

# *Mini Project: Socio-Economic Factors that Impact Potential Earning Income*

Vathana Him

November 1, 2021

## **1 Abstract**

The purpose of this mini research project is to examine future income based socio-economic factors based on the census data set gathered from the UCI Machine Learning Repository. This dataset contains many features that are intuitively considered to be predictors of income earning potential. These features are age, education, workclass, marital-status, occupation, relationship, race, capital-gains, capital-loss, hours-per-week worked, and native country. This data set will be used to train two-class classification models to predict which features will likely lead to an income of greater than fifty thousand dollars in USD and less than fifty thousand dollars in USD. Two machine learning models that will be trained and used to predict this scenario. The two machine learning models that will be used are Support Vector Machine (SVM) and Random Forest Classifiers. The result of this mini-project will be determined in two-scope: the determination of factors that influence potential income earning and the analysis of the result between SVM and Random Forest Models. The analysis of the results between the two models will be analyzed in order to understand the accuracy of the two models for this dataset, and thus, will indicate which machine learning model will be best for this dataset.

## 2 Introduction

Income inequality has been one of the most prominent social problems in America during modern times. It is often one of the main focuses of debates during midterm and presidential as well as through lenses of both the Republican and Democratic party. The increase in wealth generated for the upper-class of American household had come at the expense of the middle class as the wealth of the middle class of American families had experienced a decline since the late 1970s <sup>1</sup> Income has been one of the main contributors to wealth and financial security as it serves to protect American families from the cycle of economics ups and downs that is the characteristics of a capitalist economy.

The more income the average American families had, the more they're able to put into their savings and investments, and thus provided a financial security blanket for times of economic recessions. It can be seen time and time again that throughout the history of recessions in America, the middle class had been hit the hardest. <sup>2</sup> Examples of this can be seen through the lenses of the most recent recessions during the 2008 Fi-

nancial Crisis and the 2020 Corona Virus pandemic. These recent recessions slashed into the already thinning of the American middle class as it caused many to lose their homes and jobs, and at worst their businesses and livelihood.

As cycles of elections continue to happen, many presidential candidates and lawmakers have debated their ideas and proposed solutions to solve this growing gap of the American dying middle class. The recent polarization of American politics had left no room for this problem to be solved in a timely manner as many conservative and liberal politicians continued to enforce their own point view and thus, hindering the process of providing a solution to fix this issue for the actual victims; the American middle class families.

Therefore, the purpose of this project was to look through the main contributing factors of income through a purely logical and statistical perspective based on the machine learning methods learned in class. The methods used were the training of machine learning models to find the variables that can be seen as correlating to income with the use of support vector machine learning and random forest machine learning. This

---

<sup>1</sup>Horowitz, J. M., Igielnik, R., Kochhar, R. (2020, August 17). Trends in U.S. income and wealth inequality. Pew Research Center's Social Demographic Trends Project.

<sup>2</sup>Weller, C. (2021, January 6). The recession hits an already hollowed-out middle class. Forbes.

was aimed to understand important factors that can contribute to income and if these factors can be predicted to determine a persons potential income based on the targed variables. Finally, the comparisons of the results from these two models were also analyzed.

### **3 Methods:**

## References

- [1] Horowitz, J. M., Igielnik, R., Kochhar, R. (2020, August 17). Trends in U.S. income and wealth inequality. Pew Research Center’s Social Demographic Trends Project.
- [2] Weller, C. (2021, January 6). The recession hits an already hollowed-out middle class. Forbes.
- [3] Income inequality. Inequality.org. (2021, August 30). Retrieved November 20, 2021, from <https://inequality.org/facts/income-inequality/>.
- [4] Schaeffer, K. (2021, May 27). 6 facts about economic inequality in the U.S. Pew Research Center. Retrieved November 20, 2021, from <https://www.pewresearch.org/fact-tank/2020/02/07/6-facts-about-economic-inequality-in-the-u-s/>.