

K-Means Clustering Project Report

Project Goal

The main objective of this project is to analyze student data to identify patterns in academic outcomes (Dropout, Enrolled, Graduate). Using unsupervised learning techniques, particularly K-Means clustering, we aim to group students into meaningful clusters to better understand their academic journey and performance trends.

Methods & Techniques Used

1. **Data Preprocessing:** Cleaned and filtered the dataset, handled categorical/target variables (Dropout, Enrolled, Graduate).
2. **Exploratory Data Analysis (EDA):** Visualized distributions using pie charts and bar plots, normalized features, and performed outlier analysis.
3. **Dimensionality Reduction (PCA):** Applied Principal Component Analysis (PCA) to reduce feature space and preserve variance.
4. **Clustering (K-Means):** Implemented K-Means clustering, used Elbow method to determine optimal clusters, and evaluated results with Silhouette Score.
5. **Visualization:** Created scatter plots of clusters in reduced dimensions and visualized cluster separations for interpretability.

Conclusion

The analysis successfully grouped students into distinct clusters representing different academic outcomes. K-Means with PCA provided a clear structure in the data, highlighting patterns in dropout risk, active enrollment, and successful graduation. Silhouette score evaluation showed that the chosen cluster size was meaningful, confirming the quality of clustering. This project demonstrates how unsupervised machine learning can help in educational analytics, offering insights for universities to design better support systems for at-risk students.