Project Title:

Predicting Student Dropout Using Machine Learning

Project Goal:

To develop a machine learning model that predicts the likelihood of a student dropping out of university. Early identification of at-risk students allows timely intervention and support, potentially reducing dropout rates and improving student retention.

Dataset:

The project leverages real-world data, including information about student **demographics, academic performance, socio-economic factors, and personal habits**. Extensive **data cleaning and preprocessing** are performed to ensure data quality and consistency.

Methodology:

A variety of classification algorithms are employed to predict student dropout, including:

- **Decision Tree (86.03%)**
- **Random Forest (88.65%)**
- **Gradient Boosting (93.45%)**
- **XGBoost (92.58%)**
- **LightGBM (91.27%)**

The performance of these models is evaluated using metrics such as **accuracy, confusion matrices, and classification reports**. To further enhance model performance, **feature engineering** and **hyperparameter tuning** are conducted.

Results:

The **Gradient Boosting** model emerged as the most effective algorithm for dropout prediction, achieving the highest accuracy of **93.45%**. This model provides actionable insights into the factors significantly influencing dropout risk, aiding in the development of **targeted intervention strategies**.

Conclusion:

This project highlights the potential of machine learning to accurately predict student dropout rates and support educational institutions in identifying at-risk students. By using data-driven insights, institutions can implement timely and effective measures to reduce dropout rates.

Potential Extensions:

1. **Enhancing Model Accuracy:**
    - Incorporate external data sources, such as social media activity and student engagement metrics.
2. **Developing a User-Friendly Tool:**
    - Create a **web application** for educators to input student data and receive dropout risk predictions.
3. **Personalized Interventions:**
    - Design tailored intervention strategies based on individual student needs and their predicted risk levels.

Note: If you need access to the dataset used in this project, please contact **shbshahriar32@gmail.com**.