Predicting Student Dropout Risk

Student's Name: Pradyumn Kabra

Introduction

- Objective: Predict dropout risk using student data.

- - Detect patterns that lead to dropouts.

- - Use classification and clustering methods.

- Dataset: 500 records, 15 attributes (GPA, Attendance, etc.).

- Tools: Python, Pandas, Scikit-learn, Matplotlib.

- Methods: K-NN, K-Means Clustering.

Exploratory Data Analysis (EDA)

- Findings:

- - GPA ranges from 0.0 to 4.0.

- - Average Attendance: 75%, Median Engagement: 60.

- Patterns: Low GPA & Attendance → Higher Dropout Risk.

- Visuals used: Bar charts, frequency tables.

Methodology

- K-NN Classification:

- - Calculate distances, pick K nearest neighbors.

- - Predict based on the majority class.

- K-Means Clustering:

- - Initialize centroids → Assign points → Recompute.

- - Repeat until stable clusters form.

Results

- K-NN Results:

- - Predictions based on closest records.

- - Accuracy observed around 85%.

- K-Means Results:

- - Students divided into 3 meaningful clusters.

- - Cluster trends visible using scatter plots.

Insights and Learnings

- Takeaways:

- - Dropouts linked to GPA < 2.0, Attendance < 60%.

- - Engagement levels influence risk.

- ML tools helped identify key dropout indicators.

Challenges & Recommendations

- Challenges:

- - Missing data & inconsistent formatting.

- - Scaling features for accuracy.

- Suggestions:

- - Use more data, consider logistic regression.

- - Explore time-series engagement analysis.

Conclusion

- Summary:

- - Dropout prediction using classification & clustering.

- - Key indicators: GPA, Attendance, Engagement.

- Impact:

- - Enables early interventions.

- - Data Science helps solve real education issues.

References

- Tools: Python, Scikit-learn, Pandas, Matplotlib.

- Web Resources:

- - https://scikit-learn.org

- - https://pandas.pydata.org

- Books & Articles on Student Retention Models.