

Education plays a crucial role in shaping society. Graduates from higher education institutions have higher incomes. In 2023, U.S. employees aged 25 and over with high school diplomas made \$853 per week, while employees with master's degrees made \$1661 per week. Higher education is also connected with unemployment rates, which can affect economic growth, poverty, and societal well-being. Nonetheless, challenges such as racial disparities in graduation rates and low retention rates remain crucial in education. This thesis addresses these issues by focusing on techniques applicable to analyzing educational datasets. These techniques are also used to understand and analyze the importance of education in the political landscape.

Experimental results are divided into three parts. In the first part of the experiments, a dataset containing educational institutions in the U.S. Data was collected using Python. Using clustering techniques on this dataset, six distinct clusters were identified. The analysis of clusters highlights that universities with high tuition fees in the U.S., a small number of students, and diverse ethnic populations have higher graduation rates for all ethnic groups. The analysis further identified that universities in areas with high median household incomes and high tuition fees have better graduation rates.

In the second part of the experiments, social network techniques were used to analyze the influence of the educational background of members of parliament on the structure of the UK parliament. Key members were identified using centrality measures, with one of them, Keir Starmer, who became a prime minister in recent election. Universities like Oxford and Cambridge were the most frequently attended institutions among Members of Parliament.

In the third experiment, an early prediction model is used to assess student success. Key contributors identified are the interactions with the online learning environment and course domains (STEM or Social Sciences). The best-performing model, random forest, has an accuracy of 70% . This thesis demonstrates the application of data mining techniques to address some of the critical issues in the higher education domain.