# HW2_208742890_206661662_207188038

12/6/2025

# Introduction

**In this homework, we applied unsupervised machine learning techniques—specifically clustering algorithms—to explore and group similar data points without using labels. Our goal was to identify meaningful patterns within the dataset, evaluate different clustering methods such as K-Means and Agglomerative Clustering, and visualize the results using PCA.**

**We also analyzed cluster quality using metrics like the Elbow Method and Silhouette Score to determine the optimal number of clusters.**

# SOME OF OUR FEATURES:

## Engagement Rate

**The Engagement Rate feature measures audience interaction by calculating the ratio of likes and comments to views. It shows how actively viewers engage with the content, regardless of how many saw it.**

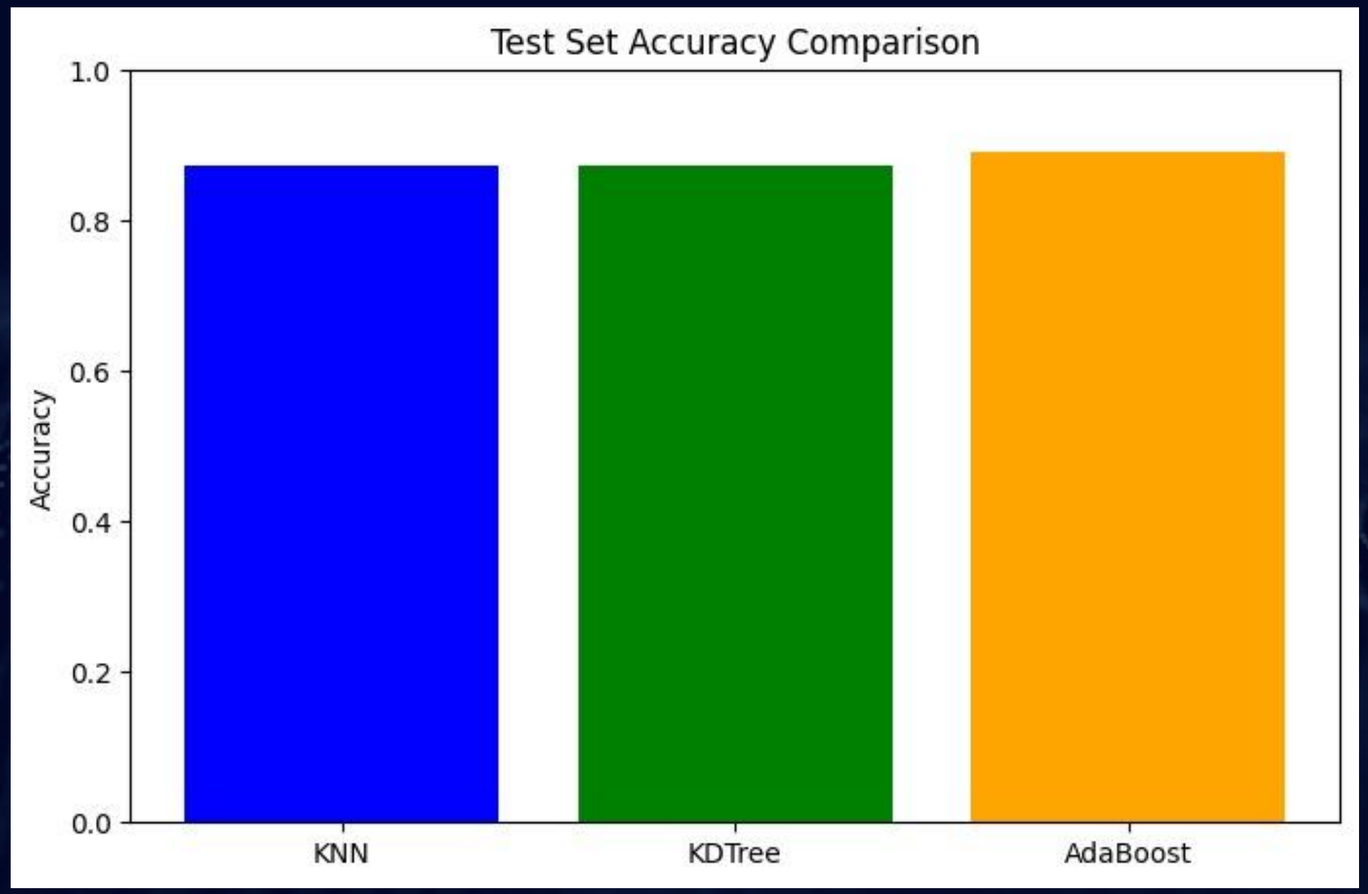
## official_licensed_views

**The feature official_licensed_views combines the number of views with indicators of whether the video is official and licensed. It gives higher weight to views of videos that are both official and licensed, reflecting more credible or professional content.**

## track_equals_album

**Songs that share the same name as their album are often lead singles or title tracks. They may carry unique characteristics in terms of popularity, importance within the album, or stylistic emphasis.**

# CLASSIFYNG

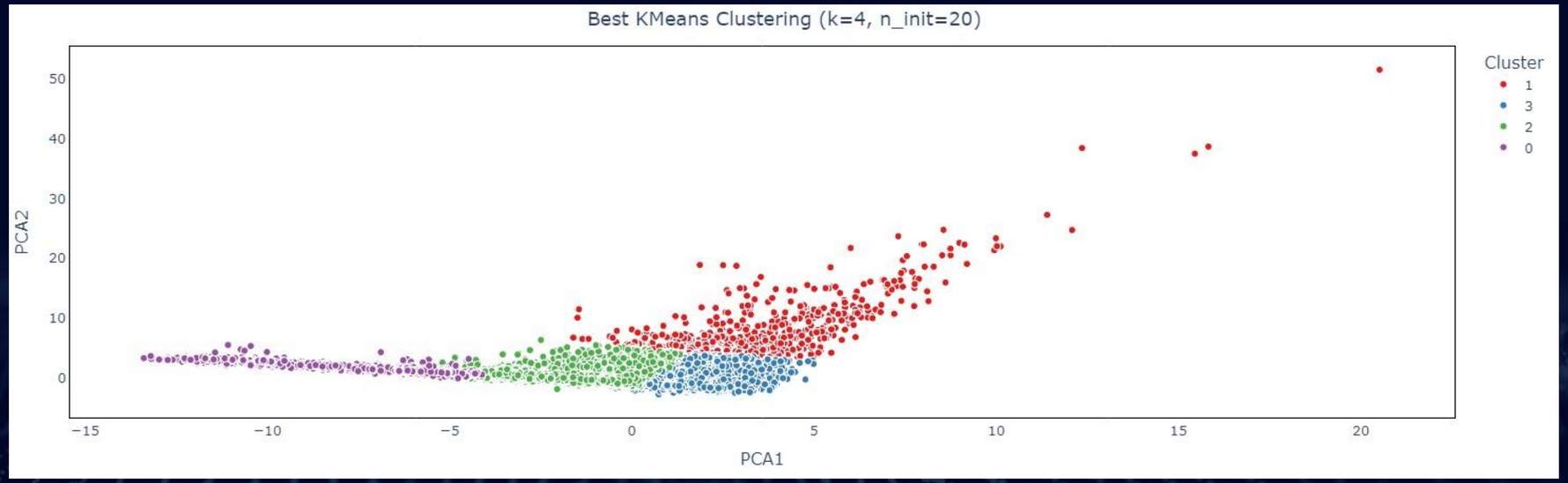



## Classifiers Accuracy Results:

**As you can see, we chose KNN, KDTree and AdaBoost as classifiers.**
**The accuracy test shows that all of the models have over 80 present accuracy, which tells us that the models went through well**

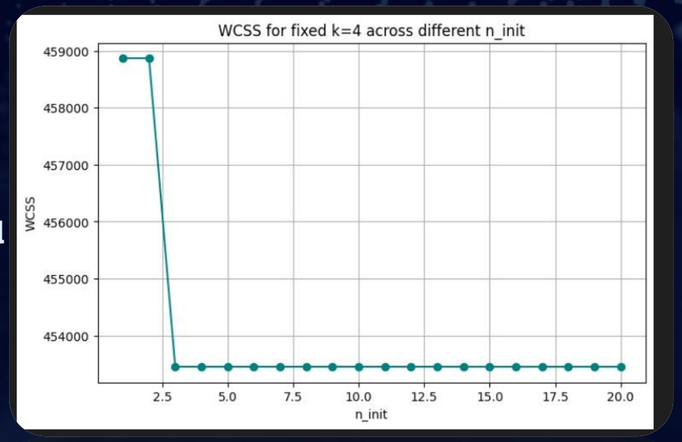
# K- MEANS ALGORITHM

## K-Means Graph Result:



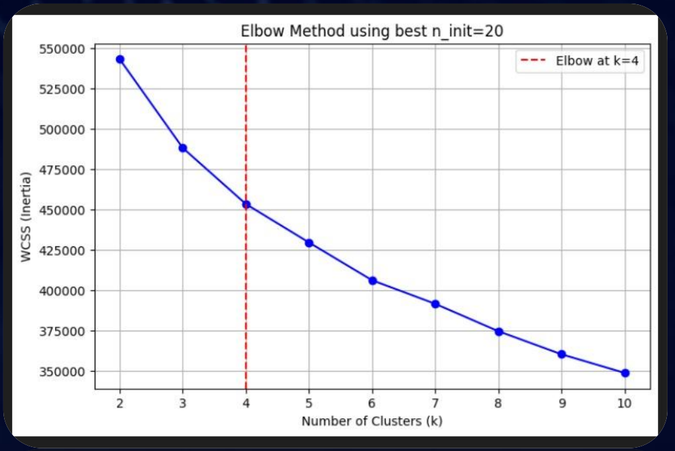

### N- PICKING

**We chose k randomly, and calculated the best n by lowest WCSS**




### K- PICKING

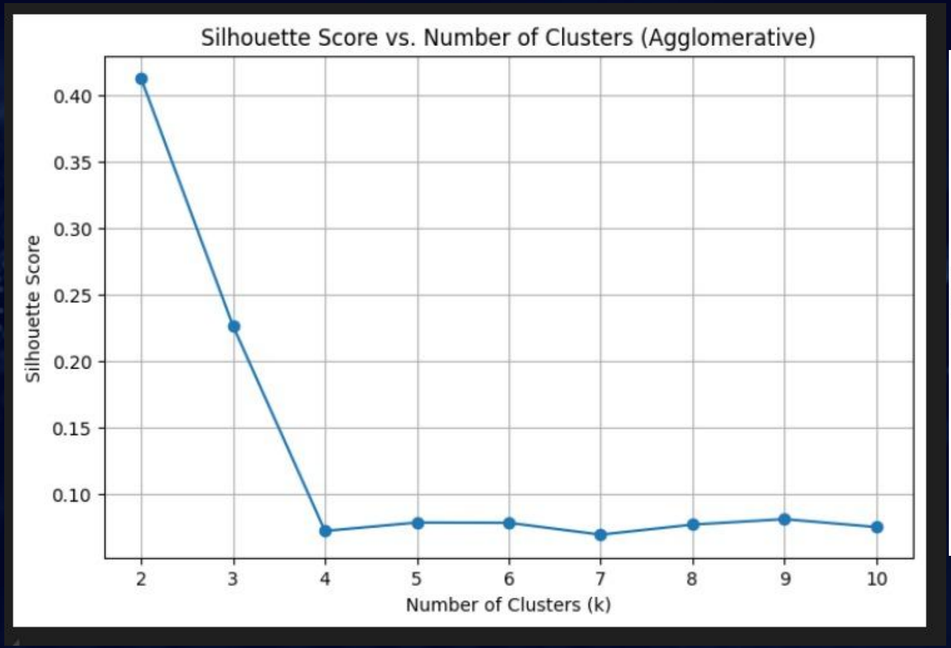
**We chose the best K using the elbow method**



# AGGLOMERATIVE CLUSTERING

## Silhouette Score:

**We Calculated this score to find the best k for agglomerative clustering**




## Agglomerative Graph Result