

PROJECT SYNOPSIS REPORT ON

“**Predict Students' Dropout and Academic Success**”

SUBMITTED TO

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Supervised By

Mr. Shivam Singh

Submitted By:

Name(s):

Akshit Rai: 2011980501

Sagar Mishra: 2011980524

Amit Aswal: 2011980501

Krrish Tapwal: 2011981212

Semester: 7th

Session: 2020-2024

Index

|  |  |  |
| --- | --- | --- |
| Sr no | Topic | Page No |
| 1 | Problem Statement | 3 |
| 2 | Title of project | 4 |
| 3 | Objective | 4 |
| 4 | Options available to execute the project | 5,6 |
| 6 | References | 7 |

# Problem Statement

Prediction of student’s performance became an urgent desire in most educational entities and institutes. This is essential to help at-risk students and assure their retention, providing excellent learning resources and experience, and improving university’s ranking and reputation. In this project, we are working on the student’s dataset which we have taken UCI that is one of the famous open-source repositories. The idea here is to build a classification model which can classify between three different categories of students based on the historical data of 4424 students.

In this dataset we have a long list of features, so we need to analyze these features through different graphs and perform hypothesis testing to understand which feature is affecting the target variable. Since it’s a classification task, so we are going to apply different Machine Learning algorithms as classifiers and to evaluate the performance of each of these algorithms we are going to use different evaluation metrics such as Accuracy score, Precision score, Recall score, and F1-score etc. The result of these evaluation metrics helps us to understand how good our models are? Does it give us the accurate prediction for future data? And the analysis on this dataset which we did with Exploratory Data Analysis (EDA) helps us to make decision and recommendations for future.

Project aim:

The aim of this project is to contribute to the reduction of academic dropout and failure in higher education, by using machine learning techniques to identify students at risk at an early stage of their academic path, so that strategies to support them can be put into place. The problem is formulated as a three-category classification task (dropout, enrolled, and graduate) at the end of the normal duration of the course.

# Title of project

“Predict Students' Dropout and Academic Success”

# Objective & Key Learning's

## Objectives:

* Analyze and preprocess the dataset: The primary objective is to thoroughly analyze and preprocess the student dataset. This involves exploratory data analysis (EDA) to understand the dataset's features and their impact on student performance.
* Feature selection and hypothesis testing: Identify the most influential features that affect student outcomes. Perform hypothesis testing to confirm the significance of these features.
* Build classification models: Develop and train machine learning classification models to categorize students into three classes: "dropout," "enrolled," and "graduate." Implement various classification algorithms and evaluate their performance using metrics like accuracy, precision, recall, and F1-score.
* Model evaluation: Assess the models' predictive accuracy and reliability. Determine which algorithms provide the most accurate predictions for future student data.
* Insights and recommendations: Provide insights into the dataset analysis and model performance. Offer actionable recommendations to help educational institutions support at-risk students and enhance their learning experience.

# Options available to execute the project

There are various options available to execute the project of Predict Students' Dropout and Academic Success data analysis.

and prediction. Some of the key options are:

* Data Preprocessing: Extensive experience in data preprocessing, including data cleaning, missing value handling, and feature transformation, is gained. This is fundamental in machine learning projects.
* Exploratory Data Analysis (EDA): EDA skills are acquired, allowing for a deep understanding of the dataset through relationship exploration, data visualization, and pattern recognition.
* Feature Selection: The project teaches the importance of identifying influential features impacting student outcomes. Techniques such as hypothesis testing, and statistical analysis are used for feature selection.
* Classification Algorithms: Hands-on experience in implementing various classification algorithms, like logistic regression, decision trees, random forests are obtained. Understanding the strengths and weaknesses of these algorithms is crucial.
* Model Evaluation: Learning how to assess model performance using metrics like accuracy, precision, recall, and F1-score enhances the ability to evaluate predictive models effectively.
* Interpretation and Recommendations: The project emphasizes the interpretation of results and the generation of actionable insights. It involves understanding the implications of model predictions and providing real-world recommendations.
* Addressing Academic Dropout: Insights are gained into how machine learning can be a powerful tool in addressing academic dropout and supporting students in higher education.
* Impact on Education: Realizing the potential of data-driven decision-making in improving education. The use of machine learning to enhance student retention and the overall learning experience becomes evident. Understanding the role of data in education institutions is a key takeaway.

The above options can be executed using various programming languages such as Python, R, and MATLAB. Each option has its own advantages and disadvantages, and the selection of the appropriate option depends on the project requirements, time constraints, and available resources.

For instance, if the project requires real-time match prediction, then a web application or mobile app deployment may be necessary. Alternatively, if the project requires more detailed analysis and visualization, then data analysis and visualization tools such as Tableau or Power BI may be more appropriate.

# References

1. Ameen, A. O., Alarape, M. A., & Adewole, K. S. (2019). STUDENTS’ ACADEMIC PERFORMANCE AND DROPOUT PREDICTION. *MALAYSIAN JOURNAL OF COMPUTING*, *4*(2), 278. https://doi.org/10.24191/mjoc.v4i2.6701
2. Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-0177-7
3. Respondek, L., Seufert, T., Stupnisky, R., & Nett, U. E. (2017). Perceived Academic Control and Academic Emotions Predict Undergraduate University Student Success: Examining Effects on Dropout Intention and Achievement. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00243
4. Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E., & Nshimyumukiza, P. C. (2022). Predicting student’s dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence*, *3*, 100066. https://doi.org/10.1016/j.caeai.2022.100066
5. Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939785
6. Patel, H. H., & Prajapati, P. (2018). Study and Analysis of Decision Tree Based Classification Algorithms. *International Journal of Computer Sciences and Engineering*, *6*(10), 74–78. https://doi.org/10.26438/ijcse/v6i10.7478
7. "Dataset." Kaggle,

https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention/data