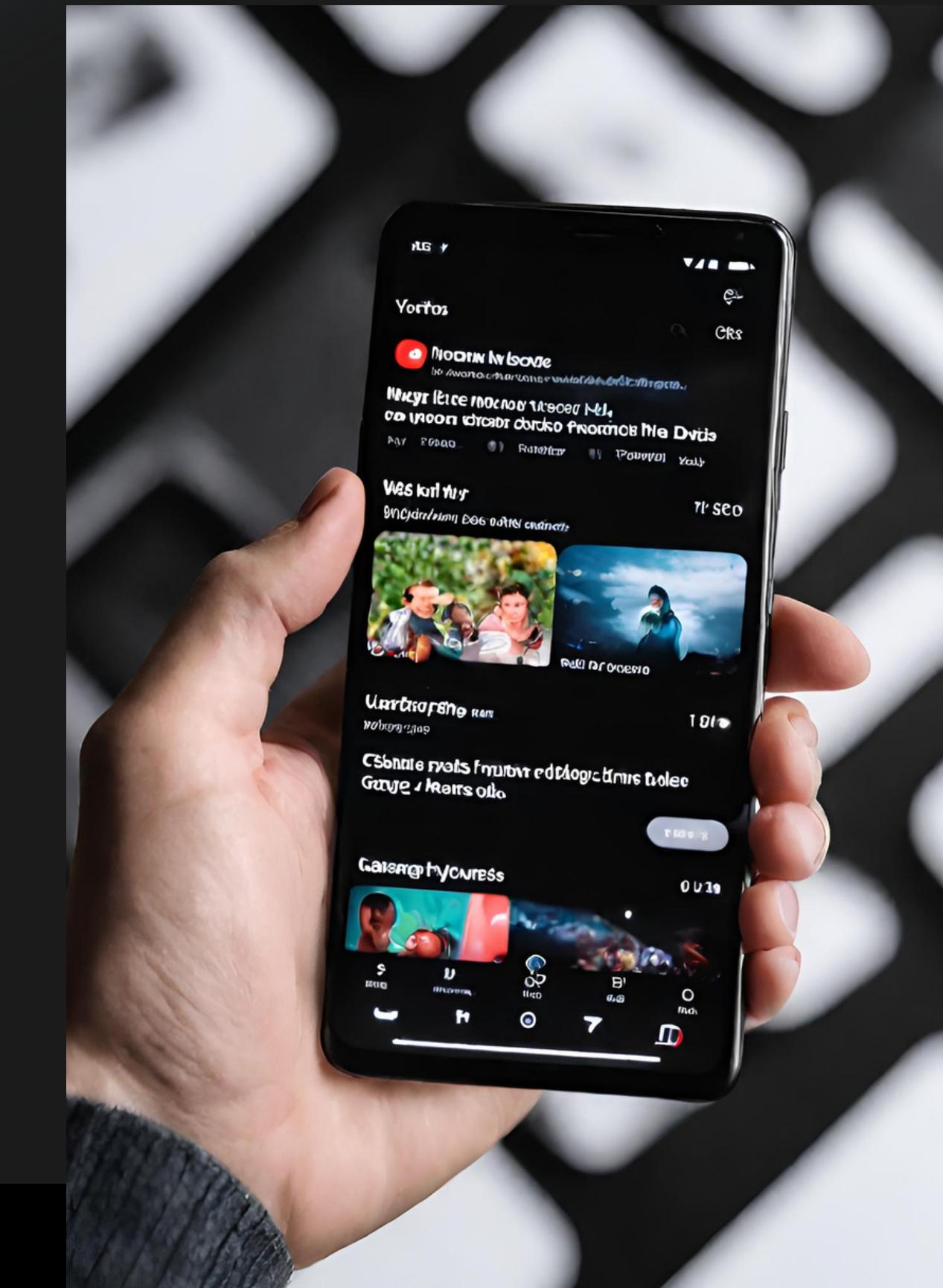


YouTube Comments Sentiment Analyzer

ARYA P G
PRATHAM K
MUGDHA S
SHREYAS S



Comment Sentiment Analyzer

It classifies the youtube
comments as positive,negative
and neutral.

Pinned by Faisal Khan
@FasBeam 2 years ago
Thank you #FasBeamers for all the love, we are 1.3 million strong now!
Like 274 Reply
▼ 23 replies

@anshbohra200 2 years ago
I spotted this car few months back in Mumbai and today i'm watching that car's review video from one of my favourite youtuber 😊
Like 351 Reply
▼ 7 replies

@hindutva5071 2 years ago
The energy in today's vlog was more in faisal...even he was excited with us ..
Like 158 Reply

@damirsethi 2 years ago

ABOUT OUR MODEL

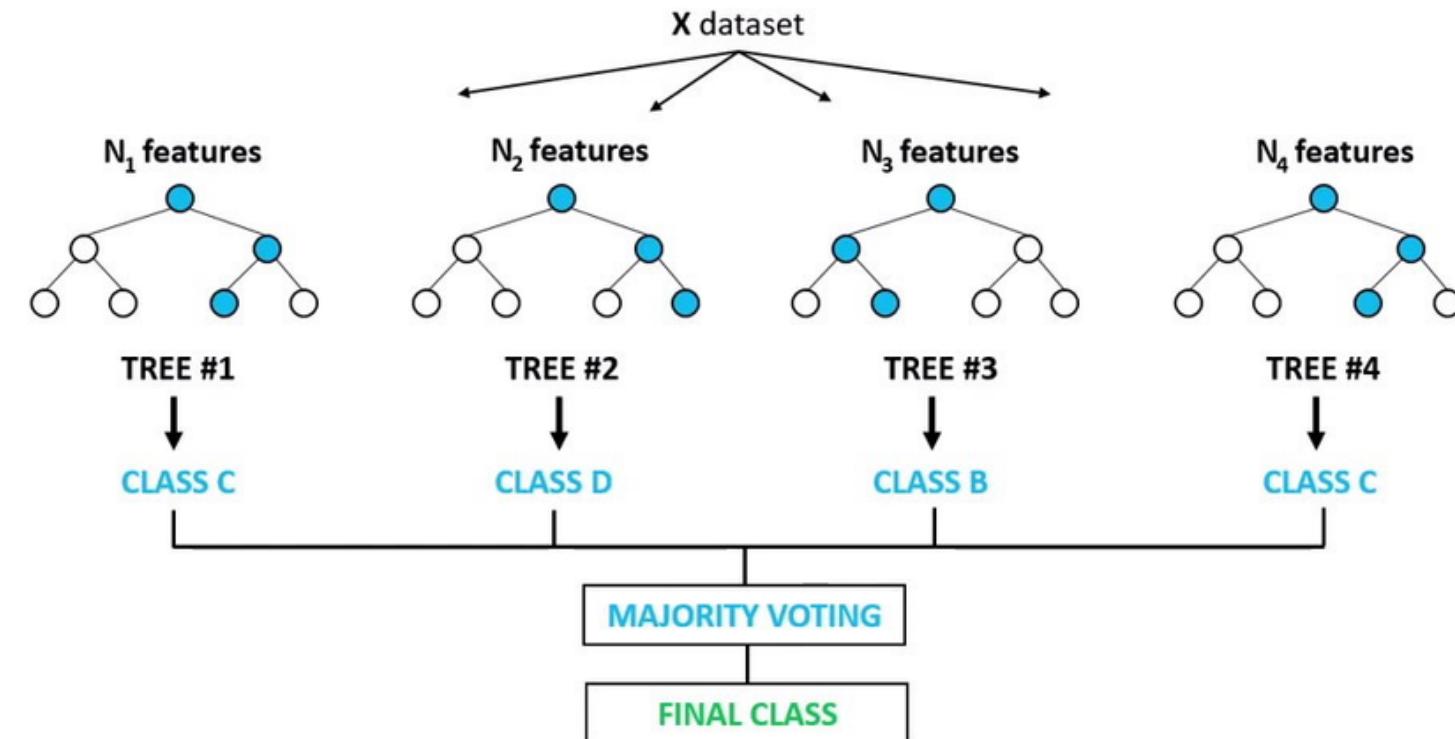
Inputs:

Youtube video link/ID

Outputs:

Sentiment of the particular
YT video's comments

Random Forest Classifier



Dataset Details

- Number of samples **varies** with respect to the youtube videos
- The independent variable: **Comments column**
- It is categorized into three classes
 - Positive +
 - Negative -
 - Neutral 😊



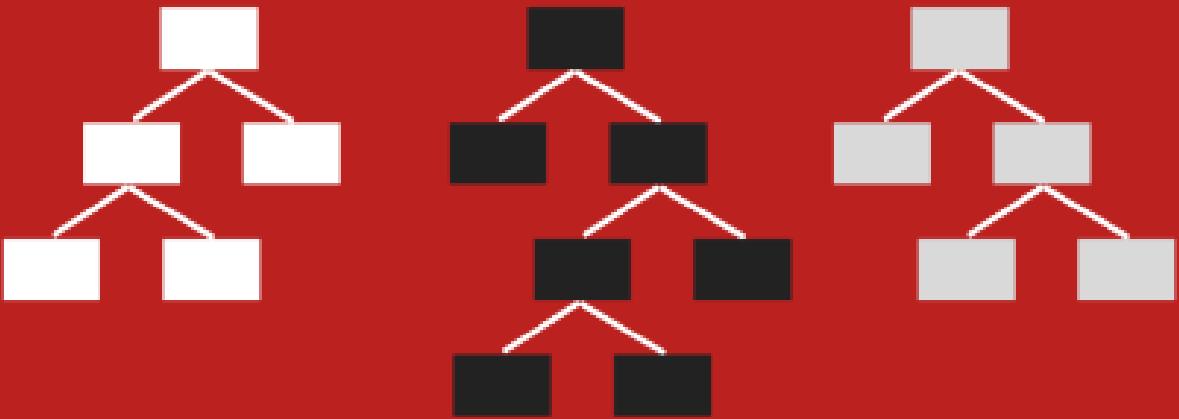
The Model used is Random Forest Classifier



Handling Non-Linearity:

- Random Forest is a powerful ensemble learning method that can effectively capture non-linear relationships and interactions between features in the data.
- This is particularly beneficial for sentiment analysis, as sentiment expression in text data often involves complex and non-linear patterns.

RANDOM FOREST



CATEGORY: SUPERVISED LEARNING

SUB-CATEGORY: CLASSIFICATION & REGRESSION



Robustness Against Overfitting:

Random Forest models are less prone to overfitting compared to some other complex models like deep neural networks.

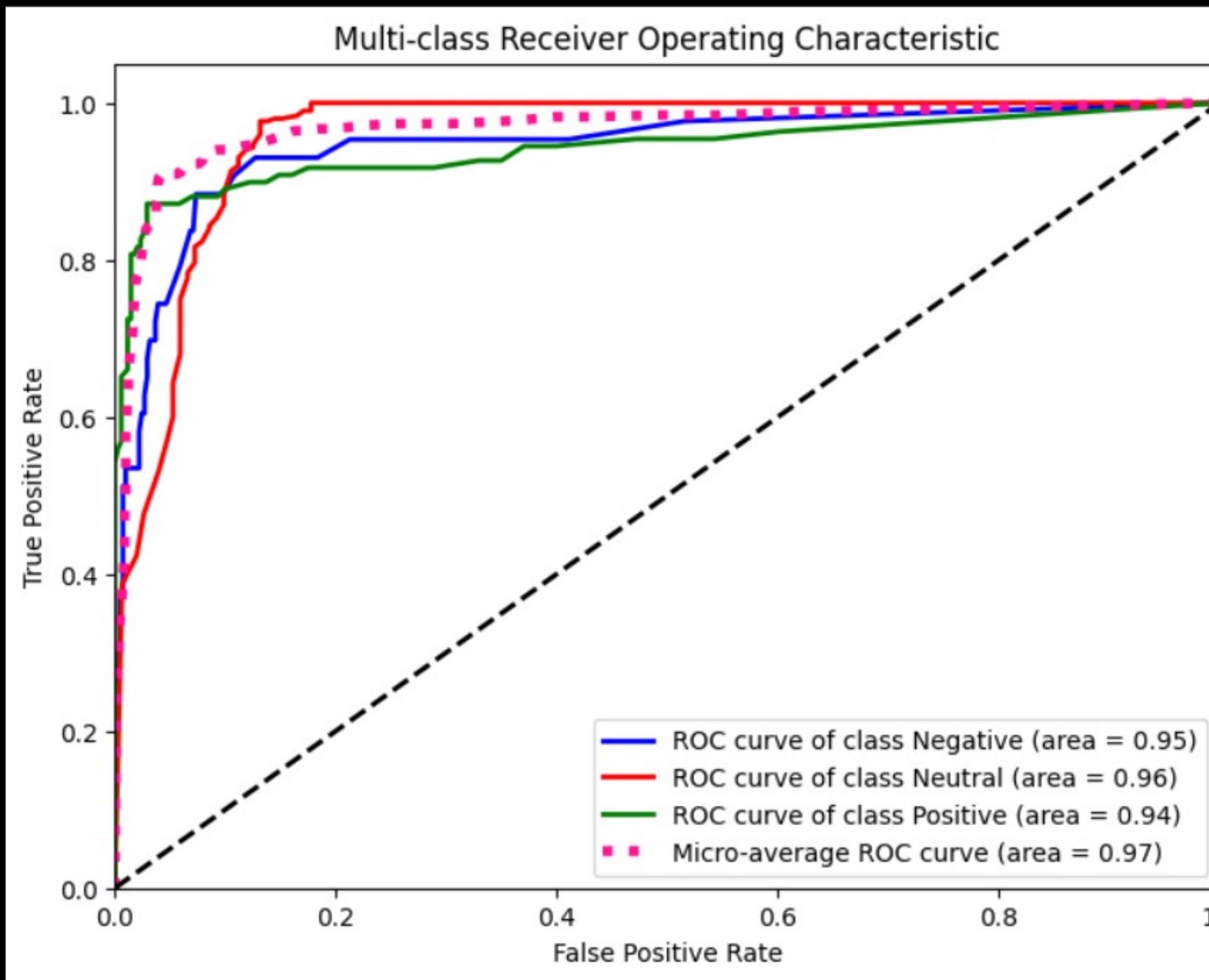
This is crucial when dealing with sentiment analysis tasks where the model needs to generalize well to unseen data and avoid memorizing noise or specific instances from the training set



Efficiency and Scalability:

Random Forest can handle large datasets efficiently and can scale well with increasing data size.

Preliminary Results



- Accuracy of the model obtained is **93%**
- Precision:**
 - Negative: 0.78
 - Neutral: 0.93
 - Positive: 0.98
- Recall:**
 - Negative: 0.54
 - Neutral: 1.00
 - Positive: 0.81
- F1-score:**
 - Negative: 0.64
 - Neutral: 0.96
 - Positive: 0.88
- AUC using the RoC curve:**
 - AUC for class Negative: 0.95
 - AUC for class Neutral: 0.96
 - AUC for class Positive: 0.94
 - Micro-average AUC: 0.97

Further PLAN



HYPERPARAMETER TUNING:

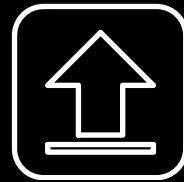
- **Class Weight (class_weight):** If your dataset is imbalanced, you can use this parameter to assign weights to different classes. Setting it to "balanced" will automatically adjust the weights inversely proportional to class frequencies.
- **Bootstrap Sampling (bootstrap):** a parameter determines whether bootstrap samples are used when building trees. It is usually set to True, but you can experiment with setting it to False to see if it improves performance.
- **Minimum Samples Leaf (min_samples_leaf):** A parameter that specifies the minimum number of samples required to be at a leaf node.



CROSS-VALIDATION AND ENSEMBLE METHODS:

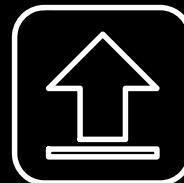
- Implementation of **k-fold cross-validation** and **stratified sampling**
- Explore ensemble methods like **Bagging and Boosting**

Further PLAN



MODEL INTERPRETABILITY AND DEPLOYMENT:

Enhance model interpretability by visualizing decision trees and utilizing **SHAP** values or **LIME** for individual prediction explanations



WEBSITE DEPLOYMENT:

- Design a minimalistic and intuitive **user interface** (UI) that allows users to effortlessly input comments for sentiment analysis. Incorporate responsive text fields and a clear, engaging submit button.
- Enhance the user experience with interactive elements for guidance and smooth **transition effects between the input and results** display to keep the user engaged and informed throughout the process.

...



Thank you!

