Sentiment Analysis of Customer Reviews Using Classical ML and Deep Learning (BERT)

Thadala Veera Venkata Ramana[1], Dr. K Swapna[2]
[1]*MSc student, Department of Information Technology and Computer Applications, College of Engineering (A), Andhra University Visakhapatnam, AP.*
[2]*Asst. Professor, Department of Information Technology and Computer Applications, College of Engineering (A), Andhra University Visakhapatnam, AP.*

***Abstract*—In the age of online shopping and social media, customers frequently express their opinions through reviews, which are valuable for businesses to understand satisfaction and improve services. This project uses automated sentiment analysis to classify these reviews as Happy, Unhappy, or Neutral. Two approaches are explored: classical machine learning using TF-IDF with models like LinearSVC and SGDClassifier, and a deep learning method using BERT, a powerful language model from Hugging Face's Transformers library.The classical models offer speed and simplicity, especially for smaller datasets, while BERT provides greater accuracy by better understanding language context. To enhance usability, a Gradio-based web interface was developed for real-time sentiment prediction. Overall, the study finds that while classical models are efficient, BERT delivers superior performance, making it suitable when accuracy is a top priority.**

***Index Terms*—Sentiment Analysis, TF-IDF, LinearSVC, SGDClassifier, BERT, Transformers, Gradio.**

I. INTRODUCTION

Sentiment Analysis is a technique used by computers to understand how people feel based on what they write. It is part of a bigger area called Natural Language Processing (NLP), which helps machines read and understand human language. With sentiment analysis, we can find out whether a person's opinion in a sentence or review is positive, negative, or neutral.
Nowadays, people share a lot of reviews and feedback online especially on shopping websites, mobile apps, and social media. For example, after buying a product, a customer might write a review saying whether they liked it or not. These reviews are important for companies because they help them understand what people like, what they don't, and how they can improve. But since thousands of reviews are written every day, it is very hard and time-

Consuming for humans to read and understand all of them. This is why we need automatic systems that can quickly read and classify these reviews.
We used two methods to do this:
1. Classical Machine Learning, where we use mathematical models like Linear Support Vector Classifier (LinearSVC) and SGDClassifier. These models are fast and easy to use.
2. Deep Learning, where we use a powerful model called BERT. BERT is better at understanding the meaning of sentences because it was trained on a large amount of text.

We also built a simple website using Gradio where users can type in a review and instantly see if the review is positive or negative. This makes it easy for anyone to test the system.
In the end, we found that while the deep learning model (BERT) gives better results, the classical models are still useful when we want faster and simpler solutions. Both methods can help companies better understand their customers and improve their products or services.

II. LITERATURE SURVEY

For the development of this sentiment analysis project, several recent research papers from 2020 to 2024 were reviewed to understand the evolution of techniques in text-based opinion mining. These works provided valuable insights into both classical machine learning models and advanced deep learning architectures like BERT for sentiment classification.

Kumar et al. [1] conducted a study on customer reviews using machine learning models such as Logistic Regression, Naive Bayes, and SVM. Their experimental evaluation showed that classical algorithms still offer competitive performance when optimized properly, especially on smaller datasets.

Devlin et al. [2] introduced BERT (Bidirectional Encoder Representations from Transformers), which has become a state-of-the-art method for deep contextual language understanding and is widely used in modern sentiment analysis applications.

Poria et al. [3] highlighted various open challenges in sentiment analysis, such as sarcasm detection, domain adaptation, and aspect-level sentiment prediction. Their work suggested that hybrid approaches, combining rule-based and neural methods, are often necessary for high accuracy.

Ayyub et al. [4] explored various features such as POS tagging and n-gram frequency, employing machine learning algorithms to enhance sentiment quantification tasks.

Ramakrishnan and Dhinesh Babu [5] improved Twitter sentiment classification by integrating BERT embeddings with a BiLSTM model and an attention mechanism. Their results confirmed that deep learning with pre-trained embeddings significantly outperforms traditional RNN-based approaches.

Xie et al. [6] applied BERT to Chinesee-commerce reviews, achieving better generalization and accuracy across multiple sentiment categories.

Wang et al. [7] proposed a weakly-supervised deep learning framework to classify sentiments where training data is sparsely labeled. Their approach showed promise in reducing manual annotation efforts.

Palani et al. [8] developed T-BERT, a hybrid model combining topic modelling with BERT, which significantly enhanced microblog sentiment analysis performance.

A 2022 IEEE review [9] emphasized the importance of combining classical methods like TF-IDF with deep learning models such as CNNs and LSTMs to improve text sentiment analysis, particularly when working with domain-specific data.

Kamyab et al. [10] presented a hybrid CNN-BiLSTM architecture, incorporating both TF-IDF and GloVe embeddings, achieving robust sentiment prediction results on product review datasets.

These studies collectively underline the effectiveness of combining classical techniques (TF-IDF, SVM) with modern deep learning models (BERT, BiLSTM) to balance performance, interpretability, and computational efficiency in sentiment analysis systems.

III.PROBLEM STATEMENT

A. Background of the Problem

In the digital age, customer reviews play a vital role in influencing product perception and business strategy. Automating the analysis of these reviews using sentiment analysis techniques helps organizations efficiently interpret user feedback at scale.

Traditional methods like Naive Bayes and SVM offer decent results but fall short in capturing deep contextual meaning. Deep learning models like BERT address this issue with improved accuracy, though they come with higher computational requirements.

B. Problem Statement

This project focuses on building a sentiment analysis system to classify customer reviews into Positive, Neutral, or Negative categories. It compares classical models such as LinearSVC and SGDClassifier with a BERT-based transformer model.

The goal is to evaluate accuracy and performance while providing an interactive user interface. This helps businesses better understand customer sentiment and make informed decisions based on real-time feedback.

IV. METHODOLOGY USED

In this project, we utilized customer review datasets obtained from open sources such as Kaggle, containing review texts and their corresponding ratings. We pre-processed the raw text data and implemented both classical machine learning and deep learning algorithms to classify the sentiments into categories like Happy, Ok, and Unhappy. Each stage in the process from raw data to prediction was carefully designed and implemented to improve model performance and interpretability.

There are multiple phases involved in the pipeline such as data collection, pre-processing, feature extraction, model training, and deployment. Every phase contributes significantly to the overall system's performance and accuracy.

a. Data Collection

For this project, we used open-source datasets consisting of product reviews and ratings. The reviews were collected in .csv format from platforms like Amazon and Flipkart. The numerical ratings (1–5) were mapped to categorical sentiment labels to simplify the classification task:

- Ratings 1–2 → Unhappy
- Rating 3 → Ok
- Ratings 4–5 → Happy

By using multiple sources and review sizes, we aimed to analyze how model performance varies across datasets.

b. Pre-processing

Pre-processing is a critical step that cleans and prepares text data for feature extraction and modelling. In this phase:

- Unwanted columns (like user ID, timestamp) were dropped.
- Reviews were lowercased and punctuation, special characters, and HTML tags were removed.
- Stop words such as "the", "is", and "was" were eliminated.
- Null or missing entries were handled by deletion or imputation.

This step ensured uniformity and reduced noise in the data, which helps in improving model training results.

c. Feature Extraction

After pre-processing, the text data was converted into numerical format using two techniques:

The two techniques are as follows

- For classical machine learning models, TF-IDF vectorization was applied at both word-level and character-level to extract meaningful statistical features.
- For deep learning models, BERT tokenizer was used to convert the reviews into embeddings compatible with the Transformer architecture.

These extracted features were used as input for both types of classifiers.

d. Model Selection and Training

Classical Models:

1. Linear Support Vector Classifier (LinearSVC)
2. Stochastic Gradient Descent Classifier (SGDClassifier)

These were trained on TF-IDF vectors.

Deep Learning Model**:**

BERT (Bidirectional Encoder Representations from Transformers**)** was implemented using the Bert-base-uncased model from Hugging Face. The model was fine-tuned on tokenized review data for sentiment classification and evaluated using accuracy, precision, recall, and F1-score.

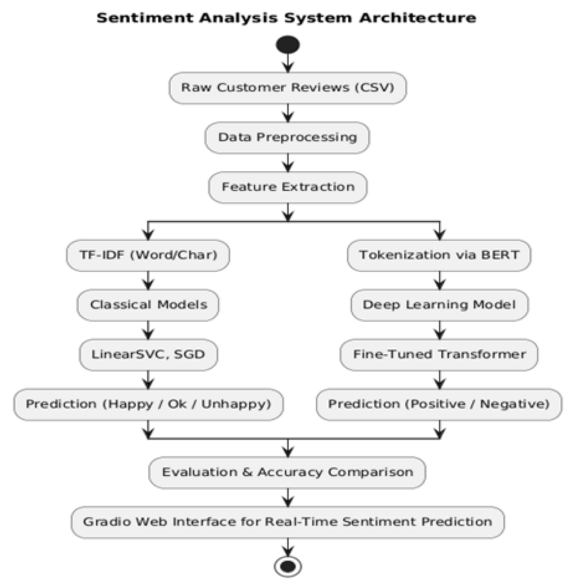

Fig-1: Sentiment Analysis Architecture

e. Sentiment Prediction Interface

To make the system interactive, we built a simple Gradio web interface. Users can input any review and receive the predicted sentiment instantly. This tool helps in real-time demonstration and user testing.

V. RESULTS AND DISCUSSION

This section presents the evaluation outcomes of the implemented sentiment analysis models using both classical machine learning and deep learning approaches. The performance was measured based on accuracy, precision, recall, F1-score, and training time using Amazon and Flipkart customer review datasets. Two primary models were compared: the

TF-IDF with Linear Support Vector Classifier (LinearSVC), and the fine-tuned BERT model.

A. Evaluation Metrics

Both models were evaluated using standard classification metrics:

- Accuracy: Overall correctness of predictions.
- Precision: The percentage of positive identifications that was correct.
- Recall: The percentage of actual positives correctly identified.
- F1-Score: The harmonic mean of precision and recall.

Model	Accuracy	Precision	Recall	F1-Score
LinearSVC	0.90	0.90	0.90	0.90
SGDClassifier	0.90	0.90	0.90	0.90
BERT (Fine-tuned)	0.94	0.94	0.94	0.94

Table 1: Shows comparison of different evaluation metrics

From the above comparison table, we can observe that BERT (Fine-tuned) outperforms both LinearSVC and SGDClassifier across all evaluation metrics. It achieves the highest accuracy of 94%, demonstrating its strong capability in understanding contextual nuances in the review data.

On the other hand, Linear SVC and SGD Classifier deliver equally strong performance with 90% accuracy and balanced precision, recall, and F1-scores. These classical models perform well and are more lightweight compared to BERT, making them suitable for scenarios with limited computational resources.

Given the results, while BERT offers superior performance, LinearSVC remains a viable and efficient choice for fast and reliable sentiment classification on moderate datasets.

B. Confusion Matrix

The confusion matrix visually represents the model's classification performance.

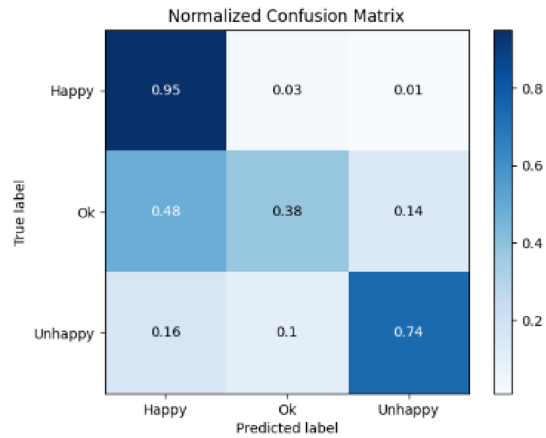

Fig-2: Confusion Matrix of Linear SVC Model

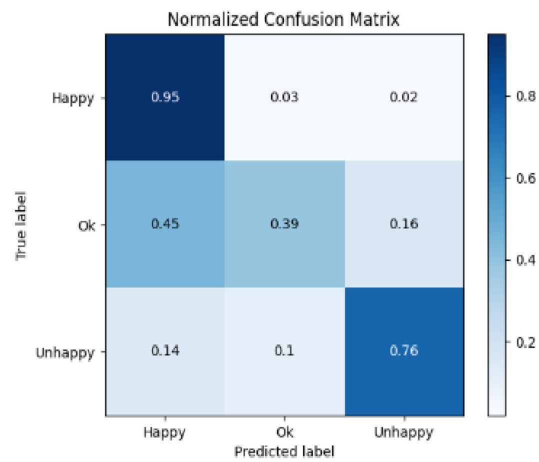

Fig-3: Confusion Matrix of SGD Model

As part of our implementation, we evaluated multiple models including LinearSVC, SGDClassifier, and BERT. After analyzing their performance, we present the metrics comparison of these algorithms in the form of a graph for better visualization and interpretation.

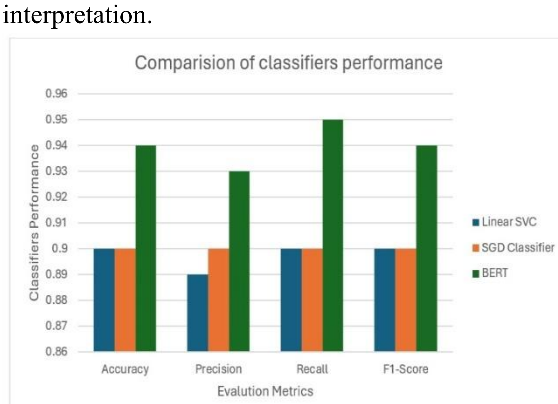

Fig-4: Graph showing different evaluation metrics of different algorithms

VI. CONCLUSION&FUTURE SCOPE

This project successfully demonstrates sentiment analysis on customer reviews using both classical machine learning (TF-IDF with LinearSVC and SGDClassifier) and deep learning (BERT). The classical models provide fast and efficient results, while BERT delivers higher accuracy by capturing deeper context. A user-friendly Gradio interface was developed to allow real-time sentiment prediction. Evaluation metrics show that the system is both accurate and practical for real-world use.

Future work can focus on improving performance and scalability by:

- Using advanced models like RoBERTa or DistilBERT.
- Supporting multiple languages.
- Deploying the system via API for integration with other platforms.
- Adding aspect-based sentiment analysis.
- Optimizing deep learning models for speed and resource efficiency.

These improvements will make the system more robust, accessible, and production-ready.

REFERENCES

[1] K. Kumar, S. K. Singh, A. Kumar, and A. S. Singh, "Sentimental Analysis of Customer's Reviews Using Machine Learning," in *Proc. of 2023 5th Int. Conf. on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp. 326–332, 2023. DOI: 10.1109/ICAC3N60023.2023.10541464

[2] J. Devlin, M. Chang, K. Lee, and K. Toutanova, [2] "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint*, arXiv:1810.04805, 2018. [Online]. Available: https://arxiv.org/abs/1810.04805

[3] S. Poria, N. Majumder, R. Mihalcea, and M. Chaturvedi, "Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 108–132, Jan. 2020.

[4] A. K. Ayyub, S. Iqbal, E. U. Munir *et al.*, "Exploring diverse features for sentiment quantification using machine learning algorithms," *IEEE Access*, vol. 8, pp. 142819–142831, 2020.

[5] S. Ramakrishnan and L. D. Dhinesh Babu, "Enhancing Twitter Sentiment Analysis using Attention-based BiLSTM and BERT Embedding," in *2023 9th International Conference on Smart Computing and Communications (ICSCC)*, Kochi, India, 2023.

[6] X. Xie, J. Cao, Z. Wu *et al.*, "Sentiment analysis of Chinese e-commerce reviews based on BERT," in *2020 IEEE 18th International Conference on Industrial Informatics (INDIN)*, Warwick, England, 2020, pp. 713–718.

[7] C. Wang, X. Yang, and L. Ding, "Deep learning sentiment classification based on weak tagging information," *IEEE Access*, vol. 9, no. 12, pp. 66509–66518, 2021.

[8] Sarojadevi Palani, Prabhu Rajagopal, and Sidharth Pancholi "T-BERT – Model for Sentiment Analysis of Micro-blogs Integrating Topic Model and BERT," *IEEE Conference Proceedings*, 2021.

[9] (Review) "A Review on Text Sentiment Analysis with Machine Learning and Deep Learning Techniques," *IEEE Xplore / IEEE Journals & Magazine*, 2022.

[10] F. Kamyab, G. Liu, and M. Adjeisah, "Attention-Based CNN and Bi-LSTM Model Based on TF-IDF and GloVe Word Embedding for Sentiment Analysis," *Applied Sciences*, vol. 11, no. 23, art. 11255, Nov 2021.

[11] Dataset Source: https://www.kaggle.com/datasets/