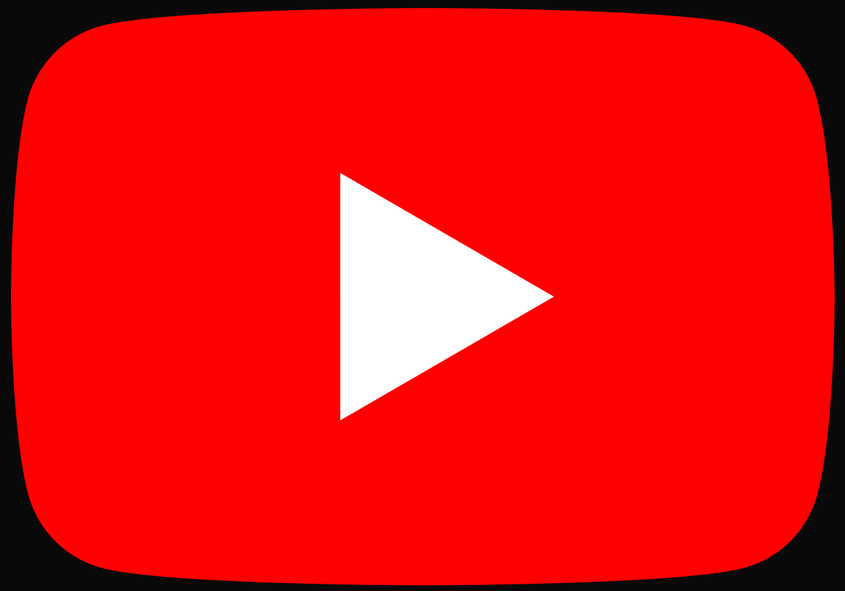
Youtube Data Analysis and Like Predictor

Submitted By - Naveen and Vallika



VIEWS

LIKES

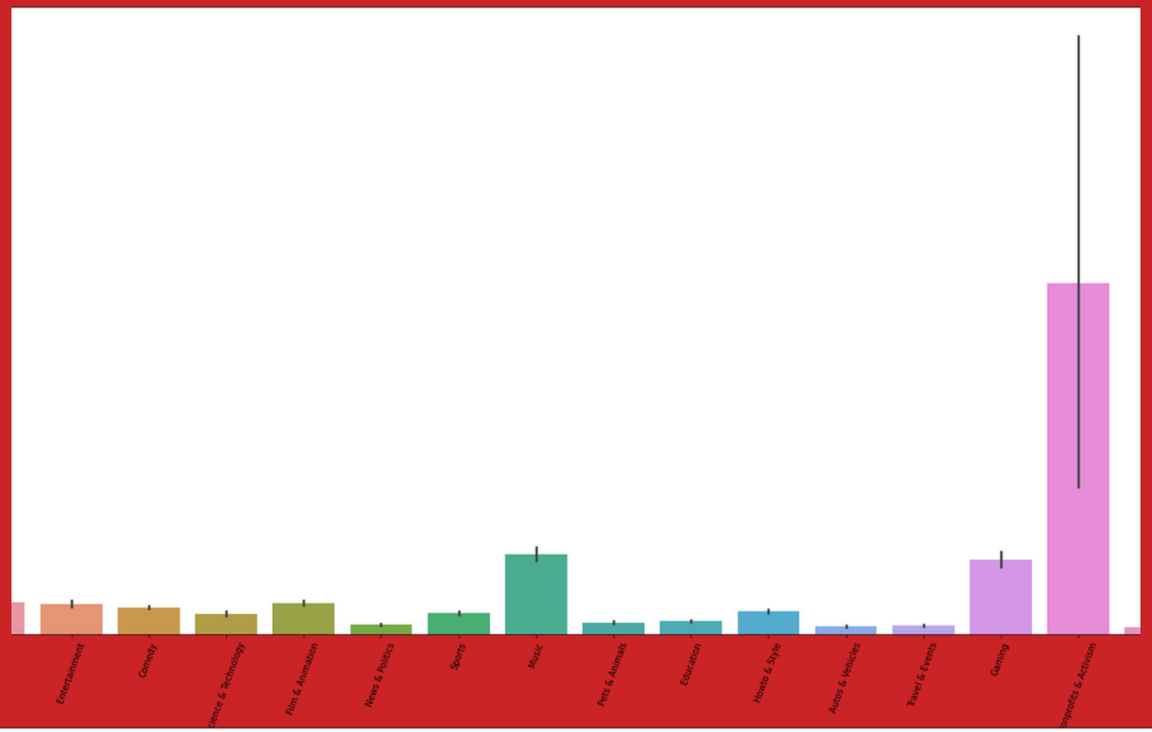
DISLIKES

INTRODUCTION

We Have Choosen this Project by seing the craze of social Media Influence on people and it was very unique , So we have Extracted the Data from Kaggle and started to Analyze that data meanwhile we thought to create a such model to predict the like on certain videos

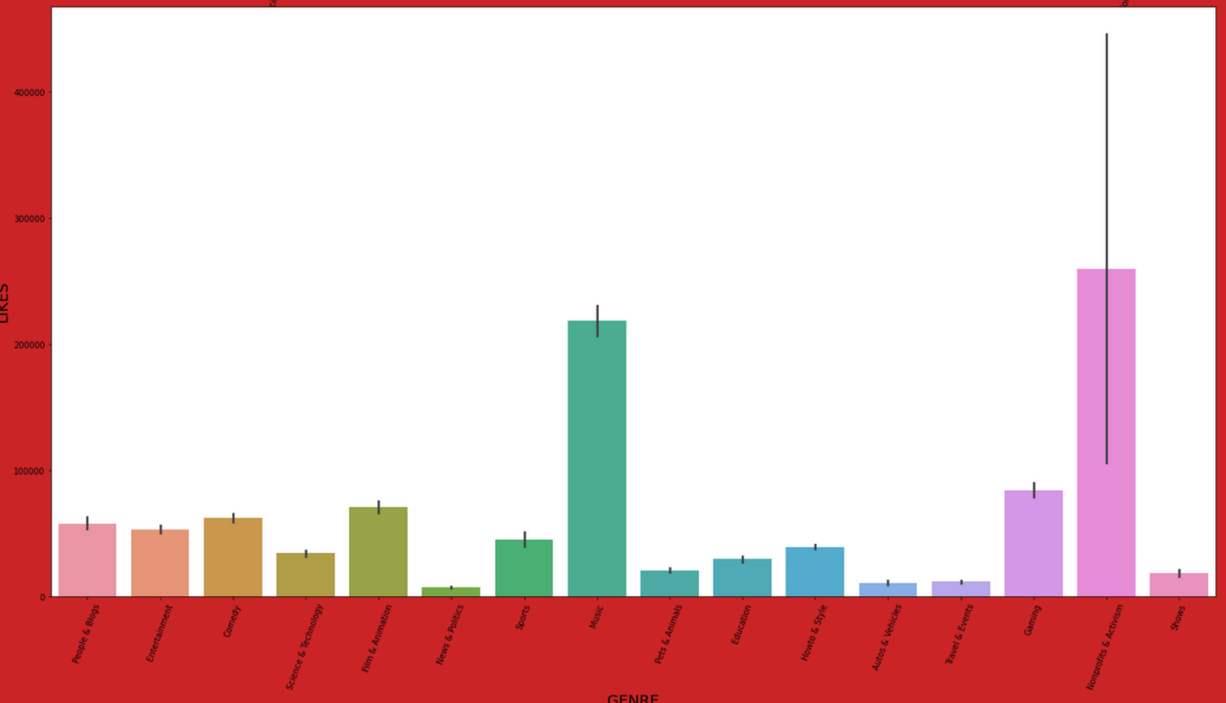
REQUIREMENTS

Pandas
Numpy
Matplotlib
Seaborn
Ipywidgets
Sklearn
Tensorflow
Keras
Venv



ANALYSIS 1

here we can see that the graph between dislikes and genre clearly indicates that nepotism and activism is most disliked video on YouTube where as pets and animal is least disliked video on YouTube

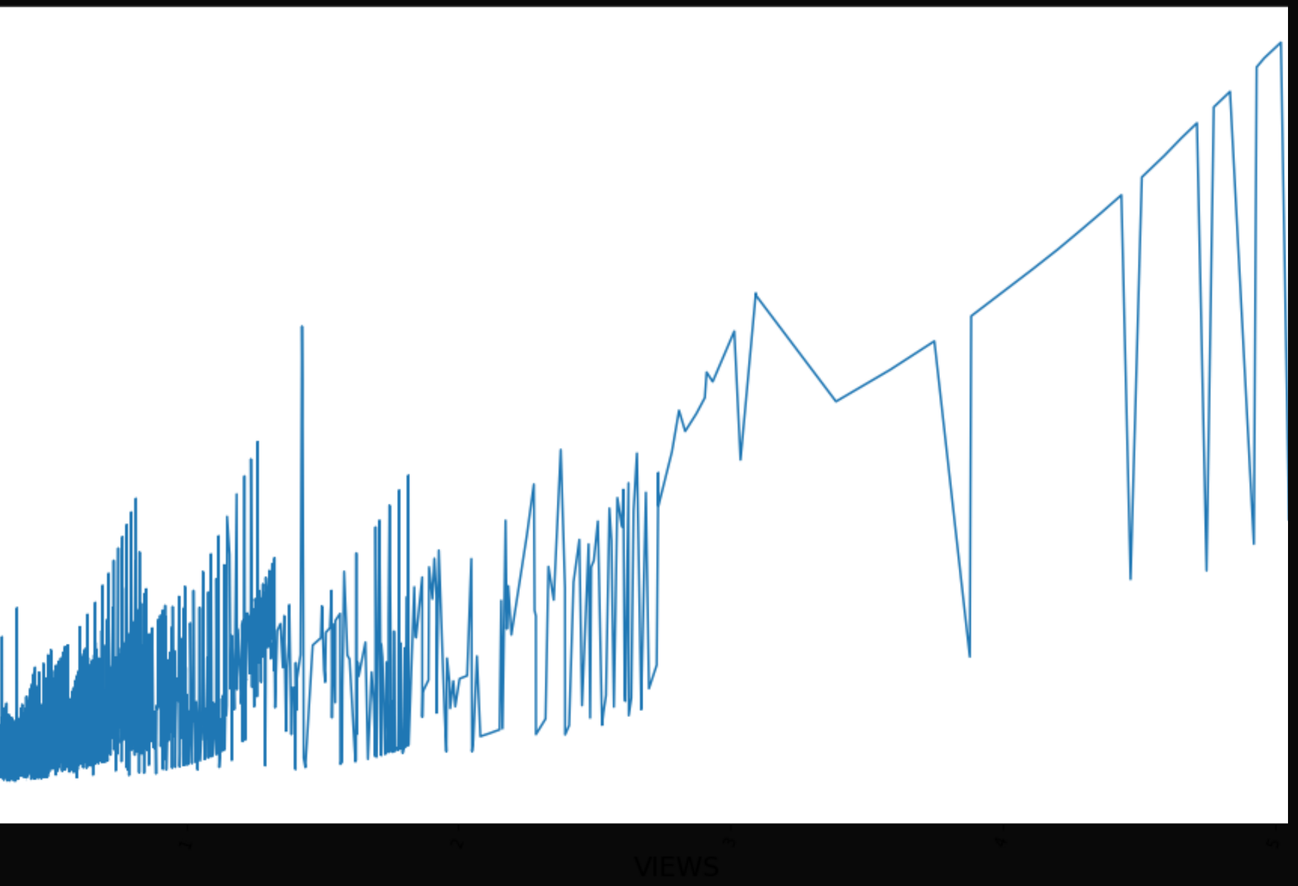


ANALYSIS 2

here we can see that the graph between likes and genre clearly indicates that nepotism and activism is most liked video on YouTube where as Music is the Second Highest Liked videos on youtube

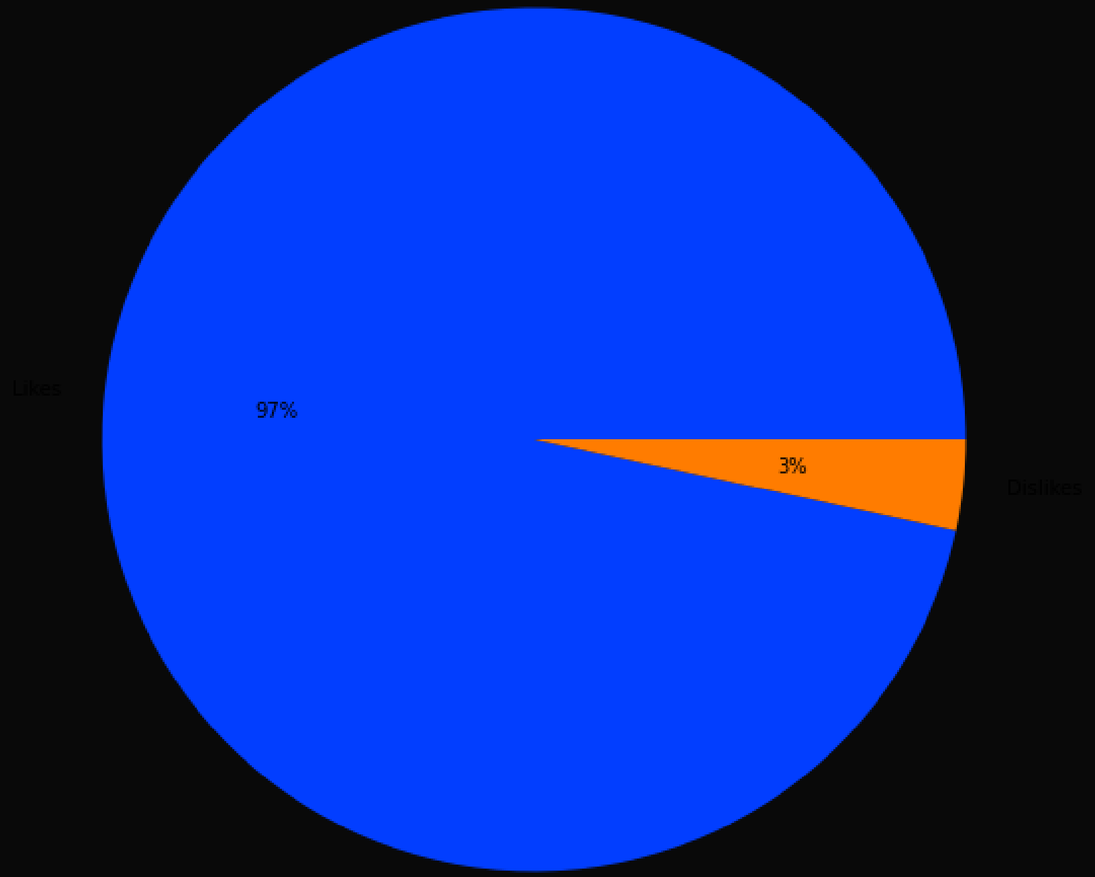
01

This Graph Indicating That As our View s are increasing our like Counts are also Increasing , this is the graph of views vs likes



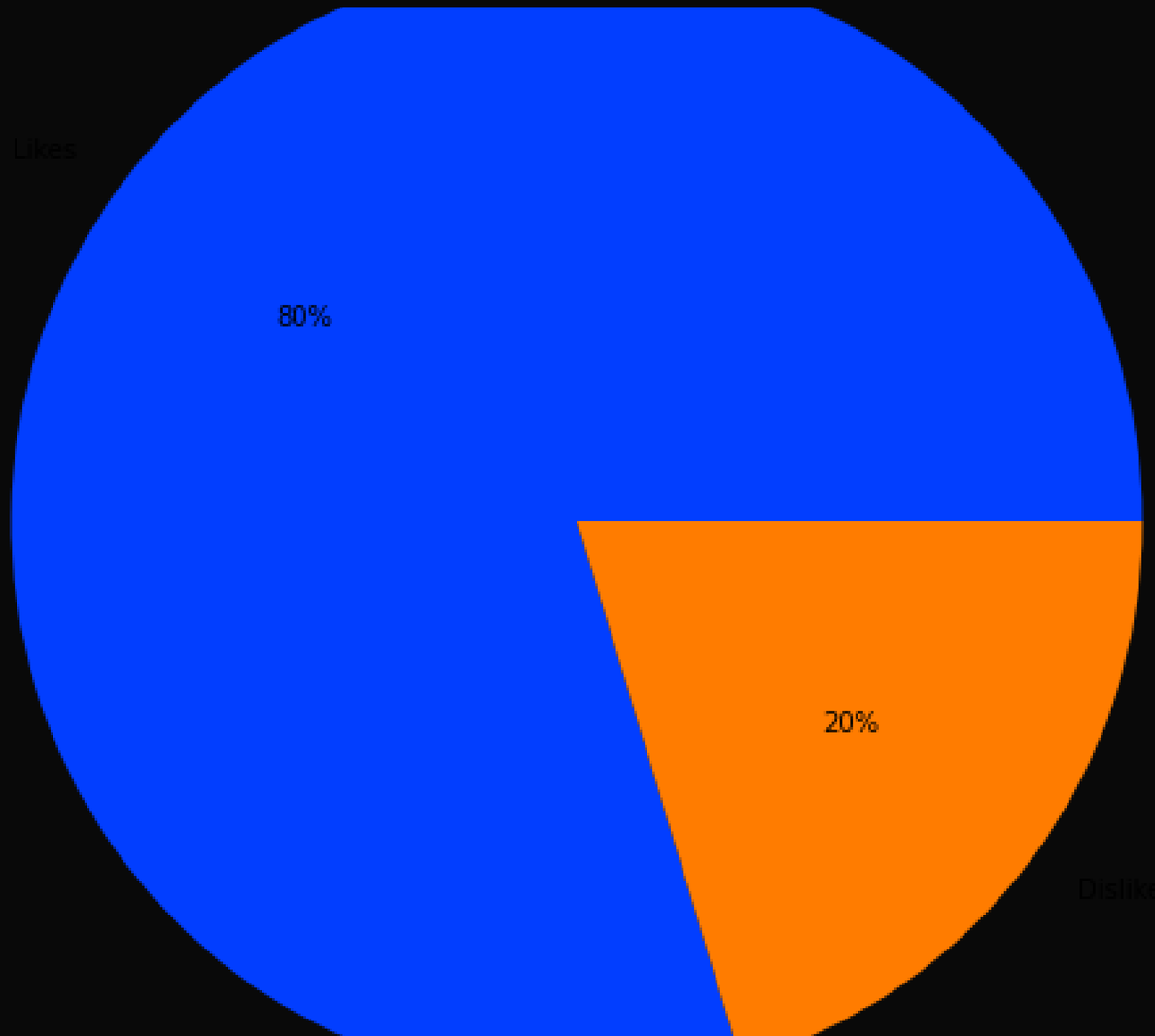
02

This Graph is Showing the Data of like vs Dislike Ratio of Music Category where we can clearly see that music category get less dislike as Compared to other Genre



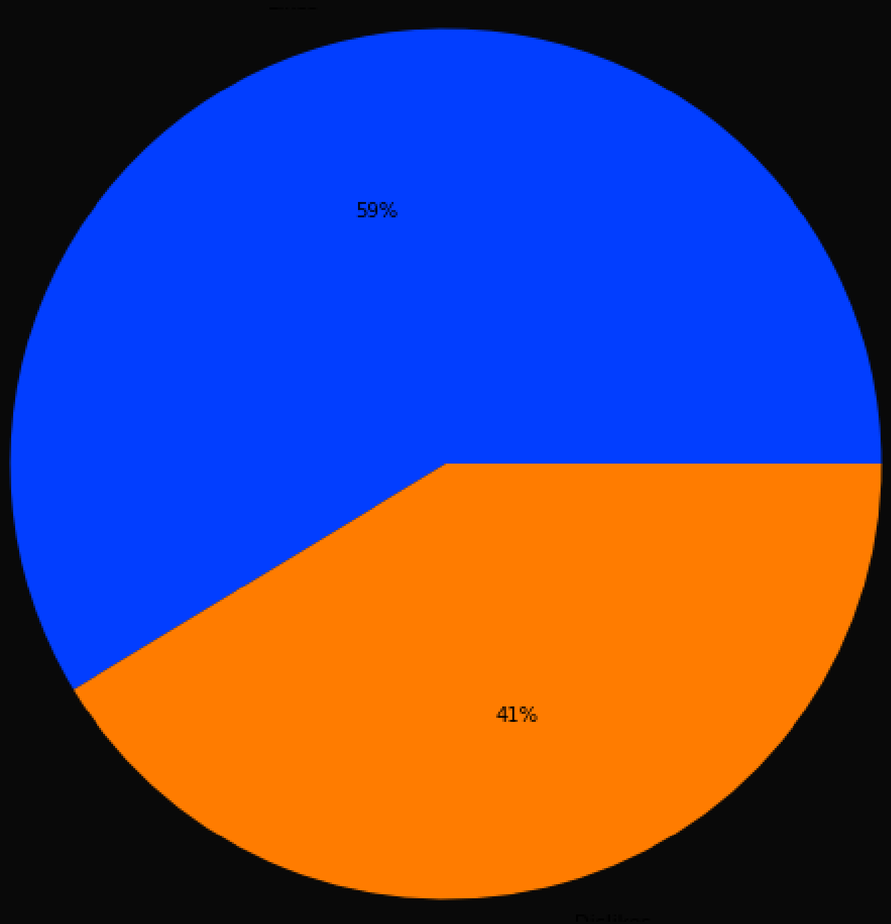
03

This Graph is Showing the Data of like vs Dislike Ratio of Nepotism and Activitism Category where we can clearly see that this category get more dislike as Compared to other Genre

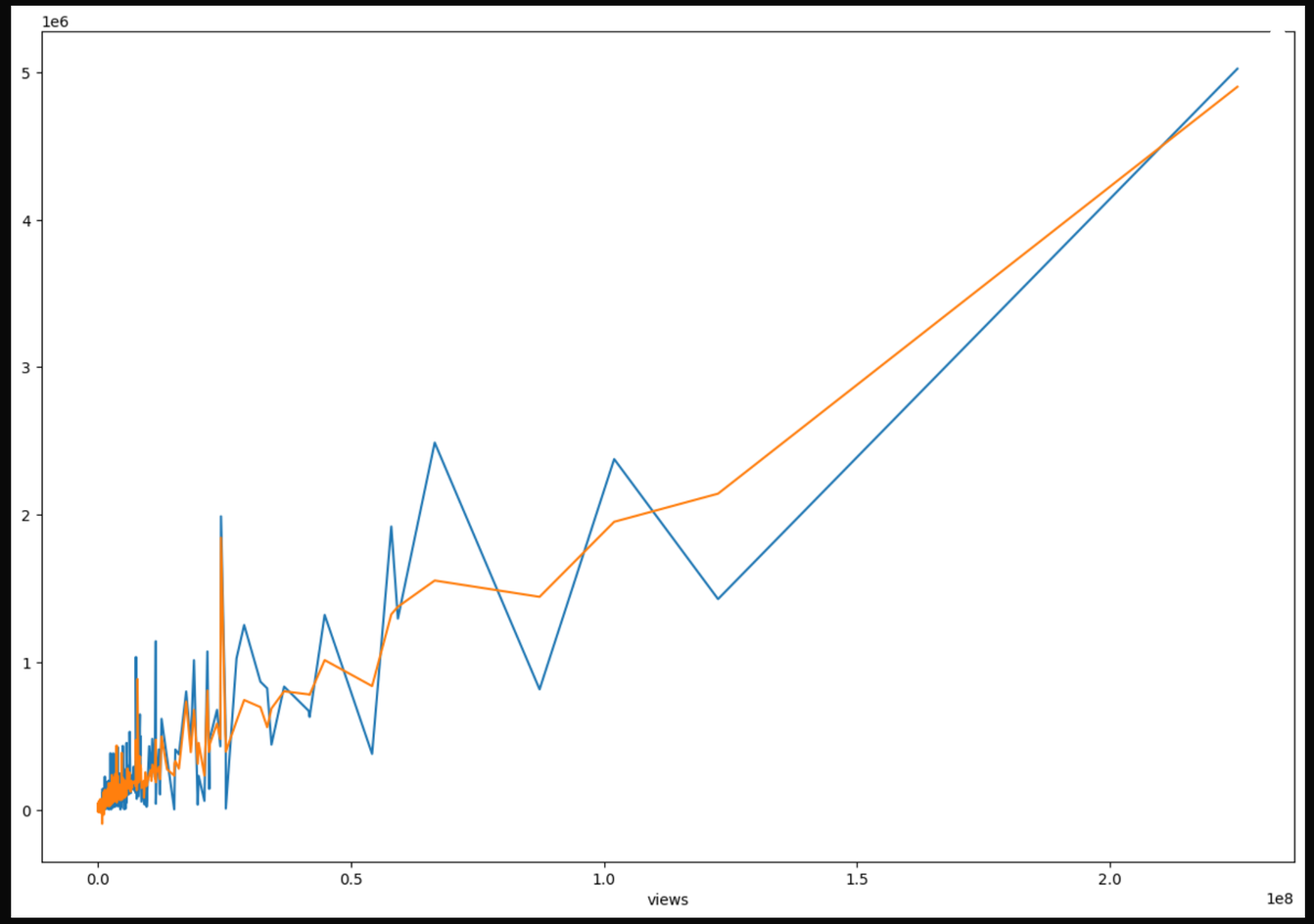


04

This is the Sample Graph of a video of category News and Politics as we can clearly see by the graph these kind of videos get more dislikes as compare to other category of videos



Linear Regression Result



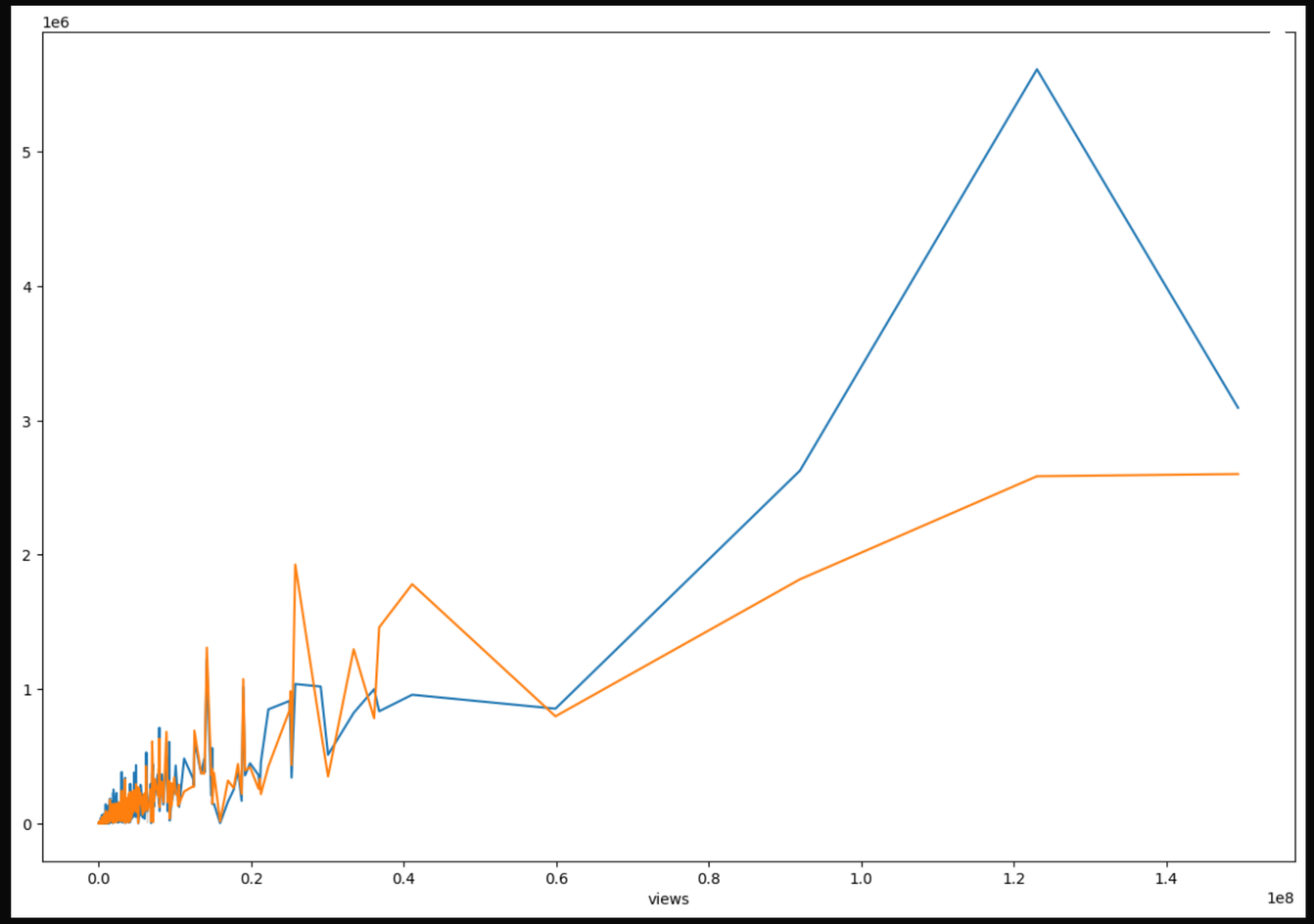
Here Blue Lines indicating our actual value and orange line is indicating our predicted value

Here we are getting a tremendous good result accuracy of 90%

For justification we have putted a graph below to see how we are predicting the values

For checking the result of any value please refer to our linear regression code there where you can select some values and you will get require result accordingly

Random Forest Regression Result



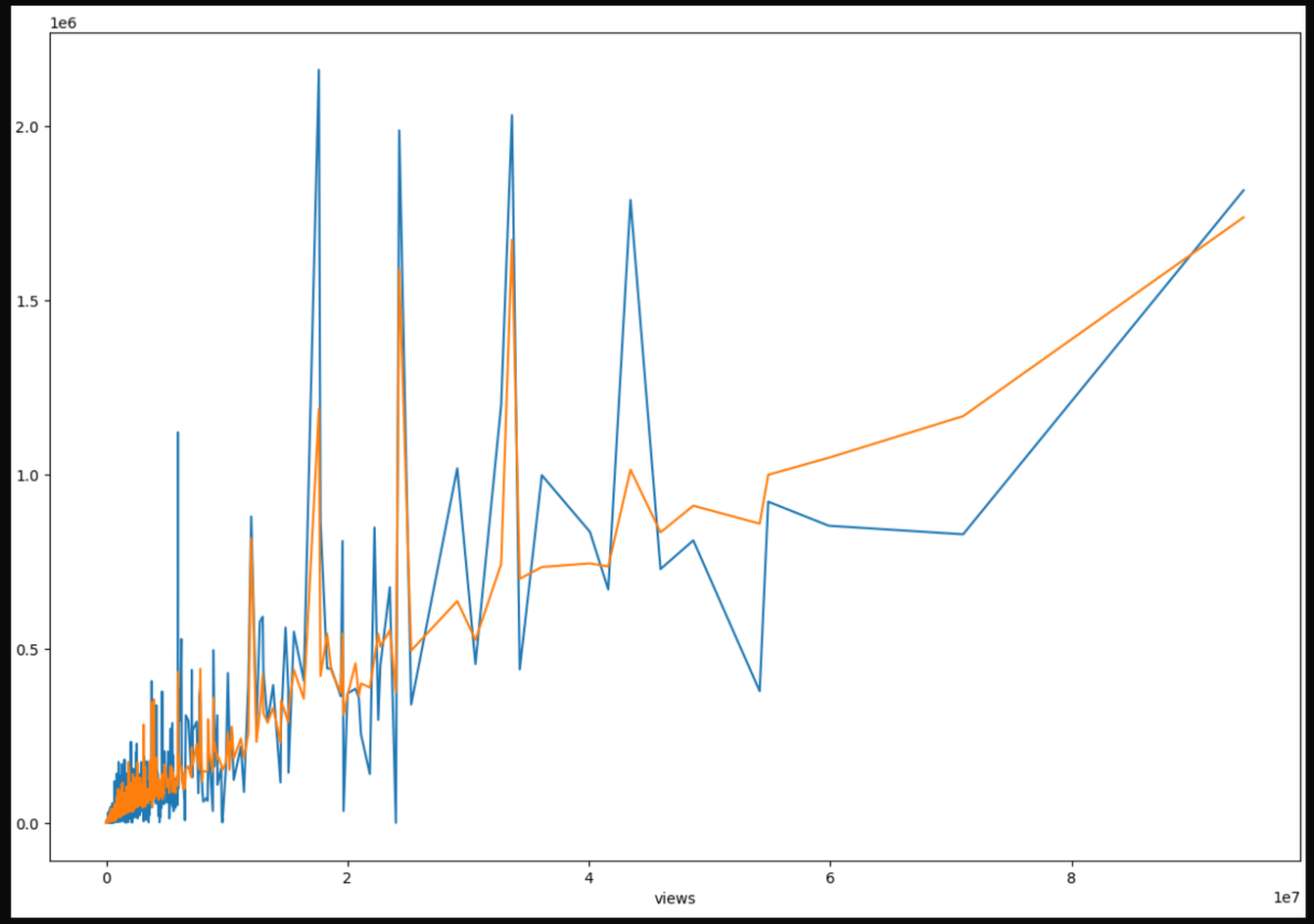
Here Blue Lines indicating our actual value and orange line is indicating our predicted value

Here we are getting a tremendous good result accuracy of 79% Which are Fluctuating Because Random Forest is based on Ensemble Technique and Decision Tree.

For justification we have putted a graph below to see how we are predicting the values

For checking the result of any value please refer to our linear regression code there where you can select some values and you will get require result accordingly

Artificial Neural Network Result



Overall by Seeing the Graph Only we Can Conclude that our Deep Learning Model is Predicting
Much better than our Random Forest Regressor

We are Getting Accuracy Score of 85% which is greater than our Random Forest Regressor and less than our Linear Regression Model

Accuracy Score is 85 %

Data Analysis Conclusion

We have explored the data

We got some good insight from the data

Views is the good input feature for our like prediction

Similarly this like comment count is also a Khud input feature for the life prediction

But the most important thing we have concluded from our data analysis that genre is the most important input feature for our like prediction

Model Conclusion

Here we have used three different models

Linear regression model

Random forest regression model

Artificial neural network model

In linear regression model we came to know that our model was very good to predict the future like counts

In random forest regression model we came to know that our model was not very good to predict the future count as compared to our linear regression model

In the artificial neural network model we came to know that by doing the hyperparameter tuning we got good result as compared to random forest regression model it was getting around 85% accuracy for increasing the accuracy of artificial neural network model we

Suggestion

Here we have tried Natural Language Processing method to analyze sentiment analysis of the title data and tags data but those data contains emojis and all the different type of characters as the basic NLP model we cannot process it and we cannot conclude any kind of sentiment analysis from the data

But for the future scope if we were able to develop some such kind of model which can analyze the behavior of emojis character and other types of languages also like we got the data in the chinese and other languages also so if we are able to make such a model who can predict good sentiment analysis out of it then we can consider that our model will have a good input feature like title tags etags and other data also which will be in the form of textual data

So here are our big data analysis role came for deep learning analysis of the thumbnail in which got the link of the thumbnail images but we were not able to and analyze that because we require a good model which should able to get some kind of analysis out of thumb nail images so it will be for the future is scope in which if we are able to get some detailing from our emails or doing some kind of sentiment analysis with the image also then we can make a very good model which can predict the like of any video if we upload it on YouTube