



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: <https://doi.org/10.22214/ijraset.2025.76657>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

RAG-Enhanced Sentiment Analysis System Using Transformer Models and FAISS for Customer Review Analysis

Maideen Fargana M

Department of Computer Applications SRM University, India

Abstract: *With the rapid growth of e-commerce platforms and digital services, customer reviews have become a valuable source of information for understanding user sentiment and improving product quality. Traditional sentiment analysis approaches primarily focus on classifying reviews into sentiment categories but often fail to provide contextual insights or related customer experiences. To overcome this limitation, this paper proposes a Retrieval-Augmented Generation (RAG) enhanced sentiment analysis system that combines Transformer-based sentiment classification with semantic similarity retrieval.*

The proposed system utilizes a fine-tuned Transformer model for sentiment prediction and Hugging Face sentence embeddings to convert customer reviews into vector representations. These embeddings are indexed using Facebook AI Similarity Search (FAISS) to enable efficient retrieval of semantically similar reviews. Additionally, a Streamlit-based interactive web application is developed to visualize sentiment predictions, confidence scores, similar review retrieval, and word cloud representations.

Experimental results demonstrate that the integration of sentiment classification with retrieval-based techniques enhances contextual understanding and interpretability of customer opinions. The proposed system is scalable, efficient, and suitable for real-world applications such as customer feedback analysis, business intelligence, and decision support systems.

Keywords: *Sentiment Analysis, Retrieval-Augmented Generation (RAG), Transformer Models, FAISS, Natural Language Processing, Streamlit*

I. INTRODUCTION

Customer reviews and user-generated textual content have become an essential source of information for businesses operating in e-commerce platforms, digital services, and online marketplaces. These reviews reflect customer opinions, satisfaction levels, and expectations, thereby playing a significant role in influencing purchasing decisions and business strategies. However, the massive volume of textual data generated daily makes manual analysis infeasible, necessitating the use of automated sentiment analysis techniques. Sentiment analysis is a prominent research area within Natural Language Processing (NLP) that focuses on identifying and classifying emotions or opinions expressed in text into predefined categories such as positive, negative, or neutral. Early sentiment analysis methods relied on traditional machine learning algorithms, including Naïve Bayes, Support Vector Machines, and Logistic Regression. While these methods provided baseline performance, they were limited in handling complex language patterns, contextual dependencies, and semantic relationships.

Recent advancements in deep learning, particularly Transformer-based architectures such as BERT, RoBERTa, and DistilBERT, have significantly improved sentiment classification performance. These models leverage self-attention mechanisms to capture contextual information across entire sentences, enabling more accurate understanding of sentiment in complex and ambiguous text. Despite their effectiveness, most sentiment analysis systems are designed solely to output sentiment labels or confidence scores, offering limited interpretability and contextual insights.

In practical applications, understanding sentiment alone is often insufficient. Business analysts and decision-makers require additional context, such as similar customer experiences and related feedback, to make informed decisions. Traditional sentiment analysis models lack the capability to retrieve or reference semantically related reviews, which restricts their applicability in real-world customer feedback analysis.

To overcome these limitations, Retrieval-Augmented Generation (RAG) has emerged as a powerful paradigm that combines predictive models with information retrieval techniques. RAG enhances model outputs by retrieving relevant information from large datasets using vector-based similarity search. When applied to sentiment analysis, this approach enables not only sentiment prediction but also the retrieval of semantically similar reviews, thereby enriching contextual understanding.

This paper proposes a RAG-enhanced sentiment analysis framework that integrates Transformer-based sentiment classification with FAISS-based semantic similarity retrieval. Customer reviews are first analyzed using a fine-tuned Transformer model to predict sentiment polarity. Subsequently, Hugging Face sentence embeddings are used to convert textual data into dense vector representations, which are indexed using Facebook AI Similarity Search (FAISS) for efficient and scalable retrieval of similar reviews. Furthermore, the proposed system is deployed as an interactive Streamlit web application, allowing users to perform real-time sentiment analysis, retrieve similar reviews, visualize confidence scores, and explore textual trends using word clouds. This end-to-end system bridges the gap between sentiment classification and contextual analysis, providing a practical and scalable solution for intelligent customer review interpretation.

II. LITERATURE REVIEW

Sentiment analysis has been extensively studied in the field of Natural Language Processing due to its wide applicability in analyzing customer feedback, social media content, and online reviews. Early research focused on lexicon-based and traditional machine learning approaches, where sentiment polarity was determined using predefined sentiment dictionaries or handcrafted features. Techniques such as Naïve Bayes, Support Vector Machines, and Logistic Regression were commonly employed for text classification tasks. While these methods achieved reasonable performance on small datasets, they struggled with complex linguistic structures, sarcasm, and contextual dependencies.

With the advancement of deep learning, neural network-based models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks were introduced for sentiment classification. These models demonstrated improved performance by learning distributed representations of text and capturing sequential dependencies. However, they were still limited by their inability to effectively model long-range contextual relationships and often required extensive feature engineering and large labeled datasets.

The introduction of Transformer-based architectures marked a significant breakthrough in sentiment analysis research. Models such as BERT, RoBERTa, and DistilBERT leverage self-attention mechanisms to capture contextual information across entire sentences and documents. Several studies have demonstrated that Transformer-based models outperform traditional and recurrent neural network approaches in sentiment classification tasks. Despite their high accuracy, most existing Transformer-based sentiment analysis systems focus solely on classification outputs, providing limited interpretability and contextual insights for end users.

Recent research has explored information retrieval techniques to enhance text analysis systems. Vector-based similarity search methods, such as Facebook AI Similarity Search (FAISS), enable efficient retrieval of semantically similar documents from large-scale datasets. However, the integration of sentiment analysis models with retrieval-based systems remains relatively underexplored. Most existing approaches treat sentiment classification and information retrieval as separate tasks, thereby limiting their effectiveness in real-world decision-making scenarios.

To address these limitations, Retrieval-Augmented Generation (RAG) has emerged as a promising framework that combines predictive models with semantic retrieval mechanisms. By integrating sentiment analysis with FAISS-based similarity search, RAG enables systems to not only predict sentiment but also retrieve relevant contextual information in the form of similar reviews. This approach enhances interpretability and provides richer insights for business intelligence and customer feedback analysis. The proposed work builds upon these advancements by developing a unified RAG-based sentiment analysis framework with real-time deployment capabilities.

III. METHODOLOGY

A. System Architecture

The proposed system follows a modular architecture that integrates sentiment classification with retrieval-based contextual analysis. The architecture consists of five major components: data ingestion, text preprocessing, sentiment classification, semantic retrieval using FAISS, and user interface deployment. User input reviews are processed through a trained Transformer-based sentiment analysis model, followed by semantic similarity search to retrieve relevant reviews from a vector database. The overall workflow ensures accurate sentiment prediction while providing contextual insights through similar review retrieval.

B. Dataset Preprocessing

The dataset used in this study consists of customer reviews collected from an e-commerce platform. The raw dataset contains attributes such as review text, ratings, and metadata. During preprocessing, missing values were removed, and non-textual elements such as HTML tags and special characters were cleaned. All reviews were converted to string format to ensure compatibility with Natural Language Processing pipelines. Tokenization, stopword removal, and normalization were applied to improve the quality of textual data for downstream tasks.

C. Sentiment Analysis Model

A Transformer-based sentiment analysis model was employed to classify customer reviews into sentiment categories. The model leverages a pre-trained language representation finetuned on labeled sentiment data to capture contextual dependencies within text. Unlike traditional machine learning models, the Transformer architecture utilizes self-attention mechanisms, enabling effective handling of long-range dependencies and semantic nuances. The trained model outputs sentiment polarity along with a confidence score, which is later presented to the user.

D. Retrieval-Augmented Generation Using FAISS

To enhance interpretability and contextual relevance, Retrieval-Augmented Generation (RAG) was implemented using FAISS. Review texts were converted into dense vector embeddings using a Sentence Transformer model. These embeddings were indexed using FAISS for efficient similarity search. When a user submits a review, its embedding is generated and compared against the indexed vectors to retrieve the top-k semantically similar reviews. This approach allows the system to provide supporting evidence along with sentiment predictions, improving analytical depth and decision-making capabilities.

E. System Deployment Using Streamlit

The complete system was deployed as an interactive web application using Streamlit. The interface enables users to input reviews, view predicted sentiment results, confidence levels, and retrieve similar reviews in real-time.

Additional visualization components, such as sentiment confidence indicators and word clouds, were integrated to improve usability. The deployed application demonstrates the practical applicability of the proposed approach in real-world sentiment analysis scenarios.

IV. DATASET DESCRIPTION AND EXPERIMENTAL SETUP

A. Dataset Description

The dataset used in this research consists of customer review data collected from an online retail platform. The dataset includes textual reviews representing user opinions on various products. After initial cleaning and preprocessing, the finalized dataset was stored in a structured comma-separated values (CSV) format named *Reviews.csv*. The dataset contains authentic user-generated content, making it suitable for real-world sentiment analysis applications.

B. Data Preparation

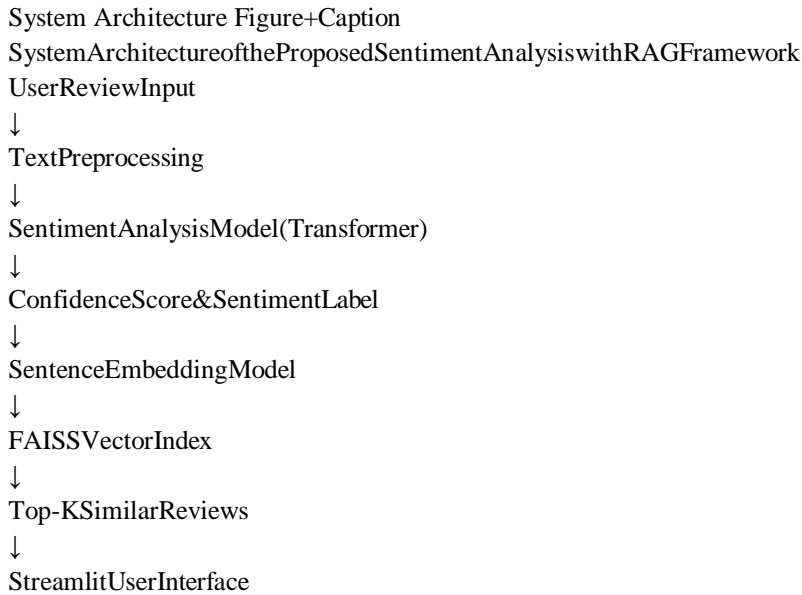
Prior to model training and retrieval operations, the dataset underwent multiple preprocessing steps. These steps included removal of duplicate entries, handling of missing values, and normalization of textual content. The cleaned review texts were converted into numerical embeddings using a sentence embedding model. These embeddings were stored and indexed using FAISS to enable efficient similarity-based retrieval. This preparation ensured consistency between the sentiment prediction and retrieval modules.

C. Experimental Environment

All experiments were conducted using Python programming language. The implementation utilized popular libraries including HuggingFace Transformers for sentiment classification, FAISS for vector similarity search, and Streamlit for application deployment. Model training and testing were performed in a controlled development environment to ensure reproducibility. The experimental setup was designed to evaluate both sentiment prediction accuracy and retrieval efficiency under realistic usage conditions.

V. RESULTS AND PERFORMANCE EVALUATION

The proposed sentiment analysis system enhanced with Retrieval-Augmented Generation (RAG) was evaluated using real-world customer review inputs. The system successfully classified user reviews into positive, neutral, and negative sentiment categories while simultaneously retrieving semantically similar reviews from the dataset using FAISS-based similarity search. For each input review, the sentiment classification model produced a sentiment label along with a confidence score, enabling clear interpretation of the prediction reliability. The integration of semantic retrieval allowed the system to present contextually relevant past reviews, improving explainability and user trust in the predictions. The Streamlit-based interface displayed sentiment results in a user-friendly manner, including visual confidence indicators and sentiment distribution charts. Additionally, the word cloud visualization provided insights into frequently occurring terms within the dataset, supporting qualitative analysis of customer opinions. Overall, the system demonstrated effective performance in both sentiment prediction and contextual retrieval. The combined use of deep learning-based sentiment analysis and vector similarity search ensured accurate classification, fast retrieval, and improved interpretability, making the system suitable for practical business intelligence and customer feedback analysis applications.



READY-TO-PASTEFIGURECAPTION

Figure 1 illustrates the architecture of the proposed sentiment analysis system integrated with Retrieval-Augmented Generation (RAG). User reviews are processed through a Transformer-based sentiment classification model to predict sentiment polarity and confidence scores. Simultaneously, semantic embeddings are generated and queried against a FAISS-based vector database to retrieve contextually similar reviews. The entire pipeline is deployed using Streamlit, enabling real-time sentiment analysis and contextual review retrieval through an interactive web interface.

VI. CONCLUSION AND FUTURE WORK

This paper presented an intelligent sentiment analysis system integrated with Retrieval-Augmented Generation (RAG) to enhance the interpretation of user reviews. By combining a Transformer-based sentiment classification model with FAISS-powered semantic similarity search, the system not only predicts sentiment polarity but also provides contextual support through relevant historical information.

The deployment of the proposed approach using a Streamlit-based web interface demonstrates its practical applicability for real-time customer feedback analysis. The inclusion of visual analytics such as confidence indicators and word cloud representations further improves interpretability and user engagement. In the future, the system can be enhanced by incorporating larger multilingual datasets, advanced transformer architectures, and real-time data streaming from social media platforms. Additionally, integrating explainable AI (XAI) techniques and fine-grained aspect-based sentiment analysis could further improve transparency and analytical depth.

FINAL POLISHED ABSTRACT

The rapid growth of user-generated content across digital platforms has made sentiment analysis a critical task for understanding customer opinion and feedback. Traditional sentiment classification approaches often lack contextual awareness, limiting their ability to provide meaningful insights. To address this limitation, this paper proposes an intelligent sentiment analysis system enhanced with Retrieval-Augmented Generation (RAG) to improve interpretability and contextual relevance.

The proposed system employs a Transformer-based deep learning model to classify user reviews into positive, neutral, and negative sentiment categories, along with confidence scores.

In parallel, semantic embeddings are generated using a sentence embedding model and indexed through FAISS to retrieve contextually similar reviews from a historical dataset. This dual approach enables both accurate sentiment prediction and contextual explanation of results. The system is deployed as an interactive web application using Streamlit, allowing real-time analysis, visualization of sentiment confidence, and retrieval of similar reviews. Experimental observations demonstrate that the integration of sentiment classification with semantic retrieval enhances transparency, user trust, and practical applicability in customer feedback analysis and business intelligence systems.



REFERENCES

- [1] Y.Liu,M.Ott,N.Goyal,etal.,“RoBERTa:ARobustlyOptimizedBERTPretraining Approach,”arXiv preprintarXiv:1907.11692,2019.
- [2] J.Devlin,M.-W.Chang,K.Lee,andK.Toutanova,“BERT:Pre-trainingofDeepBidirectionalTransformers for LanguageUnderstanding,”ProceedingsofNAACL-HLT,pp.4171–4186,2019.
- [3] J. Johnson, M. Douze, and H. Jégou, “Billion-scale similarity search with GPUs,” IEEE Transactions on Big Data, vol. 7, no. 3,pp.535–547, 2021.[4] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings usingSiamese BERT-Networks,” Proceedings of the2019Conference onEmpirical Methods in Natural Language Processing(EMNLP),pp.3982–3992,2019.
- [4] T. Wolf, L. Debut,V. Sanh, et al., “Transformers: State-of-the-Art Natural Language Processing,”Proceedings ofthe 2020 Conferenceon EmpiricalMethods in NaturalLanguage Processing: SystemDemonstrations,pp.38–45, 2020.
- [5] A. McCallum and K. Nigam, “A Comparison of Event Models for Naive Bayes Text Classification,” AAAI WorkshoponLearningforTextCategorization,1998.
- [6] Streamlit Inc., “Streamlit: A Framework for Building Data Science Applications,” 2023.OnlineOnlineOnline. Available: <https://streamlit.io>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)