****

**Assessment Report**

on

**“Predict Online Learning Completion”**

submitted as partial fulfilment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AI)**

By

Name: Aastha

Roll Number: 202401100300003

Section: A

**Under the supervision of**

“BIKKI KUMAR”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

Online learning platforms have revolutionized education, offering flexible and accessible learning environments. However, dropout rates remain a significant concern. Predictive modelling can help institutions identify learners at risk of non-completion and offer timely interventions to improve course outcomes.

**2. Problem Statement**

The aim is to develop a machine learning model that can predict whether a student will complete an online course based on their activity logs. These logs may include video views, assignment submissions, quiz attempts, and other engagement metrics.

**3. Objectives**

* To analyse learner behaviour through activity logs
* To preprocess and prepare data suitable for classification
* To build a classification model to predict course completion
* To evaluate the model using performance metrics
* To visualize results using confusion matrix heatmaps

**4. Methodology**

The methodology involves the following sequential steps:

* **Data Acquisition**: Uploading a CSV file containing user activity logs from an online course platform.
* **Data Exploration**: Viewing data structure, identifying missing values, and understanding the distribution of the target variable (completed).
* **Preprocessing**: Handling missing values, encoding categorical data (like yes/no), and splitting the dataset into training and testing sets.
* **Model Building**: Using a Random Forest Classifier for binary classification due to its accuracy and feature importance capabilities.
* **Evaluation**: Applying accuracy, precision, recall, and confusion matrix as evaluation metrics.
* **Visualization**: Generating a heatmap for the confusion matrix to visualize performance.
* **Interpretation**: Analysing evaluation metrics to understand model strengths and weaknesses.

**5. Data Preprocessing**

* **Loading Data**: Read from CSV using Pandas.
* **Missing Values**: Handled by filling with 0.
* **Target Variable**: completed was encoded to binary (1 for "yes", 0 for "no").
* **Train-Test Split**: The dataset was split (80% train, 20% test) to evaluate generalization.

**6. Model Implementation**

A **Random Forest Classifier** was used, given its ability to handle non-linear features, prevent overfitting, and provide feature importance metrics.

**7. Evaluation Metrics**

* **Accuracy**: Proportion of all correct predictions.
* **Precision**: True positives among all predicted positives.
* **Recall**: True positives among all actual positives.
* **Confusion Matrix**: Visualized using a heatmap

**8. Results and Analysis**

* **Accuracy**: 35%
* **Precision**: 50%
* **Recall**: 31%

The confusion matrix indicates strong performance in both identifying students who completed and did not complete the course, with limited false positives and negatives.

**9. Conclusion**

The project successfully demonstrated how machine learning can predict learner course completion using online activity data. This predictive capability can assist educational platforms in identifying at-risk learners early and deploying targeted interventions to improve course completion rates.

**10. References**

* Scikit-learn documentation: <https://scikit-learn.org>
* Google Collab: https://colab.research.google.com
* Seaborn documentation: https://seaborn.pydata.org
* [Include dataset citation if applicable]

