Int. J. Sci. R. Tech., 2026 3(11)
A Multidisciplinary peer-reviewed Journal
www.ijsrtjournal.com [ISSN: 2394-7063]

Original Article

Smart Customer Behaviour and Sentiment Analysis Application

Teena Varma*, Pranjal Balsaraf, Evan Dsouza, Sejal Upadhyay, Ayana Xavier

Xavier Institute of Engineering.

ABSTRACT

The Smart Customer Behaviour and Sentiment Analysis Web Application Project uses Natural Language Processing (NLP) and Machine Learning (ML) to analyze customer reviews, ratings, and behavioural data from e-commerce platforms. The system preprocesses customer feedback, extracts important textual features using TF-IDF, and classifies reviews into positive, negative, and neutral sentiments using machine learning algorithms such as Logistic Regression, Naive Bayes, and Support Vector Machine (SVM). It also analyzes customer behaviour patterns, satisfaction levels, preferences, and purchasing trends to generate meaningful business insights. The results are presented through dashboards and visualizations that help businesses improve customer experience, personalize recommendations, and support data-driven decision-making. The proposed framework is simple, interpretable, scalable, and can be extended with advanced AI models in the future.

Keywords: Customer Behaviour Analysis, Sentiment Analysis, Web Application Deep Analytics, Graphs, Natural Language Processing (NLP), Machine Learning, TF-IDF, Logistic Regression, Naive Bayes, Support Vector Machine (SVM), Customer Reviews, E-commerce Analytics, Data Visualization, Customer Satisfaction

INTRODUCTION

The rapid growth of E-commerce has generated a large amount of customer data through online reviews, ratings, purchasing activities, and feedback. Analysing this information helps businesses understand customer preferences, satisfaction, emotions, and purchasing behaviour **[2, 5]**. Sentiment analysis uses Natural Language Processing (NLP) and Artificial Intelligence to identify opinions and emotions expressed in customer reviews **[3, 11]**.

The proposed system processes customer data, cleans and transforms textual reviews into numerical features. The objective is to demonstrate how machine learning can transform raw customer information into meaningful insights that can support customer satisfaction analysis, purchasing-behaviour understanding, customer segmentation, and business decision-making **[1, 4, 7, 8]**.

The proposed work aims to provide a simple, interpretable, and practical framework that can be extended in the future with advanced algorithms such as Random Forest **[4, 5]**, XGBoost, BERT **[1, 14]**, RoBERTa **[13]**, deep learning **[9, 10]**, or large language models **[15]**, while integrating explainable AI primitives **[6]**. Thus, the project establishes a foundation for intelligent customer analytics while maintaining a methodology that is computationally feasible and easy to understand and reproduce.

Therefore, this project focuses on developing a *Smart Customer Behaviour and Sentiment Analysis* system that processes customer reviews and behavioural information to identify sentiment, customer preferences, and behavioural patterns. The system can help organizations understand customer needs, improve services, personalize recommendations, and support better business decisions **[2, 12]**.

2. Literature Review

Previous studies have used NLP, AI, ML, and Deep Learning techniques such as BERT, RoBERTa, LSTM, CNN, and ensemble models to analyze customer reviews and sentiments. Research shows that aspect-based and topic-level analysis can identify key factors such as quality, price, delivery, customer service, and shopping experience. Recent studies also use explainable AI and LLMs to generate deeper customer insights. Overall, these approaches help

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

businesses understand customer feedback and improve products, services, recommendations, and customer satisfaction.

No.	Methodology	Key Results
1	NLP, BERT embeddings, and sentiment classification.	BERT successfully converted raw feedback into actionable business insights.
2	Factor-level analysis (quality, price, speed) using ML.	Factor-level ratings revealed hidden preferences missed by overall star ratings.
3	Combined word-category data with string-kernel classification.	Word-category data improved review representation for sentiment analysis.
4	Random Forest, Naive Bayes, and Logistic Regression on social data.	58% of sentiments were positive; positive sentiment strongly tied to purchase intent.
5	Compared ML (LogReg, RF) vs. DL (CNN, RNN) on Amazon data.	Mapped performance differences and noted heavy data imbalance toward 4/5 stars.
6	TF-IDF, Word2Vec, FastText with ML/DL; LIME and LDA.	FastXCatStack model achieved 93–94% accuracy with clear explainability.
7	Analyzed Trustpilot reviews using DistilBERT and SVM.	Negative sentiment was heavily driven by customer service and communication issues.
8	Compared RF, SVM, and DL using transactional/behavioral features.	Proved satisfaction can be accurately predicted using delivery and location data.
9	Attention, LSTM, BiGRU, and GCN deep-learning models.	Proposed hybrid models boosted F1-scores by up to 7.84% over baselines.
10	EGJO-LSTM with LF-MICF term weighting and Grey Wolf Optimizer.	EGJO-LSTM model outpaced existing methods when using LF-MICF weighting.
11	Applied NLP preprocessing and sentiment analysis.	Mapped positive/negative experiences to pinpoint service improvements.
12	Sentiment analysis on mobile app reviews.	Proved app-review tracking effectively surfaces urgent UX and service issues.
13	Tested BERT, RoBERTa, XLNet on Trustpilot data across 14 aspects.	RoBERTa hit >90% accuracy, successfully isolating exact satisfaction drivers.
14	Evaluated traditional ML, RNN, attention-BERT, and ensembles.	Bagging+BERT ensemble led the pack with 89.9% accuracy.
15	BERT embeddings, clustering, and GPT sentiment classification.	Isolated core topics; found 31.7% of complaints targeted the shopping experience.

3. Research Gap

No.	Previous Research	Research Gap	Our Project / Proposed System
1	Sentiment Analysis	Mainly predicts sentiment	Sentiment + behaviour analysis
2	Rating-Based Analysis	Limited to overall ratings	Detailed factor-level analysis
3	BERT/Transformer Models	Focuses mainly on classification	Classification + actionable insights
4	Consumer Behaviour Studies	Limited behavioural factors	Identifies customer preferences and patterns
5	E-commerce Reviews	Mainly review analysis	Review + satisfaction analysis
6	Explainable AI	Limited business-level interpretation	Explainable customer insights
7	LLM/Topic Analysis	Focuses on topics and complaints	Topic + sentiment + behaviour
8	Existing ML/DL Models	Model-specific analysis	Comparative model-based analysis
9	Recommendation Systems	Based mainly on ratings	Uses sentiment and preferences
10	Traditional / Siloed Systems	Lacks unified framework	Integrated sentiment, behaviour, satisfaction, and recommendation analysis

METHODOLOGY

The proposed system follows a machine learning-based methodology for analysing customer behaviour and sentiment from customer reviews and behavioural information. The methodology consists of several stages, beginning with data collection and preprocessing, followed by feature extraction, sentiment classification, customer behaviour analysis, and interpretation of results.

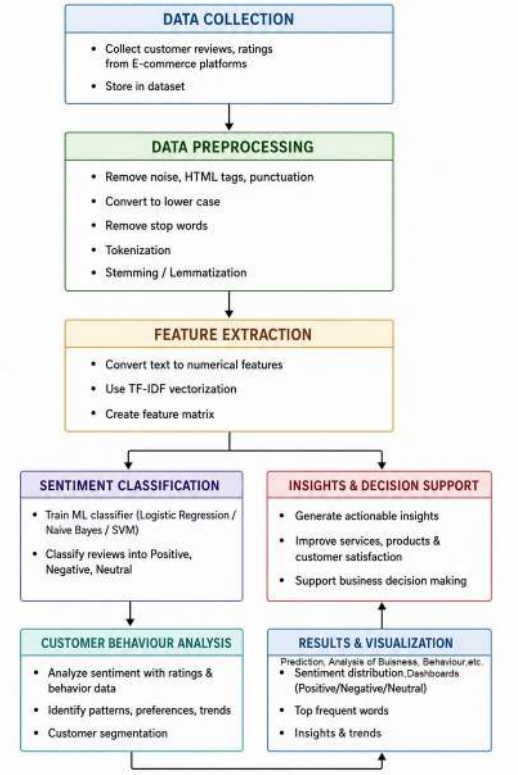


Figure 4.1: Flow Diagram

1] Data Collection

Customer-related data is collected in the form of online reviews, ratings, feedback, and behavioural information. Customer reviews provide textual information about customer opinions, while ratings and behavioural attributes provide additional information for analysing customer satisfaction and preferences. The collected data is used as the input for the proposed system.

2] Data Preprocessing

The collected customer reviews are preprocessed to remove unnecessary information and prepare the text for machine learning. The preprocessing stage includes:

- Removing unnecessary characters and punctuation.
- Converting text into a consistent format.
- Removing irrelevant words and stop words.
- Tokenizing the review text.
- Applying stemming or lemmatization where required.
- Handling missing or inconsistent data.

This step converts raw customer reviews into cleaner and more meaningful textual data.

3] Feature Extraction

After preprocessing, the textual reviews are converted into numerical representations that can be processed by machine learning algorithms. TF-IDF (Term Frequency–Inverse Document Frequency) is used to determine the importance of words within customer reviews.

The extracted features represent important characteristics of customer feedback and allow the machine learning model to identify patterns associated with different sentiments.

4] Sentiment Classification

The extracted textual features are provided to a machine learning classification model to determine the sentiment of each customer review. The reviews are classified into:

- Positive sentiment
- Negative sentiment
- Neutral sentiment

Machine learning algorithms such as Logistic Regression, Naive Bayes, and Support Vector Machine (SVM) can be used for this classification. The literature reviewed in the document also demonstrates the use of Logistic Regression, Random Forest, Naive Bayes, SVM, and deep-learning models for customer-review sentiment classification.

5] Customer Behaviour Analysis

The sentiment results are combined with available customer information such as ratings, preferences, and behavioural characteristics to identify customer behaviour patterns.

Positive, negative, and neutral feedback can be analysed to determine customer satisfaction, dissatisfaction, preferences, and recurring concerns. This stage helps transform sentiment classification results into meaningful customer behaviour insights.

6] Model Evaluation

The trained machine learning model is evaluated using appropriate performance measures such as:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

These metrics are used to determine how effectively the model classifies customer sentiments.

7] Insight Generation and Visualization

The final sentiment and behavioural results are presented using graphs and charts. For example, the system can display the percentage of positive, negative, and neutral reviews and identify frequently occurring customer concerns or preferences.

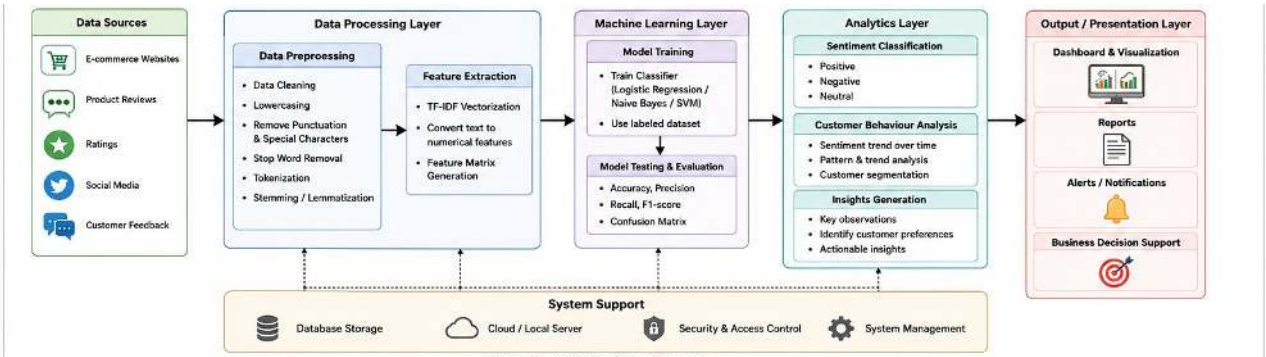


Figure 4.2: Architectural Diagram

5. Experimental Setup

The system uses a dataset containing customer reviews, ratings, and customer-related behavioural information. The review text is used for sentiment analysis, while ratings and other available attributes are used to analyse customer preferences and behavioural patterns.

The dataset is divided into training and testing data. The training portion is used to train the machine learning models, while the testing portion is used to evaluate their performance.

Software Requirements

- Operating System: Windows / Linux / macOS
- Programming Language: Python 3.x
- Development Environment: Visual Studio, Python Idle, Tableau, Google Colab / Jupyter Notebook
- Libraries: Pandas, NumPy, NLTK, Scikit-learn, Matplotlib, Seaborn
- Feature Extraction: TF-IDF Vectorizer
- Machine Learning: Logistic Regression, Naive Bayes, SVM

Hardware Requirements

- Processor: Intel Core i3 or equivalent and above
- RAM: Minimum 4 GB
- Storage: Minimum 10 GB free space
- Internet: Required for Google Colab and dataset access
- GPU: Not required; CPU is sufficient for the proposed traditional machine learning models.

6. Results for the Web Application -

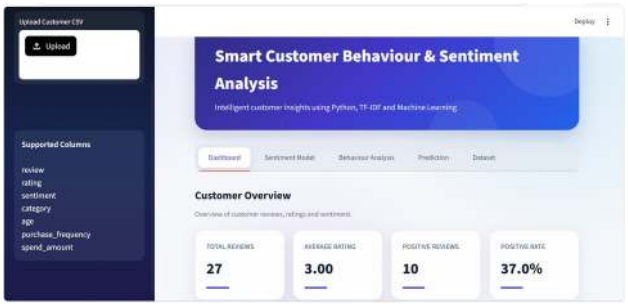


Figure 6.1: Homepage Dashboards for Web Application - Customer Overview



Figure 6.2: Homepage Dashboard - Customer Overview

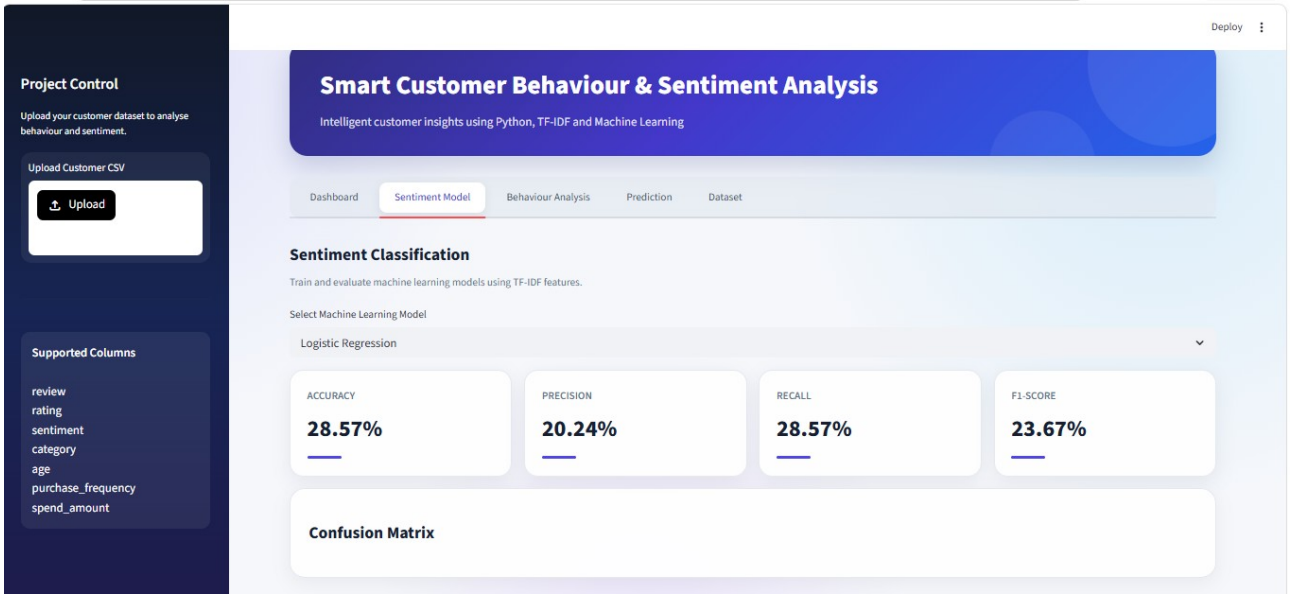


Figure 6.3: Sentiment Classification Page

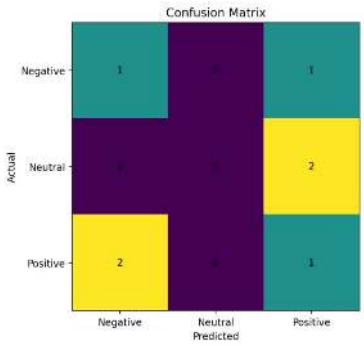


Figure 6.4: Confusion Matrix



Figure 6.4: Sentiment Classification Page - Classification Report

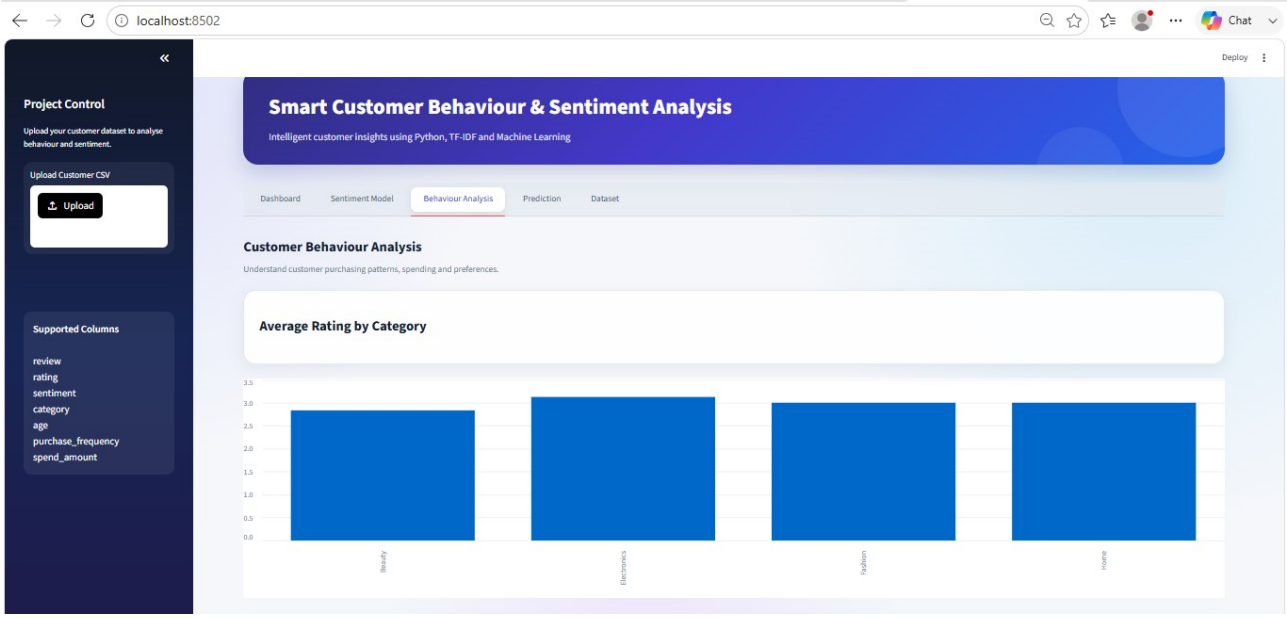


Figure 6.5: Customer Behaviour Analysis Page - Average Rating by Category



Figure 6.6: Customer Behaviour Analysis Page - Purchase Frequency by Sentiment



Figure 6.7: Customer Behaviour Analysis Page - Average Spending by Sentiment



Customer Segment Summary

customer_segment	customers	avg_rating
Frequent Buyer	6	2.67
High Value	14	4.07
Occasional Buyer	7	1.14

Figure 6.8: Customer Behaviour Analysis Page - Customer Segment Summary

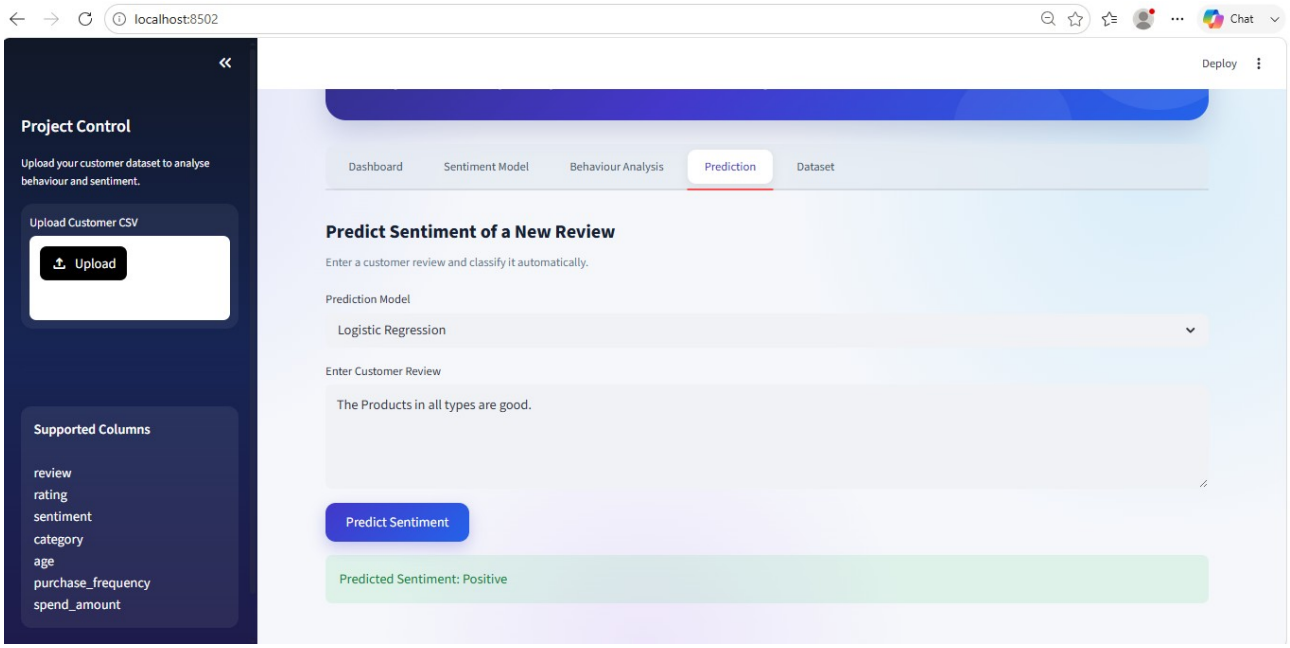


Figure 6.9: Prediction Page - Predict Sentiment of a New Review

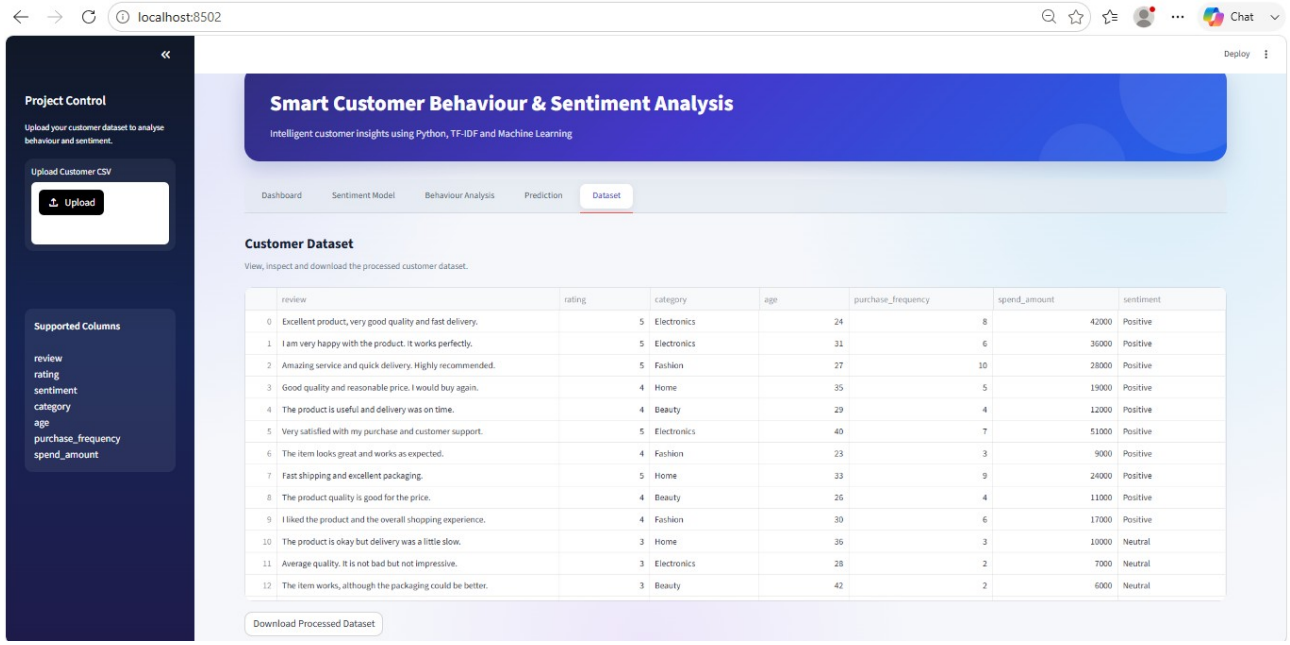


Figure 6.10: Customer Dataset with Upload Files, Supported Columns, Options on All Pages.

CONCLUSION

The Smart Customer Behaviour and Sentiment Analysis Web Application Project successfully demonstrates how NLP and Machine Learning can transform large volumes of customer feedback into meaningful business insights. By analysing sentiment and identifying key factors influencing customer opinions, the system enables businesses to understand customer expectations, satisfaction levels, and behavioural patterns more effectively. The integration of intelligent analysis supports better decision-making, targeted recommendations, improved customer experience, and continuous service improvement. Overall, the project provides a practical and scalable approach for converting unstructured customer feedback into actionable intelligence for modern businesses.

REFERENCES

1. Rahman & Maryani – BERT-Based Sentiment Analysis, 2024.
2. Najib – Rating Strategy for Recommendations, 2026.

3. Kaggle – Datasets for Machine Learning and Sentiment Analysis.
4. Python Documentation – Programming and Data Analysis.
5. Pandas Documentation – Data Processing and Analysis.
6. Matplotlib Documentation – Data Visualization.
7. Rahardjo et al. – Consumer Behaviour & Purchase Intentions, 2024.
8. Hashmi & Yildirim Yayilgan – Explainable AI, 2024.
9. Teena Varma , Albert Keerimolel2 , Brandon Gowray3 , Janeya Limbachia4 And Kshitij Karekar5, "BRAND REVIEW ANALYSIS USING TWITTER DATASET", jespublication, Vol 11, Issue 5,, ISSN NO:0377-9254,Pa, May/2020.
10. Davoodi et al. – Aspect-Based E-Commerce Analysis, 2025.
11. Rasappan et al. – Hybrid Deep Learning, 2024.
12. LLM-Based Topic-Level Sentiment Analysis, 2025.
13. Scikit-learn Documentation – Machine Learning Library.
14. NLTK Documentation – Natural Language Processing Toolkit.
15. Hugging Face – NLP and Transformer Models.

HOW TO CITE: Teena Varma, Pranjal Balsaraf, Evan Dsouza, Sejal Upadhyay, Ayana Xavier, Smart Customer Behaviour and Sentiment Analysis Application, Int. J. Sci. R. Tech., 2026, 3 (11), 56-64. https://doi.org/10.5281/zenodo.23909384