CAPSTONE PROJECT 2
Student's Dropout Prediction Project Proposal
by Srikanth Marepally PGA 36

1. Executive Summary:

The increase in academic dropout and failure rates in undergraduate colleges has become a significant concern. Early identification of students at risk can enable timely intervention, which is crucial for improving academic success and reducing dropout rates. This project aims to develop a machine learning model that predicts students at risk of academic failure based on available data, allowing educational institutions to implement targeted support strategies.

2. Problem Statement:

Background: Undergraduate colleges are increasingly facing challenges related to student retention and success. Academic dropout and failure not only impact the students future but also affect the institution's performance and reputation. Identifying students at risk early in their academic journey is vital for implementing effective support strategies that can prevent dropout and ensure academic success.

Objective: This model will analyze various factors known at the time of enrollment, The objective of this project is to develop a predictive model using machine learning techniques to identify students at risk of academic failure at an early stage of their academic path.

Scope: It involves analyzing, visualizing, and processing student data to build a predictive model. The model will be designed to identify at risk students with high accuracy and will be deployed to provide actionable insights for academic institutions.

3. Data Sources:

- We will use a dataset that contains information on parent's qualification, previous qualification, nationality, etc**.**

4. Methodology:

Preprocessing:

- The dataset will be processed to handle missing values, duplicates, outliers in the data. Visualization, Feature selection, Feature engineering, Encoding, and Scaling will be performed.

Model Building:

- All the ML and DL Classification models are built.

Model Evaluation:

- The models will be evaluated based on metrics such as accuracy, precision, recall, F1-score, confusion matrix, ROC curve, and AUC.

Model Selection:

- The models will be trained and tested on the dataset using cross-validation to ensure reliable performance. The best-performing model will be selected for deployment.

Model Deployment:

- After evaluation, the selected model will be deployed on hugging face

5. Expected Outcomes:

- To gain insights from the data.
- To make a good model that can make accurate predictions.
- To successfully deploy the model.

6. Tools and Technologies:

- Jupyter notebook: For data preprocessing, EDA.
- Gradio and Hugging face: It is used for deployment

7. Risks and Challenges:

- Data preprocessing: Making sure that the raw data is cleaned and preprocessed.
- Reverse encoding: As all my features are encoded I have to reverse encode it for EDA

8. Conclusion:

This dropout prediction project has the potential to significantly reduce dropouts and failure rates in undergraduate institutions. By leveraging data driven insights and predictive modeling, educational institutions can take proactive measures to support students at risk, thereby enhancing academic success and retention rates.