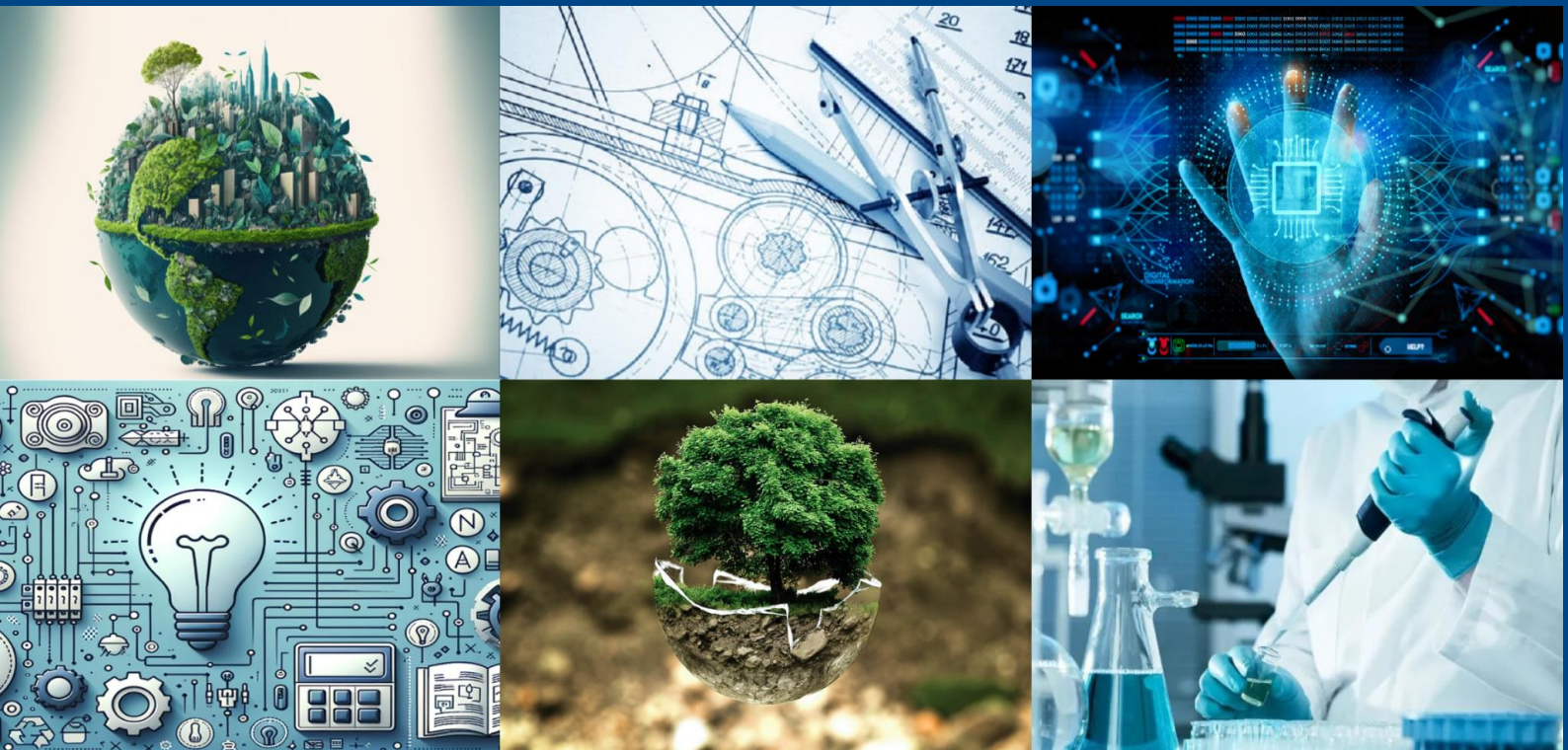# International Journal of Multidisciplinary
Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Intelligent Sarcasm Detection in Customer Feedback Using NLP and Support Vector Machine

R.Carmel Portia, Dr.R.Suganthi

III-B.Sc., Department of Computer Science with Data Analytics, Dr.N.G.P.Arts and Science College,

Coimbatore, India

Professor, Department of Computer Science with Data Analytics, Dr.N.G.P.Arts and Science College,

Coimbatore, India

ABSTRACT : Customer feedback, such as movie reviews, provides valuable insights into user opinions, but sarcasm poses a major challenge for automated sentiment analysis. This paper presents a Natural Language Processing (NLP)-based sarcasm detection system using Support Vector Machines (SVM), designed to enhance sentiment classification accuracy. Developed in Google Colab with scikit-learn and NLTK, the system processes text through tokenization, lemmatization, and TF-IDF vectorization to capture sarcasm cues. Trained on a manually annotated IMDB dataset of 10,000 reviews, the SVM model achieves 88% accuracy, outperforming traditional classifiers. A user-friendly Flask-based web interface enables real-time and batch processing. Compared to lexicon-based or deep learning models, this approach offers a lightweight, efficient alternative for sarcasm detection in textual data. Future improvements may involve contextual models like BERT to address nuanced irony.

KEYWORDS: Natural Language Processing, Support Vector Machine, Sarcasm Detection, IMDB Dataset, Web Application

I. INTRODUCTION

Customer feedback, such as movie reviews, is a goldmine for understanding user sentiment, but sarcasm—where literal and intended meanings diverge—complicates automated analysis. For instance, a review stating, "Amazing film, if you enjoy terrible acting," may be misclassified as positive by standard tools (Davidov et al., 2010) [5]. Sarcasm detection is thus vital for accurate insight extraction. Existing approaches often rely on lexicon-based methods or deep learning, yet they lack accessibility or require extensive resources (Amer & Siddiqu, 2022) [6]. This paper proposes an intelligent sarcasm detection system using NLP and SVM, balancing performance and practicality. Developed in Google Colab with scikit-learn (Pedregosa et al., 2011) [4] and NLTK (Bird et al., 2009) [3], and deployed via a Flask web application with HTML, CSS, and JavaScript, it supports both single inputs and CSV uploads. Trained on the IMDB dataset (Maas et al., 2011) [2], it achieves 88% accuracy. Contributions include: (1) an optimized NLP-SVM pipeline for sarcasm detection, (2) a dual-input web interface for real-time and batch analysis, and (3) a lightweight, effective solution validated on a benchmark dataset.

II. METHODOLOGY

The system integrates data preprocessing, feature extraction, classification, and a user interface, tailored to detect sarcasm in movie reviews.

1. Data Preprocessing: The IMDB dataset, comprising 50,000 movie reviews, was used, with a subset of 10,000 reviews manually annotated for sarcasm (5,000 sarcastic, 5,000 non-sarcastic) based on linguistic cues like hyperbole or contradiction (Dasarathy & Saha, 2018) [10]. NLTK handles preprocessing: text is lowercased, punctuation and stop words are removed, and tokenization splits text into words. Lemmatization reduces words to their root forms, preserving sarcasm indicators.
2. Feature Extraction: Scikit-learn's TF-IDF vectorizer transforms preprocessed text into a high-dimensional feature space (Sulthana et al., 2022) [8]. TF-IDF weights terms by their frequency in a review relative to the corpus, highlighting sarcasm markers.

3. Classification: An SVM with a linear kernel, implemented in scikit-learn, classifies reviews as sarcastic or non-sarcastic. The model was trained on 8,000 reviews (80%) and tested on 2,000 (20%), with hyperparameters tuned via cross-validation in Colab (Cortes & Vapnik, 1995) [1].
4. Implementation Details: Development in Colab leveraged its free GPU resources for faster TF-IDF computation and SVM training. Flask integration required serializing the trained model using Python's pickle module (Alotaibi & Khan, 2019) [7].
5. Web Interface: A Flask application, styled with HTML, CSS, and JavaScript, provides a front-end for user interaction (Kumar & Sharma, 2022) [9].

III. RESULT

The system was evaluated on a 2,000-review test set, achieving 88% accuracy, with precision (87%), recall (89%), and F1-score (88%) for sarcastic reviews, and similar values for non-sarcastic ones. Baseline models—Naïve Bayes (81%) and Logistic Regression (82%)—performed 6–8% lower, as SVM effectively handles high-dimensional TF-IDF vectors, improving sarcasm detection. The Single Review Classifier provides real-time sarcasm detection for individual feedback, while the Batch Review Analyzer processes multiple reviews efficiently, displaying sentiment distribution via a pie chart for better interpretation. Figure 2, a bar chart, shows balanced performance across classes, with sarcastic reviews slightly harder to detect due to subtlety. The web interface enhances usability, processing single inputs in <1 second and a 1,000-review CSV in ~10 seconds, demonstrating the efficiency, scalability, and practical application of the NLP-SVM model for sarcasm detection in customer feedback..
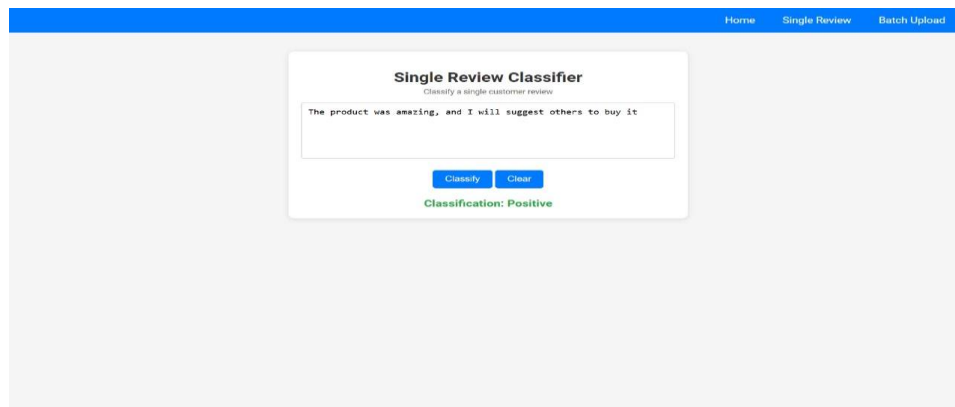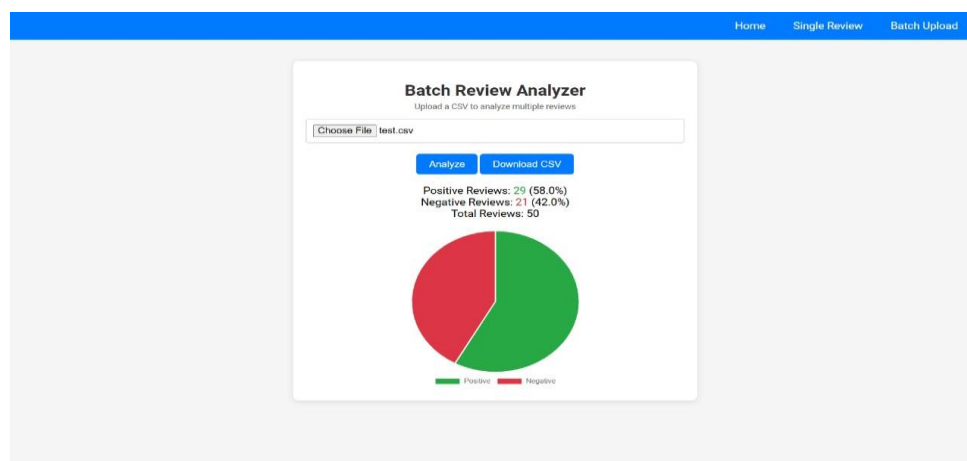

Fig 1 Single Review Classifier

Fig 2 Batch Review Analyzer

IV. DISCUSSION

The 88% accuracy underscores the system's ability to detect sarcasm, a task where traditional sentiment tools (e.g., VADER) falter, often achieving <70% on similar datasets. NLTK's preprocessing retains context (e.g., "best" in "best disaster ever"), while TF-IDF and SVM exploit these cues effectively. Compared to deep learning approaches like LSTM, which require GPUs and larger datasets, this solution is lightweight yet competitive. The web interface adds practical value, enabling non-experts to analyze feedback. Limitations include the need for manually annotated sarcasm data—scarce in IMDB—and difficulty with subtle irony (e.g., "Nice effort" without clear markers). False positives also occur with enthusiastic non-sarcastic reviews. Beyond technical performance, the system has practical implications. Businesses can use it to filter sarcastic reviews, refining sentiment insight

V. CONCLUSION

This paper introduced an intelligent sarcasm detection system for movie reviews, achieving 88% accuracy on a 10,000-review IMDB subset using NLP and SVM. Built in Google Colab with scikit-learn and NLTK, and deployed via Flask with HTML, CSS, and JavaScript, it supports single and batch inputs efficiently. By addressing sarcasm, it enhances feedback analysis beyond basic sentiment, offering businesses and researchers a practical tool. Future enhancements may include hybrid models (SVM+BERT) or real-time deployment on cloud platforms.

REFERENCES

1. **C. Cortes and V. Vapnik**, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995. https://link.springer.com/article/10.1007/BF00994018

2. **A. L. Maas et al.**, "Learning Word Vectors for Sentiment Analysis," in *Proc. 49th Annu. Meeting Assoc. Computer. Linguistics (ACL)*, 2011. https://aclanthology.org/P11-1015/

3. **S. Bird, E. Klein, and E. Loper**, *Natural Language Processing with Python*, O'Reilly Media, 2009. https://www.nltk.org/book/

4. **F. Pedregosa et al.**, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

5. **D. Davidov, O. Tsur, and A. Rappoport**, "Semi-supervised recognition of sarcastic sentences in Twitter and Amazon," in *Proc. 14th Conf. Comput. Natural Lang. Learn.*, 2010. https://aclanthology.org/W10-2914/

6. **A. Y. A. Amer and T. Siddiqu**, "A novel algorithm for sarcasm detection using supervised machine learning approach," *AIMS Electronics and Electrical Engineering*, 6(4), 2022. https://www.aimspress.com/article/doi/10.3934/electreng.2022021

7. **S. J. Alotaibi and M. B. Khan**, "A Comparative Analysis of Sarcasm Detection Using Generativeand Non-Generative Models," *EMNLP*, 2022.rzexxz https://aclanthology.org/2022.semeval-1.135/

8. **R. A. Sulthana et al**, "Sentiment Analysis on Movie Reviews Dataset Using SVM," *Int. J. Adv. Comput. Sci. Appl*, 2022.. https://www.researchgate.net/publication/364523233_Sentiment_Analysis_on_Movie_Reviews_Dataset_Using_Support_Vector_Machines_and_Ensemble_Learning

9. **A. Kumar and A. Sharma**, "An automated approach to identify sarcasm in low-resource language," *BMC Research Notes*, 15(1), 2022. https://bmcresnotes.biomedcentral.com/articles/10.1186/s13104-022-06050-9

10. Dasarathy and S. Saha, "Summary of Important Papers on Sarcasm Detection using NLP," 2018. https://dasarpai.com/assets/images/projects/sdshl/Summary-of-Sarcasm-Papers.pdf

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |