Recommendation System with Sentiment-Aware Ranking and Demand Forecasting

Presented by: Dipanjan Saha | Arkajyoti Sarkar | Debjit Chatterjee
Saikat Mukherjee | Srinjoy Dhar

Guided By: Dr. Lewlisa Saha

Plan of talk

- Introduction
- Literature Survey
- Problem Statement
- Objectives
- Methodology
- Results
- Conclusion
- References

Introduction

- Traditional recommendation systems often fail to capture the dynamic nature of customer preferences and market trends in fashion e-commerce.
- This project introduces an integrated framework combining user behavior, sentiment analysis, and demand forecasting to deliver smarter recommendations.
- By leveraging this multi-dimensional approach, the system enhances customer satisfaction, drives engagement, and ensures efficient inventory management.

Literature Survey

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

Literature Survey

Reference	Techniques Used (IT based)	Dataset Used	Performance Metric/Inference Drawn
Singh et al. [4]	Time-Series Forecasting, ARIMA	E-commerce sales transaction logs	Accurately forecasted product-wise demand over 6 months
Zhao et al. [5]	User-Based & Item-Based Collaborative Filtering	MovieLens, Amazon Reviews	UBCF outperformed IBCF in cold-start scenarios
Patel et al. [6]	Exponential Smoothing, LSTM	Fashion product sales logs	Predicted seasonal sales with 90% accuracy

Problem Statement

Research gap:

- Most systems ignore product reviews, missing customer satisfaction insights.
- Lack of demand analysis causes outdated or irrelevant recommendations.

Problem statement:

Develop a recommendation system that:

- Personalizes suggestions based on user history.
- Capture customer feedback from textual product reviews.
- Responds to shifting product demand and inventory constraints.

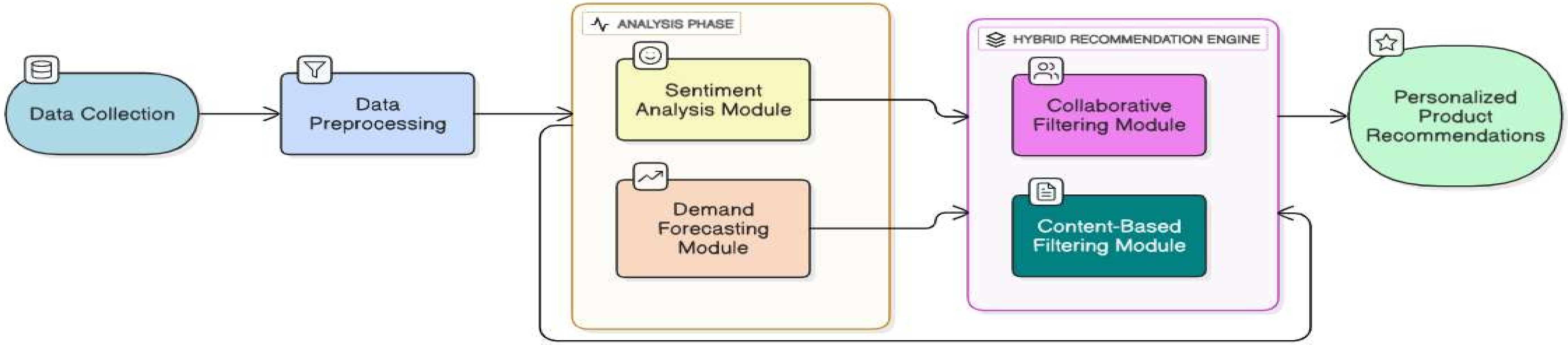
Objectives

- Develop a recommendation engine based on purchase history and ratings for personalized suggestions.
- Use sentiment analysis to refine recommendations with insights from positive reviews.
- Predict future demand with time-series analysis of sales trends.
- Prioritize trending, high-demand products in recommendations.
- Include features like “Trending Now” and “Low Stock Warning” for actionable insights.
- Validate the system with real-world data to improve relevance and satisfaction.
- Ensure scalability for adapting to new data and features.

Methodology Overview

Proposed Methodology

- Sentiment Analysis
- Demand forecasting
- Recommendation System



Data Collection

Source: Kaggle

Data Types:

- Clothing Review Dataset

Usage in Project:

- Rating-based recommendation system
- Sentiment analysis from review text
- 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.

	Customer ID	Customer Name	Customer Age	Gender	Purchase Date	product_id	Product Category	product_name	Quantity	Payment Method	Review Text	Rating	sub_category
0	46251.0	Christine Hernandez	37.0	Male	12-11-2020 13:13	778036.0	Clothing	Slim Fit Tank Top with Denim Fabric	1.0	PayPal	Absolutely wonderful - silky and sexy and comf...	4.0	Tank Tops
1	13593.0	James Grant	49.0	Female	05-05-2020 20:14	905147.0	Clothing	Fitted Jacket with Breathable Fabric	2.0	PayPal	Love this dress! it's sooo pretty. i happene...	5.0	Jackets
2	13593.0	James Grant	49.0	Female	05-05-2020 20:14	938121.0	Clothing	Fitted Jacket with Breathable Fabric	2.0	PayPal	Love this dress! it's sooo pretty. i happene...	5.0	Jackets
3	28805.0	Jose Collier	19.0	Male	31-03-2021 09:50	763149.0	Clothing	Textured T-Shirt with Cotton Fabric	1.0	PayPal	I had such high hopes for this dress and reall...	3.0	T-Shirts
4	28805.0	Jose Collier	19.0	Male	02-07-2020 02:54	708904.0	Clothing	Acid Wash Cargo Pants with Cotton Fabric	1.0	Credit Card	I love, love, love this jumpsuit. it's fun, fl...	5.0	Pants

Data Collection

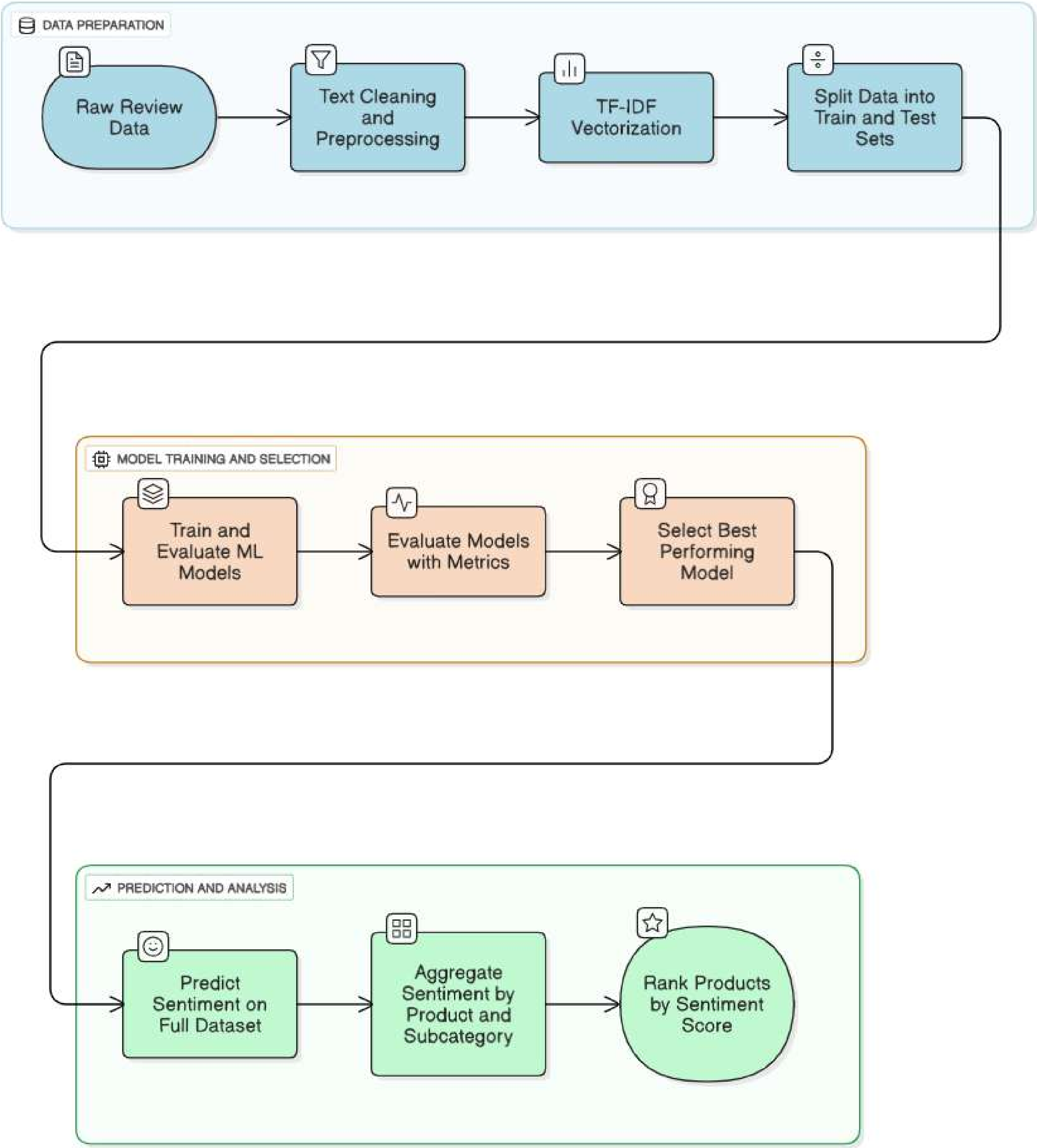
Source: Kaggle

Data Types:
- Clothing Description Dataset

Usage in Project:
- Content-based filtering using textual descriptions and attributes
- Enriching recommendation results with product metadata
- Highlighting product details in the recommendation UI

	product_id	product_name	product_brand	product_category	sub_category	product_description	size	color	price	material	gender
0	636938	Vintage Cargo Pants with Eco-Friendly Fabric	Versace	Clothing	Pants	Classic and modern design with premium materials.	XS, S, M, L	Green	7223.49	Wool	Unisex
1	923314	Printed Tank Top with Lightweight Fabric	Nike	Clothing	Tank Tops	Designed for both casual and formal occasions.	XS, S, M, L	Blue	3808.87	Cotton	Unisex
2	520072	Slim Fit T-Shirt with Eco-Friendly Fabric	Levi's	Clothing	T-Shirts	Trendy and versatile clothing piece.	S, M, L, XL	Pink	13756.42	Denim	Women
3	671796	Slim Fit T-Shirt with Cotton Fabric	Nike	Clothing	T-Shirts	Perfect fit with high-quality fabric.	M, L, XL	Blue	11130.30	Denim	Men
4	821136	Ripped Tank Top with Fleece Fabric	Zara	Clothing	Tank Tops	Perfect fit with high-quality fabric.	S, M, L	Pink	5640.68	Silk	Women

Sentiment Analysis Workflow Diagram



Sentiment Analysis

Steps:

1. **Text Preprocessing**
   - **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.

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.

Sentiment Analysis

Steps:

3. Model Selection:

- Logistic Regression, Multinomial Naive Bayes, Support Vector Machine (SVC), Decision Tree Classifier, Random Forest Classifier, K-Nearest Neighbors (KNN), Gradient Boosting Classifier, Extra Trees Classifier (ETC)
- 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.

4. Sentiment Scoring:

- 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.**

5. Product Ranking:

- Rank products globally by sentiment scores to highlight top-rated items reflecting strong customer satisfaction.
- Perform category-wise ranking to identify top-performing products within specific sub-categories.

Outcome:

- **Top N products** overall based on sentiment score.

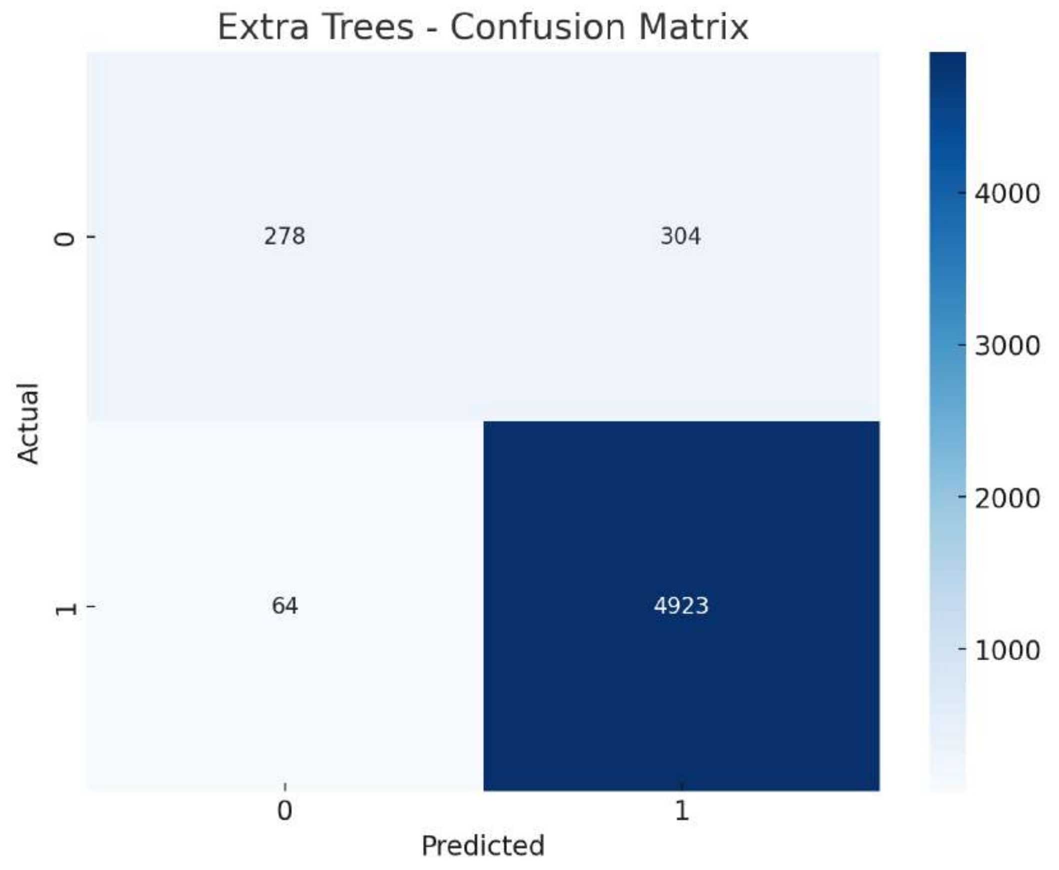
Sentiment Analysis Results

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

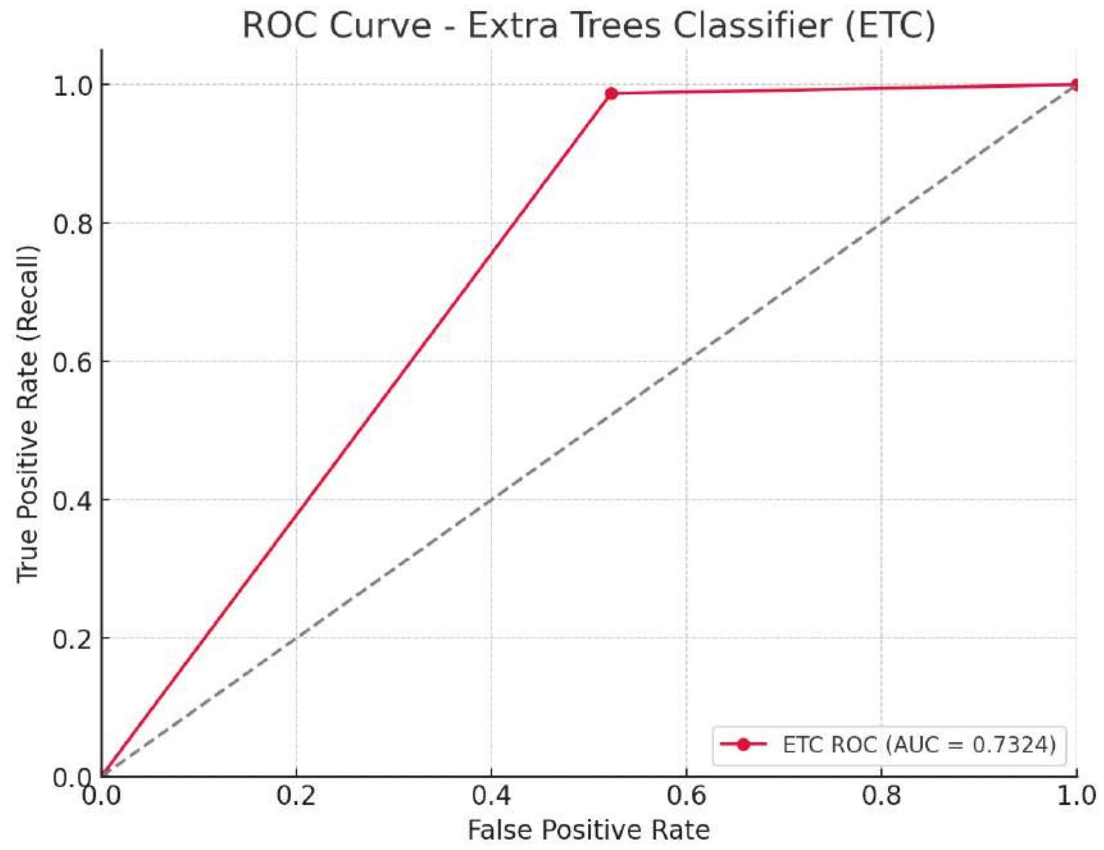
We chose the Extra Trees Classifier (ETC) for its speed and efficiency, fast training, parallel processing, and robustness against overfitting, making it ideal for real-time sentiment analysis. ETC also delivered high precision (94.2%), minimizing false positives in positive sentiment detection.

Sentiment Analysis Results

Confusion Matrix for ETC (Extra Tree Classifier)



ROC Curve for ETC



Performance table for ETC

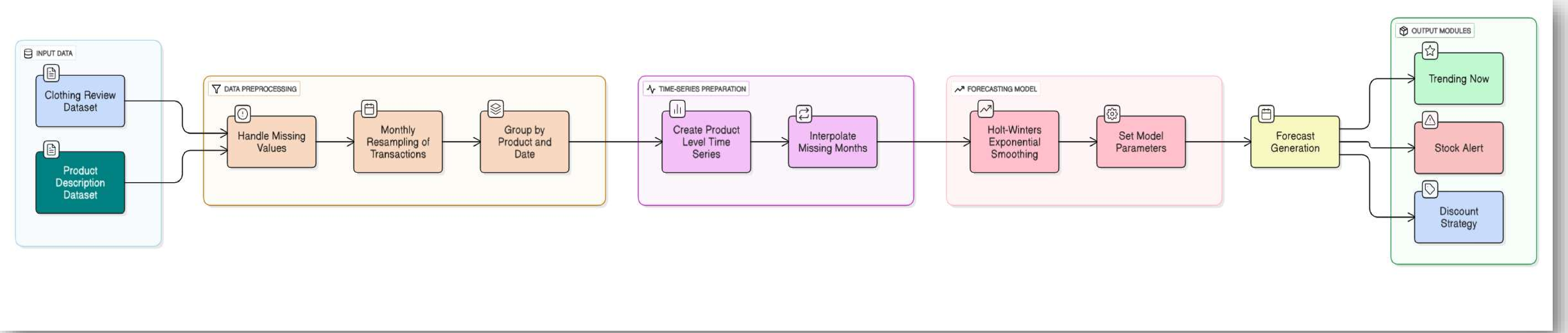
Model	Accuracy	Precision	Sensitivity	Specificity
Extra Trees	0.9339	0.9418	0.9224	0.9375

Sentiment-Based Product Rankings

The sentiment analysis was extended to rank products based on their sentiment scores within each category. The top products identified for each category are as follows:

Product ID	Product Name	Sentiment Score	Overall Rank
123902	Vintage Blazer with Luxury Fabric	0.9872	1
207193	Ripped Dress with Luxury Fabric	0.9747	2
840307	Textured Shirt with Premium Fabric	0.9730	3
224131	Acid Wash Hoodie with Cotton Fabric	0.9718	4
345402	Acid Wash Shirt with Premium Fabric	0.9701	5
955280	Printed Shirt with Eco-Friendly Fabric	0.9623	6
566768	Acid Wash T-Shirt with Stretch Fabric	0.9600	7
348872	Vintage Jacket with Eco-Friendly Fabric	0.9556	8
724429	Embroidered Cargo Pants with Moisture-Wicking Fabric	0.9551	9

Demand Analysis Workflow Diagram



Demand Forecasting Workflow

This workflow illustrates the complete pipeline used for demand forecasting. It begins with input data from clothing reviews and product descriptions, followed by preprocessing steps such as handling missing values, resampling, and grouping by product and date. A time-series is then created for each product, with missing months interpolated to ensure continuity.

Demand Analysis

Data Preprocessing:

- Aggregate sales monthly for trend clarity.
- Handle missing data with interpolation and short-series adjustments.

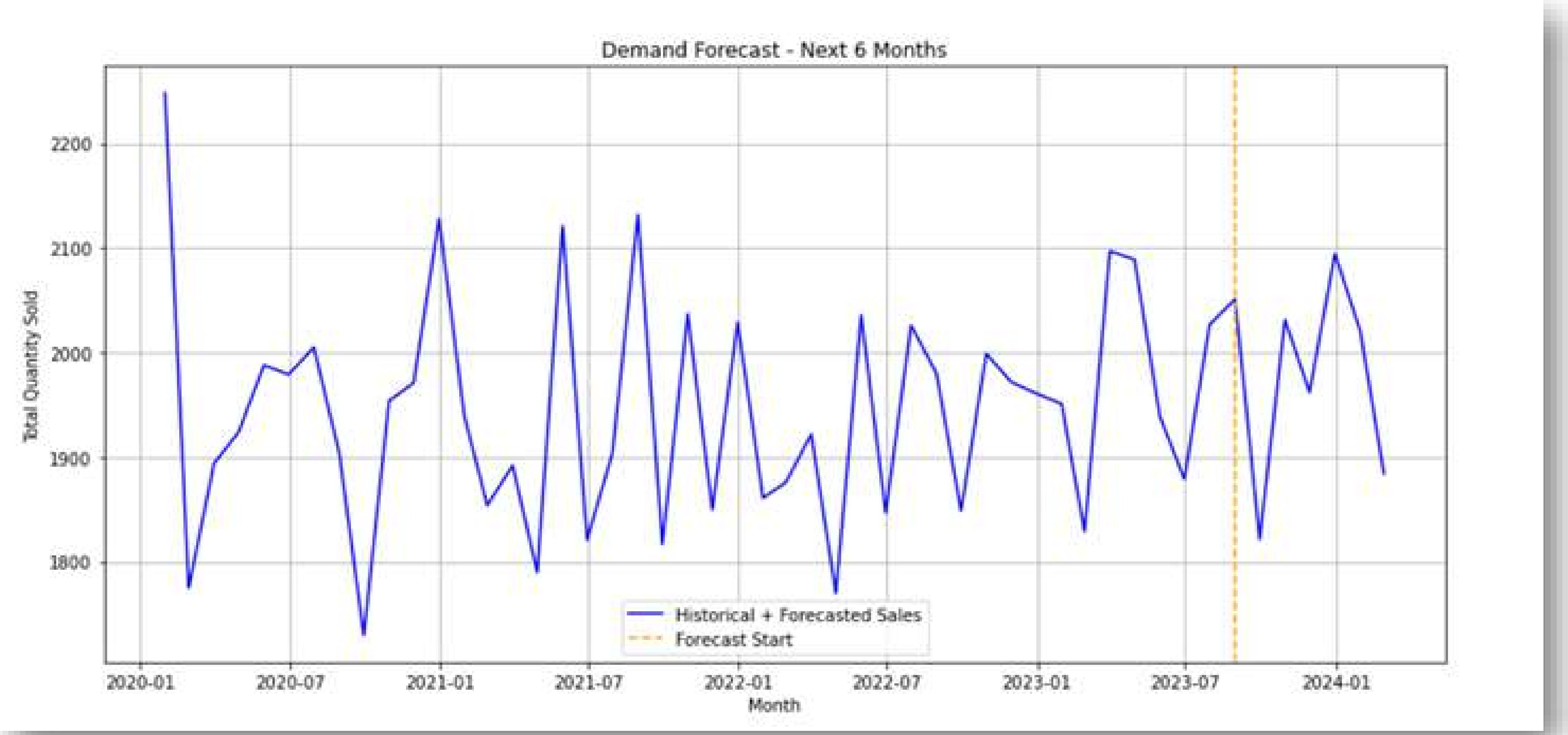
Forecasting Model:

- Use Holt-Winters Exponential Smoothing to capture trends.
- Predict 6-month demand with dynamic seasonal adjustments.

Applications:

- **Trending Now:** Showcase top 10 best-sellers.
- **Low Stock Warning:** Highlight high-demand items for proactive planning.
- **Discount Strategy:** Apply tiered discounts to low-demand products for clearance.

Demand Analysis Results



Plot showing historical and forecasted sales. The orange dashed line marks the forecast start point, highlighting projected demand trends till early 2024.

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

Forecasted Sales for the Next 6 Months
Projected monthly sales volume shows moderate seasonal variation, peaking in December 2023 and dipping slightly by February 2024.

Demand Analysis Results

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

High-Demand Products

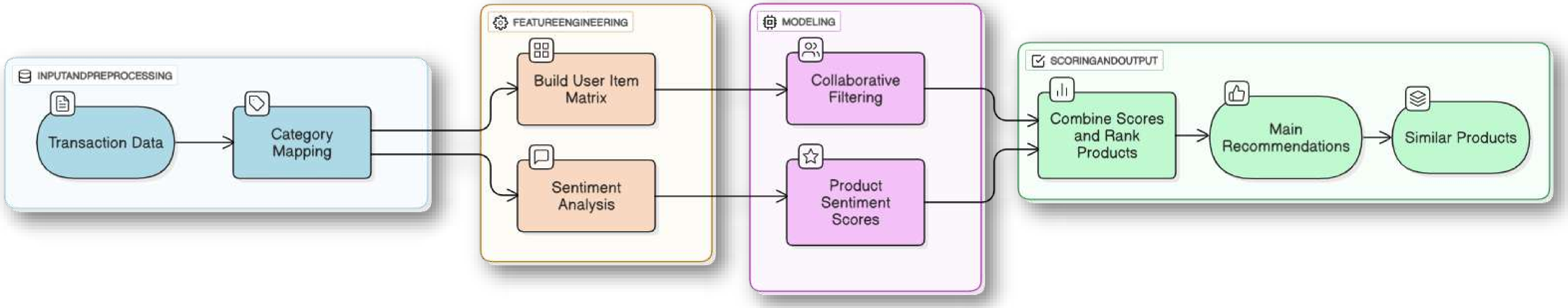
Top-performing clothing items predicted to have the highest demand, led by distressed t-shirts and embroidered hoodies.

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

Low-Demand Products

Low-selling items identified for clearance strategy with suggested discounts to boost sales and reduce inventory holding costs.

Recommendation Workflow



Product Recommendation Workflow

This workflow outlines the complete pipeline of the recommendation engine. It begins with transaction data preprocessing and category mapping, followed by feature engineering steps such as building a user-item interaction matrix and performing sentiment analysis. In the modeling phase, collaborative filtering and sentiment scores are used to generate product relevance. The scores are then combined and ranked to produce the final output, **personalized recommendations** and **similar product suggestions**—delivering both precision and variety in user experience.

Product Recommendation System

1. Collaborative Filtering

- Build a User-Item Rating Matrix using historical user ratings across products.
- Apply K-Nearest Neighbors (KNN) to predict how much a user might like unrated products based on similar users' preferences.

2. Content-Based Filtering

- Extract detailed product features such as category, brand, specifications, and descriptions.
- Use TF-IDF vectorization or one-hot encoding to transform these features into numerical form.
- Compute item-item similarity to recommend products similar to those the user liked before.

3. Hybrid Scoring

- Calculate a combined recommendation score by weighting collaborative and content-based results:

  final_score = 0.6 * collaborative + 0.4 * content
- Filter the top-N products per user based on this hybrid score to deliver personalized recommendations.

4. Sentiment Integration

- Incorporate average predicted sentiment scores derived from customer reviews for each product.
- Merge sentiment data with the top-N recommendations and re-rank products by descending sentiment score to prioritize products that customers genuinely like and trust.

Final Recommendations (Example)

Output for Customer ID 46251:

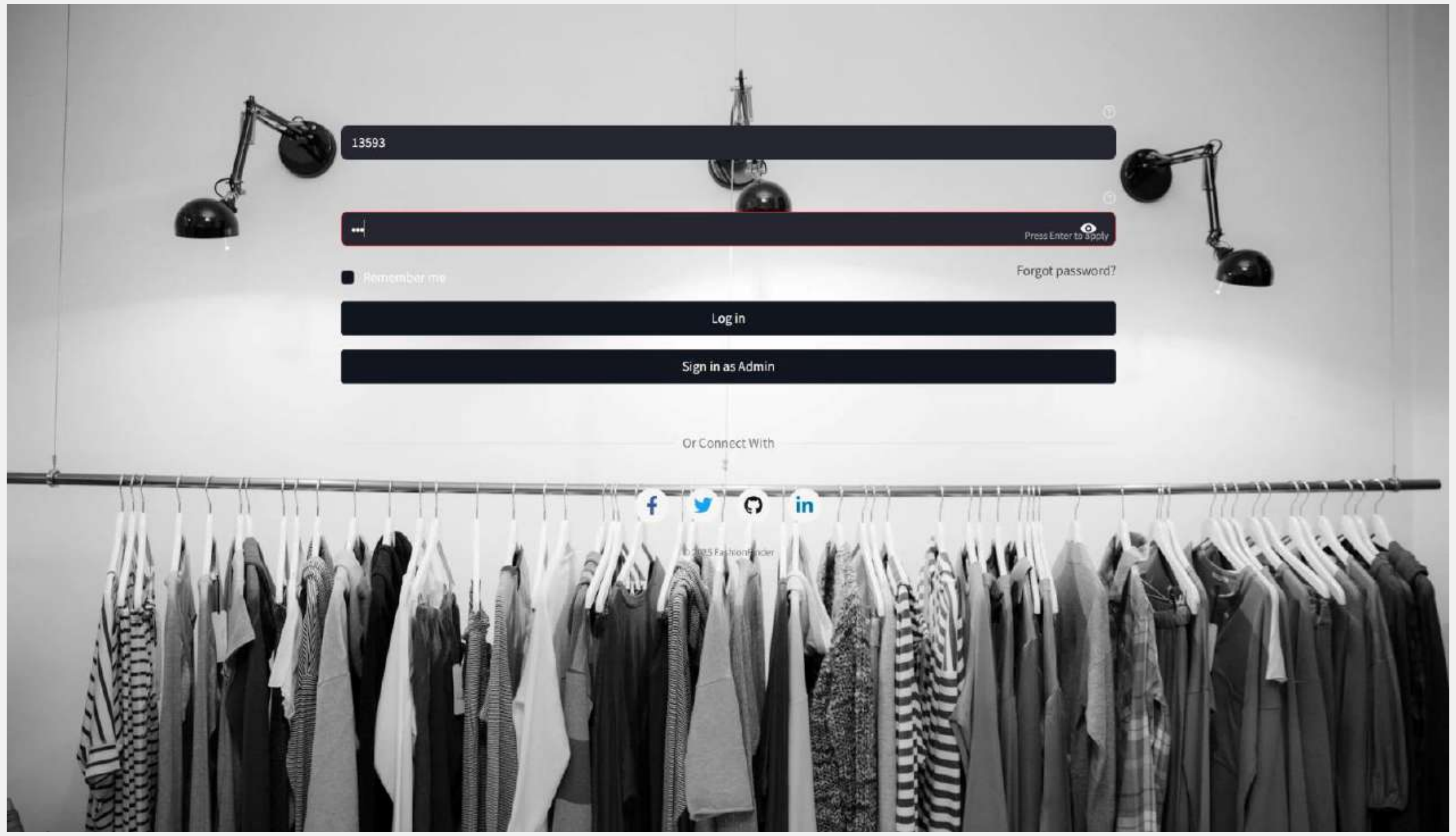
- **Hybrid Recommendations:**
  - Oversized Blazer with Premium Fabric (Adidas)
  - Oversized Jacket with Stretch Fabric (Zara)

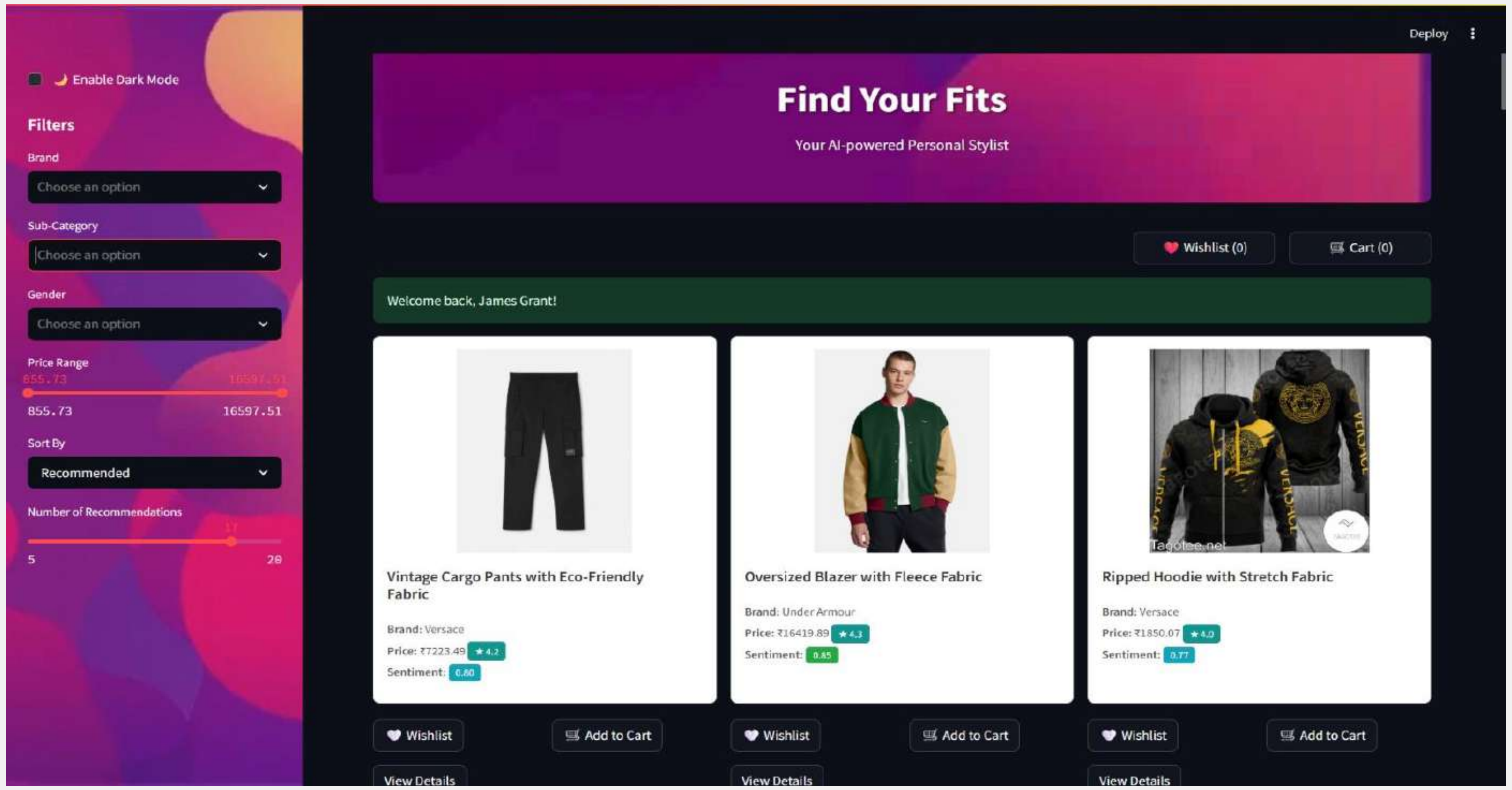
- **User-Based Recommendations:**
  - Acid Wash T-Shirt with Denim Fabric (Levi's)
  - Ripped Tank Top with Denim Fabric (Levi’s)

- **Item-Based Recommendations:**
  - Oversized Blazer with Premium Fabric (Adidas)
  - Oversized Jacket with Stretch Fabric (Zara)

In the final recommendations, it will show hybrid recommendations with ranking products by descending sentiment score to prioritize highly liked items.

Frontend Implementation





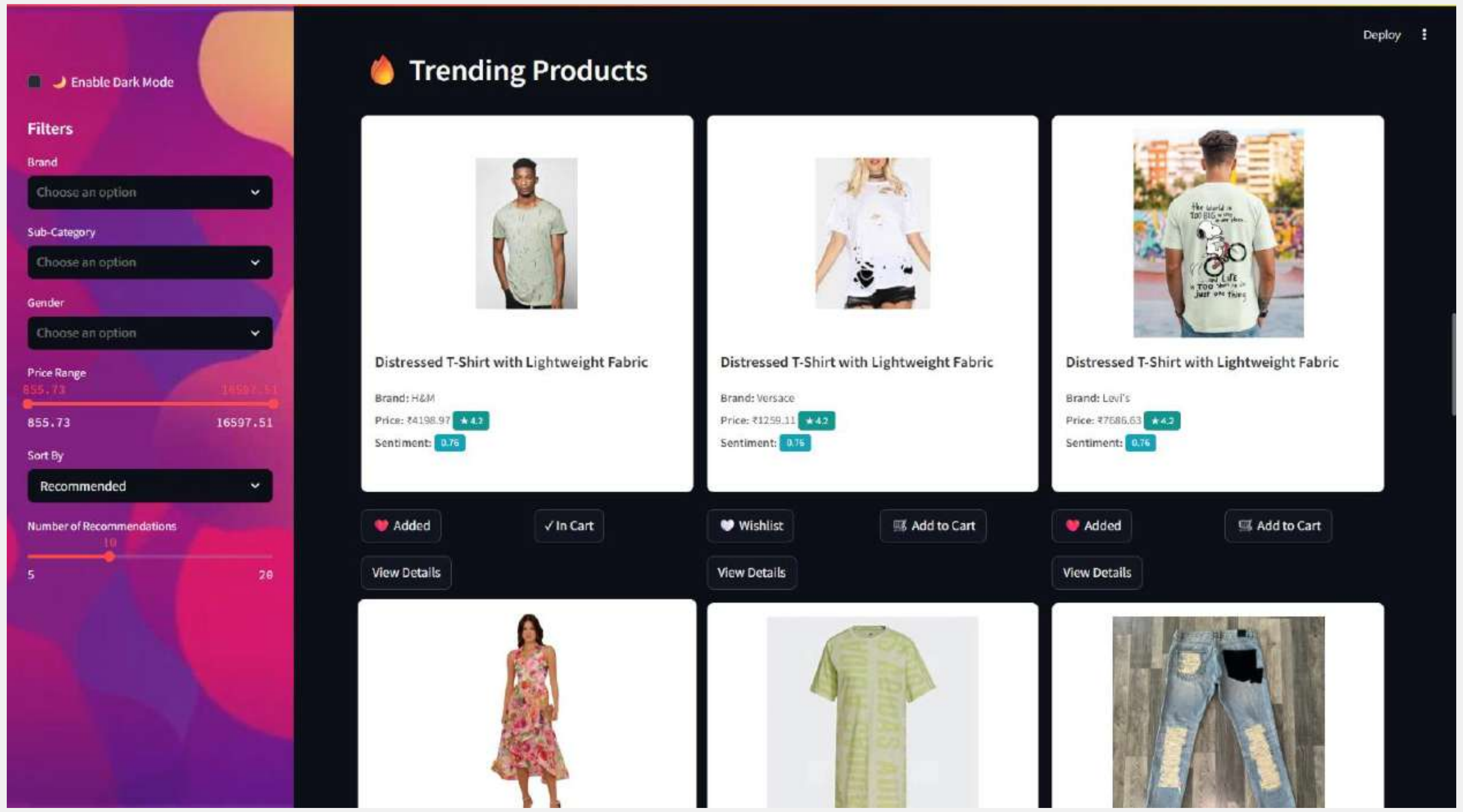
Login Page

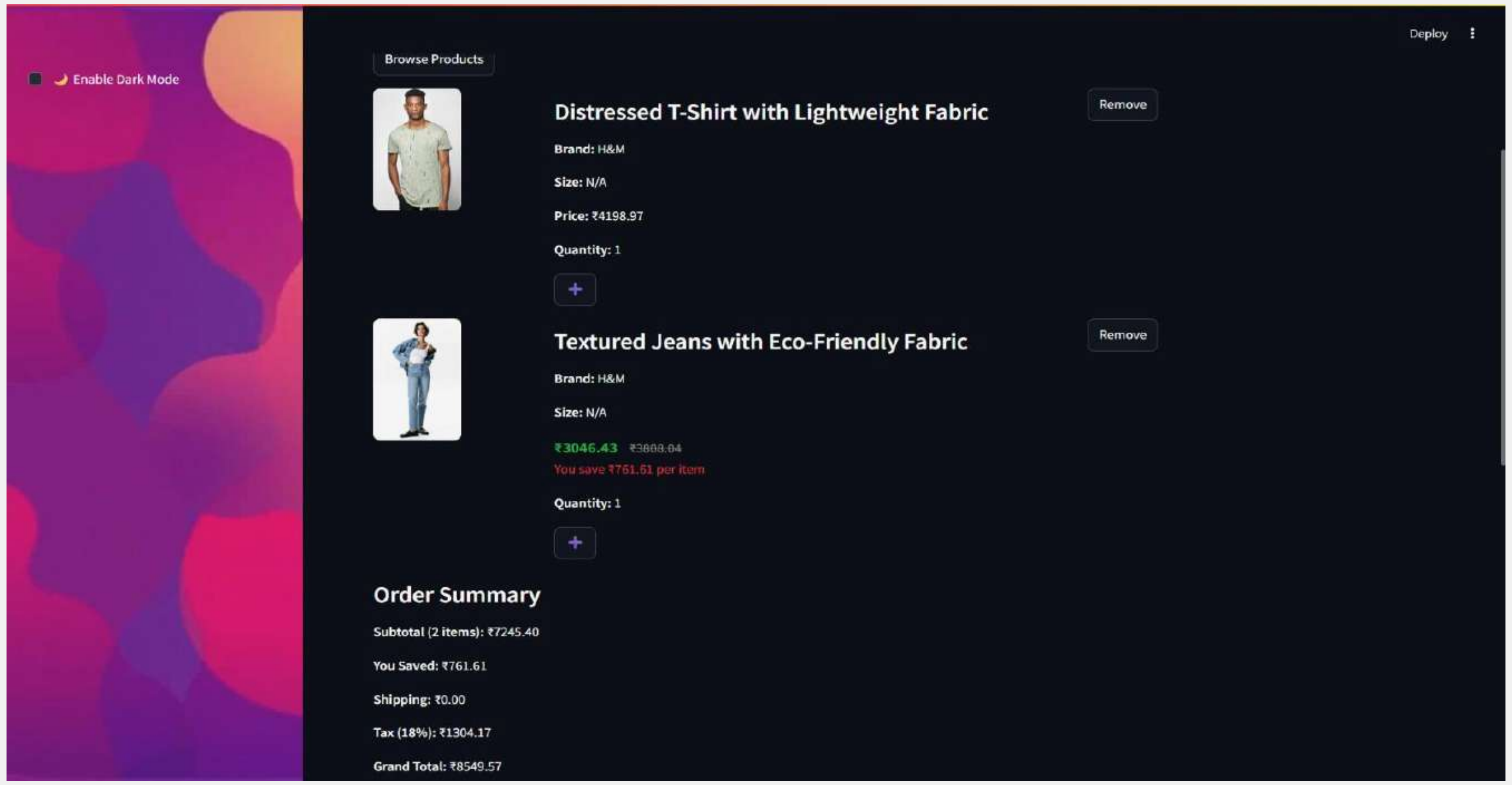
Clean and minimal login screen with social logins and admin access, styled for a fashion-first feel

Personalized Recommendations

Personalized outfit suggestions based on user data, with a friendly greeting and seamless shopping options.

Frontend Implementation





Trending Products
AI-curated trending items with filters, sentiment scores, and ratings. Supports quick cart and wishlist actions.

Cart & Order Summary
Cart view with item details, live price updates, and auto-calculated order summary (tax + discount).

Conclusion

- Integrated a **multi-strategy recommendation system** tailored for fashion & clothing retail.
- Combined **collaborative filtering** with **sentiment analysis** to understand customer preferences.
- Incorporated **time-series forecasting** to predict product demand and trends.
- Enabled features like:
  - "**Trending Now**" product sections
  - "**Low Stock Warning**" badges for high-demand items
- Benefits for businesses:
  - Improved **inventory planning** & **stock optimization**
  - Targeted **promotions for low-demand products**
  - Enhanced **customer engagement**
- Unified machine learning, NLP, and forecasting into a **cohesive architecture**.
- Foundation for future enhancements like:
  - **Real-time personalization**
  - Seamless **e-commerce integration**

Overall, delivers **data-driven insights** that elevate both user experience and business performance.

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. Singh, P., & Gupta, A. (2024). ARIMA-based demand forecasting model for e-commerce retail. *Journal of Retail and Distribution Management, 52*(4), 355–372.
5. 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
6. 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

Thank You

Welcoming questions to further explore the topic.