DATA SCIENCE IN EDUCATION SECOTR

PROBLEM: DATA SCIENCE TO CURB STUDENT DROPOUT RATES IN EDUCATION SECTOR

The student dropout rate in Malaysia poses a significant issue, adversely impacting students' prospects and tarnishing the education system's reputation. Such dropouts are often subjected to unemployment, low income, and reliance on social aid, deepening educational inequities particularly among disadvantaged students. Leveraging data science could be a transformative approach to this complex problem, offering predictive tools for potential dropout rates, guiding institutions to devise effective interventions. This would not only uplift individual students but also bring wide-ranging societal benefits and enhance the overall educational landscape.

Methodology

- Problem Understanding
- Data Collection and Cleaning
- Data Exploration
- Data Modeling
- Results Interpretation

Sample dataset from Kaggle: https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention

Dataset

STUDENT DEMOGRAPHICS

Information about students' age, gender, socioeconomic background, and geographic location.

ACADEMIC RECORDS

Details like grades, test scores, and participation in extracurricular activities.

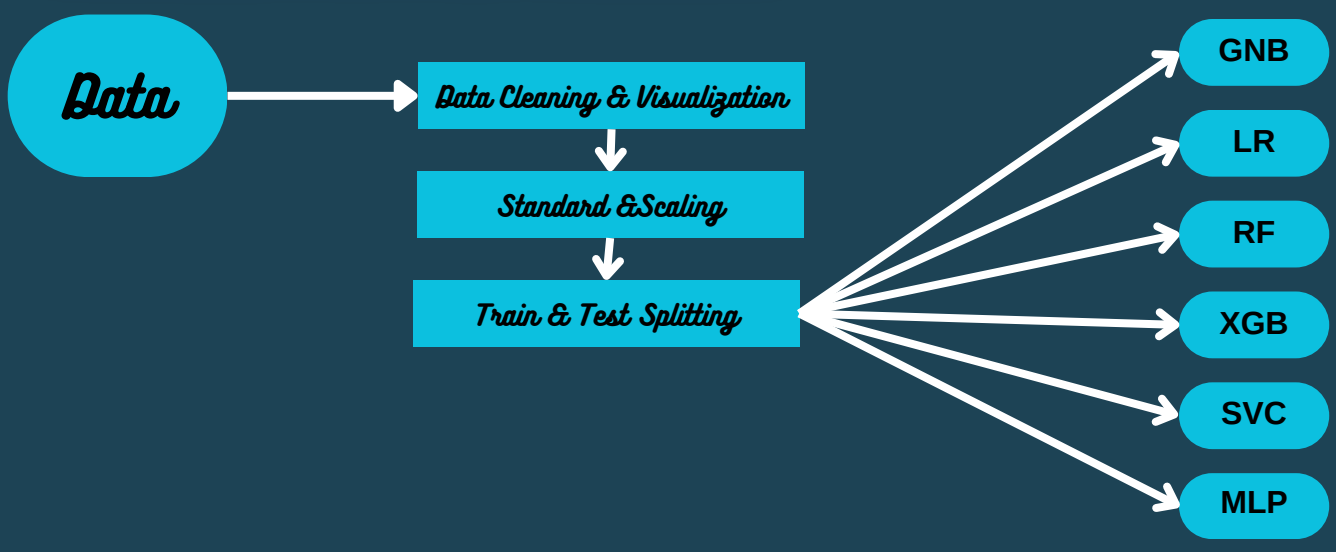
BEHAVIOURAL DATA

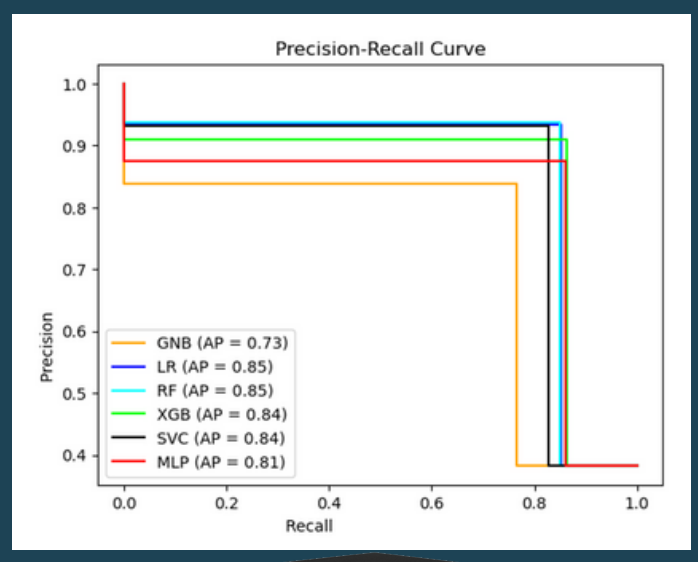
Student attendance records, instances of disciplinary action, etc.

ENVIRONMENTAL FACTORS

Data on family background, school environment, and peer influence.

VISUALIZATION AND MODELING





Conclusion

The precision-recall scores of the tested models using the Kaggle dataset indeed confirm the potential of data science in curbing student dropout rates in the education sector. The effectiveness of Logistic Regression and Random Forest models, with precision-recall scores of 0.85, underscore the power of data science in predicting and addressing the dropout issue. Similarly, the XGBoost Classifier and Multi-Layer Perceptron models, scoring 0.84 and 0.81 respectively, highlight that a variety of data science methods can provide valuable insights to tackle this challenge. Although the Gaussian Naive Bayes model scored lower at 0.73, it still represents a meaningful application of data science in this context. Thus, these results collectively emphasize the capacity of data science to make significant strides in mitigating student dropout rates and enhancing the overall educational landscape.