

PROJECT REPORT

MOVIE REVIEW SENTIMENT ANALYSIS

Prepared by: Muhammad Ahmed Naeem

APPROACH

Approach

1. Data Preprocessing:

- Text Cleaning: Lowercasing, removing HTML tags, stopwords
- Tokenization: Split text into words
- Feature Extraction: Used TF-IDF for sentiment context.

2. Model Training:

- Use SVM, LOGISTIC REGRESSION AND MULTI-NOMIAL NAIVE-BAYES FOR OPTIMAL MODEL TRAINING

3. Evaluation:

- Metrics: Accuracy, f1-score
- Confusion Matrix: Visualized positives/negatives values.

CHALLENGES FACED:

The main issue was to tokenize on the large dataset for long time processing and models comparison analysis

TECHNIQUES AND IMPROVEMENTS:

- Use stemmer and pre-process logic to easily find the negative and positive reviews of datasets and its improvement of datasets.