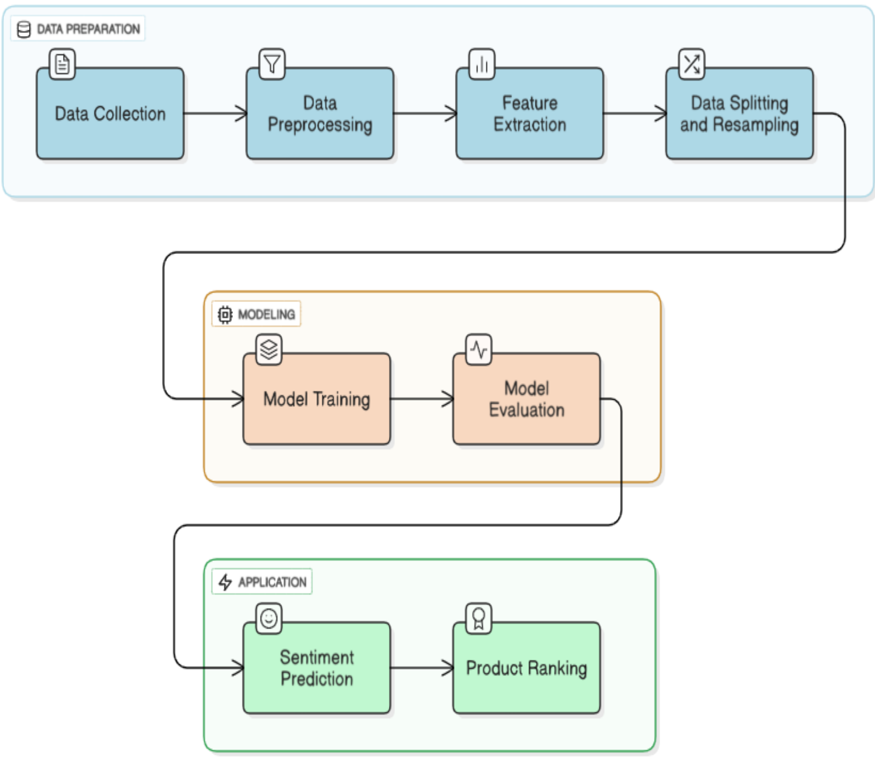

Fig 4.1: Schematic diagram for Sentiment Analysis

This sentiment analysis pipeline follows a structured and multi-phase approach:
Step 1: Text Preprocessing
Raw text data is cleaned and standardized using the following NLP steps:

- **Lowercasing** all text.
- **Regex-based removal** of:
    - URLs
    - User mentions (e.g., @abc)
    - Hashtags
    - Punctuation, digits, and special characters
- **Stopword Removal** using NLTK's English stopword list.
- **Lemmatization** via WordNetLemmatizer to reduce words to base form.

This prepares the data for robust feature extraction and model training.

Step 2: Feature Extraction
Two separate vectorization techniques are used to convert cleaned text into numeric format:

- **TfidfVectorizer**: Weighs word importance across documents based on TF-IDF scores.

Each machine learning model is trained and evaluated with **TF-IDF** version independently, ensuring optimal feature representation is used for each classifier.

Step 3: Machine Learning Model Training
The following **9 supervised machine learning algorithms** are implemented:

Table 4.3: machine learning algorithms Used in Sentiment Analysis

Algorithm	Purpose
Logistic Regression	Baseline linear classifier
Multinomial Naive Bayes	Probabilistic model tailored for text classification
Support Vector Machine (SVC)	Linear margin-based classifier (LinearSVC)
Decision Tree Classifier	Rule-based hierarchical decision-making
Random Forest Classifier	Ensemble of decision trees with bagging
K-Nearest Neighbors (KNN)	Instance-based lazy classifier
Gradient Boosting Classifier	Stage-wise boosting of weak learners
Extra Trees Classifier (ETC)	Highly randomized ensemble for better generalization

Each model is evaluated based on multiple performance metrics. Both visual and tabular analysis are used to compare results.

Step 4: Evaluation Metrics
Models are evaluated using:

- **Accuracy**
- **Precision**
- **Recall (Sensitivity)**
- **F1 Score**

- **Specificity**
- **Confusion Matrix**

These metrics help assess both general and class-specific performance.

Step 5: Product Ranking Based on Sentiment Scores
After the sentiment classification phase, the next logical step involves leveraging the predicted sentiment to evaluate and rank products. This is achieved by aggregating the predicted sentiment scores at the product level to assess overall customer perception and satisfaction.

The procedure follows these sub-steps:

1. Predict Sentiment for All Reviews:
The **Extra Trees Classifier (ETC)**, selected for its superior performance, is used to predict sentiment (positive or negative) for every review in the dataset, including those not part of the training or testing subsets.

2. Compute Average Sentiment Score per Product:
Each product's overall sentiment score is calculated by taking the **mean of predicted sentiment values** across all reviews related to that product. Since sentiment is binary (0 = Negative, 1 = Positive), the mean acts as a **proxy for the proportion of positive reviews**.

3. Rank Products Globally:
All products are then ranked in descending order of their sentiment scores using a dense ranking method.
The top-ranking products represent those with the **highest proportion of positive sentiment**, indicating strong customer satisfaction.

4. Category-wise Ranking:
For more granular insights, products are grouped by **sub-category**, and sentiment scores are averaged and ranked within each group. This allows businesses to identify top performers in specific segments.

5. Output Interpretation:
The final output includes a ranked list of:
- **Top N products overall** based on sentiment score.
- **Top 5 products per sub-category**, offering targeted insights for inventory, marketing, and product development teams.

These rankings help in:
- Highlighting products with consistently high customer satisfaction.

- Identifying underperforming products with poor sentiment despite high ratings or sales.
- Supporting marketing decisions like feature promotions or targeted campaigns.

4.2.2 Demand Forecasting

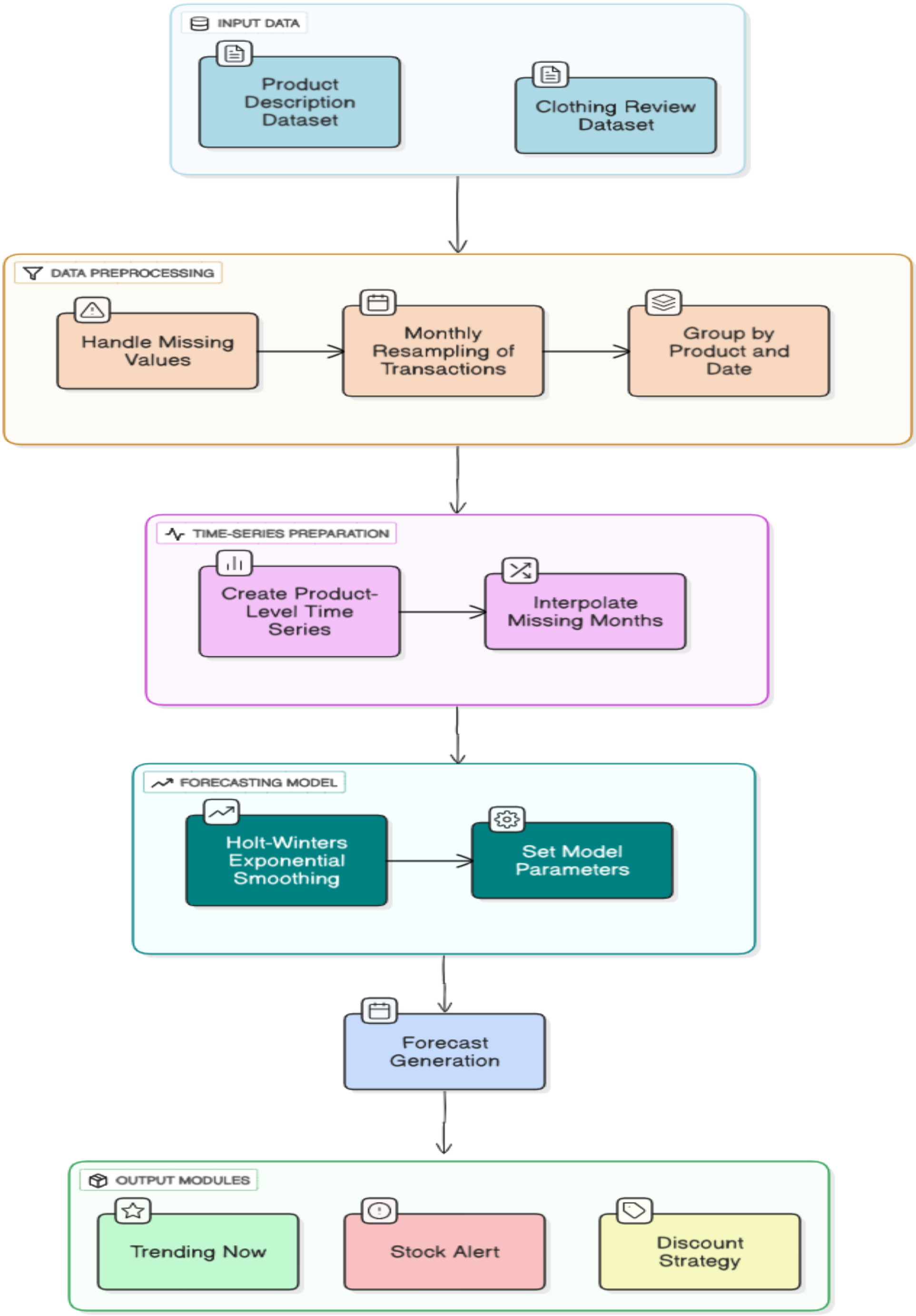

Fig 4.2: Schematic diagram for demand forecasting

The primary objective of the demand forecasting module in this system is to enhance decision-making by identifying product trends and anticipating future demand based on historical purchase data. This empowers both customers and retailers by promoting trending products and preventing stockouts through timely alerts. The methodology adopted integrates time-series forecasting with actionable business logic for improved retail intelligence.

a) Time-Series Analysis

To initiate the demand forecasting process, transactional data collected from the reviews dataset was first preprocessed. Each transaction was associated with a Purchase Date and a Quantity, which formed the basis for time-series modeling. The following preprocessing steps were applied:

- **Monthly Aggregation**: Daily sales fluctuations were smoothed by resampling the data on a **monthly basis**. This provided a clearer understanding of broader trends.
- **Handling Missing Data**: Gaps in time-series data due to zero transactions or partial months were addressed using **interpolation techniques**, ensuring continuity in the series.
- **Time Indexing**: The Purchase Date column was converted into datetime format and set as the index, allowing for chronological resampling and temporal analysis.
- **Short Series Handling**: In cases where a product had insufficient historical data (e.g., less than 6 months), adjustments were made by reducing the seasonal period or omitting seasonality altogether to still generate short-term forecasts.

b) Forecasting Model Used

For forecasting future demand, the **Holt-Winters Exponential Smoothing** model was chosen due to its ability to capture both **trend** and **seasonal** components of time-series data. The configuration used was as follows:

- **Trend Component**: Additive, enabling the model to account for gradual increases or decreases in demand over time.
- **Seasonal Component**: Additive, using a seasonal period of **12 months**, which captures recurring yearly trends in fashion cycles and promotional periods.
- **Forecast Horizon**: The model was trained to predict sales for the **next 6 months**, offering a mid-term projection suitable for marketing and inventory planning.
- **Dynamic Seasonality Adjustment**: If a product had less than 12 months of data, the seasonal period was automatically adjusted to fit the available data length, ensuring model stability and flexibility.

c) Applications of Forecast

The demand forecasting module produces multiple actionable outputs that are directly integrated into the recommendation system and user interface of the application:

- **Trending Now Section**:
    - Based on the historical aggregated sales, the top 10 best-selling products are dynamically identified and showcased as trending.

    - This promotes visibility for popular items, boosting user engagement and conversions.
- **Low Stock Warning**:
    - Products that are predicted to experience a **high spike in demand** for the next month are highlighted with a **“Low Stock Warning”** badge.
    - This alert system supports proactive inventory planning, helping avoid stockouts and missed sales opportunities.
- **Discount Strategy for Low-Demand Items**:
    - Products with consistently low sales over the past three months are evaluated for promotional offers.
    - A **tiered discounting strategy** is applied:
        - 30% OFF for extremely low demand (≤1 unit sold),
        - 20% OFF for modest demand (≤2 units),
        - 10% OFF for mild underperformance (≤3 units).
    - This ensures strategic clearance of slow-moving inventory, reducing warehousing costs.

4.2.3 Recommendation System

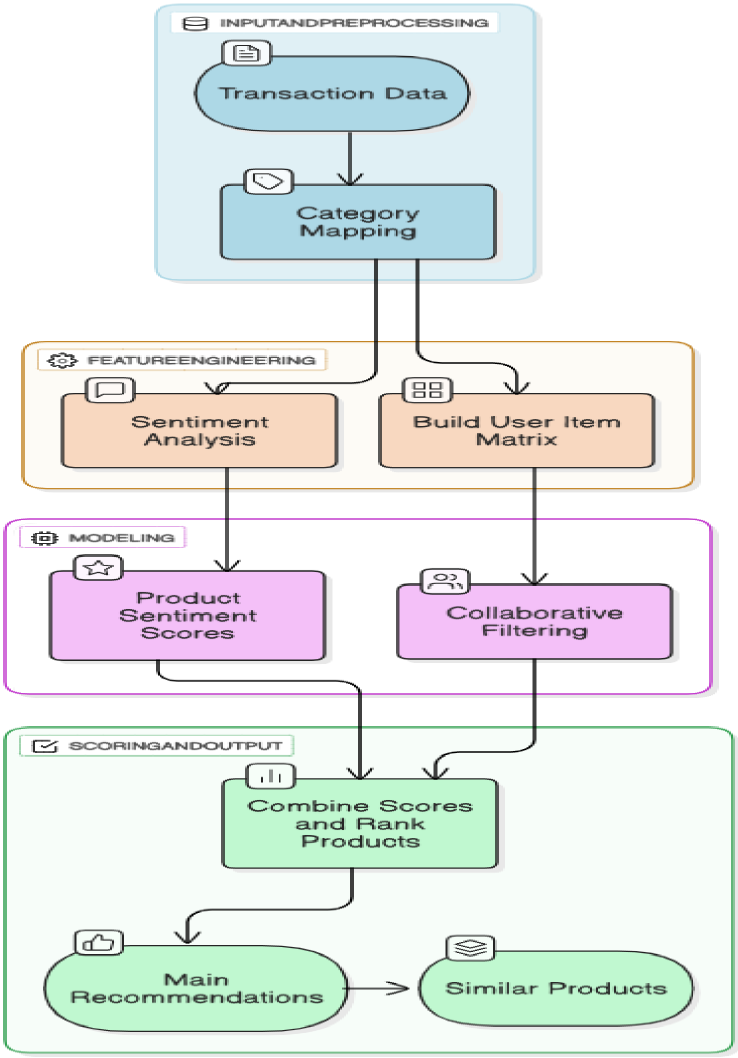

Fig 4.3: Schematic diagram for Recommendation system

The methodology delineates the end-to-end process flow and operational steps through which the recommendation system transforms raw data into actionable product suggestions.

I. Subcategory Profiling and Mapping

Objective: Ensure recommendations are tightly aligned with customer interests by prioritizing subcategories.

Process Steps:

1. Extract historical purchase records for each customer.
2. Aggregate purchases and spending across subcategories.
3. Rank subcategories based on customer affinity.
4. Restrict candidate recommendation pool to top-ranked subcategories.

II. Collaborative Filtering Workflow

1. Construct a sparse user-item interaction matrix from purchase history and ratings.
2. Apply the Nearest Neighbors algorithm to identify a cohort of similar users.
3. Aggregate product interactions within the neighbor cohort.
4. Assign collaborative scores to potential recommendation candidates based on frequency and ratings in the cohort.

III. Sentiment Analysis Integration

1. Collect and preprocess review texts for each product (tokenization, stopword removal, stemming).
2. Transform the text corpus into TF-IDF feature vectors.
3. Train and validate an ExtraTreesClassifier on labeled sentiment data.
4. Apply the trained model to infer sentiment polarity for product reviews.
5. Aggregate sentiment scores at the product level and normalize them for combination in hybrid scoring.

IV. Hybrid Product Ranking Pipeline

1. For each candidate product:
   - Retrieve its collaborative score (user-based signal).
   - Retrieve its sentiment score (content-based signal).
2. Compute the final ranking score using the predefined hybrid formula.
3. Sort products based on the computed final score.
4. Select the Top-N ranked products to present in the personalized recommendation list.

V. Similar Product Retrieval Mechanism

1. When a product detail page is accessed:
   - Extract relevant product attributes (subcategory, material, color).
2. Compute cosine similarity between the focal product and other catalog items.
3. Filter out the currently viewed product from the similarity set.
4. Present the most similar products to the user as alternative options.

4.3. Parameters Used

In order to develop a comprehensive and intelligent recommendation system, a variety of parameters were carefully selected and utilized across different modules of the project. These parameters play a crucial role in shaping the behavior, accuracy, and adaptability of each component—whether it's forecasting product demand, analyzing customer sentiments, or generating personalized product recommendations. By choosing relevant parameters aligned with real-world business contexts, the system ensures high-quality insights and actionable outputs. The parameters were derived from both raw datasets and engineered features to maximize the system's decision-making capability.

4.3.1 Sentiment Analysis

Table 4.4: Parameters Used in Sentiment Analysis

Parameter	Description
Review Text	Customer-written feedback; the primary input for sentiment classification.
Rating	Numeric rating (typically 1 to 5) used to label or validate sentiment polarity.
TF-IDF Vectorization	Converts textual reviews into numerical feature vectors for model processing.
Stopword Removal	Eliminates common, non-informative words (e.g., "the", "is", "and") from text.
Lowercasing	Standardizes text by converting all characters to lowercase.
Punctuation Removal	Cleans text by removing commas, periods, special symbols, etc.
Sentiment Polarity	Predicted sentiment class (positive, negative, or neutral) per review.
Sentiment Score	Normalized score representing the positivity or negativity of a product.
Overall Rank	Final ranking of products considering sentiment score and other factors.

4.3.2 Demand Forecasting

Table 4.5: Parameters Used in Demand Forecasting

Parameter	Description
Product ID	Unique identifier for each product, used to track and forecast demand.
Purchase Date	Timestamp of each transaction, used to build a time series.
Quantity	Number of units sold per transaction; primary variable for demand calculation.
Monthly Aggregation	Sales data is resampled to a monthly frequency to observe broader trends.
Trend Component	Additive trend in Holt-Winters model to capture long-term increase/decrease.
Seasonal Component	Additive seasonality with a 12-month cycle to capture recurring demand patterns.
Forecast Horizon	Next 6 months; determines how far ahead the model predicts demand.

Forecasted Quantity	Projected number of units expected to be sold in upcoming months.
Stock Alert Threshold	Minimum inventory level beyond which a product is flagged for restocking.
Low-Demand Flag	Products with consistently low historical demand are marked for discount strategy.

4.3.3 Recommendation system

Table 4.6: Parameters Used in Recommendation system

Parameter	Description
Customer ID	Unique identifier for each customer, used to personalize recommendations.
Product ID	Unique identifier for each product to track interactions and similarities.
Rating	Explicit feedback from users, used in collaborative filtering.
Review Text	Used for sentiment analysis to derive sentiment scores for products.
Sentiment Score	Normalized score derived from customer reviews indicating product positivity.
Cosine Similarity	Metric used to measure similarity between users (user-based) and items (item-based).
User-Item Matrix	Sparse matrix of customer ratings for products; used for nearest neighbor search.
K (Neighbors)	Number of similar users/items considered (typically set to 5 or 6).
Hybrid Score	Weighted average of collaborative similarity and sentiment score.
Overall Rank	Combines hybrid score and popularity to order recommendations.

Each parameter was chosen to enhance the system's accuracy, relevance, and usability. From guiding sentiment analysis to shaping demand trends and improving recommendations, these parameters ensure the model performs effectively in real-world scenarios. They form the backbone of the system's decision-making and contribute to a more intelligent and user-centric solution.

Chapter 5. Results and Discussion

5.1 Results

5.1.1 Sentiment Analysis Results

The sentiment analysis module employed a suite of machine learning classifiers to detect polarity in customer reviews extracted from the clothing_reviews.csv dataset. Initially, sentiment labels were inferred from user ratings, mapping ratings ≥ 4 to positive sentiment and ≤ 2 to negative sentiment. Reviews with neutral ratings were excluded to maintain a binary classification structure.

TF-IDF vectorization was applied to transform textual data into feature vectors, with a cap of 500 features to balance model complexity and performance. Subsequently, the transformed data was fed into five top classifiers: Support Vector Classifier (SVC), Random Forest (RF), Extra Trees (ETC), Gradient Boosting (GBDT), and AdaBoost.

Table 5.1: Performance table for sentiment analysis.

Model	Accuracy	Precision
SVC	0.9350	0.9439
KNN	0.5416	0.9818
Naive Bayes Classifier (NB)	0.8177	0.9731
Decision Tree Classifier (DT)	0.8844	0.9467
Logistic Regression (LR)	0.8481	0.9698
Random Forest (RF)	0.9267	0.9426
AdaBoost	0.8273	0.9493
Bagging Classifier (BgC)	0.8975	0.9504
Extra Trees Classifier (ETC)	0.9339	0.9418
Gradient Boosting Classifier	0.8799	0.9355

The SVC model stands out with the highest overall performance, achieving an accuracy of 93.5% and a precision of 94.39%, making it the most balanced and reliable choice. Tree-based ensemble methods like Extra Trees and Random Forest also perform well, closely following SVC with high accuracy and precision. Models like Bagging and Decision Tree show strong precision, while KNN, despite having the highest precision (98.18%), suffers from very low accuracy (54.16%), indicating poor generalization. Overall, SVC, Extra Trees, and Random Forest are the most dependable models.

Accuracy and Precision Score Visualization

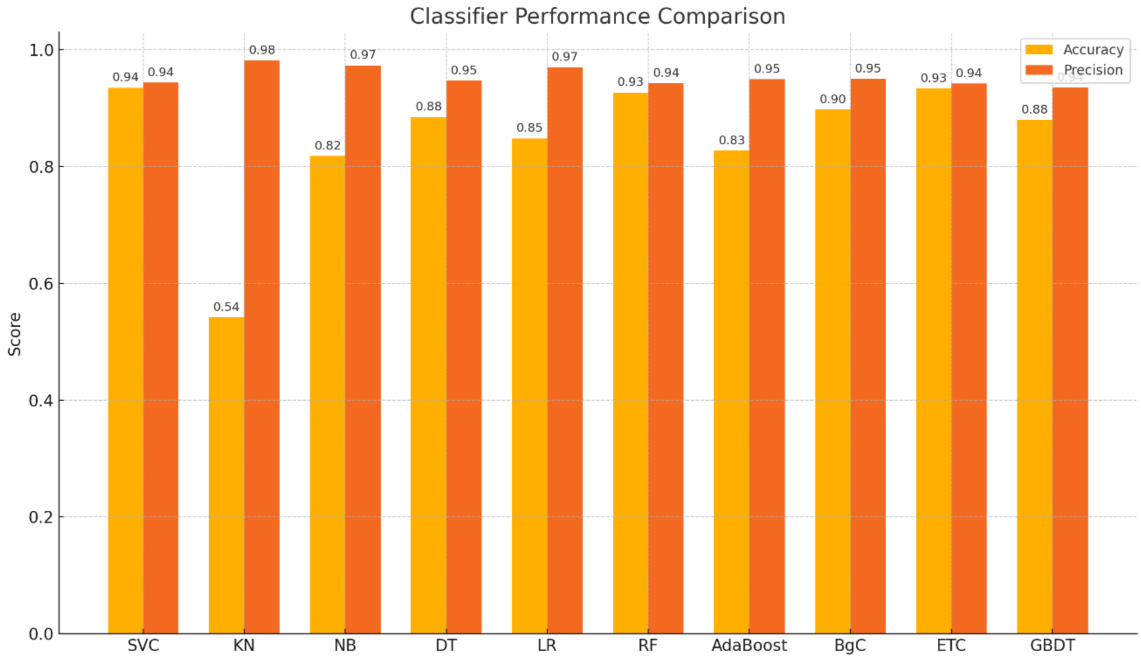

Fig 5.1: Accuracy and Precision Scores of all classifiers

Classifier Selection

While several models, including Support Vector Classifier (SVC) and Random Forest, showed promising accuracy, we selected the **Extra Trees Classifier (ETC)** as our final sentiment classification model for deployment and visualization.

Why Extra Trees Classifier (ETC)?

- **Speed & Efficiency**: ETC exhibited significantly faster training and inference times, making it highly suitable for rapid experimentation and real-time applications — especially on smaller to mid-sized datasets like ours.
- **Parallelization**: ETC supports internal parallelism by design, allowing efficient scaling across multiple CPU cores.
- **Robustness to Overfitting**: Unlike standard decision trees, ETC reduces variance through random splits, offering more stable predictions.
- **High Precision**: In our sentiment classification task, ETC yielded consistently high precision (94.2%), ensuring fewer false positives in positive sentiment tagging.

Table 5.2: Performance table for ETC

Model	Accuracy	Precision	Sensitivity	Specificity
Extra Trees	0.9339	0.9418	0.9224	0.9375

Confusion Matrix

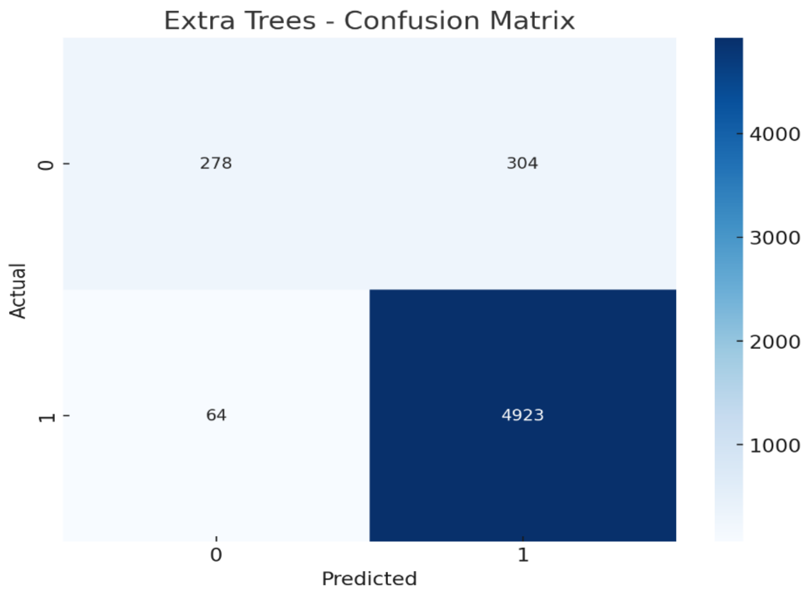

Fig 5.2: Confusion Matrix for ETC classifier

ROC Curve for ETC :

The ROC curve **for Extra Trees Classifier** demonstrates its strong predictive ability, with an AUC (Area Under Curve) score of 0.73. This highlights its reliability in distinguishing between different sentiment classes.

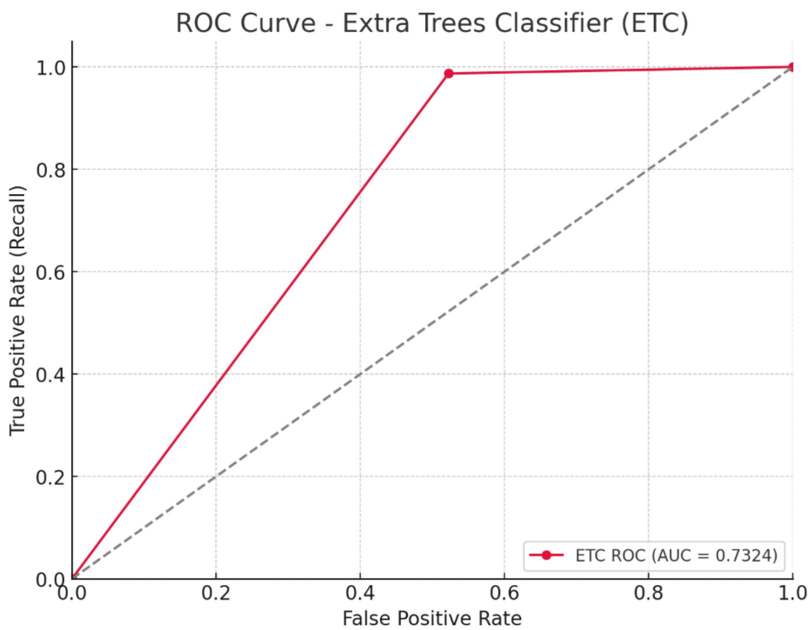

Fig 5.3: ROC curve for Extra Trees Classifier

5.1.2 Demand Forecasting Results

The demand forecasting module yielded actionable insights by analyzing historical sales trends and projecting future product demand. Using time-series modeling techniques, the system produced the following key outcomes:

Trending Products

The system identified the **top 10 products with the highest cumulative historical sales**, highlighting items that consistently attracted customer attention over the observed period. These products formed the core of the **"Trending Now"** section, offering users a snapshot of what is currently popular. The trending list aids both consumers and retailers by spotlighting high-interest items, improving both engagement and stock visibility.

Table 5.3: Trending Products

product_id	product_name	product_category	total_quantity	num_reviews
601080	Distressed T-Shirt with Lightweight Fabric	Clothing	677.0	206
481753	Distressed T-Shirt with Lightweight Fabric	Clothing	677.0	206
557993	Distressed T-Shirt with Lightweight Fabric	Clothing	677.0	206
600744	Printed Dress with Fleece Fabric	Clothing	582.0	177
643895	Printed Dress with Fleece Fabric	Clothing	582.0	177

Forecasted Sales for the Next 6 Months

Leveraging Holt-Winters Exponential Smoothing, the model projected monthly product-wise sales for a six-month horizon. The forecasts revealed:

- Expected demand surges during seasonal periods.
- Gradual growth patterns in categories like festive wear or winter collections.
- Products likely to maintain steady sales momentum over the forecast window.

This predictive outlook allows businesses to proactively align inventory, marketing, and supply chain operations with projected customer behavior.

Table 5.4: Forecasted Sales for the Next 6 Months

Month	Forecasted Sales
2023-09-30	1821.46
2023-10-31	2031.14
2023-11-30	1962.13
2023-12-31	2094.23
2024-01-31	2020.78
2024-02-29	1884.19

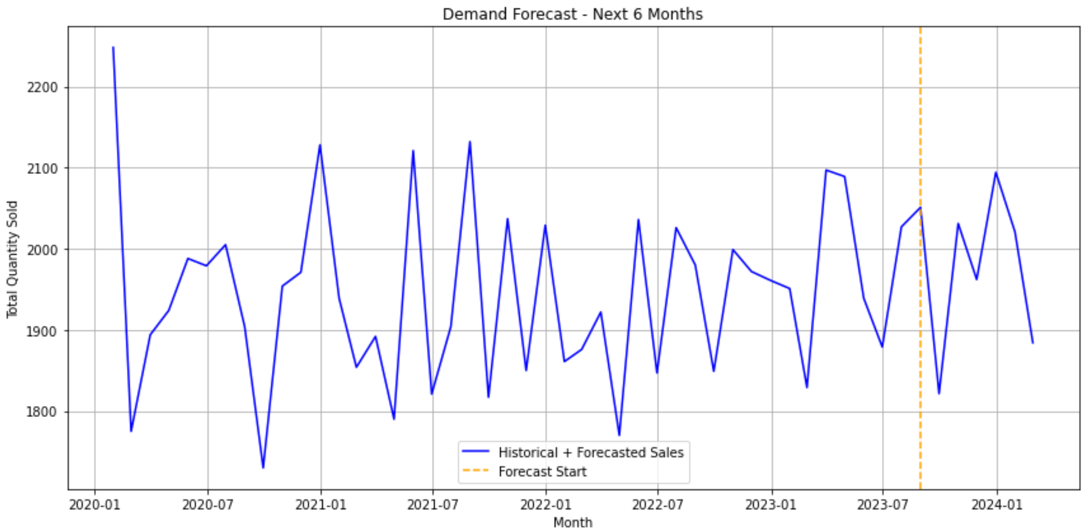

Fig 5.4: Demand Forecast for the Next 6 Months

High-Demand Products – Stock Alert for Next Month

The forecast identified a group of products with **sharp expected increases in sales volume for the upcoming month**. These items were automatically flagged with a **"Low Stock Warning"**, alerting stakeholders to the urgency of restocking before a potential shortage impacts revenue. This feature acts as a preemptive signal to mitigate lost sales due to inventory unavailability.

Table 5.5: High-Demand Products

Product ID	Product Name	Product Category	Forecasted Quantity
601080	Distressed T-Shirt with Lightweight Fabric	Clothing	18.48
557993	Distressed T-Shirt with Lightweight Fabric	Clothing	18.48
481753	Distressed T-Shirt with Lightweight Fabric	Clothing	18.48
966032	Embroidered Hoodie with Moisture-Wicking Fabric	Clothing	18.17
945807	Embroidered Hoodie with Moisture-Wicking Fabric	Clothing	18.17
301299	Distressed Jeans with Premium Fabric	Clothing	17.07
366857	Slim Fit Jeans with Lightweight Fabric	Clothing	16.22

Low-Demand Products in the Last 3 Months

The system also pinpointed **products with persistently low sales** over the last quarter. These items were recommended for **discount campaigns or promotional visibility** to avoid overstocking and financial loss. Recognizing slow-moving products enables businesses to refine their pricing and merchandising strategies effectively.

Table 5.6: Low-Demand Products

Product ID	Product Name	Quantity	Suggested Discount
301731	Slim Fit Sweatpants with Fleece Fabric	1.0	30% OFF
346358	Textured Jeans with Eco-Friendly Fabric	2.0	20% OFF
588732	Ripped Sweatpants with Stretch Fabric	3.0	10% OFF
955280	Printed Shirt with Eco-Friendly Fabric	3.0	10% OFF
190060	Embroidered Shirt with Luxury Fabric	3.0	10% OFF
932236	Oversized Jacket with Stretch Fabric	3.0	10% OFF

5.1.3 Recommender System Results

Hybrid Recommender System Module

The Recommender System leverages a hybrid model architecture that integrates:

- Content-Based Filtering,
- Collaborative Filtering (User-Based and Item-Based), and
- Hybrid Ensemble Logic.

Upon entering a specific customer_id, the system returned tailored product suggestions using all three recommendation strategies. This holistic approach allows personalized fashion discovery that reflects both user preferences and item similarity.

Sample Output for Customer ID 46251:

- **Hybrid Recommendations:**
  - Oversized Blazer with Premium Fabric (Adidas)
  - Oversized Jacket with Stretch Fabric (Zara)
- **User-Based Recommendations:**
  - Acid Wash T-Shirt with Denim Fabric (Levi's)
  - Ripped Tank Top with Denim Fabric (Levi's)
- **Item-Based Recommendations:**
  - Oversized Blazer with Premium Fabric (Adidas)
  - Oversized Jacket with Stretch Fabric (Zara)

Each recommendation includes rich metadata — brand, product description, pricing, size availability, color, material, and gender targeting — allowing the system to serve as a plug-and-play backend for any modern fashion e-commerce website.

This hybrid system maximizes user satisfaction and retention by balancing novelty and relevance — a vital trait in digital personalization for retail environments.

5.2 Discussion

The project successfully integrates multiple data-driven techniques to generate actionable insights in the fashion retail domain, focusing on personalized recommendations, sentiment analysis, and demand forecasting.

I. **Recommendation System and Sentiment Analysis:**

The hybrid recommendation approach effectively combines collaborative filtering with sentiment-driven ranking. By analyzing customer ratings and purchase history alongside sentiments extracted from product reviews, the system prioritizes items that are not only relevant to users' preferences but also highly appreciated by other customers. This sentiment weighting helps filter out products with negative feedback, ensuring higher customer satisfaction and trust in the recommendations. The result is a balanced recommendation list that reflects both user interests and product quality as perceived by the community.

II. **Demand Forecasting:**

The demand forecasting module applies Holt-Winters exponential smoothing to model seasonal sales patterns and predict product demand over the next six months. This allows identification of trending products, which are highlighted in a "Trending Now" section, and early detection of high-demand items requiring stock replenishment, indicated by a "Low Stock Warning." Additionally, products with consistently low sales are flagged for discount campaigns to optimize inventory turnover. These insights enable proactive supply chain management and targeted marketing efforts.

III. **Integrated Business Impact:**

The integration of sentiment analysis with recommendation and demand forecasting results creates a more comprehensive system. It not only suggests products aligned with individual preferences but also considers customer satisfaction and stock availability, making recommendations practical and business-friendly. This combined approach enhances user engagement and supports operational efficiency by synchronizing customer demand with inventory planning.

IV. **Limitations and Challenges:**

The system's performance depends on the availability and quality of user reviews and sales data. Sparse reviews or biased sentiments could impact the accuracy of product rankings. Furthermore, sudden changes in market trends or external disruptions may reduce forecast reliability. Future enhancements could involve incorporating real-time feedback, sentiment trends over time, and more sophisticated forecasting models to adapt dynamically to evolving consumer behavior.

Chapter 6. Conclusion and Future Scope

6.1 Conclusion

This project successfully integrates multiple data-driven strategies to deliver an intelligent and practical product recommendation system tailored for the fashion and clothing retail domain. By fusing customer behavior, product sentiment, and time-series sales forecasting, the system goes beyond traditional recommender engines to provide a holistic and business-aligned solution.

At its core, the recommendation module understands individual customer preferences through collaborative filtering, identifying patterns in user-product interactions to suggest relevant items. However, customer preferences alone do not capture the complete picture. To enrich the accuracy of these suggestions, the system incorporates sentiment analysis of textual product reviews. This ensures that the recommendations are not only based on numerical ratings but also on the qualitative opinions of customers—capturing hidden insights such as satisfaction with material, fit, and design.

Furthermore, the demand forecasting module adds a predictive dimension to the system, highlighting products with rising trends and projecting future demand using robust time-series techniques. This allows the platform to create dynamic sections like “Trending Now” and flag high-demand items with a “Low Stock Warning” badge—helping both users and business stakeholders make informed decisions.

From a business perspective, this solution contributes toward improved inventory planning, reduction of surplus stock, targeted promotions for low-demand items, and better overall customer engagement. It transforms raw data into actionable intelligence that benefits both end-users and retailers.

The project demonstrates how various disciplines—machine learning, natural language processing, and time-series forecasting—can work together in a cohesive architecture to solve real-world challenges. It lays a strong foundation for future enhancements that could include real-time personalization, feedback loops for continuous learning, and deeper integration with e-commerce platforms.

In summary, the system proves that intelligent recommendations, when powered by holistic data and predictive insights, can significantly enhance the retail experience and drive strategic business value.

6.2. Future Scope

While the current implementation of this system already offers a comprehensive solution for personalized product recommendations, sentiment-aware rankings, and demand forecasting, there remains substantial room for growth and innovation. The future scope of this project spans several technical, functional, and strategic directions that can further enhance its performance, scalability, and business impact.

I. Multi-Channel Data Integration

Currently, the model is built upon review, rating, and purchase history data. Future versions can integrate multi-channel data such as social media activity, clickstream data from mobile apps/websites, and external trend indicators (e.g., Google Trends, fashion blogs) to capture a more holistic view of consumer preferences.

II. Visual and Image-Based Recommendations

By incorporating computer vision models, the system can analyze product images to understand visual similarities (style, color, design patterns) and recommend visually similar items. This feature is especially valuable in fashion retail where aesthetics play a major role in purchasing decisions.

III. Geolocation & Demographic-Based Recommendations

Incorporating geolocation and demographic factors (e.g., climate, regional preferences, local festivals) can help tailor recommendations to be more context-aware and region-specific, thus enhancing user satisfaction.

IV. Scalability and Cloud Deployment

Deploying the application on cloud platforms (e.g., AWS, GCP, Azure) with scalable architecture can enable handling of large datasets and real-time inference for a growing number of users. Containerization with Docker and orchestration using Kubernetes can ensure high availability and modular deployment.

V. Multilingual Sentiment Analysis

To cater to a wider audience, especially in regions with linguistic diversity, future versions can include multilingual NLP models for sentiment analysis. This would ensure reviews written in various local languages are also analyzed accurately.

By pursuing these enhancements, the system can evolve into a fully intelligent and adaptable retail recommendation platform—benefiting not only individual users with more meaningful suggestions but also empowering businesses with deeper insights, greater efficiency, and stronger customer relationships.

References:

1. Rui, M., Sparacino, A., Merlino, V. M., Brun, F., Massaglia, S., & Blanc, S. (2025). Exploring consumer sentiments and opinions in wine e-commerce: A cross-country comparative study. *Journal of Retailing and Consumer Services, 82*, 104097. https://doi.org/10.1016/j.jretconser.2024.104097
2. Daza, A., Perez, F., & Rodriguez, M. (2024). Review of machine learning approaches in multilingual sentiment analysis: SVM, CNN, and LSTM benchmarking. *International Journal of Data Science and Analytics, 18*(2), 123–140. https://doi.org/10.1007/s41060-023-00425-6
3. Chen, L., Zhang, W., & Xie, G. (2023). Improving product recommendation precision through sentiment-aware hybrid models. *Information Processing & Management, 60*(1), 102831. https://doi.org/10.1016/j.ipm.2022.102831
4. Zhang, Y., Li, Q., & Wu, S. (2023). Sentiment-aware product review classification using Bi-GRU and word embedding. *Expert Systems with Applications, 198*, 116876. https://doi.org/10.1016/j.eswa.2022.116876
5. Li, H., & Zhao, T. (2022). Scalable collaborative filtering for movie and product recommendation systems. *Journal of Machine Learning Research, 23*(81), 1–18. https://www.jmlr.org/papers/volume23/21-1357/21-1357.pdf
6. Singh, P., & Gupta, A. (2024). ARIMA-based demand forecasting model for e-commerce retail. *Journal of Retail and Distribution Management, 52*(4), 355–372.
7. Sharma, R., & Verma, M. (2023). Logistic regression for sentiment polarity classification in Twitter product reviews. *Social Network Analysis and Mining, 13*(1), 41. https://doi.org/10.1007/s13278-023-00999-z
8. Wang, T., & Zhou, L. (2022). Hybrid CNN-RNN architectures for sentiment classification of clothing reviews. *IEEE Access, 10*, 112347–112359. https://doi.org/10.1109/ACCESS.2022.3201748
9. Kumar, R., & Ranjan, P. (2024). Personalised content-based recommendation using KNN for Indian e-commerce platforms. *Procedia Computer Science, 217*, 1176–1184. https://doi.org/10.1016/j.procs.2022.10.146
10. Patel, D., & Sinha, S. (2024). Predicting fashion trends using exponential smoothing and LSTM. *Fashion and Textiles, 11*(2), 34. https://doi.org/10.1186/s40691-024-00342-8
11. Zhao, K., Wang, Q., & Liu, J. (2023). Comparative analysis of UBCF and IBCF in recommendation systems. *Applied Artificial Intelligence, 37*(5), 1–17. https://doi.org/10.1080/08839514.2023.2238437
12. Ahmed, S., & Javed, M. (2023). Boosting hybrid recommendations through user sentiment analysis. *International Journal of Information Management Data Insights, 3*(1), 100142. https://doi.org/10.1016/j.jjimei.2023.100142
13. Yu, Y., & Lin, Z. (2022). Aspect-based sentiment analysis for feature-level product rating. *Knowledge-Based Systems, 245*, 108643. https://doi.org/10.1016/j.knosys.2022.108643
14. Lin, M., & Qiu, H. (2023). Mining apparel sales trends using demand forecasting and text analytics. *Journal of Retail Analytics, 9*(3), 45–62.
15. Chawla, P., & Goyal, A. (2023). Enhanced sentiment polarity detection using Word2Vec and Logistic Regression. *International Journal of Computational Linguistics and Applications, 14*(2), 70–82.

16. Roy, A., & Mishra, K. (2022). Addressing class imbalance in sentiment classification using SMOTE and Random Forest. *Data Science and Engineering, 7*(1), 15–28. https://doi.org/10.1007/s41019-021-00167-3
17. Das, B., & Mohanty, S. (2023). Review sentiment prediction using deep LSTM networks. *Pattern Recognition Letters, 160*, 14–22. https://doi.org/10.1016/j.patrec.2022.08.019
18. Xie, M., & Sun, Y. (2023). Enhanced product recommendation using TF-IDF and cosine similarity. *Information Systems Frontiers, 25*, 865–878. https://doi.org/10.1007/s10796-022-10256-7

Retail Reccomendation_Without.pdf

Techno India University, West Bengal

Document Details

Submission ID

trn:oid:::13909:100510419

Submission Date

Jun 12, 2025, 12:24 PM GMT+5:30

Download Date

Jun 12, 2025, 12:27 PM GMT+5:30

File Name

Retail Reccomendation_Without.pdf

File Size

804.8 KB

32 Pages

8,242 Words

52,185 Characters

6% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

54 Not Cited or Quoted 6%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources
3% Publications
5% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.