

SENTIMENT ANALYSIS

Movie reviews

BY Rokayya Aly

Review ----> ML model ----> Positive vs Negative

ML model

- Data
- Preprocessing
- Logistic Regression Model

TOOLS

- **NLTK** : (Natural Language Toolkit). **Stopword Removal:** Filtering out common words ("the," "is"...) that don't contribute to sentiment. **Tokenization:** Splitting text into words or sentences. **Corpus and Datasets:** Provides datasets like Movie Reviews for training sentiment analysis models.
- **Sci-kit learn:** for data preprocessing, model training, and evaluation. **Text Vectorization (Feature Extraction):** Converts text into numerical. CountVectorizer → Converts words into a frequency matrix. TF-IDF (Term Frequency - Inverse Document Frequency) → Gives importance to words based on their frequency in a document vs. the whole dataset.
- **Matplot**

- **Wordcloud** : Visualizing Important Words: Helps identify the most commonly used words in positive and negative sentiment texts. Useful for exploratory data analysis (EDA). Helps in selecting important words that might be useful
- **Streamlit**: makes it easy to build interactive web application. It is widely used to create real-time sentiment analysis apps. **Integration with ML Libraries** → Works with Scikit-learn, NLTK, TensorFlow, etc. **Real-time Analysis** → Allows users to input text and get sentiment predictions instantly. **Interactive Widgets** → Add text input, buttons, sliders, and graphs with one line of code.
- **Numpy**
- **Pandas**

Summary of the Code

- 1. Load Data** → Read the IMDB dataset.
- 2. Preprocess Text** → Remove stopwords, punctuation, and lowercase the text.
- 3. Visualize Data** → Generate WordClouds for positive/negative sentiment.
- 4. Feature Engineering** → Convert text into numerical vectors using TF-IDF.
- 5. Train Model** → Use Logistic Regression to classify reviews as positive/negative.
- 6. Evaluate Model** → Generate a confusion matrix to check model performance.
- 7. Save Model** → Store the trained model and vectorizer using Pickle.

Thank You