PRESENTATION OF

"MACHINE LEARNING" TRAINING

AND PROJECT ON

"YOUTUBE ADVIEW PREDICTION"

Submitted By:

Name: Trisha Sahu

Class: BE CSE 5th

Semester

Roll No: SG19357

Submitted To:

Mrs. Meenu Bhagat

ORGANISATION PROFILE

Internship Studio is a platform developed to help students build their profiles by providing them the right exposure to develop the required skills in the respective domain.

- ❖ Encouraging students to work on projects & learn from the professionals.
- ❖ Infusing a learning spirit through the best of best mentorship.
- ❖ Filling the gap between bookish knowledge and practical knowledge by providing training + internship

Instructor: Kashish Kumar

TOPICS

Training Week 1

Introduction to Statistics

Summary Statistics

Probability

Permutations and Combinations

Discrete Probability Distributions

Continuous Probability
Distributions

Inferential Statistics

Training Week 2

Basics of Python

Programming

Advanced Python

Programming

Python Libraries: Numpy

Python Libraries: Pandas

Python Libraries: Matplotlib

Training Week 3

Introduction to Machine Learning

Python Libraries: Sklearn

Linear Regression

Logistic Regression

Decision Tree and Random Forest

Ensemble Techniques

Naïve Bayes and SVM

Unsupervised Learning

Key ML Algorithms – KNN

Neural Network and Deep Learning

Internship Week 1

ML Internship Project Problem Statement

Internship Week 2

ML Internship Project Submission

TECHNOLOGY AND CONCEPTS

Machine Learning

In classic terms, machine learning is a type of artificial intelligence that enables self-learning from data and then applies that learning without the need for human intervention.

Linear Regression

Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. It's used to predict values within a continuous range, (e.g. sales, price) rather than trying to classify them into categories (e.g. cat, dog).

There are two main types:

- 1. Simple Regression
- 2. Multiple Regression

Support Vector Machine

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate.

Decision Tree

Decision tree analysis involves making a tree-shaped diagram to chart out a course of action or a statistical probability analysis. It is used to break down complex problems or branches. Each branch of the decision tree could be a possible outcome.

Artificial Neural Network (ANN)

An artificial neural network (ANN) is the piece of a computing system designed to simulate the way the human brain analyzes and processes information. It is the foundation of artificial intelligence (AI) and solves problems that would prove impossible or difficult by human or statistical standards. ANNs have self-learning capabilities that enable them to produce better results as more data becomes available.

PROJECT PROBLEM STATEMENT

Youtube advertisers pay content creators based on adviews and clicks for the goods and services being marketed. They want to estimate the adview based on other metrics like comments, likes etc.

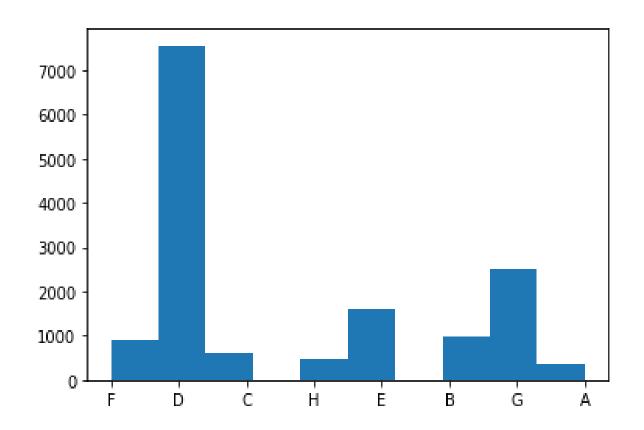
The problem statement is therefore to train various regression models and choose the best one to predict the number of adviews. The data needs to be refined and cleaned before feeding in the algorithms for better results.

STEPS FOR ADVIEW PREDICTION

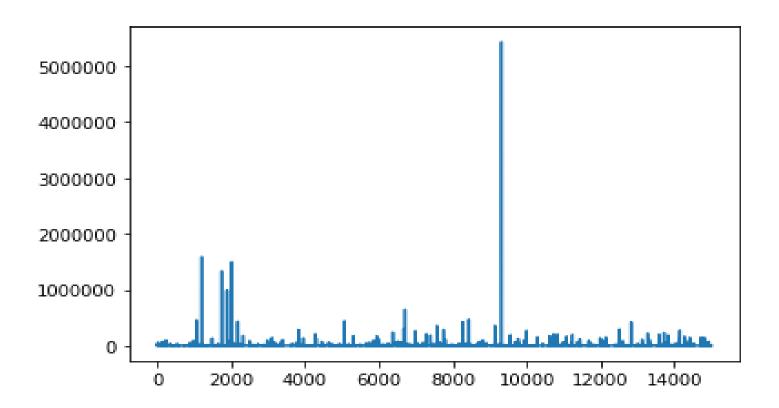
- 1. Import the datasets and libraries, check shape and datatype.
- 2. Visualise the dataset using plotting using heatmaps and plots. You can study data distributions for each attribute as well.
- 3. Clean the dataset by removing missing values and other things.
- 4. Transform attributes into numerical values and other necessary transformations
- 5. Normalise your data and split the data into training, validation and test set in the appropriate ratio.
- 6. Use linear regression, Support Vector Regressor for training and get errors.
- 7. Use Decision Tree Regressor and Random Forest Regressors.
- 8. Build an artificial neural network and train it with different layers and hyperparameters.

VISUALIZATION

Histogram of "Category" column



Histogram of "adview" column



Heatmap which shows the co-relation of all columns with each other



RESULT

Algorithm	Linear	Random	Decision tree	Support vector	ANN
	Regression	forest		machine	
Mean Absolute	3707.37800	3274.69029	3059.31079	3707.37800	3304.26489
Error	5824529	66905504	2349727	5824529	4606637
Mean Squared	835663131.	644433788.	1226286165	835663131.	829552666.
Error	1210335	0361483	.4118853	1210335	7955565
LITOI					
		05005 7004	0.501.0.000		
Root Mean Squ		25385.7004	35018.3689	28907.8385	28801.9559
ared Error	7573986	6376795	713254	7573986	5433679

CONCLUSION

Best Model

From the training dataset by applying all algorithms for train the model, we found that "Random Forest Regressor" algorithm has less root mean squared error as compared to other algorithms. As we know model having less root mean squared error is more perfect. So here for prediction of test dataset we use "Random Forest" algorithm.

CERTIFICATE OF TRAINING + INTERNSHIP



CERTIFICATE

OF INTERNSHIP

THIS IS TO CERTIFY

Trisha Sahu

for completing 5 weeks of internship in Machine Learning with the title of YouTube adview prediction at Internship Studio from 16th July to 20th August. We wish you all the best for your future endeavours.

Kashish Kumar

Kashish Kumar Internship Mentor Aniket Maheshwari

Aniket Maheshwari

Founder & CEO

Certificate ID: ISMLI5304

Date:-20/08/2021

THANK YOU!!!