Text

Description automatically generated

**Activity based**

**Project Report on**

**DWDM**

**Submitted to Vishwakarma University, Pune**

**Under the Initiative of**

**Contemporary Curriculum, Pedagogy, and Practice (C2P2)**

**By**

**Arun Adhikari**

**SRN No : 202100406**

**Roll No : 38**

**Div : G**

**Third Year Engineering**

**Department of Computer Engineering**

**Faculty of Science and Technology**

**Academic Year**

**2023-2024**

**Contents**

1. **Problem Statement**
2. **Objectives**
3. **Data Set**
4. **Data Source**
5. **Decision Tree Diagram**
6. **Code**
7. **Conclusion**

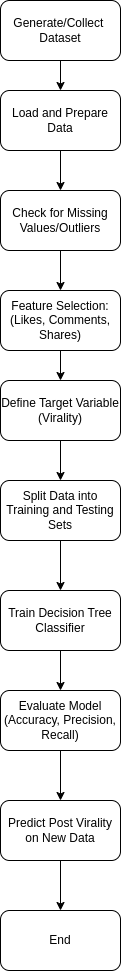
**Problem Statement**

Use Decision Tree classification in the social media domain to predict post virality based on engagement metrics like likes, shares, and comments

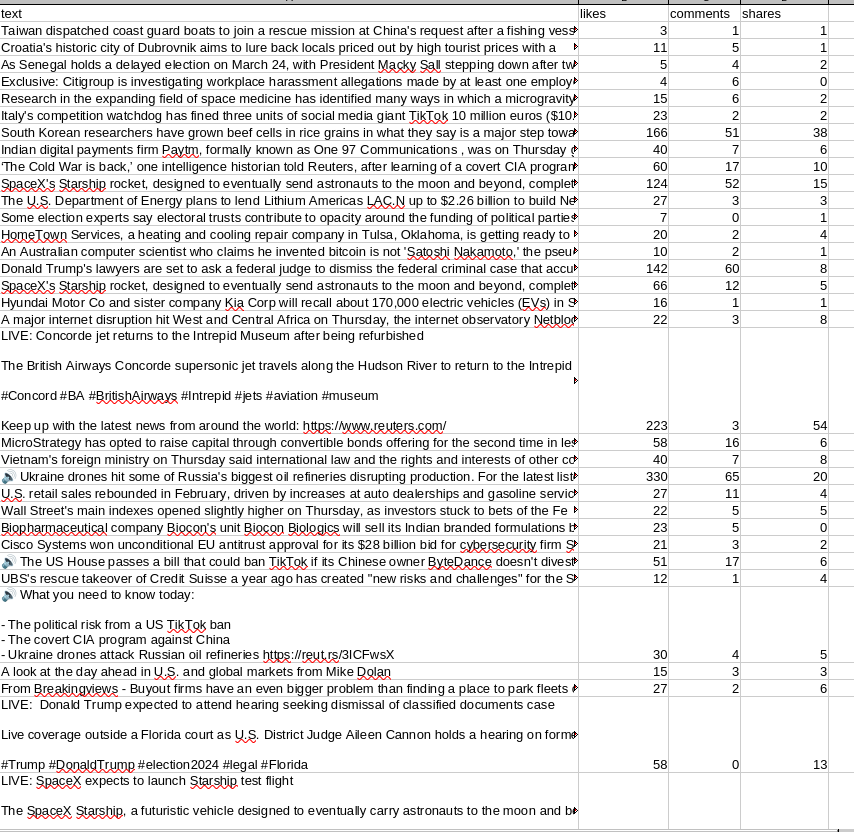
**Objectives**

**Data Collection and Preprocessing**: Gather and preprocess data from social media platforms, focusing on key engagement metrics such as likes, shares, comments, and other relevant features.

* **Feature Selection**: Identify the most influential engagement metrics that contribute to the virality of posts.
* **Model Development**: Develop a Decision Tree classification model that uses engagement metrics to predict whether a post will go viral.
* **Evaluation**: Assess the performance of the Decision Tree model using appropriate metrics such as accuracy, precision, recall, and F1-score.
* **Optimization and Fine-tuning**: Optimize the model by tuning hyperparameters and comparing performance with alternative models (if necessary) to ensure the highest predictive accuracy.
* **Insights and Recommendations**: Provide insights into which engagement metrics have the most significant impact on post virality and offer actionable recommendations for content optimization strategies.

****

**Dataset:**

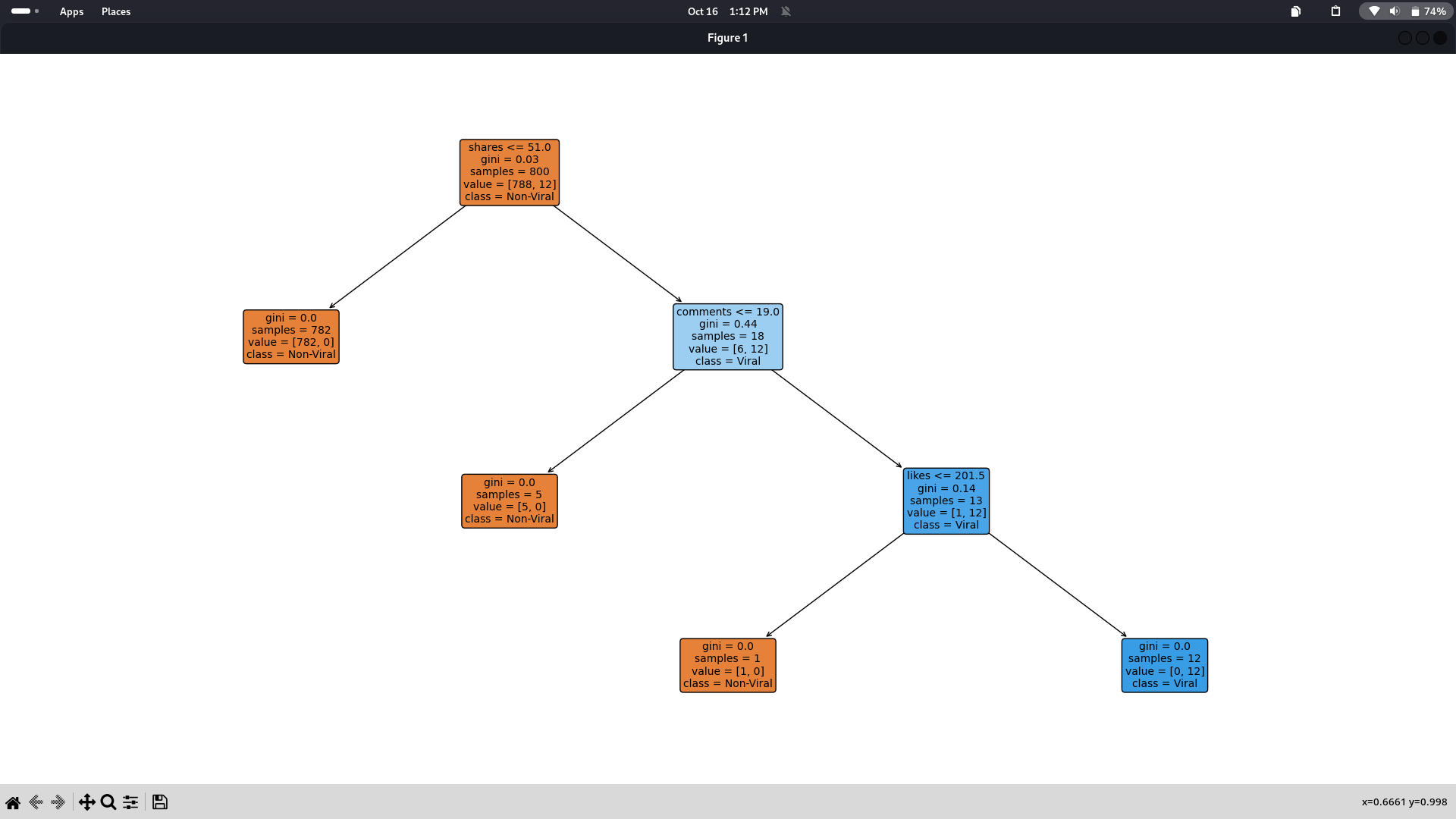
****

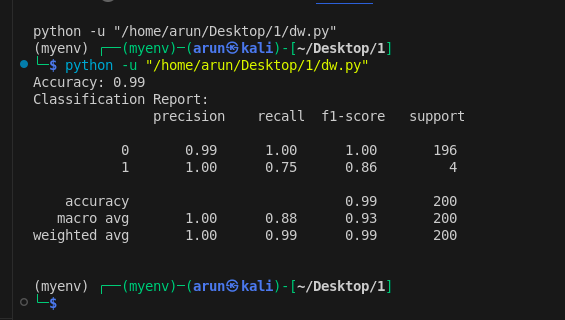
* text: The content of the post (this may require text processing if included in the model).
* likes: Number of likes the post received.
* comments: Number of comments the post received.
* **shares: Number of shares the post received.**

### **Data Source:**

The data for this project was sourced from **Kaggle**, a well-known platform for sharing datasets and conducting data science competitions. The dataset contains a comprehensive collection of social media posts along with key engagement metrics, including likes, shares, comments, and additional features relevant to measuring post performance. Kaggle datasets provide a reliable foundation for analyzing patterns and trends in social media engagement, offering rich, diverse data that is critical for training and validating the Decision Tree classification model used in this project. The dataset underwent preprocessing steps to ensure quality and relevance, enabling accurate predictions of post virality.

**Decision Tree Diagram:**



****

**Code:**

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.tree import DecisionTreeClassifier

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.metrics import classification\_report, accuracy\_score

file\_path = 'reuters.csv'

data = pd.read\_csv(file\_path)

data['Viral'] = ((data['likes'] > 200) & (data['shares'] > 50) & (data['comments'] > 20)).astype(int)

X = data[['likes', 'comments', 'shares']]

y = data['Viral']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

dt\_classifier = DecisionTreeClassifier(random\_state=42)

dt\_classifier.fit(X\_train, y\_train)

y\_pred = dt\_classifier.predict(X\_test)

accuracy = accuracy\_score(y\_test, y\_pred)

classification\_rep = classification\_report(y\_test, y\_pred)

print(f"Accuracy: {accuracy:.2f}")

print("Classification Report:\n", classification\_rep)

from sklearn.tree import plot\_tree

import matplotlib.pyplot as plt

fig, ax = plt.subplots(figsize=(25,15))

plot\_tree(dt\_classifier,

feature\_names=X.columns,

class\_names=['Non-Viral', 'Viral'],

filled=True,

proportion=False,

rounded=True,

precision=2,

fontsize=10,

ax=ax)

plt.subplots\_adjust(left=0.05, right=0.95, top=0.95, bottom=0.05)

plt.show()

**Conclusion:**

This project successfully explored the use of Decision Tree classification to predict the virality of social media posts based on engagement metrics such as likes, shares, and comments. By analyzing these key indicators of user interaction, we developed a model capable of classifying whether a post is likely to go viral. The results highlight the predictive power of engagement metrics in determining the potential reach of content, offering valuable insights for content creators and marketers.

The Decision Tree model provided interpretable results, revealing the most influential factors that contribute to post virality. By understanding the significance of various engagement metrics, users can strategically optimize their content to maximize visibility and engagement. The project lays the foundation for future improvements, such as incorporating additional features and exploring more advanced machine learning techniques, to further enhance prediction accuracy and applicability across diverse social media platforms.