****

**Assessment Report**

on

**“Predict Student Dropout”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AI)**

By

Name : Himanshu Singh

Roll Number : 202401100300125

Section: B

**Under the supervision of**

“ABHISHEK SHUKLA”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

This project focuses on predicting student dropout using supervised machine learning. The goal is to identify students at risk of dropping out based on attributes like attendance, grades, and participation. Early identification can allow institutions to intervene and provide academic support.

**2. Problem Statement**

To classify students as potential dropouts based on historical data including attendance, academic grades, and class participation. This can help in proactive measures and better educational planning.

**3. Objectives**

* Preprocess the dataset to prepare it for modeling.
* Train a classification model to predict student dropout risk.
* Evaluate the model using classification metrics.
* Visualize the confusion matrix for better interpretability.

**4. Methodology**

**Data Collection:** A CSV file containing student data was used.

**Data Preprocessing:**

* Encoded the categorical target label ('dropout\_risk') using LabelEncoder.
* Split the dataset into training and testing sets.

**Model Building:**

* Used Random Forest Classifier.
* Trained the model on 80% of the dataset and tested on the remaining 20%.

**Model Evaluation:**

* Evaluated accuracy, precision, recall, and F1-score.
* Generated and visualized the confusion matrix.

**5. Data Preprocessing**

The dataset was cleaned and preprocessed as follows:

* Categorical values (dropout risk) were encoded into numerical format.
* Features and target were separated.
* Data was split into training (80%) and testing (20%).

**6. Model Implementation**

Random Forest Classifier was used due to its robustness and good performance on tabular data. The model was trained and predictions were made on the test set.

**7. Evaluation Metrics**

* **Accuracy**: Measures overall correctness of the model.
* **Precision**: Indicates the percentage of predicted dropouts that were actually correct.
* **Recall**: Measures how many actual dropouts were correctly identified.
* **F1 Score**: Harmonic mean of precision and recall.
* **Confusion Matrix**: Visualized to identify prediction errors.

**8. Results and Analysis**

The confusion matrix provided visual insight into model performance. The classifier showed a good balance between identifying actual dropouts and avoiding false positives. The metrics suggested the model was effective in predicting student dropout.

**9. Conclusion**

The project successfully demonstrated how machine learning can be used to predict student dropout. With relatively simple features, the model was able to make effective predictions. Further enhancement can be done by incorporating more diverse student data and testing other classification models.

**10. References**

* scikit-learn documentation
* pandas documentation
* seaborn visualization library
* Online resources on dropout prediction using ML







