# Project Report: Multimodal Rating Analysis with NLP

## 1. Introduction

This project focuses on classifying text-based data by integrating Natural Language Processing (NLP) techniques with structured data analysis. The objective is to compare multiple machine learning models and determine the best-performing one for classification tasks.

## 2. Technologies Used

- Python
- NLP
- Machine Learning Algorithms
- Pandas
- Scikit-learn library
- Matplotlib
- NLTK
- Numpy

## 3. Methodology

### 3.1 Data Preprocessing

- Loaded data.

- Data Exploration & Cleaning

- Performed text cleaning, tokenization, and sentiment extraction.

- Applied TF-IDF transformation to text features.

- Encoded categorical variables using one-hot encoding.

### 3.2 Model Selection

The following models were tested:

- Random Forest Classifier

- Gradient Boosting Classifier

A combined approach combining structured features and text-based TF-IDF features , sentiment feature and one hot encoded column features was implemented to enhance model performance.

### 3.3 Training and Evaluation

- Used **train-test split (80-20%)** to separate data.

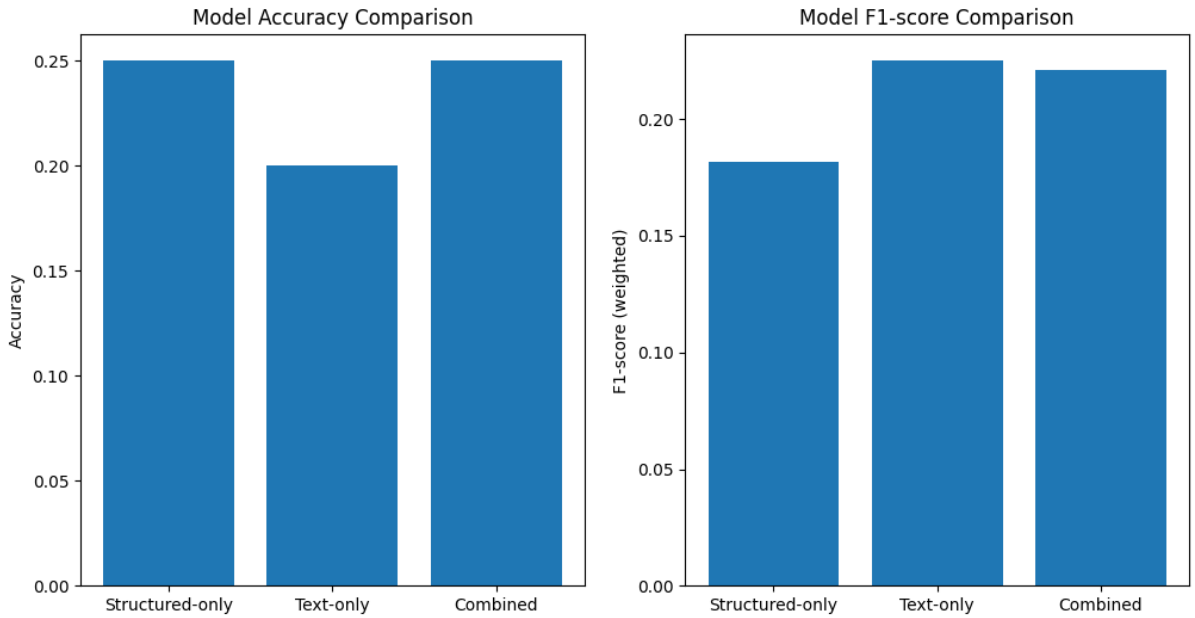
- Evaluated models on **Accuracy, F1-score, and Classification Report**.

### 4. Conclusion

- NLP techniques combined with structured data features improve classification accuracy.

- Random Forest Classifier is the recommended model due to its good performance.

### 5. Visualizations for Model Accuracy Comparison



## 6. Visualizations for NLP