

Sentiment Analysis Project Report

Introduction

Sentiment analysis is a Natural Language Processing (NLP) technique used to determine whether text expresses a positive or negative emotion. This project demonstrates a simple sentiment analysis model built using Python. The objective is to preprocess text data, convert it into numerical features, train a machine learning model, and evaluate its performance.

Dataset Description

A small dataset simulating customer reviews or tweets was created. Each text sample is labeled as positive or negative. This dataset helps demonstrate how sentiment classification works in real-world applications such as product reviews and social media analysis.

Data Preprocessing

Text preprocessing is an important step in NLP. The text is converted to lowercase, punctuation is removed, and stopwords are filtered out. This cleaning process ensures that only meaningful words remain for analysis, improving model performance.

Feature Engineering

TF-IDF (Term Frequency-Inverse Document Frequency) is used to convert text into numerical form. This technique highlights important words while reducing the importance of common words. The transformed data becomes suitable for machine learning algorithms.

Model Implementation

A Logistic Regression classifier is trained on the processed dataset. The dataset is split into training and testing sets to evaluate performance. The model learns patterns associated with positive and negative sentiments.

Evaluation

Model accuracy and classification metrics are used to measure performance. A confusion matrix visualizes prediction results. The model demonstrates strong performance for a simple dataset and

shows the effectiveness of NLP techniques.

Conclusion

This project successfully demonstrates sentiment analysis using Python and NLP. The workflow includes preprocessing, feature extraction, model training, and evaluation. The same approach can be applied to larger real-world datasets such as tweets or customer reviews for business insights.