International Centre for Education and Research (ICER)
VIT – Bangalore

(Review 0)

Sentiment Analysis Using Python and Machine Learning

Submitted by

Devjyot Singh Sidhu

(24MSP3075)

Abstract

Sentiment analysis is an essential tool in natural language processing that determines the sentiment of a given piece of text. This project focuses on analysing sentiments from textual data using Python and machine learning techniques. The IMDB movie review dataset, containing 50,000 reviews, was used to build and evaluate a model that classifies reviews as positive or negative. We employed text preprocessing, feature extraction using the Bag of Words approach, and trained a Naive Bayes classifier. The model achieved 85% accuracy, demonstrating its effectiveness in distinguishing between positive and negative reviews. However, limitations were observed in handling sarcasm and ambiguous text, which could be addressed in future work.

Problem Statement

Sentiment analysis is widely used in industries to gauge customer satisfaction, monitor brand reputation, and derive actionable insights. Companies rely on user-generated data, such as reviews and social media posts, to understand customer sentiments and improve their products or services.

References

1. Shaukat, Z., Zulfiqar, A.A., Xiao, C. et al. Sentiment analysis on IMDB using lexicon and neural networks. SN Appl. Sci. 2, 148 (2020)

2. Pang, B., & Lee, L. (2004). "A Sentimental Education: Sentiment Analysis Using Machine Learning Techniques." Presented at the Association for Computational Linguistics (ACL).

3. Zhang, L., Wang, S., & Liu, B. (2018). Deep Learning for Sentiment Analysis: A Survey. Wiley Interdisciplinary Reviews.

4. "Sentiment Analysis Using Machine Learning and Deep Learning Models on IMDB Dataset." IEEE Xplore, 2024.

5. "A Literature Review on Application of Sentiment Analysis Using Machine Learning." SSRN Papers, 2023.

6. "Sentiment Analysis: Machine Learning Approaches Comparison." IEEE Xplore, 2024.

7. Tumasjan, A., et al. (2010). "Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment." Proceedings of the International Conference on Web and Social Media (ICWSM).

8. "Optimization of Sentiment Analysis Using Machine Learning Classifiers." SpringerOpen, 2023.

9. Kaggle Dataset: *https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews*