Abstract:

Introduction & Research Question:
Through our study, we will look into the micro-contributors of unemployment, determine the importance amongst them with respect to the unemployment rate, analyze the strength and direction of the correlation with the same. We will then compare the features correlation with that in the literature available.

Literature Review:
We reviewed the existing work on the factors contributing to unemployment other than the well-known economic factors, these factors were mainly related to health and social issues. For our study we selected three factors namely Cardiovascular Disease (CVD), Neurological Disorders, and Interpersonal Violence. From the paper titled "Effects of Health and Education on Labor Force Participation" by the Australian Government, 82% people with no CVD contribute to the labor force where 64% with heart condition do the same. Similarly, for Mental Disorders it is 80% without while only 39% with the condition (Laplagne, P., Glover, M., & Shomos, A. (2007). Effects of Health and Education on Labor Force Participation). Furthermore, According to Legal Momentum, an advocacy group, in the United States, the victims of domestic violence lose an average of 137 hours of work a year. Intimate partner violence causes victims to lose the equivalent of 32,000 full-time jobs each year (Domestic Violence and the Workplace - Workplace Fairness. Workplace Fairness, 2021).

Dataset:
The data for the disease prevalence was collected as part of the The Global Burden of Disease (GBD) project. It is published by both the researchers at the Institute of Health Metrics and Evaluation (IHME) and the 'Disease Burden Unit' at the World Health Organization (WHO), which was created in 1998. The IHME continues the work that was started in the early 1990s and publishes the Global Burden of Disease study.
The data retrieved for this study was for the United States from the year 1990-2016, the features selected are CVD, Neurological disorders, Interpersonal violence per 100,000 persons. Furthermore, the data on unemployment rate was fetched from the The Bureau of Labor Statistics database.

Methodology:
We loaded the data using the open_csv library into the pandas dataset. Both the datasets were merged into one. Since there were no missing values for data between 1990-2016, therefore there was no need for interpolation. Min-Max normalization was performed such that the minimum and maximum values of the features became 0 and 1, respectively. Once the dataset was pre-processed, it was split into train and test sets as well as input and target variables were separated. As a result, we got X_train, y_train, X_test, and y_test while the number of observations for train and test dataset were 21 & 6 respectively.

Two machine algorithms were run on the dataset namely OLS Linear Regression and Random Forest Regression, where the accuracy for the latter was high i.e., 0.94. The mean_absolute_error was 0.17 whereas mean_squared_error came out to be 0.05. The data

was also analyzed by plotting using the MATLAB Library and generated the correlations between the issue/disease and the unemployment rate.

Analysis and Results:

From the features participating in the data modeling, Interpersonal Violence stood out as most important, followed by CVD and then Neurological diseases that contributed in the unemployment rate.

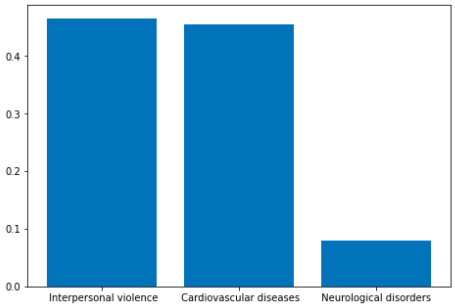

Figure x.x: Feature Important Plot

Furthermore, upon analyzing the data via plotting and generating correlations, we attained the following insights:

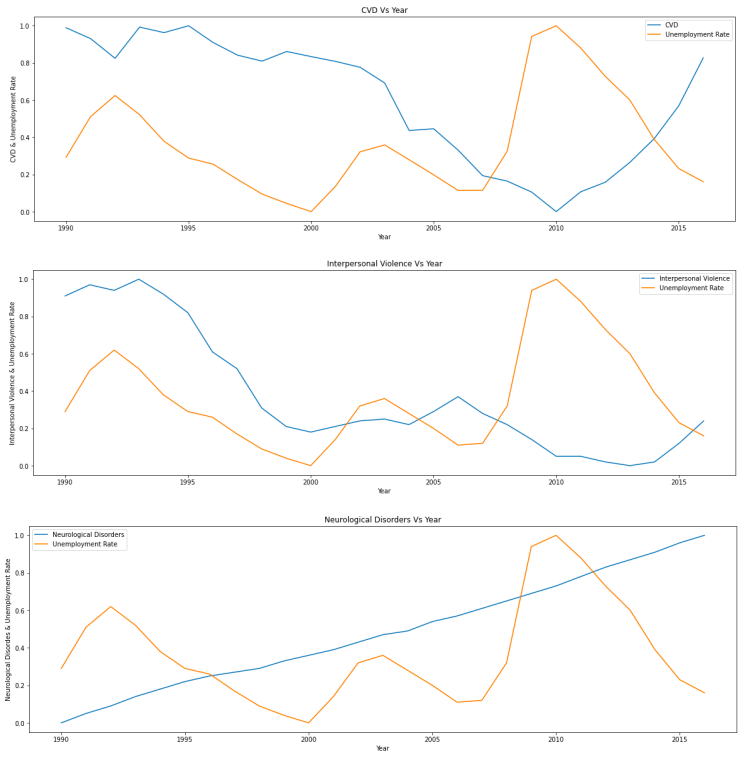

Figure x.x: Trends of Cardiovascular Disease (CVD), Neurological Disease, and Interpersonal Violence occurrences per 100,000 persons and Unemployment Rate between 1990-2016

The resultant correlation scores for occurrences of Cardiovascular Disease, Neurological Disease, and Interpersonal Violence with unemployment were -0.52, 0.23, and -0.1 respectively.

Conclusion:

Through our machine learning based data modeling, it could be seen that Interpersonal Violence stood out from other two issues i.e., CVD and Neurological Diseases, where the later one contributed the least in unemployment.

Moreover, the correlation between unemployment and CVD as well as Interpersonal Violence is negative which means that with less people contributing in the labor force results in less people getting CVD, similarly, there is a slightly lesser chance of interpersonal violence for people who are unemployed. However, although weak correlation, people having Neurological Disorder are less likely to be employed.

Our results vary from the literature we reviewed, where with rise in diseases e.g., CVD and Neurological disorder increases the unemployment, and the impact is stronger for Neurological disorder, similarly, for interpersonal violence, due to increased interpersonal violence, the employment rate lifts up, while we were expecting the same for United States like in Australia, similarly the interpersonal violence related results for United States to match the data by Legatum Momentum, however it was not case.

Future Work and Enhancements:

Our data modeling lacked comprehensiveness since the observations count required to train the model was quite low. Furthermore, our features can be added e.g., Arthritis and physical disabilities could be added to the dataset for analysis and modeling.

References:

http//www.workplacefairness.org. (2021). *Domestic Violence and the Workplace - Workplace Fairness*. Workplace Fairness. https://www.workplacefairness.org/domestic-violence-workplace

Laplagne, P., Glover, M., & Shomos, A. (2007). Effects of Health and Education on Labour Force Participation. SSRN Electronic Journal. Published. https://doi.org/10.2139/ssrn.1018889