URL to Program Website: https://www.honorstransfercouncil.org/the-conference-1

250-Word Proposal Abstract

Analyzing the Relations of Education Level, Income, and Occupational Stability using Data Science: Linear Regression and KNN Classification Algorithm

Throughout the past century, it has become a nationwide concept that pursuing higher education and attaining a degree leads to sustainable jobs in the work field. However, statistics gathered by the United States Census Bureau show that total college enrollment peaked in 2010 and has since declined by 11.61%. Similarly, enrollment in California has decreased by 8.4% since 2010 (Hanson). This raises the question of whether higher education is still a reasonable investment for career life and challenges the traditional concept. This study analyzes the relations between education, income, and occupation sustainability by visualizing linear regression and performing hypothesis testing using the K-Nearest Neighbors (KNN) Algorithm and T-paired Testing. The "Short-Term Occupational Employment Projections" dataset published by the Employment Development Department of California was analyzed on the interactive computing platform Jupyter Notebook using Python. The variables "Entry Level Education" and "Median Hourly Wage" had a correlation coefficient of 0.7, indicating a strong and positive association (higher education = more income). Occupational stability was measured using true or false for positive employment growth. Given attributes of an unclassified job, KNN decides which binary the occupation belongs to depending on its similarity to pre-classified jobs. By calculating the accuracy of two KNN models, one with wage attribute, and one without, T-paired Testing was used to find the significant P-value of 0.013 (< 0.05). The significance of adding the attribute to the model indicates that occupational stability is correlated to wage and consequently, higher education is still a reasonable investment.

Works Cited

Andrade, Chittaranjan. "The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives." *Indian Journal of Psychological Medicine*, vol. 41, no. 3, 2019, p. 210, https://doi.org/10.4103/IJPSYM.IJPSYM_193_19.

California Employment Development Department. *Short-Term Occupational Employment Projections*. Created 14 Aug. 2019, updated 27 Mar. 2024, https://data.ca.gov/dataset/short-term-occupational-employment-projections.

Christopher, Antony. "K-Nearest Neighbor." *Medium*, The Startup, 1 Feb. 2021, medium.com/swlh/k-nearest-neighbor-ca2593d7a3c4.

Hanson, Melanie. "College Enrollment & Student Demographic Statistics." *Education Data Initiative*, 31 Aug. 2024, educationdata.org/college-enrollment-statistics.

25-Word Proposal Description

Analyzing the relations between occupations' entry-level education and employee wage versus their stability and employment projections for 2025 using California's Employment Development Department's government data.