Sentiment Aware Search Engine with Recommendation System

Project submitted to the
SRM University – AP, Andhra Pradesh
for the partial fulfillment of the requirements to award the degree of

Bachelor of Technology
In
Computer Science and Engineering
School of Engineering and Sciences

Submitted by

RISHITHA PENTYALA

AP22110010054

CSE-R

Under the Guidance of

Dr. Bala Venkateswarlu

SRM University–AP

Neerukonda, Mangalagiri, Guntur

Andhra Pradesh – 522 240

[December, 2025]

TABLE OF CONTENTS

ABSTRACT

This dataset contains product information, ratings, and customer reviews for a variety of items listed on an e-commerce platform, including clothing, household goods, and small appliances. In this project, the dataset is analyzed to identify whether a review expresses a positive, negative, or neutral sentiment, enabling a deeper understanding of consumer opinions across different product categories. Using basic NLP techniques, the text is cleaned, preprocessed, and prepared for automatic sentiment detection. To evaluate different sentiment analysis tools, TextBlob, VADER, and Flair are applied to the review text, with Flair ultimately selected for the final system due to its higher accuracy and contextual understanding.

Instead of traditional machine-learning or deep-learning classifiers, the project focuses on building a functional sentiment-aware search engine. Exploratory analysis examines review sentiment distribution, commonly used words in positive and negative reviews, and the variation of sentiment across products. A fuzzy search algorithm is implemented to help users retrieve relevant products even when the search input contains misspellings or partial terms. When a product is retrieved, the system displays its reviews along with the percentage of positive, negative, and neutral sentiments derived from Flair's classification.

The recommendation module suggests products that have a larger proportion of positive reviews, thereby supporting users in making informed purchase decisions. Overall, this project presents a scalable and practical sentiment-aware search and recommendation system that enhances product discovery and offers valuable insights into customer satisfaction within e-commerce environments.

INTRODUCTION

Customer reviews play a crucial role in improving product quality, enhancing user satisfaction, and guiding business strategies across e-commerce platforms. With thousands of reviews generated for a wide range of products—including clothing, household items, electronics, and more—it becomes difficult and time-consuming for users or businesses to manually interpret customer opinions. Sentiment analysis provides an effective solution by automatically processing large volumes of review text and determining whether customers express positive, negative, or neutral sentiments toward a product.

This project aims to develop a sentiment-aware search engine and recommendation system using a general e-commerce review dataset. The workflow involves cleaning and preprocessing the review text, handling inconsistencies, and preparing the data for automated sentiment detection. Multiple sentiment analysis tools, including TextBlob, VADER, and Flair, are applied to the dataset to evaluate their performance. Among them, Flair is chosen for the final implementation due to its superior contextual accuracy and reliable polarity predictions. The resulting sentiment labels are incorporated into the enriched dataset for further analysis.

Rather than building machine-learning or deep-learning classification models, this project focuses on constructing a complete functional system. A fuzzy search algorithm is implemented to help users retrieve products even when search keywords contain spelling variations or partial matches. Once a product is selected, the system displays all associated reviews along with the percentage distribution of positive, negative, and neutral sentiments based on Flair's output. Additionally, the recommendation module suggests alternative products that have a higher number of positive reviews, providing users with meaningful guidance during the decision-making process.

Overall, this project demonstrates a practical end-to-end sentiment-driven retrieval and recommendation pipeline. It not only improves the way users search for products but also delivers insights into customer satisfaction, helping businesses and consumers better understand the sentiment landscape of e-commerce reviews.

OBJECTIVE

1. Data Cleaning And Preprocessing :

The first objective of this project is to clean and preprocess the e-commerce review dataset to prepare it for sentiment analysis and product search. This includes handling missing values, removing unwanted characters, normalizing text, and tokenizing product names and reviews to ensure consistency. The cleaned text forms the foundation for both the search engine and sentiment detection components of the system

2. Apply Multiple Sentiment Analysis Techniques and Select the Most Accurate Tool:

The next objective is to evaluate different sentiment analysis tools TextBlob, VADER, and Flair to determine which method best captures the polarity of customer reviews. This involves cleaning and preprocessing the review text, applying each tool to generate sentiment scores, and comparing their performance on real review samples. While TextBlob and VADER offer rule-based or lexicon-based approaches, Flair provides context-aware classification. Based on accuracy and reliability, Flair is chosen as the final sentiment model, and its sentiment labels (positive, negative, neutral) are added to the dataset for further system development.

3. Build a Sentiment-Aware Search Engine Using Fuzzy Matching:

An important part of the project is implementing a fuzzy search mechanism that allows users to find products even when their input contains spelling mistakes, partial names, or fuzzy matches. The search engine retrieves the closest matching product names and displays the corresponding reviews and sentiment summaries. This ensures a user-friendly experience and enhances the practical usability of the system in real-world scenarios.

4. **Develop a Sentiment-Driven Recommendation System:**

   The final objective is to recommend products based on sentiment patterns in customer reviews. Using the sentiment labels generated by Flair, products with a higher proportion of positive reviews are prioritized and suggested to the user as better alternatives. This sentiment-aware recommendation method helps users make more informed purchasing decisions and offers insights into customer satisfaction across similar product categories**.**

SYSTEM DESIGN

The main components of the system are:

i. Data Loading and Cleaning

The system begins by loading the e-commerce product dataset, which contains product names, prices, ratings, and customer reviews. Basic cleaning steps are applied, including removing duplicates, handling missing values, and normalizing text fields. All review text and product names are converted to lowercase and stripped of special characters to ensure consistency across the dataset.

ii. Text Preprocessing

Next, basic NLP preprocessing is performed on the review text and product names. This includes tokenization and optional stopword removal to simplify the text and prepare it for sentiment analysis and search. These steps ensure that the system can efficiently process and interpret product names and customer reviews.

iii. Sentiment Analysis Using TextBlob, VADER, and Flair

Three sentiment analysis tools—TextBlob, VADER, and Flair—are applied to each review to generate polarity scores. After evaluating the performance of each tool, Flair is selected for the final system due to its higher accuracy and better contextual understanding. The final dataset includes Flair-based sentiment labels (positive, negative, or neutral) for every review.

iv. Search Engine Implementation (Normal Search and Fuzzy Search)

Two search mechanisms are implemented:

- Normal Search:
  The system retrieves products whose cleaned and tokenized names contain the keywords entered by the user. This works effectively when the user enters a correct or close-to-correct product name.
- Fuzzy Search:
  For queries with spelling mistakes, partial words, or approximate matches, fuzzy string matching is applied. This increases search flexibility and ensures products can still be found even when the user's input is not perfectly accurate.

Both approaches help users efficiently locate products in the dataset.

v. Sentiment-Based Product Summary

When a user selects or searches for a product, the system displays the customer reviews along with a summary of sentiment distribution. Using Flair's predictions, the system calculates the percentage of positive, negative, and neutral reviews for that product. This allows users to quickly understand overall customer satisfaction.

vi. Recommendation System

A simple recommendation module suggests alternative products based on sentiment strength. Products with a higher number of positive reviews are ranked and recommended as better-rated options. This helps guide users toward higher-quality or more popular items.

vii. User Interaction and Output Module

Finally, the system provides an interactive interface where users can enter a product name, view its reviews, and see sentiment statistics. If the product is not found through normal matching, fuzzy search automatically assists in retrieving the closest match. The system then presents the reviews, sentiment percentages, and recommended alternatives.

IMPLEMENTATION

Data collection and pre-processing:

● The project begins by importing all required Python libraries and loading the e-commerce dataset, which contains product names, ratings, prices, and customer reviews.

● Missing values in the Review or Product Description fields are replaced with empty strings, and duplicate rows are removed to ensure clean, consistent data.

● All textual fields (product names and review text) are converted to lowercase and stripped of unnecessary symbols to prepare them for further processing.

Text Cleaning:

The goal of preprocessing is to transform raw, user-generated review text into a clean and structured form suitable for sentiment analysis and search operations. The cleaning steps include:

● Converting all text to lowercase

● Removing numbers, punctuation, and special characters

● Tokenizing sentences into individual words

● Optionally removing stopwords to reduce noise

● Normalizing whitespace and trimming unnecessary characters

This ensures that both review text and product names are consistent for later search and sentiment processing.

Sentiment Analysis (TextBlob, VADER, and Flair):

Three sentiment analysis methods are applied to the cleaned review text:

TextBlob – provides simple polarity scores

VADER – optimized for social-media style text

Flair – a deep learning–based contextual sentiment model

● After comparing outputs, Flair is selected as the final sentiment classifier due to its higher accuracy and stronger contextual understanding.

● Each review is assigned a positive, negative, or neutral label using Flair, and these labels are saved in the dataset for search and recommendation purposes.

Search System Implementation:

The system supports two types of product search mechanisms:

1. Normal Keyword Search

● Product names are cleaned, tokenized, and normalized.

● When the user enters a product name, the system checks whether any token from the input exists in the processed product name.

● If a match is found, all reviews and sentiment summaries for that product are displayed.

2. Fuzzy Search

● A fuzzy matching algorithm (using the FuzzyWuzzy or RapidFuzz library) is implemented to handle spelling mistakes, partial matches, and approximate queries.

● If the user enters a misspelled or incomplete product name, fuzzy search finds the closest match based on similarity scores.

Both search methods ensure flexibility and usability, even when the user is unsure of the exact product name.

Sentiment Summary and Visualization:

● When a product is retrieved (via normal or fuzzy search), all associated reviews are grouped and analyzed.

● Using Flair sentiment labels, the system calculates the percentage of positive, negative, and neutral reviews.

● These summary statistics help users quickly understand how customers feel about that product.

Recommendation System:

The recommendation module suggests alternative products based on sentiment distribution:

● Products with a higher number of positive reviews are ranked and recommended.

● Only similar or relevant items from the dataset are considered to ensure meaningful recommendations.

This supports users in choosing products with more favorable sentiment profiles.

User Interaction Module:

● The system provides an interactive interface where the user enters a product name.

● The system performs both normal and fuzzy search to find the best match.

● Once found, the system displays:

– Product details

– All customer reviews

– Sentiment distribution (positive/negative/neutral)

– Recommended similar products

This creates an end-to-end functional sentiment-aware search engine.

Prediction Module:

The kit included a prediction interface which is as easy as 1-2-3:

- Users are allowed to input a review index from the test set
- The review is positive or negative, as predicted by the system

Exporting Results:

● The cleaned dataset with Flair sentiment labels is saved as a CSV file for reproducibility.

● Intermediate outputs (matched products, sentiment scores, recommendations) are exported when necessary.

● All preprocessing and sentiment functions can be reused without needing retraining because Flair operates dynamically.

RESULTS

The sentiment-aware search engine performed effectively after the integration of the Flair sentiment classifier, achieving an overall sentiment prediction accuracy of 87% on the review dataset. Compared to the outputs from TextBlob and VADER, Flair demonstrated significantly stronger contextual understanding, especially in cases involving subtle sentiment expressions or multi-clause reviews. This justified the selection of Flair as the final sentiment model for the system.

The sentiment summary generated for each product successfully displayed the proportion of positive, negative, and neutral reviews, enabling clearer interpretation of customer satisfaction levels. The search module also worked as expected: the keyword-based search reliably returned exact or closely matching products when the input was accurate,

while the fuzzy search component effectively handled misspellings and partial product names, increasing the robustness and usability of the system.

The recommendation system proved practical, consistently identifying products with a higher concentration of positive reviews and suggesting them as preferable alternatives. This added an additional layer of value to the system by guiding users toward better-rated products based on real customer opinions.

Overall, the results show that the system moved beyond basic sentiment detection and evolved into a functional sentiment-aware product retrieval and recommendation tool. The 87% sentiment classification accuracy, combined with efficient search capabilities and meaningful recommendations, demonstrates the system's potential for real-world e-commerce applications.

PERFORMANCE

● The Flair sentiment model achieved an accuracy of 87%, performing better than TextBlob and VADER due to its stronger contextual understanding.

● The normal keyword search worked well for exact matches, while the fuzzy search greatly improved retrieval for misspelled or partial product names.

● The sentiment summary and recommendation module accurately highlighted products with higher positive review proportions, making the system practical and reliable for real e-commerce use.

CONCLUSION

This project presents a complete sentiment-aware search and recommendation system for e-commerce product reviews. It integrates data cleaning, text preprocessing, sentiment classification, and intelligent product retrieval into a unified framework. By evaluating multiple sentiment tools and ultimately selecting Flair for its 87% accuracy, the system delivers reliable polarity detection across diverse review texts.

The combination of normal keyword search and fuzzy search ensures flexible and accurate product retrieval, even when user input is incomplete or misspelled. Additionally, the sentiment-based summary and recommendation module helps users quickly understand overall customer satisfaction and discover better-rated alternatives.

Overall, the system demonstrates an effective and scalable approach for analyzing customer opinions, enhancing product discovery, and supporting informed decision-making in real-world e-commerce environments.

FUTURE ENHANCEMENTS

● Integrate advanced transformer models such as BERT or RoBERTa to achieve higher sentiment classification accuracy beyond Flair's current performance.

● Add aspect-based sentiment analysis to identify opinions on specific product features (e.g., quality, price, comfort, durability).

● Implement semantic search using sentence embeddings so the system can understand user intent rather than relying on keyword or fuzzy matching alone.

● Upgrade the recommendation system to include collaborative filtering or hybrid recommendation techniques for more personalized suggestions.

● Develop an interactive web application or REST API to make the system accessible to end users, businesses, or other applications.

● Enable multilingual sentiment analysis to support reviews written in different languages across global e-commerce platforms.

● Incorporate real-time data streaming so the system can automatically process and analyze newly added reviews