

Social Media Sentiment Analysis

AI & Data Science Internship Project — Logicbox IT Solutions

1. Project Overview

This project was developed as part of a three-month internship at **Logicbox IT Solutions**. The objective was to apply **Natural Language Processing (NLP)** and **Machine Learning (ML)** techniques to classify social media content (tweets) into sentiment categories: **Positive** and **Negative**.

The project simulates a real-world use case where sentiment analysis supports digital marketing teams in measuring audience engagement and optimizing content strategies.

2. Dataset Description

- **Source:** Sentiment140 Dataset (Kaggle)
 - **Sample Size Used:** 10,000 tweets (balanced)
 - **Columns Used:** text, sentiment
 - **Labels:**
 - 0 = Negative
 - 4 = Positive*(Neutral class excluded for binary classification)*
-

3. Data Preprocessing

The following preprocessing steps were applied:

- Removal of:
 - Mentions (e.g., @user)
 - Hashtags
 - URLs
 - Non-alphabetic characters
- Lowercasing of text
- Tokenization and stopword removal
- Word stemming using Porter Stemmer

- TF-IDF vectorization of cleaned text
-

4. Technologies Used

- **Programming Language:** Python
 - **Libraries:**
 - Pandas, NumPy
 - NLTK (Natural Language Toolkit)
 - Scikit-learn
 - Matplotlib, Seaborn
 - **Platform:** Google Colab
 - **Version Control:** GitHub
-

5. Model Summary

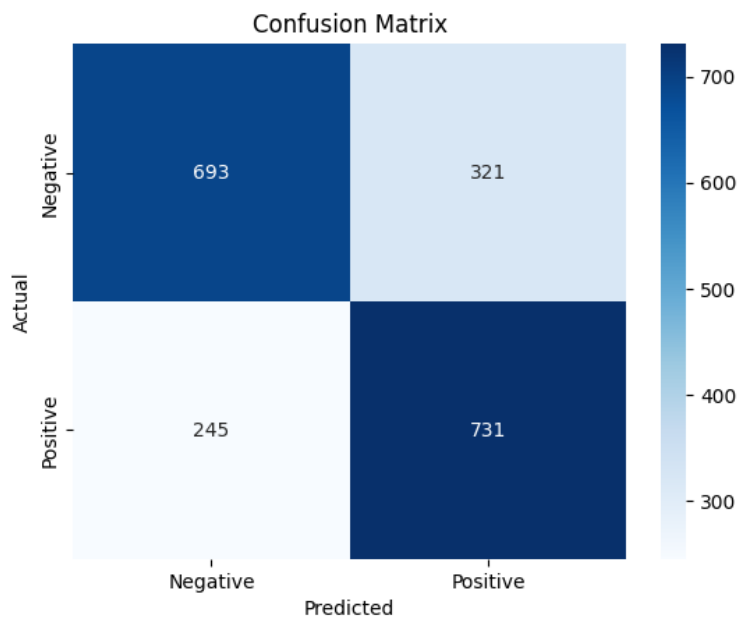
- **Model Used:** Logistic Regression
 - **Text Representation:** TF-IDF (Top 5,000 Features)
 - **Split:** 80% Training / 20% Testing
 - **Evaluation Metrics:**
 - Accuracy
 - Precision
 - Recall
 - F1-Score
 - Confusion Matrix
-

6. Model Performance

Final evaluation on the test dataset (1,990 samples):

- **Accuracy:** 71.56%
- **F1-Score:** ~0.72
- **Precision (Positive):** 0.69
- **Precision (Negative):** 0.74

Confusion Matrix:



	Predicted Negative	Predicted Positive
Actual Negative	693	321
Actual Positive	245	731

7. Internship Context

This project was carried out under the internship at **Logicbox IT Solutions**, a company focused on digital marketing, application development, and customer engagement solutions. The project reflects an internal contribution exploring how AI/ML techniques can be applied to analyze public sentiment and support marketing strategy.

8. Contact

For any queries or collaborations, please connect via

GitHub : <https://github.com/Mayankojhaa>.

LinkedIn : <https://www.linkedin.com/in/mayankojha2228>.