2.⁠ ⁠Background Research

The relationship between demographic factors and income has been extensively studied using machine learning techniques. Chakrabarty and Biswas utilized the UCI Adult Dataset and applied Gradient Boosting to classify individuals into income brackets (>50K or ≤50K) (Navoneel Chakrabarty, 2018). They achieved a high prediction accuracy of 88.16%, highlighting education and employment features as critical determinants of income. This work emphasizes the value of advanced machine learning algorithms in uncovering patterns in socioeconomic data. Similarly, various machine learning algorithms, including decision trees, to analyse the Adult Census Income Dataset (Chet Lemon, 2018). His research underscored the importance of education and occupation as significant factors in predicting higher income levels, with a focus on improving data preprocessing techniques to enhance model performance. Lemon, Zelazo, and Mulakaluri also investigated the dataset using Naïve Bayes, Logistic Regression, and Decision Trees. They found Decision Trees to be the most effective, identifying education and hours worked per week as the most influential predictors of income. These studies collectively demonstrate the critical role of education in income prediction while highlighting gaps in evaluating proportional differences across educational tiers. This underscores the need for further research into the nuanced relationship between education and income.

Why RQ is of Interest

Despite numerous studies utilizing the UCI Adult Dataset, there remains a gap in fully understanding the nuanced impact of education on income levels. Prior research confirms education as a critical factor in income prediction, but limited work has been done to evaluate the proportional differences across specific educational tiers. This research aims to address this gap by investigating the question: *Is there a difference in the proportions of income across different levels of education among adults in the USA?* Addressing this question is vital for crafting policies aimed at reducing income inequality and enhancing economic mobility. Future directions include integrating longitudinal data to analyze trends over time and refining models to account for complex, non-linear relationships between education and income.

# Bibliography

Ali, A., April 2020. *Adult Census Income Dataset Analysis,* s.l.: s.n.

Chet Lemon, C. Z. K. M., 2018. *Predicting if income exceeds $50,000 per year based on 1994 US Census Data with Simple Classification Techniques,* s.l.: s.n.

Navoneel Chakrabarty, S. B., 2018. *A Statistical Approach to Adult Census Income Level Prediction,* s.l.: s.n.