Data Science in Finance and Economics

https://www.aimspress.com/journal/dsfe

DSFE, 4(2): 309–332
DOI:10.3934/DSFE.2024013
Received: 25 March 2024
Revised: 03 May 2024
Accepted: 27 May 2024
Published: 19 June 2024

Research article

Prediction of minimum wages for countries with random forests and neural networks

Matthew Ki and Junfeng Shang*

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH, 43403, USA

* **Correspondence:** Email: jshang@bgsu.edu.

Abstract: Minimum wages reflect and relate to many economic indexes and factors, and therefore is of importance to mark the developmental stage of a country. Among the 195 countries in the world, a handful of them do not have a regulated minimum wage mandated by their governments. People debate as to the advantages and disadvantages of imposing a mandatory minimum wage. It is of interest to predict what these minimum wages should be for the selected nations with none. To predict the minimum wages, motivations vary with the specific country. For example, many of these nations are members of the European Union, and there has been pressure from this organization to impose a mandatory minimum wage. Open and publicly available data from Excel Geography are employed to predict the minimum wages. We utilize many different models to predict minimum wages, and the random forest and neural network methods perform the best in terms of their validation mean squared errors. Both of these methods are nonlinear, which indicates that the relationship between the features and minimum wage exhibits some nonlinearity trends that are captured in these methods. For the method of random forests, we also compute 95% confidence intervals on each prediction to show the confidence range for the estimation.

Keywords: random forests; neural networks; deep learning; minimum wages; prediction; excel geography data

JEL Codes: J31, C45, C13, C88

1. Introduction

A minimum wage is the lowest remuneration that employers can legally pay their employees. The purpose of minimum wages is to protect workers against unduly low pay and to help ensure a fair and equitable share of fruits of progress to all, and a minimum living wage to all who are employed and in need of such protection (International Labor Organization, 1996-2024). Minimum wages also play an important role in overcoming poverty and reducing inequality.

The topic of minimum wages is not without debate by economists and policy makers. One debate is whether minimum wages actually benefit low-income workers. Regulations are only as good as their enforcement. Many employers may pay under-the-table to evade taxes or may not report employment at all, avoiding the minimum wage law entirely. According to Eurofound, the European Union agency for the improvement of living and working conditions (Eurofound, 2018), non-compliance with the minimum wage was an issue in Germany and the Netherlands. The lack of compliance has called for a government inspection that has yet to begin. Another argument is that, even where minimum wages are enforced, many may face heavy labor or social security taxes. Employers may also offset higher wages by reducing benefits or laying off workers. This circles back to the problem that minimum wages are trying to solve. By incorporating these regulations and increasing minimum wages, employers may end up employing less as a result, thus paradoxically increasing wealth inequality.

There have been many studies researching this topic with mixed results, for example, in Romich J (2017) and MaCurdy T (2015).

Currently, what is typically done to determine minimum wages in Europe are a tripartite consultation between the government, the unions, and the employers. In Bulgaria, Estonia, Hungary, Latvia, Lithuania, and Spain, these social partners were able to agree on the minimum wage, and the government followed the recommended level. Another way of determining minimum wage levels is with the involvement of independent expert committees. This is what is done in countries like France, Germany, Ireland, Malta, and the UK (Eurofound, 2018). In Estonia, there were discussions arguing that an increase in the minimum wage would artificially increase wage levels without actual growth in productivity. Ultimately, the trade unions disagreed and recommended to increase the minimum wage. In Belgium, the Netherlands, and Malta, indexation mechanisms have been used to determine the level of minimum wage in 2018. Depending on the country, the determination of minimum wages can be influenced by the percent change in contract wages expected, cost of living, and the retail price index.

Interestingly, there are a handful of nations with no minimum wages. In ten such countries of interest, seven of which are in Europe, while the other three are Singapore, Rwanda, and United Arab Emirates. We are not debating whether the advantages of a minimum wage outweigh its disadvantages. Rather, the interest is to predict an estimate and a range considering a random variation of what these countries' minimum wages should be based on their other economics and related characteristics with the statistical methods.

Minimum wages can reduce inequality, and a measure of income inequality is the Gini index. The Gini coefficient incorporates the detailed shares data into a single statistic which summarizes the dispersion of income across the entire income distribution as shown in United States Census Bureau (2024). The Gini coefficient ranges from 0, indicating perfect equality (where everyone receives an equal share), to 100, perfect inequality (where only one recipient or group of recipients receives all the income).

We focus on the prediction of minimum wages. While it is worthwhile to relate minimum wages to the predictors, we predict the minimum wages of the ten countries with no current minimum wage regulations. The main methods adopted here are random forests and neural networks, which serve as better tools in prediction.

Section 2 introduces the backgrounds of related countries. Section 3 describes the data utilized for the analysis. Section 4 uncovers structure in the data with exploratory data analysis. Section 5 presents the method of random forests and the analysis results. Section 6 presents the method of neural networks and the analysis results. Section 7 further discusses the analysis results for the random forest and neutral network approaches. Section 8 concludes and discusses.

2. Related country backgrounds

The following presents the breakdown of the regulatory climate of ten countries of interest regarding wage allocation.

2.1. European countries

As of this writing, the following countries in Europe have no minimum wage. Their respective Gini coefficients (Gini Index , World Bank Estimate) are reported as well.

- Austria - Gini: 30.2
- Denmark - Gini: 27.7
- Finland - Gini: 27.7
- Iceland - Gini: 26.1
- Norway - Gini: 27.7
- Sweden - Gini: 29.3
- Switzerland - Gini: 33.1

Note that the Gini coefficient serves as a measure of inequality, and a larger value implies a severe inequality exists and also indicates a necessity of a minimum wage. These coefficients are between 20 and 40, indicating that there is adequate equality in these nations. Instead of a minimum wage system, the determination of wage is a matter of agreement between the employer and employee. Many of these agreements are in the form of collective agreements, where on one side is a trade union or a collective bargaining unit and an employer, company, or employers' organization on the other side. In Denmark, for example, trade unions are entitled to conclude collective agreements with employers. The rules of industrial action are based on extensive case laws from the Danish Labour Court.

These nations are more developed and have a higher life expectancy and GDP per capita. Sweden, Denmark, Switzerland, Norway, and Finland are amongst the top ten countries with highest quality of life (USNews, 2024). This ranking is achieved through broad accesses to food, housing, education, healthcare, and employment, in addition to job security, political stability, individual freedom, and environmental quality. Given their relatively low Gini index score, there does not seem to be major income inequality.

Austria, Denmark, Finland, and Sweden are members of the European Union. In 2019, the President of the European Commission, Ursula von der Leyen, promised to ensure that every worker in the Union has a fair minimum wage that allows for a decent living, increase in equality, and reduced poverty. Many Nordic countries including Denmark and Sweden responded negatively from politicians, trade unions, and employers. Despite the friction, there is a push from the European Union for a minimum wage which will affect countries with none at the moment. Many countries in the European Union stressed that pay gaps or pay inequalities have arisen as a consequence of the application of different minimum wages to the focus on applying lower rates of minimum wage to promote youth employment in the aftermath of the financial crisis Eurofound (2018). We propose an estimate for what the minimum wage would be in these seven European countries.

In Austria in January 2017, the government asked social partners to negotiate the implementation of a cross-sectional monthly minimum wage of 1500 Euros and to present a solution by mid–2017. The social partners agreed to this which prevented statutory implementation by the government.

Amongst all the nations in the European Union with minimum wages, the range is rather large. The variation ranges from 260 Euros in Bulgaria to 2000 Euros in Luxembourg. This suggests that the predictions amongst these nations may be more spread out.

2.2. Rwanda

Rwanda is a predominantly rural landlocked country in Central Africa. Its leading sectors include energy, agriculture, trade, and hospitality. The minimum wage here was set over 40 years ago in 1973 at 100 Rwandan Francs (RF), or to the equivalent of 0.00087 United States dollars. Since then, none of its ministers has pushed for an increase. According to the President of the Socialist Labour Party, Jean Baptiste Rucibigango, minimum wage should be increased from 100 RF to at least 750 RF. In 2012, a survey conducted by the University of Dar-es-Salaam, Kigali Independent University and University of Amsterdam (Taarifa, 2017) showed that the minimum wage in Rwanda should be 450 RF. As part of that survey, it showed that, "26 percent of Rwandan workers earned less than 150 RF per hour, another 24 percent earn between 150 RF and 450 RF, 29 percent earn between 450 RF and 1350 RF. The remaining 21 percent earn more than 1350 RF per hour." The nation's Gini index score 43.7 indicates that there is a big income gap. This can also indicate that there is social and political instability in a nation.

The fraction of workers currently covered by minimum wages in sub-Saharan Africa is small, but according to Bhorat et al. (2017), wage regulation will become increasingly relevant. This study confirms our analysis here that higher minimum wages are associated with higher levels of GDP per capita. Specifically in the region that Rwanda is in, the nations here with low income have relatively higher minimum wages than middle or upper income countries. Many nations in this area have complex minimum wage schedules, and on average there are high levels of noncompliance among covered workers. There are studies which state that, by and large, introducing and raising the minimum wage may have small negative or no employment impacts. There are also studies where increasing the minimum wage has substantial negative effects on employment.

2.3. Singapore

Singapore is an island country in Southeast Asia. 75% of the country's GDP comes from the services sector and it employs about 80% of the eligible workforce. The government here does not prescribe minimum wages. The Ministry of Manpower, the ministry of the Government of Singapore responsible for the formulation and implementation of policies related to the workforce in Singapore, argues that a competitive pay structure determined by supply and demand of labour will motivate employees to work harder. In 2018, Singapore was ranked 149 in an Oxfam index of 157 countries based on efforts to reduce inequality between rich and poor. There is much debate as to whether a federal regulated minimum wage would close this gap. Singapore's Gini index score 45.9 indicates that there is a big income gap, which is a goal for incorporating a minimum wage.

2.4. United Arab Emirates

The United Arab Emirates (UAE) is a country in West Asia in the Middle East. Hydrocarbons play an important role in its economy, with 30 percent of the UAE's GDP directly in the oil and gas industry. The nation's Gini index score 26 indicates adequate equality. The workers in the UAE's manufacturing sector are among the lowest paid worldwide according to the IMD World Competitiveness Centre's Labour Market Index. In 2012, an average hourly compensation for manufacturing workers in the UAE was 12.4 United Arab Emirates Dirhams or 3.38 United States dollars. Unlike in the European countries, it is illegal to form labour unions making it hard for employees to exercise any control over their pay.

3. Data description

The data are extracted from "Excel Geography" and can be found in Ki M (2024), and the features for each country are downloaded. There are currently, as of this writing, 195 countries in the world. It should be noted that we are only analyzing countries, not territories. This means that territories like Guam, Puerto Rico, and the Gaza Strip are excluded from this analysis. After excluding the missing data, the remains consist of 162 countries with 35 attributes (including minimum wage) for each. Each feature is continuous and included:

- Latitude
- Longitude
- Distance from Equator
- Gross Domestic Product (GDP)
- Gross Domestic Product per Capita (GDP divided by population)
- Population
- Urban Population
- Population Density (Population divided by country's area)
- Birth Rate (Birth rate, also known as natality, is the total number of live human births per 1000 population for a given period divided by the length of the period in years.)
- Infant Mortality (Infant mortality is the death of an infant before his or her first birthday. The infant mortality rate is the number of infant deaths for every 1000 live births.)
- Fertility Rate (The total fertility rate in a specific year is defined as the total number of children that would be born to each woman if she were to live to the end of her child-bearing years and give birth to children in alignment with the prevailing age-specific fertility rates.)
- Maternity Mortality Ratio (The maternal mortality ratio (MMR) is defined as the number of maternal deaths during a given time period per 100,000 live births during the same time period. It depicts the risk of maternal death relative to the number of live births and essentially captures the risk of death in a single pregnancy or a single live birth.)
- Death Rate (The number of people per thousand who die in a particular area in a year)
- Physicians per Thousand
- Life Expectancy
- Out of Pocket Health Expenditure (Out of Pocket Health Expenditure as a percent of how much paid OOP over total cost)
- Consumer Price Index (CPI)
- Consumer Price Index Change

- Gasoline Price
- Minimum Wage
- Total Tax Rate
- Unemployment Rate (The unemployment rate represents the number of unemployed people as a percentage of the labor force (the labor force is the sum of the employed and unemployed).)
- Pop Labor Force Participation (Percent of people participating in labor force)
- Literacy (Percent of population that is literate)
- Phones (Phones per thousand people)
- Area (Area in square miles)
- Coastline (Coastline in miles)
- Agriculture Land (Percent of land used for agriculture)
- Forested Area (Percent of land forested)
- Carbon Dioxide Emissions
- Armed Forces Size
- Agriculture (Percent of labor in agriculture industry)
- Industry (Percent of labor in the industrial or manufacturing industry)
- Service (Percent of labor in the service industry)
- Net migration

In the collection of this data set, there are 34 countries that had many missing attributes. For example, Vatican City has scarce data when it comes to certain features, like life expectancy, fertility rate, GDP, and many others. We omit these observations from the data analysis.

The data are collected from a number of different sources. The most significant source is Microsoft's geography database, which can be accessed through the Excel application. The data are collected from 2015 to 2019. The population and GDP figures are from 2019. Another data source is a public data set on the website Kaggle, which provides data on the coastline miles, land area, and region. This is an older data set from 2006, however land area and region used should not have changed significantly from 2006. Finally, the remaining data used is to fill in any omissions contained in the last two sources. Various other sources are used to fill in the data.

4. Exploratory data analysis

Before we predict minimum wages, it is helpful to understand the structure of our data. A correlation analysis can detect linear associations between the 34 features, and a principal components analysis bi-plot provides a low-dimensional representation of the data.

4.1. Correlation analysis

The Pearson correlation coefficient is a measure of linear correlation between any two features denoted by x and y in the data. The formula is given here:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

The values of the correlation coefficient can take falls between negative one and one inclusive. A coefficient of one means that there is a positive linear relationship, and negative one means that there is a negative linear relationship.

The correlation coefficient is calculated between each pair of features in the data. The resulting plot is given in Figure 1. The R package that we use to make this graph is called “corrplot”, made by Wei and Simko (2021). The blue denotes a positive linear correlation, while red denotes a negative linear correlation.

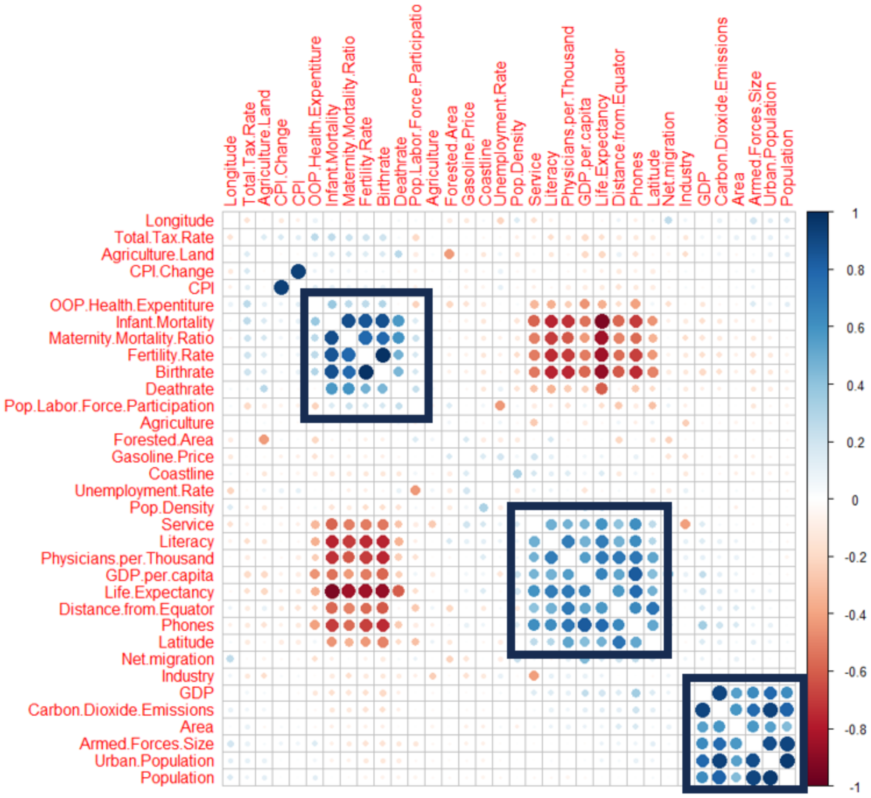

Figure 1. Correlation table for 34 features of 162 countries.

From Figure 1, there appears to be three main positively correlated groupings of variables denoted by the three rectangles. There are also two negatively correlated groups in the data denoted in red in Figure 1. Features in the upper left rectangle are positively correlated with one another, while negatively correlated to the group in the second rectangle going left to right. Then there is a group in the lower right rectangle of the figure that are positively correlated to one another, but not very correlated to the rest of the data.

More importantly, with the existence of correlation between the features as shown here, some of them play an insignificant role in predicting minimum wages, providing a good support for why the random forest model removes unimportant features and still achieves a significant result in the prediction; the random forest method and analysis will be discussed in Section 5.

A summary of the results and a grouping of the features are given by Table 1.

Group A can be categorized as the poor healthcare group. It is understandable that higher infant mortality tends to lead to a higher death rate. Countries with high infant mortality rates are partially a

Table 1. Subgroupings of features from the correlation table.

Group A	Group B	Group C
Out of Pocket Healthcare Expenditure	Service	GDP
Infrant Mortality	Literacy	Carbon Dioxide Emissions
Maternity Mortality Ratio	Physicians per Thousand	Area
Fertility Rate	GDP per Capita	Armed Forces Size
Birthrate	Life Expectancy	Urban Population
Deathrate	Distance from Equator	Population
	Phones	
	Latitude	

function of poverty. Lack of clean water, poor sanitation, malnutrition, endemic infections, and poor or nonexistent primary health care services will effect the life of a newborn before their first birthday. To put it in perspective, the United States' infant mortality rate is 5.6, meaning that for every thousand live births, 5.6 on average will die by their first birthday. The rate in Nigeria, 75.7 (International Trade Association, 2024), is more than 13 times that of the US. The healthcare of Nigeria, according to the International Trade Administration, is "still underdeveloped and lacks modern medical facilities. The country's healthcare indicators are some of the worst in Africa."

Group B is negatively correlated with group A from Table 1. The features in group B are a little more diverse. Overall, they are driven by the traits of good healthcare, economic prosperity, and more educated nations. The obvious variable in the group is life expectancy. It would make sense that life expectancy would be negatively related to the lack of abundance and quality of healthcare within a nation. Another variable is the number of mobile phones per thousand. Mobile phones are an indicator of a better quality of life because access to them can provide educational, healthcare, and financial services to regions with no access. For example, M-pesa (Vodafone Group PLC, 2024) was Africa's first mobile payments service providing more than 51 million African customers with financial services like sending and receiving money. The literacy rate is also in this group and can be loosely related to the amount of education people in a country have. It is also correlated with an abundance of service jobs which usually require literacy as a prerequisite. GDP per capita is a global measure for evaluating the economic prosperity of nations. It is not surprising that it is correlated with literacy and phones.

Variables in group C are correlated with one another, but not with any others outside this group. The dominant trait in this group is population size. It intuitively makes sense that nations with greater population have greater gross domestic product, carbon dioxide emissions, more land area, and a bigger military. It is important to note that gross domestic product is the total value of goods and services for final use made by residents of the country. GDP per capita divides GDP by a nation's population and measures a country's economic output per person. A country may have a large GDP, but when divided by a large population it may have lower economic output per person than a nation with lower GDP and lower population. Economists often use GDP to analyze a country's domestic productivity, and use GDP per capita to determine how prosperous countries are.

These three groups can also be clearly exhibited in conducting the principal components analysis (PCA) in the next subsection.

4.2. Principal components analysis

PCA is a dimension reduction method that helps visualize the data in an informative way. It finds a low-dimensional representation of a data set while containing the most variation possible. Each of the dimensions in PCA is a linear combination of the number of features in the design matrix $\boldsymbol{X}$. There exists an eigen decomposition of $\mathbf{X}^T\mathbf{X}$ such that $\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{V}^T$ (James et al. , 2021).

We have $\boldsymbol{D} = diag[\lambda_1, \lambda_2 ... \lambda_p]$ with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$, denoting the non-negative eigenvalues of the matrix $\mathbf{X}^T\mathbf{X}$, and the eigenvectors v_j (columns of $\boldsymbol{V}$) are called the principal component directions of $\boldsymbol{X}$. The first principal component direction v_1 has the largest sample variance amongst all normalized linear combinations of the columns of $\boldsymbol{X}$. Subsequent principal components have decreasing variance subject to being orthogonal to the earlier ones.

PCA is performed on the data, and Figure 2 is the result. We used the R packages "factoextra" (Kassambara and Mundt , 2020) and "ggplot2" (Wickham , 2016) to create the plots.

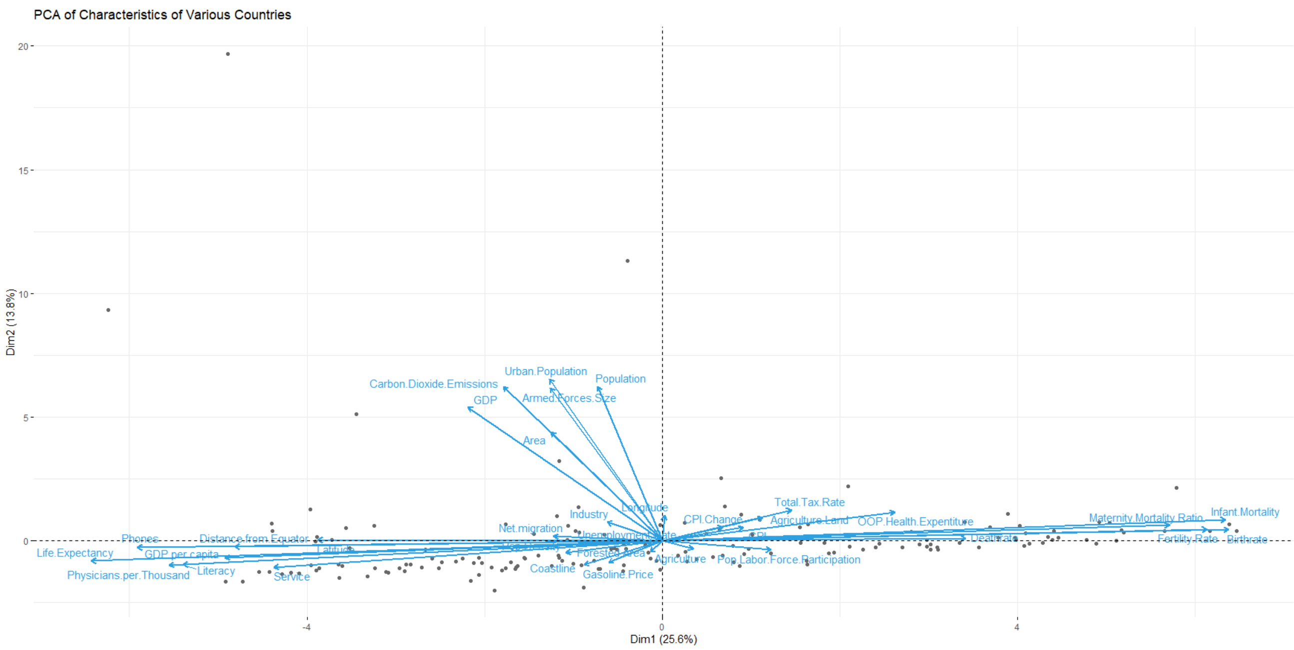

Figure 2. Principal components analysis on the country data with feature labels.

The PCA graphs show a two-dimensional representation of the data. Each dot represents a different nation, and each vector represents a different feature. The first two components explain 39.4 percent of the variance. There are revealing groupings within the data that are similar to the results from the correlation table. Recall the three groups revealed by the correlation table. The same three groupings appear in the PCA plot in Figure 2. The x-axis can be thought of as the life expectancy or quality and abundance of healthcare axis. Nations towards further to the left tend to have higher GDP per capita, life expectancy, and physicians per thousand. Countries to the right tend to have the opposite, with higher infant mortality and death rates. The y-axis is dominated by the third group trend of population. The loading vectors here of population, urban population, armed forces size, carbon dioxide emissions, area, and GDP are located close to one another, indicating correlation, but are far away from the variables at play driving the x-axis, indicating that they are uncorrelated with them.

Figure 3 is the same as Figure 2 except with country labels instead of feature labels. The PCA bi-plot presents the positions of individual countries. What immediately pops out in Figure 3 is the distance that China, the United States, India, and Russia are from the rest of the countries. The positions of these countries are predominantly in the upper left-hand quadrant of the graph. China and the United States seem to stand out in terms of long life expectancy and high GDP per capita in combination with high population and armed forces size. The latter is not surprising since China has the largest population at a size of 1,413,142,846 people and the United States ranks third with a population of 334,994,511 people.

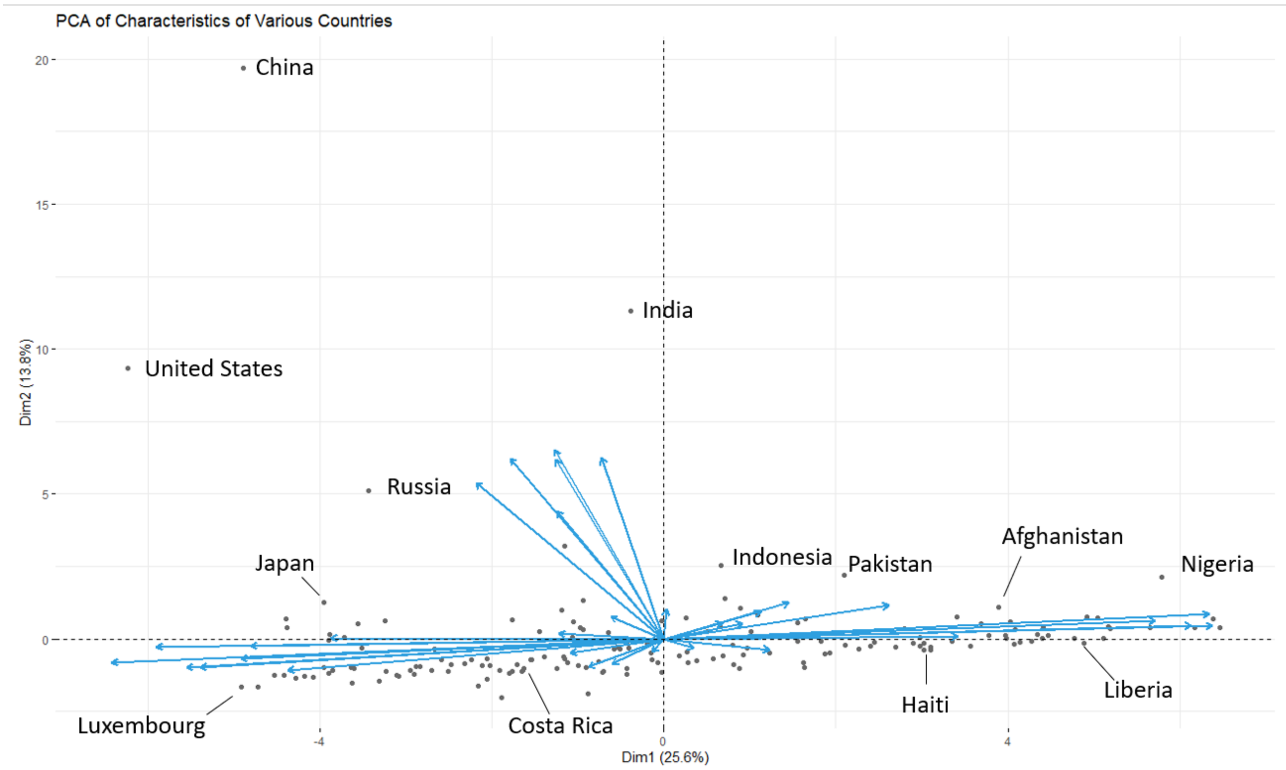

Figure 3. Principal components analysis on the country data with country labels.

5. Methodology and analysis of random forests for prediction of minimum wages

5.1. Methodology of random forests

Random forests are an extension of bagging trees. Bagging trees can be thought of as bootstrap aggregation, where we take repeated samples from the single training set and generate B different bootstrapped training sets. From each of these training sets we fit B regression trees, and the predictions for the response variable denoted by $\hat{f}^{*1}(x), \hat{f}^{*2}(x), ..., \hat{f}^{*B}(x)$ are averaged over the B samples (James et al. , 2021), and the average is written as

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{*b}(x). \tag{1}$$

The drawback to bagged trees are that the trees are correlated to one another. Each time a split in a tree is considered, a random sample of m predictors is chosen as candidates to be split from the full set of predictors. Only one of those m predictors is allowed for the split and a new sample of m predictors is taken at each subsequent split. The number of predictors considered at each split is approximately equal to the square root of the total number of predictors.

Random forests decorrelate the trees by forcing each split of the regression tree to consider only a subset of the predictors, which results in many different kinds of trees. Averaging over a more diverse set of trees leads to a larger reduction in variance of predictions, which improves the performance in predicting. Based on this advantage, the utility of random forests for prediction is prioritized.

The algorithm for random forests starts by drawing B bootstrap samples of size N from the training data and then grows a tree by picking the best splitting point from a subset of the predictors and splitting the node into two nodes. Such an algorithm is done for all B bootstrap samples, and the prediction is derived by averaging over the B samples in the term of a regression response denoted by $f(x)$. The prediction estimate as in Equation (1) is utilized to calculate the prediction, though the bootstrap samples are generated using the random forest algorithm.

Next, we would like to calculate standard errors for our predictions from the model. We use a method adopted by Wager et al. (2014). Their work highlights estimates for bagged predictors that can be computed from bootstrap samples. The goal is to obtain the variance of the prediction once we make B large enough to eliminate the bootstrap effects. One of the two methods proposed and adopted here is called the infinitesimal jackknife estimate, and the estimated variance for the prediction is written as

$$\hat{V}_J^{\infty} = \frac{B-1}{B} \sum_{i=1}^{B} (\bar{f}^*_{(-i)}(x) - \bar{f}^*(x))^2, \tag{2}$$

where $\hat{V}_J^{\infty}$ is the jackknife after bootstrap estimate. $\bar{f}^*_{(-i)}(x)$ is the average of estimated $f^*(x)$ over all the bootstrap samples except the ith sample. $\bar{f}^*(x)$ is the mean of all the estimated $f^*(x)$. The infinitesimal jackknife is an alternative to the jackknife where we look at what happens to a statistic when we individually down-weight each observation by an infinitesimal amount.

We can obtain the estimated standard error of our prediction, which is the square root of the estimated variance in Equation (2). We then compute a 95% confidence interval for the prediction with the estimated prediction and the estimated standard error denoted by $\hat{se}$, which is produced by the infinitesimal jackknife method in Equation (2). The confidence interval can be calculated as Gaussian confidence intervals

$$\hat{f}(x) \pm z_{\alpha/2}\hat{se}, \tag{3}$$

where $\hat{f}(x)$ denotes the predicted value of $f(x)$, which here are minimum wages, α denotes the significance level for the confidence interval, and $z_{\alpha/2}$ is the upper percentile of the standard normal distribution. We then apply this confidence interval to our estimates for the minimum wages in Expression (3), reflecting the confidence for estimating the minimum wages. From the confidence interval, we use $100(1 - \alpha/2)\%$ confidence to express that the true minimum wage should be within the range of confidence interval.

5.2. Results of analysis

In the process of choosing the best model to predict minimum wages, we split the data into a training and validation set. We allocate 80% of the data to the training set and the remainder to the validation set. Simple linear regression, multiple linear regression, decision trees, and boosted trees are also considered.

The random forest approach performs the best among these tree methods, and second best just behind neural networks.

The statistic we use to compare the models is root mean square error (RMSE). The equation for RMSE is given below. We have

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}},$$

where y_i is the actual observed response in the validation set, and $\hat{y}_i$ is the predicted response from the model. Ideally, the lower the RMSE is, the better the model is at predicting the actual observed response with a better accuracy.

To perform our analysis, we use the R package "ranger" created by Wright and Ziegler (2017). For the random forest approach, a model with 1000 trees performs the best as the test RMSE tended to level off at around this threshold. More predictors are included in a tree may generate a tree with a greater depth. When all the available predictors are used in the random forests, an RMSE of 1.265 is obtained. This can be interpreted to mean that the average prediction of our model is \$1.265 off from the true minimum wage.

Although the random forests are rather complex, we can determine which variables are more important to the model. Figure 4 depicts the relative importance of all the variables. Variable of importance is a measure of by how much removing a variable decreases accuracy. We can see from Figure 4 that GDP per capita plays a big role in the prediction results followed by the number of phones per thousand people, life expectancy, infant mortality, and maternity mortality ratio.

Since some of the predictors are not important in the random forests prediction, we run the model with a subset of the predictors. The predictors we drop are below the black line in the variable importance graph in Figure 4. Interestingly, the RMSE drops from the prior value of 1.265 to 1.091. This means that GDP per capita, phones per thousand, life expectancy, infant mortality, maternity mortality ratio, percent of labor in the service sector, net migration, percent of labor in the agriculture industry, and literacy rate play a big role in predicting minimum wages. This suggests that the features of developed nations like high GDP per capita and high life expectancy command a higher minimum wage. Recall from the correlation section that many of these features are positively and negatively correlated with one another. We run a model with only GDP per capita, infant mortality, and net migration since these are all in different groups in Table 1. The corresponding RMSE is a little higher at 1.114. Since the goal of the prediction is to minimize prediction accuracy, we run analysis on the model with the lowest RMSE, which is the model with the nine predictors above the black line in Figure 4. The training and validation sets are then combined to perform the final minimum wage predictions.

The ten countries of interest are run through the model and the standard errors are computed. A 95 percent Gaussian confidence interval is calculated by adding and subtracting 1.96 times the standard error to the prediction. We have the confidence interval

$$\hat{y} \pm z_{\alpha/2}\hat{se}, \tag{4}$$

where $\hat{y}$ is the predicted value and $\pm z_{0.025}\hat{se}$ represents that the middle $1-\alpha/2$ probability of the standard normal distribution is contained within ± 1.96 standard deviations from the mean when $\alpha = 0.05$. The results can be summarized in Table 2. For example, Denmark's point estimate minimum wage is \$10.20, meaning the random forest model predicts Denmark's minimum wage to be this number. The 95% confidence interval for Denmark is \$8.67-\$11.73, which means that confidence intervals created by this

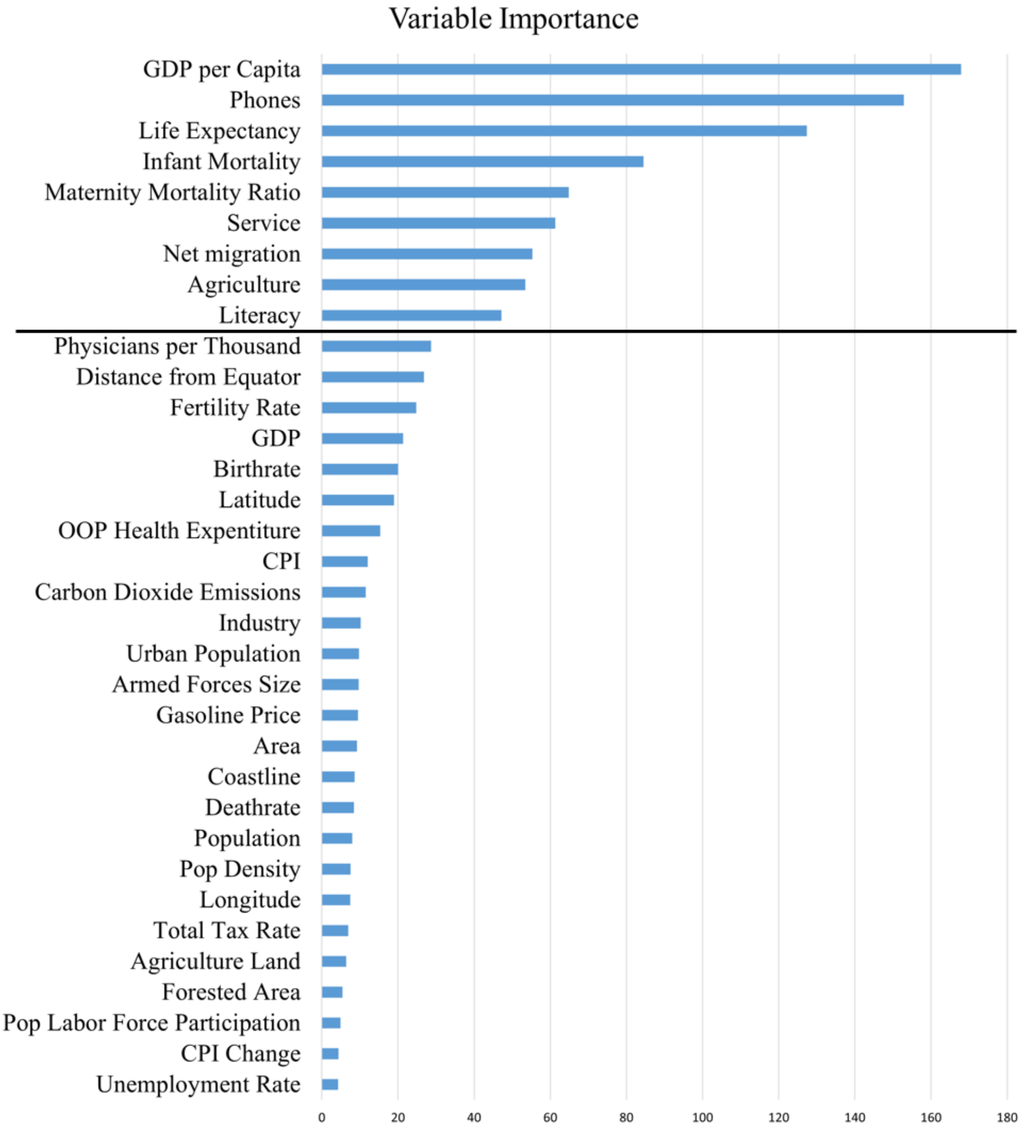

Figure 4. Variable importance from random forest model.

method would trap the true minimum wage prediction 95% of the time, and we are 95% confident that the true minimum wage is within \$8.67–\$11.73.

The following results are then translated back to their original county's currency in Table 3. For instance, Finland's native currency is the Euro, and 1 United States dollar is approximately 0.92 Euros. All the outputs in Table 2 are simply converted to the nation's currency in Table 3. Finally, a bar chart with the 95 percent confidence interval in USD is given in Figure 5. The grey bar denotes the point estimate of the random forest model, and the whiskers represent the 95% confidence interval. For example, Norway's grey bar stops at \$9.46, meaning that the point estimate is this number and the whiskers represent the 95% confidence interval, which are at \$8.69 and \$10.23. The interpretation of this is that the confidence intervals created by this random forest model would trap the true minimum wage 95% of the time, and we are 95% confident that the true minimum wage is within \$8.69–\$10.23. Figure 5 also displays that the model is more certain about the Rwanda, Singapore, and Sweden predictions, and

Table 2. Predicted minimum wage results in USD.

Country	Predicted Min Wage in USD	Standard Error	95% Confidence Interval
Austria	\$8.98	\$0.95	\$7.13-\$10.83
Denmark	\$10.20	\$0.78	\$8.67-\$11.73
Finland	\$7.96	\$0.99	\$6.02-\$9.90
Iceland	\$10.00	\$0.96	\$8.12-\$11.88
Norway	\$9.46	\$0.39	\$8.69-\$10.23
Sweden	\$9.46	\$0.41	\$8.66-\$10.26
Switzerland	\$9.33	\$0.69	\$7.98-\$10.68
Rwanda	\$0.28	\$0.04	\$0.20-\$0.35
Singapore	\$6.94	\$0.39	\$6.19-\$7.70
United Arab Emirates	\$5.36	\$0.73	\$3.94-\$6.78

Table 3. Predicted minimum wage results in native country's currency.

Country	Currency	Predicted Min Wage	Standard Error	95% Confidence Interval
Austria	Euro	€ 8.27	€ 0.87	€ 6.57-€ 9.98
Denmark	Danish Krone	70.05 kr	5.35 kr	59.55 kr-80.55 kr
Finland	Euro	€ 7.33	€ 0.91	€ 5.55-€ 9.12
Iceland	Icelandic Króna	1372.2 kr	131.67 kr	1114.12 kr-1630.28 kr
Norway	Norwegian Krone	99.98 kr	4.15 kr	91.83 kr-108.13 kr
Sweden	Swedish Krona	98.23 kr	4.25 kr	89.89 kr-106.58 kr
Switzerland	Swiss Franc	8.39 F	0.61 F	7.19 F-9.61 F
Rwanda	Rwandan Franc	349.41 RF	50.82 RF	249.8 RF-449.03 RF
Singapore	Singapore Dollars	\$9.31	\$0.51	\$8.3-\$10.33
United Arab Emirates	UAE Dirhams	19.67 Dh	2.66 Dh	14.44 Dh-24.9 Dh

less certain about other nations like Iceland and Denmark. A nation with a tighter confidence interval indicates a smaller variation, i.e., the model is more certain that the process of creating this interval would trap the true minimum wage 95% of the time.

6. Methodology and analysis of neural networks in prediction of minimum wages

6.1. Methodology of neural networks

Neural networks arose in the late 1980s and have recently seen a certain amount of hype for many success stories in image and video classification, speech and text modeling, and others. More importantly, they consist of a layered structure that resembles the human brain and looks like neurons. Through neurons, a nonlinear relation between the predictors and response can be built up, and such a relationship provides a complex and flexible connection between them. High-dimensional predictors can work very well using neural networks. From the neuron-like relationship between the predictors and response, the response variable can be effectively classified or predicted. We start our discussion of neural networks with how this method works, and then we further introduce how the gradient is

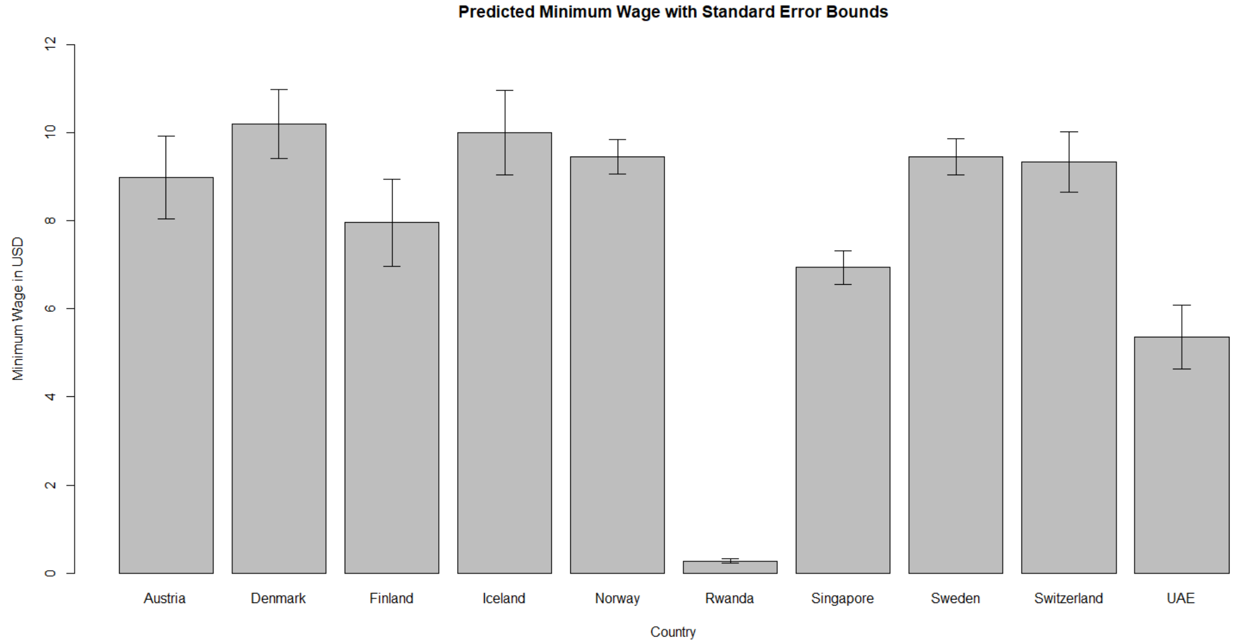

Figure 5. Random forest minimum wage predictions with 95% confidence interval in USD.

computed for optimization, which is for the estimation in neuron network modeling.

Neural networks, or in our case more specifically feedforward networks, can be seen as a network made up of 3 types of layers. There is an input layer working with data directly from the design matrix, hidden layers, and output layers. How the network learns, or optimizes the cost function, is by gradient descent. Further, how it calculates the gradient is called backpropagation. We now briefly provide an overview on how this process works, as shown in Hastie et al. (2009), Goodfellow et al. (2016), etc.

Let us start with the method of gradient optimization. A gradient generalizes the idea of the derivative where the derivative is with respect to a vector. The gradient of f is the vector containing all partial derivatives, denoted $\nabla_x f(x)$. The directional derivative in the direction $\mathbf{u}$ (unit vector) is the slope of the function f in the direction $\mathbf{u}$. To minimize f, we would like to find the direction in which f decreases the fastest. We have

$$\min_{\mathbf{u},\mathbf{u}^T\mathbf{u}=1} \mathbf{u}^T \nabla_x f(x) = \min_{\mathbf{u},\mathbf{u}^T\mathbf{u}=1} ||\mathbf{u}||_2 ||\nabla_x f(x)||_2 cos\alpha,$$

where α is the angle between $\mathbf{u}$ and the gradient. This can be simplified to $\min_{\mathbf{u}} cos\alpha$. This is minimized when $\mathbf{u}$ points directly uphill, and the negative gradient points directly downhill.

The idea behind gradient optimization is the optimization when conducting modeling and estimation and to find a local minimum. Imagine a terrain with many peaks and valleys. We start at an initial point somewhere in the terrain, perhaps on a hill, and we wish to take a path down the hill to the bottom. The process described above is the algorithm that will compute the direction of greatest descent. We then take a step in this direction and repeat the process of computing the next direction of greatest decent. This will find a local minimum of the cost function because the algorithm stops when the gradient is a vector of all zeros, at which a minimum is reached. Consider a hill with a bowl-shaped base but beyond the lip of the bowl, the terrain descends even more. In this case, the algorithm described above will stop at the bottom of the bowl and not consider the lower terrain beyond the bowl. In many cases, local minimums are sufficient to estimate models. In what follows, the minimum of an associated cost function is calculated using the gradient optimization.

The next step is how to calculate the gradient for feedforward neural networks, and it can be conducted by backward propagation. It is useful to introduce the model of a feedforward neural network at this point. The first hidden layer flowing from the input layer to the first hidden layer of the neural network is depicted below (James et al. , 2021):

$$f(x) = B_0 + \sum_{k=1}^{K} B_k g(W_{k0} + \sum_{j=1}^{p} W_{kj} X_j),$$

where B_0, B_k, W_{k0}, and W_{kj} are the model parameters, K is the number of hidden units, p is the number of features (columns) in the design matrix, g is the activation function and varies with some options, X_j of the jth feature is the design matrix, and W_{kj} is the weight matrix for the kth hidden unit and jth feature.

The cost function associated with this network is given by

$$\min \frac{1}{2} \sum_{k=1}^{K} (y_{ik} - f_k(x_i))^2,$$

where K is the number of hidden layers and y_{ik} is the observed response for the kth layer.

For ease of computation, let us look at a single hidden layer of this cost function, $R_i(\theta) = \frac{1}{2} \sum_{k=1}^{K} (y_{ik} - f_k(x_i))^2$. θ represents all the parameters in the neural network model. The computation that follows is to find the direction to move θ to decrease the cost function $R_i(\theta)$. This is equivalent to finding the gradient of $R_i(\theta)$, evaluated at some current value. Finally, the optimization is to estimate θ for the minimum cost $R_i(\theta)$, and then the neural network model is estimated at this estimated θ. We wish to compute the gradient with respect to the parameters B_k and W_{kj}. It is helpful to expand $R_i(\theta)$ and denote certain parts of the expansion by $f_\theta(x_i)$ and z_{ik}. Then we have

$$R_i(\theta) = \frac{1}{2}(y_i - f_\theta(x_i))^2, \;\; f_\theta(x_i) = B_0 - \sum_{k=1}^{K} B_k g(z_{ik}), \text{ and } z_{ik} = W_{k0} + \sum_{j=1}^{p} W_{kj} X_j.$$

We wish to compute $\frac{\partial R_i(\theta)}{\partial B_k}$ and $\frac{\partial R_i(\theta)}{\partial W_{kj}}$. Notice that this is very easy to do with the chain rule of calculus. Then we have

$$\frac{\partial R_i(\theta)}{\partial B_k} = \frac{\partial R_i(\theta)}{\partial f_\theta(x_i)} \frac{\partial f_\theta(x_i)}{\partial B_k} \text{ and } \frac{\partial R_i(\theta)}{\partial B_k} = -(y_i - f_\theta(x_i))g(z_{ik}).$$

Likewise, we have

$$\frac{\partial R_i(\theta)}{\partial W_{kj}} = \frac{\partial R_i(\theta)}{\partial f_\theta(x_i)} \frac{\partial f_\theta(x_i)}{\partial g(z_{ik})} \frac{\partial g(z_{ik})}{\partial z_{ik}} \frac{\partial z_{ik}}{\partial W_{kj}} \text{ and } \frac{\partial R_i(\theta)}{\partial W_{kj}} = -(y_i - f_\theta(x_i)) B_k g'(z_{ik}) X_{ij}.$$

The next step is to apply gradients to update the weights/parameters in the neural networks to minimize the cost function.

Gradient descent defines the updates in a general way. Specifically, in the neural networks here, we can compute the updates with respect to by B_k and W_{kj} as

$$B_k^{(r+1)} = B_k^{(r)} - \gamma \sum_{1}^{n} \frac{\partial R_i(\theta)}{\partial B_k^{(r)}} \text{ and } W_{kj}^{(r+1)} = W_{kj}^{(r)} - \gamma \sum_{1}^{n} \frac{\partial R_i(\theta)}{\partial W_{kj}^{(r)}}.$$

where γ is the learning rate, a positive scalar determining the size of the step. From the rth step, the new updates at the $(r + 1)$th step are calculated, and another new point is computed. The parameters/weights converge when every element of the gradient is zero. In summary, the concept of gradient descent is repeatedly making a small move toward better configurations. It is worth noting that this optimization may fail to find a global minimum when there are multiple local minima or plateaus. This is not usually a problem for neural networks as long as they correspond to significantly low values of the cost function.

6.2. Results of analysis

To run the neural network, we use the R package "neuralnet" created by Fritsch et al. (2019). There are many adjustments one can make when fitting a neural network. The number of hidden layers, choice of activation function, how many units per layer, and what kind of output layer are a few of the tuning parameters. For our analysis, we use the validation set RMSE to tune the number of hidden layers. For the activation function, we choose to use the ReLu activation function because it is pretty standard in fitting networks. Finally, since our response is continuous and not categorical, we decide to use a linear output layer, which is standard for a network like ours.

To determine how many hidden layers to utilize, we fit the model to many different numbers of hidden layers and compute the validation RMSE. The results are given in Table 4. We can see that initially, as we add layers, the RMSE seems to decline. But, after coming to three hidden layers, the RMSE jumps back up and levels off.

Table 4. Hidden layer corresponding RMSE.

Hidden Layers	RMSE
1	1.789
2	1.342
3	0.925
4	1.255
5	1.011
10	1.012

Since the model seems to perform best with three hidden units, we incorporate this model. Next, we choose to use twenty units or neurons per hidden layer as a good balance between too few and too many units or neurons. Both extremes have mixed results with RMSE, and twenty works the best for the number of observations in our design matrix. It is worth noting that the random forest's RMSE is 1.091, so the neural network model with three hidden layers performs better than the random forest by a small margin. Again, the neural network model performs the best out of the six models we explore.

A graphical visual of the full model is displayed in Figure 6. To create this image we use the R package "NeuralNetTools" created by Beck MW (2018). We can see that there are five total layers of the neural network model. Going from left to right, we have first the input layer. Here we have 34 predictors or features from the nations with minimum wages here. Every node of the input layer has a corresponding weight parameter with each of the twenty nodes of first hidden layer. Each of the twenty nodes of the first hidden layer has a corresponding weight parameter with each of the twenty nodes of the second hidden layer, and this continues to go through the third hidden layer. Finally, the twenty nodes of the final hidden layer are connected to the output node to give a numeric estimate for

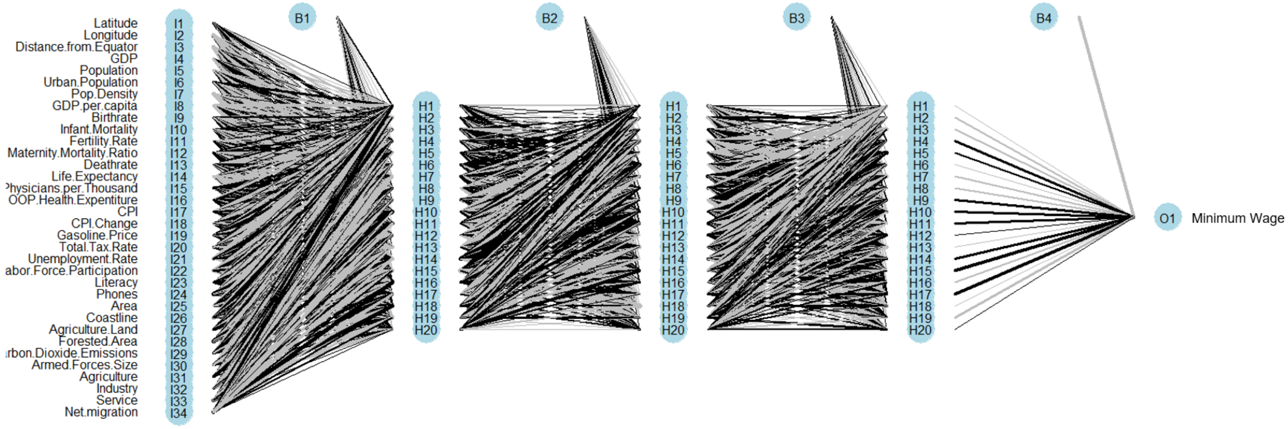

Figure 6. Neural network architecture.

the predicted minimum wage. Along the way, the $B_1, ..., B_4$ along the top of the diagram represent the biases or intercepts. In total there are 272,061 parameters in this network. In conclusion, the predictors or features in the nodes start in the leftmost layer and each node of this layer is connected to each node on the first hidden layer by weight parameters. Biases are introduced at each hidden layer while the predictors in the data flow through the network from left to right. In the final output layer, there is only one node which reports the predicted minimum wage value. After the model is trained, a nation we are interested in predicting minimum wage for is run through the network with all the trained weights and biases to generate a minimum wage output.

The minimum wage point estimates of our model are provided in Table 5. For example, Singapore's point estimate or predicted minimum wage in United States dollars is $6.52. This result translates back to Singapore dollars of $8.62.

Table 5. Neural network results.

Country	Predicted Min Wage in USD	Native Currency	Converted
Austria	$5.71	Euro	€ 5.18
Denmark	$8.51	Danish Krone	57.59 kr
Finland	$6.37	Euro	€ 5.78
Iceland	$10.09	Icelandic Króna	1377.38 kr
Norway	$10.56	Norwegian Krone	107.43 kr
Sweden	$11.14	Swedish Krona	111.23 kr
Switzerland	$10.51	Swiss Franc	9.45 F
Rwanda	$0.96	Rwandan Franc	1199.01 RF
Singapore	$6.52	Singapore Dollars	$8.62
United Arab Emirates	$4.59	UAE Dirhams	16.84 Dh

7. Further interpretation of analyzed results

Recall that the two best models were a neural network and a random forest with respective RMSEs of 0.925 and 1.091. Although the neural network performs better, the results from the two models are not too far apart, see Figure 7. We can see that for some countries like Iceland, Singapore, and the United Arab Emirates, the predictions are close to each other. However, for other nations like Austria, the predictions are farther apart. Since the neural network model has a lower RMSE, we lean more in favor of those results.

The neural network model may perform better because it is more complex than random forests in the sense that there are 272,061 parameters in total. Random forests are essentially an average of bagged trees where a subset of predictors are considered at each split of the decision tree. Compared to neural networks, they are not as complex and may not capture as much non-linearity as neural networks.

It is worth noting the advantages and limitations of both methods in the context of our data. Even a modest sized neural network has a large number of parameters to optimize. In our case, the number of parameters to optimize are over 1000 times the observations in the training set. Many times, when training these types of models, it is advantageous to have a very big training set. This is limited in our analysis because there are only a small finite number of countries in the world. There is also the possibility for the neural network to overfit, but the training error is impressive. We also assume that the missing data we have was missing at random. This is so that we use the observed data as a good representative of the complete data. If this assumption is violated, then this may impact the reliability of the predictions. Finally, neural networks may be very flexible and can catch nonlinear complex relationships but at the cost of model interpretability. Policy makers may be interested in how the model makes decisions. Neural networks are very complex and hard to interpret. It is easier to interpret random forests because they are less flexible, as shown in Figure 4, and we are able to obtain a variable of importance plot for a good interpretation.

Figure 8 shows the relationship between the predicted minimum wage from the random forests model and GDP per capita. We can see that, going left to right, the GDP per capita is increasing while the predicted minimum wage tends to increase along with the prediction. One dominant trend is that the countries with higher GDP per capita tend to have higher minimum wages. In fact, GDP per capita and minimum wage have a correlation coefficient of 0.8823, which is a very strong positive linear association. GDP per capita and the predicted minimum wage also have a similar correlation coefficient of 0.8583. Recall from the correlation table section in Table 1 that GDP per capita is positively linearly associated to other features like percent of labor force in the services industry, literacy rate, physicians per thousand, life expectancy, and phones per thousand. These traits can be generalized to nations with good healthcare, more economic prosperity, and more educated people. The main drivers of minimum wage prediction are these traits in nations. In addition, recall that GDP per capita is negatively correlated with group A in Table 1. The opposite can be said about nations with high infant mortality, fertility rate, and other features that can be generalized to be nations with poor of lack of access to good healthcare. The worse the healthcare is, the lower the predicted minimum wage tends to be.

In addition, the random forest model with the nine variables performs better than the model with the full predictors. The nine predictors from the model in order of importance are as follows:

1. GDP per Capita
2. Phones per Thousand

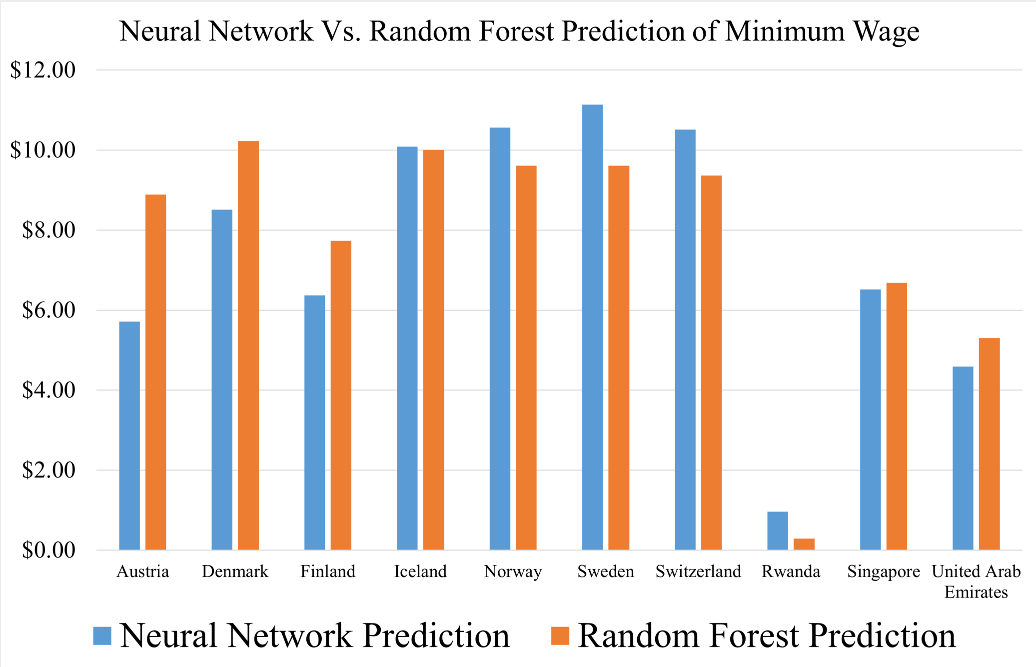

Figure 7. Predicted minimum wage between two models.

3. Life Expectancy
4. Infant Mortality
5. Maternity Mortality Ratio
6. Percent of Labor Force in the Service Sector
7. Net Migration
8. Percent of Labor Force in the Agriculture Sector
9. Literacy Rate

These features are the most important for the random forest model to make predictions. Removing any one of these variables would potentially decrease the accuracy of the prediction. They are in order of importance, showing that GDP per capita plays a major role in minimum wage prediction. Phones per thousand, life expectancy, percent of labor in the service sector, and literacy rate are all correlated with one another and form the characteristics of developed nations group in the prior section in Table 1. These traits are key at predicting minimum wages.

These important predictors may help inform policy decisions. Currently how minimum wages are determined in some nations are a tripartite consultation between the government, unions, and employers. Some countries take into account the cost of living and certain indices like the retail price index. What our analysis shows is that characteristics like GDP per capita and life expectancy can serve as additional criteria in determining minimum wages. A thorough analysis of the current and future trend of GDP per capita and life expectancy can help calculate the corresponding minimum wage.

7.1. European countries

We begin by looking at the European nations that do not have a minimum wage and are in the European Union. This includes Austria, Denmark, Finland, and Sweden. The confidence intervals vary

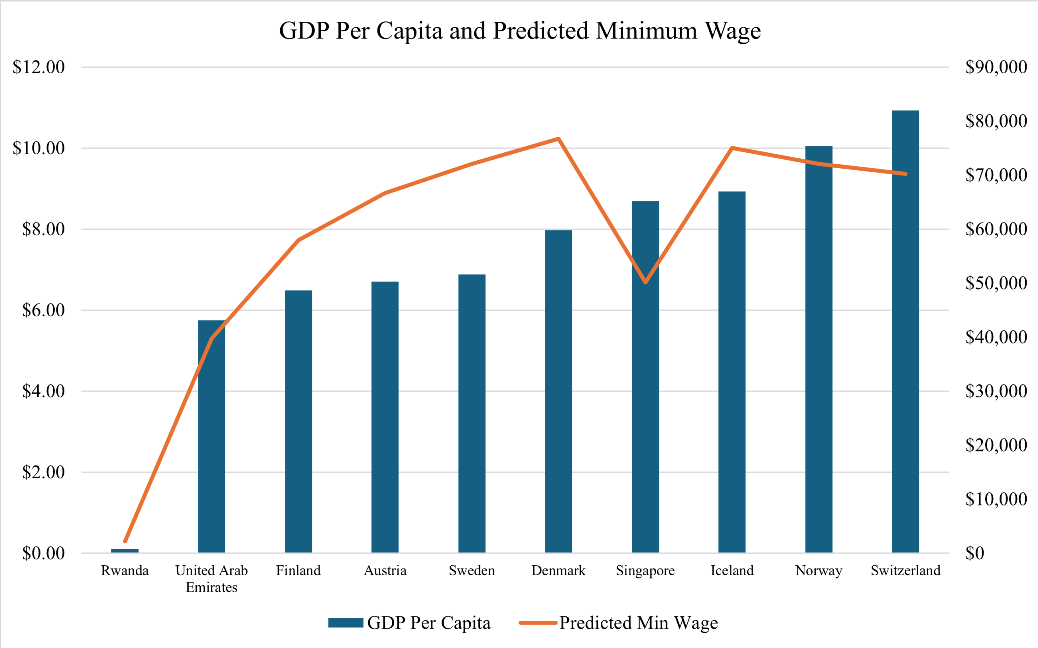

Figure 8. GDP per capita and predicted minimum wage.

for each country. The random forest model seems more certain about the Sweden prediction giving low standard errors. The rest are a little wider.

Between the neural network and random forest models, the results differ. For Austria, the prediction by the deep learning algorithm is significantly lower than the other model at $5.71 USD. This is outside the range of the random forest's confidence interval. Recall that Austria's government called to implement a cross-sectional monthly minimum wage of 1500 Euros. This translates to about 9.375 Euros an hour. This is consistent with the random forest model as the 95% confidence interval covers a range of 6.57 Euros and 9.98 Euros. The neural network model predicts a slightly lower minimum wage for Denmark at $8.61 USD. Similar results are readily seen for Finland. The neural network model predicts a slightly higher minimum wage of $11.14 for Sweden compared to the random forest's $9.46.

Recall that the range of minimum wages in the European Union is large. In fact, there is a range of 1740 Euros amongst these nations. Our highest prediction is Denmark at $10.20, and the lowest is Finland at $7.96. This creates a range of about 2.03 Euros. Although the range is large, our results are well within the range of the typical minimum wages of surrounding nations in the European Union.

Looking beyond the countries in Europe, Iceland's confidence intervals are the widest. This suggests that the random forest model is more uncertain about this prediction. However, the Iceland results between models are very close to each other.

7.2. Rwanda

Rwanda's prediction with the random forest model is $0.28 United States dollars, or 349.41 Rwandan Francs. The 95% confidence interval is between 249.8 RF and 449.03 RF. Since 100 RF is not within this range, we can see that the minimum wage set forty years ago may need to be updated. Jean Baptiste Rucibihango's estimation of 750 RF may be too high, as it is out of the range of our confidence interval. However, the neural network's prediction is even higher at $0.96 USD, or 1199.01 RF. The results of the survey (Taarifa, 2017) conducted by the University of Dar-es-Salaam, Kigali Independent University, and University of Amsterdam show that the minimum wage should be 450 RF, just at the upper edge limit of the random forests model.

Although our models provide quite a bit of spread between potential minimum wages, they are both consistent in the fact that the minimum wage of Rwanda set over 40 years ago at 100 Rwandan Francs may need to be updated since both models predict wages significantly higher than the one in the present.

7.3. Singapore

The Singapore prediction with the random forest model is \$6.94 United States dollars, or \$9.31 Singapore dollars. The confidence interval has a range of \$1.51 USD. The neural network's prediction is well within that confidence interval at a prediction of \$6.52 United States dollars or \$8.62 Singapore dollars. The consistency between models suggests that if Singapore should decide to incorporate a minimum wage, these results may serve as a starting point.

7.4. United Arab Emirates

The United Arab Emirates random forest prediction is \$5.36 United States dollars, or 19.67 UAE Dirhams. The 95% confidence interval is between 14.44 Dh and 24.9 Dh. The neural network's prediction is within this range at \$4.59 United States dollars, or \$16.84 UAE Dirhams. If the UAE decides to incorporate a minimum wage, it would likely increase the earnings for manufacture workers whose average compensation is 3.38 USD or 12.4 Dh.

8. Conclusions

Various methods are used in terms of prediction accuracy for minimum wages. The two top performers are random forests and neural networks. Both of these methods are nonlinear, suggesting that a nonlinear algorithm is the best at predicting minimum wage.

The random forest model is more interpretable with 95% confidence intervals. The width of confidence intervals varies between countries, and the largest interval is for Iceland at a range of \$3.76 United States dollars.

The neural network model varies somewhat to the other model. Between the models, their predictions seem to be more consistent with Iceland, Singapore, and Norway. The biggest difference is the Austria estimate, by about \$3 USD.

We do not claim that our methods model the true data generating process with 100% accuracy, but we wish to provide a starting point for minimum wage discussion. The RMSE of neural networks is 0.925, meaning that the model is on average about \$0.925 USD off from the true prediction. We simply provide a ball park estimate for what the true minimum wages for these countries should be.

There are a number of future directions for research into this area. Our data have all quantitative or numerical variables. One may wish to include qualitative data, such as survey data. Questions like overall level of happiness are more qualitative in nature, and running a survey from each country may add to the results. Next, we choose to omit the missing data of the 33 nations with scarce data. This leads to our assumption that the data are missing at random. Another approach could be to impute this data. Although this introduces artificial data, it may help remove our assumption. Also, further regularization of the neural network model to reduce overfitting can be done to potentially improve the test RMSE. Finally, another future direction would be to collect time series data. This can provide information on how the predictors relate to minimum wage through time and how to forecast minimum wage from past data.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research received no funding.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

Beck MW (2018) NeuralNetTools: Visualization and Analysis Tools for Neural Networks. *J Stat Softw* 85: 1–20. https://doi.org/10.18637/jss.v085.i11

Bhorat H, Kanbur R, Stanwix B (2017) Minimum Wages in Sub-Saharan Africa: A Primer. *World Bank Res Obser* 32: 21–74. https://doi.org/10.1093/wbro/lkw007

Eurofound (2018) Statutory minimum wages 2018. Available from: https://web.archive.org/web/20200831133204/https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef18005en.pdf.

Fritsch S, Guenther F, Wright MN (2019) neuralnet: Training of Neural Networks. R package version Available from: https://CRAN.R-project.org/package=neuralnet.

Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press. Available from: http://www.deeplearningbook.org.

Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed), Springer-New York. Available from: https://doi.org/10.1007/978-0-387-84858-7.

International Labor Organization (1996–2024) How to define a minimum wage? Available from: https://www.ilo.org/global/topics/wages/minimum-wages/definition/WCMS_439072/lang–%en/index.htm.

International Trade Association (2024) Nigeria-Country Commercial Guide. Available from: https://www.trade.gov/country-commercial-guides/nigeria-healthcare.

James G, Witten D, Hastie T, et al. (2021) An Introduction to Statistical Learning. Springer, New York.

Kassambara A, Mundt F (2020) factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. Available from: https://CRAN.R-project.org/package=factoextra.

Ki M (2024) Country Data Set. Available from: https://www.kaggle.com/datasets/matthewki/country-data-set.

MaCurdy T(2015) How Effective Is the Minimum Wage at Supporting the Poor? *J Polit Econ* 123: 497–545. https://doi.org/10.1086/679626

Romich J (2017) Is Raising the Minimum Wage a Good Idea? Evidence and Implication for Social Work. *J Storage* 62:367-370. https://doi.org/10.1093/sw/swx033

Taarifa (2017) For 40 Years, Rwandan Labourers Earn Rwf100. Who Is Blocking Minimum Wage Increase? Available from: https://taarifa.rw/for-40-years/.

Gini Index (World Bank Estimate) (2024). Available from: https://genderdata.worldbank.org/indicators/si-pov-gini/.

United States Census Bureau (2024) Gini Index. Available from: https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/gini-index.html.

USNews (2024) Quality of Life. Available from: https://www.usnews.com/news/best-countries/rankings/quality-of-life.

Vodafone Group PLC (2024) M-PESA. Available from: https://www.vodafone.com/about-vodafone/what-we-do/consumer-products-and-services/m-pesa.

Wager S, Hastie T, Efron B (2014) Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jacknife. *J Mach Learn Res* 15: 1625–1651.

Wei T, Simko V (2021) R package 'corrplot': Visualization of a Correlation Matrix. Version 0.92. Available from: https://github.com/taiyun/corrplot.

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. Available from: https://ggplot2.tidyverse.org.

Wright MN, Ziegler A (2017) Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J Stat Softw* 77: 1–17. https://doi.org/10.48550/arXiv.1508.04409