

**Project Report**

* **Project Title**: Classify whether a learner will complete an online course based on activity logs.
* **Student Name** Adarsh Srivastava
* **Roll Number**: 202401100300013
* **Course**: Introduction to artificial intelligence (Mr Bikki Kumar)
* **Date**: 22-04-2025

**1. Introduction**

This project aims to create a prediction model based on the activity logs given in the dataset. This project predicts if the student have completed the course.

**2. Problem Statement**

Develop a predictive model to classify whether a learner will complete an online course based on their activity logs, including videos watched, assignments submitted, and forum posts. Leverage machine learning algorithms to analyse patterns and behaviours, providing an accurate prediction of success or failure. Your model will unveil key insights into learner engagement, helping online education platforms identify at-risk students and take proactive steps to boost completion rates.

**3. Methodology**

To create a prediction model for a student, the following approach is used:

1. **Data Collection**:
   * The dataset contains information such as timestamps, page views, unique visitors, and bounce rates.
   * Data is loaded from a CSV file for processing.
2. **Data Preprocessing**:
   * Cleaning and formatting data (e.g., converting timestamps, handling missing values).
   * Removing outliers that may affect analysis.
3. **Data Analysis**:
   * Identifying trends in values using statistical measures.
   * Analyzing values and student habits.
   * Calculating key performance metrics like forum posts.
4. **Visualization**:
   * Generating graphs to illustrate trends in page views and visitor patterns.
   * Visualizing search space reduction and user behavior insights.
5. **Interpretation & Insights**:
   * Extracting meaningful conclusions from the analysis.
   * Suggesting improvements based on observed trends.

**4. Code Implementation**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

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

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import accuracy\_score

# Load the student csv

df = pd.read\_csv("online\_learning.csv")

df

# check for null values

df.isnull().sum()

df.shape

#convert yes or no to 1 and 0

df['completed'] = df['completed'].map({'yes': 1, 'no': 0})

print(df)

#print a kdeplot for visualization

fig, axes = plt.subplots(2, 2, figsize=(15, 10))

axes = axes.flatten()

for idx, column in enumerate(df.columns):

    if idx < len(axes):

        sns.kdeplot(data=df, x=column, ax=axes[idx])

        axes[idx].set\_title(column)

plt.tight\_layout()

plt.show()

#creating a prediction model, testing and training

y = df[['videos\_watched', 'assignments\_submitted', 'forum\_posts']]

X = df.drop(['completed'], axis=1)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(

    X, y, test\_size=0.4, random\_state=42)

ml = LinearRegression()

ml.fit(X\_train, y\_train)

a = ml.predict(X\_test)

a

#printing the accuracy

from sklearn.metrics import r2\_score

target\_columns = ['videos\_watched', 'assignments\_submitted', 'forum\_posts']

for i, col in enumerate(target\_columns):

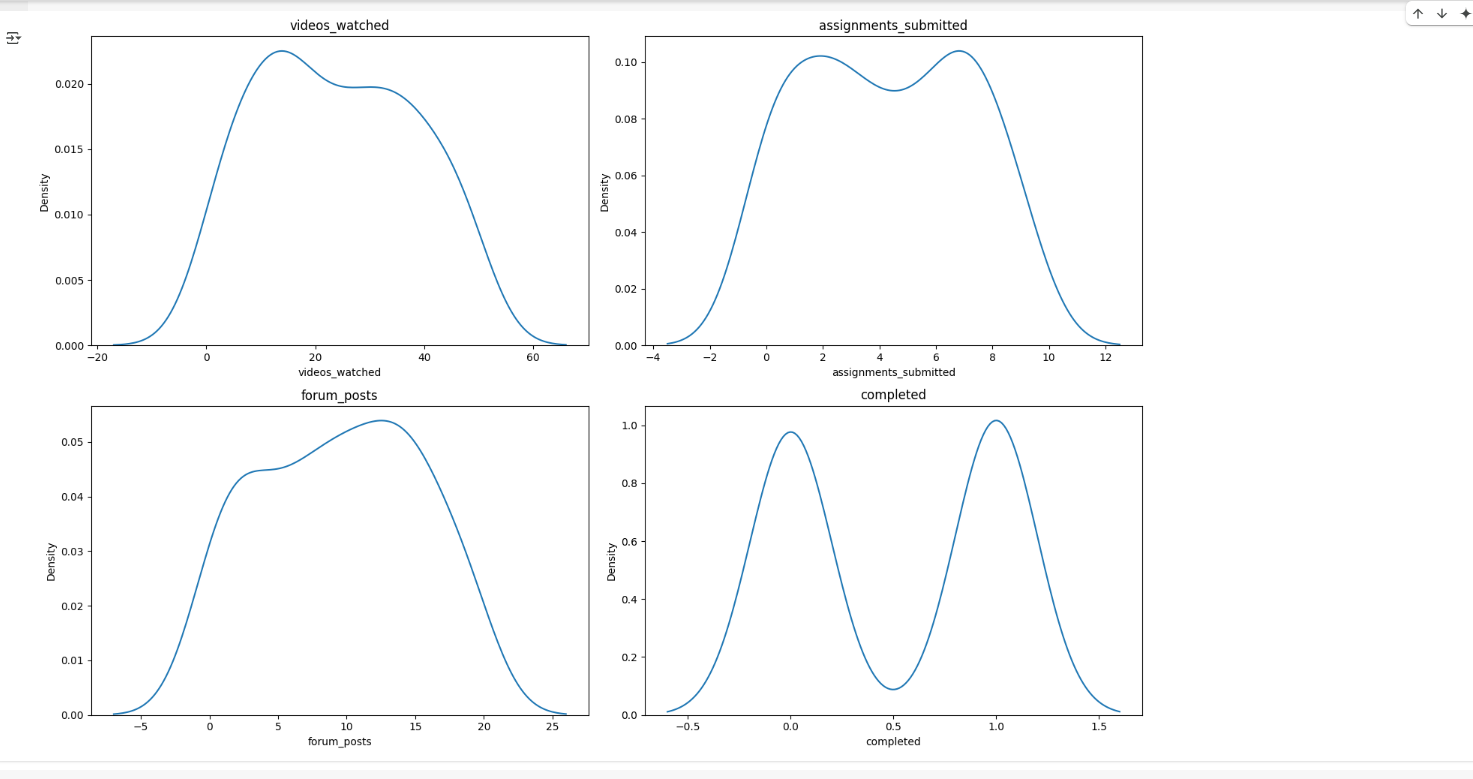
    r2 = r2\_score(y\_test[col], a[:, i])

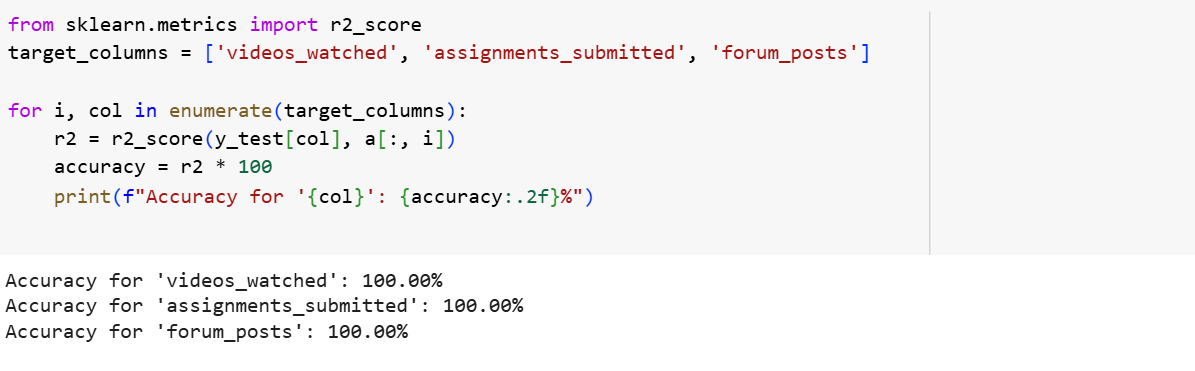
    accuracy = r2 \* 100

    print(f"Accuracy for '{col}': {accuracy:.2f}%")

**5. Results and Output**

The output is visualized by graphs after cleaning and analysing the data





**8. References/Credits**

* Excel Datasheet to analyse the data
* Python libraries like pandas, numpy, matplotlib, sklearn modules.
* Python project by Adarsh Srivastava