Journal of Information Systems and Telecommunication
Vol.11, No.2, April-June 2023, 163-175
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

An analysis of Covid-19 pandemic outbreak on Economy using Neural Network and Random Forest

Md. Nahid Hasan[1*], Tanvir Ahmed[1], Md. Ashik[1], Md. Jahid Hasan[1], Tahaziba Azmin[1], Jia Uddin[2*]

[1].Department of Computer Science and Engineering, Brac University, 66 Mohakhali, Dhaka-1212
[2].Artifical Intelligence and Big Data Department, Endicott College, Woosong University, Daejeon, South Korea

Received: 23 Feb 2022 / Revised: 15 Jun 2022 / Accepted: 19 Jul 2022

Abstract

The pandemic disease outbreaks are causing a significant financial crisis affecting the worldwide economy. Machine learning techniques are urgently required to detect, predict and analyze the economy for early economic planning and growth. Consequently, in this paper, we use machine learning classifiers and regressors to construct an early warning model to tackle economic recession due to the cause of covid-19 pandemic outbreak. A publicly available database created by the National Bureau of Economic Research (NBER) is used to validate the model, which contains information about national revenue, employment rate, and workers' earnings of the USA over 239 days (1 January 2020 to 12 May 2020). Different techniques such as missing value imputation, k-fold cross validation have been used to pre-process the dataset. Machine learning classifiers- Multi-layer Perceptron- Neural Network (MLP-NN) and Random Forest (RF) have been used to predict recession. Additionally, machine learning regressors-Long Short-Term Memory (LSTM) and Random Forest (RF) have been used to detect how much recession a country is facing as a result of positive test cases of covid-19 pandemic. Experimental results demonstrate that the MLP-NN and RF classifiers have exhibited average 88.33% and 85% of recession (where 95%, 81%, 89% and 85%, 81%, 89% for revenue, employment rate and workers earnings, respectively) and average 90.67% and 93.67% of prediction accuracy for LSTM and RF regressors (where 92%, 90%, 90%, and 95%, 93%, 93% respectively).

Keywords: Multi-Layer Perceptron (MLP); Long Short-Term Memory (LSTM), Random Forest, Economic Recession, Machine learning (ML), Covid-19.

1- Introduction

An economic forecast directly affects the financial institutions during a pandemic outbreak caused by contagious organisms. Any wrong decision taken during this time may have significant adverse effects. As a result, detection of different economic sectors plays a vital role in finance. Over the decades, multiple pandemic outbreaks have taken place. For instance, the Spanish Flu (1918-1919), which is estimated to have infected and killed millions of people causing a severe impact in economic sectors. Asian Flu (1957-1958) which has also killed millions of people worldwide, depleting the economy across the globe. SARS coronavirus (2003) caused respiratory illness and killed 774 people creating a huge negative impact on human health as well as the economy. Swine Flu (2009) caused the death of 150000 to 575000 people, depleting the stock markets, tourism, food as well as transportation industries incurring a considerable loss. And lastly the novel Coronavirus (COVID-19) (2019-present) causing the death of 6.3 million people till now and destroying the economy throughout the world [1]. Significant numbers of countries have already gone through a financial crisis after a pandemic outbreak. For example, the SARS outbreak hit many Asian countries' economies and took millions of lives [2]. The economic costs from a global disease go beyond the direct damages incurred in disease-inflicted countries' affected sectors. The disease spreads rapidly through countries across networks linked to worldwide travel. Any economic shock to one country spreads quickly to other countries through expanded trade and financial ties related to globalization. Infectious diseases are likely to increase the global cost. As a result, it is essential to take the early initiative to

✉ **Jia Uddin**
jia.uddin@wsu.ac.kr

revive the economy. This paper provides a method for predicting the economy during a pandemic outbreak. In particular, the techniques that have been used are examined, the experiments that have been conducted are reviewed, and directions of future work from the perspective of machine learning are considered [3]. This model will capture crucial linkages across different economic sectors, thereby comprehensively measuring disease-related costs.

Millions of livelihoods around the globe are affected by the economic instability created by the pandemic. Everyone fights a longstanding battle against microorganism, which causes the pandemic and affects people's livelihood throughout the world. However, such pandemics are a problem of an entirely different nature and demand an unparalleled response scale. To respond to the imminent challenges to the long-term impact on the country's economy, machine learning techniques are used. Machine learning techniques can detect economic crises which help the public and private sector leaders, as well as the policymakers, plan a better solution to combat the financial crisis. Previously, Keogh-Brown et al. described that due to SARS virus outbreak 3.7% loss in GDP (US$ billion), 23.1% loss in export and trade (US$ billion), and 0.86% loss in tourism (US$ billion) sector in Hong-Kong was recorded [4]. Canada loses 3.2-6.4% in GDP (US$ billion), 1% in Growth (US$ billion), 5.2% in export and trade (US$ billion), 0.03% in tourism (US$ billion) and 6.25% in Airline (US$ billion) [4].

On the other hand, in 2020, in the USA, due to the Covid-19 outbreak, different economic sectors were affected. Those sectors are affected by increasing covid-19 positive test cases daily, which causes a national economic recession. Earnings received by workers will decrease by increasing positive test cases compared with January 2020. Falling workers' wages suggests recession which causes unemployment and downward pressure on wages. Manufacturing sector workers earnings decreased 41%, Retail Trade workers' earnings decreased 36%, Transportation and Warehousing decreased 30% and Health Care, and Social Assistance decreased 31%.

Employment level is also affected by this pandemic situation. Employment level or employment rate is defined as the number of people engaged in productive economic activities. Manufacturing employees' level decreased 41%, Retail Trade employees' level decreased 38%, Transportation and Warehousing decreased 32% and Health Care, and Social Assistance decreased 31% compared with January 2020.

These sector-wise downturns of employment rate and worker wages reflect the USA's national economic recession. Moreover, net Revenue for all businesses is also decreasing simultaneously. The maximum employment rate was 1% at the end of January 2020. Still, with the increase of positive test cases, it gradually decreased and within the first week of May 2020, the rate decreased 39% compared with January 2020. Also, workers' wages decreased 38% within the first week of May 2020. However, revenue for all small businesses started increasing from the last week of January 2020, and the maximum value was 14%. Suddenly, it started decreasing, and at the end of March 2020, revenue decreased 49% compared with January 2020.

The paper aims at defining robust financial crisis predictors. A link between a covid-19 pandemic and its impact on people's revenue, earnings, and employment is found. The chosen data-set consisting of different sectors revenue, earnings, and employment helped to find how a pandemic affects the economy. In comparison, this paper's uniqueness lies in the variables commonly correlated with the novel covid-19 pandemic. Overall, this paper contributes to decision-makers and leaders detecting the economic crisis during or after a pandemic and introducing measures to eliminate or dampen a crisis entirely.

2- Literature Review

A number of researches have been conducted due to the rise and fall of the economy during the pandemic. These studies relate how a pandemic can influence a country's economy, culture, people and others.

Smith et al. has shown the UK's economy's impact based on the Computable General Equilibrium (CGE) model [5]. This research estimates that only pandemic influenza can minimize GDP by 0.3%, 0.4%, and 0.6% for mild, moderate, and severe cases, respectively. Additionally, losses of different sectors are 1.5% in domestic output, 2% in household consumption, 3% in exports, 2.5% in imports, 2% in government consumption as calculated from data-set 2003 supply chain in the UK. This paper presented that a large portion of the economy was damaged due to the pandemic.

Keogh-Brown et al. suggest macro-economic outcome after SARS outbreak, through affected countries and their economic sectors evaluated by economic indicators [4]. Researchers calculated losses of various sectors such as GDP, growth, exports and trade, tourism, food and travel. In Hong-Kong, 3.7% loss in GDP, 4.75% loss in Growth, 23.1% loss in export and trade, and 0.86% loss in tourism is recorded according to their databases. This research gives a clear picture of affected sectors of individual countries and why stakeholders could develop solutions to be aware of the next pandemic outbreak.

Fernandez-Delgado et al. compared several classification algorithms to predict the economic crisis [6]. The results showed that multiple models demonstrated different detection results. Hence, deciding which one to be selected was a challenge. Lastly, the random forest family algorithm showed the best results in the early warning of the economic crisis.

In recent years, bankruptcy forecasts have been made by machine-learning using standard statistical approaches [7]. To obtain improved failure–detection solutions of problems, analysis has been done using the mathematical and machine learning techniques [8]. In particular, the datasets, financial partnerships, country of origin, and the timeline of the analysis were used, and the results and implementation were compared with several different detection accuracy techniques [3-8].

This paper aims to detect and forecast the financial status of institutions or individuals by using machine learning algorithms [3]. If this can be done, human sufferings can be reduced by taking the best decision and the financial crisis can be overcome. Here, tools such as neural networks, decision trees, etc., have been extensively studied to predict financial crises. Use MLP, Random Forest (RF) [36] to apply machine learning methods, such as pattern classification techniques, single classification techniques and soft classification techniques. We have seen some problems that are not widely discussed in the literature in the observation. When different datasets are used for different training and evaluation sets and cross-validation, more accurate results can be obtained.

It is difficult to make crucial decisions due to the vulnerability of the emerging coronavirus outbreak. The deep learning and fuzzy detection approach are proposed by Fong et al. on different outcomes of Coronavirus and its effect [9]. The current events and possible actions were described using the Composite Monte-Carlo simulation system. Fong et al. addressed the daunting challenges of predicting the fate of an outbreak correctly effectively using the availability of dataset, the layout for picking the best predictive model and finely tuning each model's parameter [9].

The statistical data were benchmarked in a paper by Bluwstein et al. (2020), used in many machine learning models, such as decision trees, random forests, large-scale random trees, SVM [38] and artificial neural networks [10]. The finding of the paper, except for human decision-making agencies, all machine learning models performed better than logistic regression [10]. The best performing machine learning model (a tree that is too random) can correctly predict the global financial crisis of 2007-2008 and provide outstanding signals in countries and regions with different economic realities and achievements [10].

In [11], Nyman gives an actual forecasting scenario where they use a small number of financial variables for the detection of the economic recession. Two estimation techniques- ordinary least squares regression and random forest machine learning were utilized. Random forests noisy, non-linear, high-dimensional detection can be tackled using machine-learning models [12-13]. The author obtains qualitatively similar results for the UK and USA through the random forest algorithm's predictive power is more efficient for the USA. Finally, the author says the machine learning approach is very efficient for forecasting horizons and providing better information.

Car, Z. et al. [14] built a model that can detect the spread of covid-19 infection to predict its impact. They used a 51- day data set to train a multi-layer perceptron (MLP) neural network, and chose to use a grid search algorithm for hyperparameter optimization. After performing k-fold cross-validation, they found that the accuracy of positive confirmed cases was 94%, the accuracy of recovered patients was 78%, and the accuracy of deceased patients was 98%. This is an excellent model for the detection of covid-19 infection.

To date, many models have been developed with different domain knowledge to predict the economic crisis. Logistic regression (Ohlson 1980) and factor analysis are some of the conventional statistical approaches (West 1985) [15-16]. Some of Artificial Intelligence methods such as artificial neural network ANN (Atiya 2001), Support vector machine (SVM) (Min and Lee 2005) (Shin et al. 2005), Bayesian networks (Sarkar and Sriram 2001) (Sun and Shenoy 2007) and many integrated machine learning techniques (Fedorova) et al. 2013) (Abellán and Mantas, 2014), several hybrid methods are widely used to predict economic crises [17-22].

3- Proposed Methodology

This section discusses constructing the models to determine the impact of economic indicators in a country due to a pandemic. Firstly, we start with a description describing where we have collected datasets, structure, etc. Then, how we process the data-set to fit into the models. Then, the implementation of algorithms for detecting recession due to this covid-19 pandemic. Figure 1 presents detailed architecture of the proposed model.

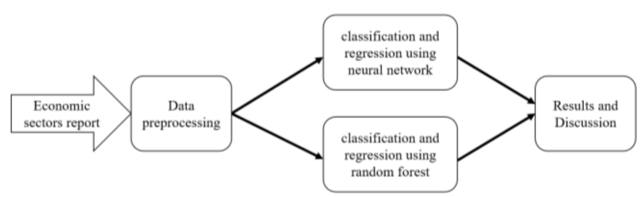

Fig. 1 Architecture of proposed model

3-1- Data Description

The Opportunity Insights Team built a publicly available database [23]. The database tracks economic activity at a granular level in real-time using anonymized data from private companies. They build the database using private sector data to show the Economic Impacts of COVID-19.

They studied how COVID-19 affects the economy by examining heterogeneity in its influences. The data-set used in this research is collected from a publicly available archive, operated by the Opportunity Insights Economic Tracker. The archive contains data from leading private companies from credit card providers to payroll firms to offer a real-time image of indicators such as employment rates, consumer spending, and job postings across counties, industries, and income groups. The data-set is split into employment rate, worker's earnings, and national revenue of various economic sectors. During this research, the dataset contained the data of 239 days for COVID-19 positive cases, small business revenue, and employment rates among low-income workers. Moreover, the chosen data-set, as published, is organized as time-series data. Besides, this dataset contains data points of 239 days, which is split into five folds.

3-2- Data Preprocessing

The dataset has three different sorts of information: employment rate, worker's earnings, and national revenue of various economic sectors. Firstly, the dataset was time-series data, so we converted it to fit the model. As we are classifying, is there any recession due to the impact of covid-19? If any value from the employment rate of all sectors is negative, then there exists a recession; otherwise, not. For this purpose, we are adding one extra feature into the database depending on the overall employment rate, which will be the attribute column for classifying recession in employment rate and the same for others.
In the real-world data-set, missing data points are not unexpected. In this paper, used data-sets even have missing data points. We used statistical imputation to fill up the gaps of missing data points to reduce biases. Although there are varieties of techniques for imputation, we have used mean values for each feature. Replacing missing data points with mean is simple; besides, it does not introduce many biases in the data.

$$Mean = \frac{x_1+x_2+x_3+\cdots+x_n}{n} \tag{1}$$

Each data set is divided into five random folds used to train and validate the model. Here, folding is used to train the model (k-1) times, and the rest will be used to verify the model, where k = 5. This process continued unless each fold was used as a test set for the model, which gave us the advantage of having a generalized model. In addition, because the data points are limited; therefore, this K-fold cross-validation technique has helped us use each data point as a training set, and a test set [24]. To determine the cross-validation result, we calculate the average of the R^2 scores.

$$R^2 = \frac{1}{5}\sum_{n=1}^{5} R_k^2 \tag{2}$$

3-3- Implementing Classifier in Economic Recession Detection Model

This section discusses how we detect economic recession due to pandemic outbreaks. We first use MLP (multilayer perceptron) to detect the issue first and then Random Forest classifier. Both algorithms help us determine the economic recession a country might face due to this covid-19 effect.

3-3-1 Multi-layer Perceptron Classifier

In Artificial Neural Network, Multi-layer Perceptron (MLP) is a fully connected feed-forward neural network that mimics the human brain to build up a machine learning model. Multi-layer Perception (MLP) is a class of Feed-forward Artificial Neural Network (ANN). MLP has three kinds of a layer, such as input layer *(i)*, which has an equal number of neurons, same as data-set features, hidden layer j_0 *to* j_n and output layer *(k)* consist of a single neuron. Signals flow from left to right layer by layer for computing output of each neuron, which is named as Forward Propagation of Function Signals [25]. Again, to minimize errors of the network, an error signal in the output neuron propagates backward layer by layer referred to as Backward Propagation of Error Signal.
To generate a network that can produce a higher accurate result, choosing hyper-parameters is a crucial factor. For the different values of hyper-parameters, the network will show different performance. For instance, if a hyper-parameter like the number of iterations is excessive then the network might face an over-fitting problem. Additionally, depending on the learning rate, the network will converge slowly or quickly. We applied the Grid Search Algorithm to find best-fitted hyper-parameters such as weights, learning rate, number of hidden layers, etc., for network convergence [26]. Besides, we set a few hyper-parameters constant such as stochastic gradient descent, which is a solver for weight optimization, a maximum number of iterations is 100, activation function *(relu)*, an initial learning rate 0.1. However, Summation of inputs and connected links weights will pass through *relu* activation function. *Rectified Linear Unit (relu)* function shows better convergence performance and is computationally efficient since it maps only the value *max (0, z)* where *z* is *z>0*. This function is also differentiable. Hence, the total equation of output from the *i-th* neuron is,

$$O(j) = relu\left(\sum_i^n w_{j_i}\, y_i + b_i\right) \tag{3}$$

Here, *relu* is the activation function, W_{ji} is the weights of connected links, y_i are inputs, and b_i is the bias in this

case it is 1. This process of computing output is the same for all hidden layer and output layer neurons.
An activation function decides whether it will fire the output of that particular neuron or not. Now, η is the learning rate, which decides how quickly the network will converge and we put $\eta=0.1$ so that machine can learn slowly and perform with higher accuracy in the long run. Hence updated *Delta rule*,

$$w_{new} = w_{old} + (0.1 * \delta_k * y_j) \quad (4)$$

Where δ_k is Local Gradient, and y_j is the input of *k-th* neuron or output from *j-th* neuron, w_{old} is the previous iteration's weight. After that, equations for calculating δ is different for hidden and output layers. If *k-th* is a neuron of the output layer,

$$\delta_k = \emptyset'(\sum_k w_{kj} y_i)(\partial_k - y_k) \quad (5)$$

Here, δ_k is the desired output from the *k-th* neuron, and ϕ' is the first derivative of the activation function. Again δ_j for *j-th* neuron of a hidden layer,

$$\delta_j = \emptyset'(\sum_j w_{ji} y_i)(\sum_{c \in k} \delta_c w_{cj}) \quad (6)$$

Where *c* is the set of next hidden/output layer neurons and y_i is a set of input layer neurons or hidden layer neurons, but it is the previous layer of *j-th* neuron's layer. In this way, MLP model training will continue until it meets stopping criteria. When the Average square error (loss function) change is sufficiently small for per epoch, then the training process will stop. Another measure is, after each period, the MLP model will be tested for Generalization, and if this generalization performance is suitable only then, training will stop.
After putting the values mentioned above, we have got a satisfactory result from the network. However, to check this network's robustness, we have applied the k-fold *(k=5)* cross-validation technique. After applying this technique, some of the results dropped, and some showed more accuracy. Nevertheless, this difference is shallow, so this network fulfilled our exception for detecting the issue that we wanted to solve. Finally, this is a simplex network having the ability to solve a complex problem like detecting financial crisis with lower training time.

3-3-2 Random Forest Classifier

The random forest classifier is a collection of projected trees, where each tree is subordinate to independently evaluated random vectors, with comparable transport within the random forest with one another tree. It, too accomplishes the proper speed required and productive parameterization within the process. The random forest classifier bootstraps random tests where the expectation with the most elevated vote from all trees is chosen. The distinction of each tree is ensured due to the taking after qualities. To begin with, each tree training within the test employs random subsets from the beginning training tests. Besides, the ideal part is chosen from the unpruned tree nodes' arbitrarily chosen features. Thirdly, each tree develops without limits and ought not to be pruned at all. In this model, we have implemented a random forest classifier for classification that uses an ensemble learning approach to detection [27], which uses several decision trees during the training process and average individual tree detection outputs. Random Forest efficiently runs on massive data-sets, can handle thousands of input variables without variable deletion, produces significant variable for forecasts, creates an internal unbiased measure of generalization error as forest growth increases, has an adequate method to estimate lost data, and maintains accuracy where a large proportion of data is lacking [12,28]. The chosen RFC's key objective is the power of the individual decision tree and the relation between base trees [12]. Random forest classifier comprises various individual classification trees, where each tree may be a classifier given diverse weighted classification. The output of the classification determined the overall classification. It builds each tree by part number of features for each part without pruning [29].

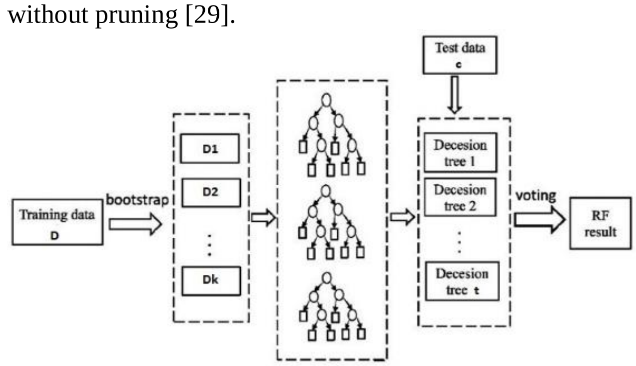

Fig. 2 Random Forest model

The classifier of Random Forest functions as consequently [30]: we choose C subsets from which to construct the training set T and initial training data D using bootstrap sampling

$$T = D_1, D_2, D_3, \dots D_c \quad (7)$$

The algorithm automatically generates the decision tree for C models with a random vector θc for each subset:

$$M = M_1, M_2, M_3, \dots M_c \quad (8)$$

The random vectors θc, $c = 1,2......C$ are both distinct and distributed identically. Each decision tree evolves freely without pruning so that all the trees are combined to get a forest.

$$r(T,c), c = 1,2,3 \ldots C \tag{9}$$

For determining the classification of a new input variable t, the final vote of each established decision tree model is accurate. The result for classification is:

$$R(t) = \sum_{i=1}^{c} I(ri(t) = Z) \tag{10}$$

Where R(t) denotes the product of the classification is; ri(t) denotes the outcome of the decision tree classification; Z is the target group, and I *(ri(t)=Z)* stands for the characteristic function.

n-estimators are the number of trees in the forest. Having a few distinctive forms of trees with different depths and sizes boosts the generalization of the n-estimators' trees that the algorithm needs to construct. If the estimator is n, it gives n other decision trees. Using n-estimators, we build several trees within a Random Forest before aggregating the detection. We want to make the computational expense while taking the trees as it delays code to run.

Using the criterion Gini-impurity, we determine the feature of a tree that has to split the parameters. It measures the quality of each split. We use a max-depth to see how much further the tree has to be expanded to each node until we get to the leaf node. We set the max-depth to run the option for the risks of over-fitting this model. We observe the impact of the max-features hyper-parameter. We realize that the random forest selects a few random tests to explore the main break from the functions. We can see that the execution of the first increments demonstrates that the number of max features increases.

Nonetheless, the training score keeps expanding after a certain point. However, the test score saturates and begins diminishing the conclusion, implying that the show starts to over-fit. Ideally, the general execution of the demonstration near six of the highest highlights is the most notable. In general, the ideal number of total features tends to be similar to this value. To decide whether the algorithm will avoid further splitting, we use min samples to define the minimum number of records present in each node. If the split number is less than n, there will be no further split. We use max-leaf nodes to expand the tree in a best-first manner resulting in a relative reduction in impurity. We use max-leaf nodes to grow the tree in a best-first way resulting in a relative decrease in impurity. The random state makes it simple for others to imitate results if given the same training data and parameters. For sampling data points, we use the bootstrap process.

The algorithm randomly selects many rows with replacement to construct the trees using bootstraps once we provide the Random Forest Classifier model's training data. If the bootstrap option is set to False, there will be no random sampling, and the entire data set will be used to build the trees. We use oob-score as it is much quicker because it gathers all of the trees' observations and finds the highest score on each observation base's trees that did not use that observation to train. Oob-score is a cross-validation technique similar to a leave-one-out validation technique in which a model's generalized approximate output is trained on n-1 data samples. We set n-jobs to -1 will often lead to faster processing. If we use -1, there is no limit on how much computing resources the code can use n-jobs helps the program know how many processors it can use. The default value of 1 means that only one processor can be used. We set the logging output to be verbose, which gives us continuous feedback on what the model is doing as it is processed. This parameter defines the verbosity of the construction method of the tree. We use false Warm Start for recursive feature collection, and false Warm Start suggests that other features will gain in value as we drop such features, and it will be repeated used. It is often used in regression models with backwards exclusion and is not often used in classification models.

3-4- Implementing Regression in Economic Recession Detection Model

This section covers how much economic recession a country might face due to a pandemic outbreak in an individual economic sector. First, we will discuss how RNN implementation helped then Random Forest regressor.

3-4-1 Long term short-term memory

In this model, we used the long-term short-term memory (LSTM) method to estimate the economic impact of the COVID-19 outbreak on different regions of the United States [31,37]. When dealing with time series, we used LSTM modelling, an in-depth learning method that is useful when trying to model time series. In the LSTM unit, there are four functions in this model, which are 3 *Sigmoids* (f, i and o below) and *Tanh* (c function below). It is mentioned that the coefficient of deviation may be a general feature of all functions in the learning model, which can be set or measured in advance during the training process. In order to help it adjust to the various situations of each case, the bias can be used to model calibration. A common LSTM cell is formed of three gates-*input gate, output gate, and forget gate* [32].

In the LSTM structure, the first layer is called the forget gate, which selects the information to forget. It can produce any value between 0 (completely forget) to 1 (usefully) [33].

$$f_t = \sigma(W_f(h_{t-1}, x_t) + b_f) \quad (11)$$

The next step of the algorithm is to determine the new input which needs to be added.

$$i_t = \sigma(W_i(h_{t-1}, x_t) + b_i) \quad (12)$$

After that, the algorithm determines the new candidate value of neural cells.

$$\check{C} = \tanh(W_c(h_{t-1}, x_t) + b_c) \quad (13)$$

After computing above equations, it states the new cell states by computing C_t.

$$c_t = f_t \times C_{t-1} + i_t \times \check{C}_t \quad (14)$$

Lastly, the output function will predict the value. This layer is called detection layer and the detection will then

$$o_t = \sigma(W_0(h_{t-1}, x_t) + b_0) \quad (15)$$
$$h_t = o_t \times \tanh(C_t) \quad (16)$$

We add the LSTM layer and later integrate a few Dropout layers to prevent overfitting. We integrate the LSTM layer with 50 units which is the dimensionality of the output space. We define the dropout layers 0.2. This indicates that it will reduce the number of layers by 20%. After that, we merge the dense layer, which determines the output of 1 unit. Next, we used a very common optimizer, the Adam optimizer. Then, we fit the LSTM model to run at 100 epochs and 32 batch sizes.
We will conduct a k-fold on the data to improve results' complexity. K -fold is a cross-validation technique that requires the creation of various models on subsets of the data set. This approach would be very beneficial in achieving the desired predictive precision standard.

3-4-2 Random Forest Regressor

The random forest regression (RFR) incorporates a vast collection of decision trees trained together to produce a more precise and reliable final forecast [12]. A regression tree is ordered from the leaf to the root node based on some parameters. RFR consists of a supervised learning algorithm for predicting output target feature average by bootstrap aggregation or bagging of independently built decision trees [34]. Bootstrap aggregation or bagging is used for lowering variance error sources of independently built decision trees. We first randomly select points from the training data set of employment rate, worker income and revenue of different sectors. Then we created a bootstrap sample of the random data with replacement and created the root node, eventually forming a decision tree. After that, we created 200 different decision trees from it. To estimate an economic recession, each decision tree forecasts a data point value for worker income, employment rate and revenue and finally assigns the average new data point over all the expected values. Forest-random regression trees use a greedy top-down method to define ideal recursive divisions in binary nodes.

$$Min(SSE) = \left(\sum_{i=1}^{n}(y_i - y_s)^2\right) \quad (17)$$

Here SSE=Sum Squared error, $y_t =$ output target feature data, and $y_s =$ terminal node output target feature mean
And y_s is calculated by,

$$y_s = \frac{1}{m}\sum_{i=1}^{m}(y_t) \quad (18)$$

Here y_s is the mean terminal node, m=number of observations in the terminal node, and y_t output target feature data.
In a Random Forest regression algorithm, tree bagging consists of predicting the output feature of an independently built decision tree by calculating the arithmetic mean,

$$y_p = \frac{1}{k}\sum_{i=1}^{k}(y_s) \quad (19)$$

Here $y_p =$ mean output target feature detection, k=number of independently built decision trees, and y_s =independently built decision trees output feature detection.
Unlike other machine learning techniques, Random Forest regressor only needs to set two parameters to construct a detection model: the number of regression trees and the random state, which is 200 regression trees and the random state is set at 0. For accuracy, we calculated the mean absolute error, which gives the average of error in detection and mean squared and root mean squared error to check how close the detection is to the actual value. Lastly, we cross-validation to check how well random forest regression performs on test data.

4- Result and Discussion

In this section, we discuss the achieved results in detail by following the described methodology. We will also analyze essential factors for achieving these results.

4-1- Result of Classifiers

We choose to use two different classifiers to detect a country's recession due to a pandemic. Now we will discuss results found from multi-layer perceptron and random forest algorithm described in section -4.1. The formulas for calculating Table 1 values for MLP and Random Forest classifiers are equations no. 20, 21, and 22.

$$Accuracy = \frac{Correctly\ predicted\ class * 100\%}{Total\ testing\ class} \quad (20)$$

$$Sensitivity = \frac{TP}{TP+FP} \quad (21)$$

In equation 21, *TP = true positive* and *FN = false negative.*

$$Specificity = \frac{TN}{TN+FP} \quad (22)$$

In equation 22, *TN = true negative* and *FP = false positive.*

4-1-1 Multi-layer perceptron

The MLP classifier demonstrated an accuracy of 81% for the employment level, 89% for worker earnings, and 95% for national revenue as shown in Table 1. To achieve these results, we have used the initial learning rate (alpha) 0.1, *relu* activation function, three hidden layers with different units of neurons, and others.

Table 1: Results from MLP classifier (CV=cross validation

Section	Accuracy	Sensitivity	Specificity	Accuracy (cv)
Employment level	81%	100%	0%	82%
Worker Earnings	89%	100%	0%	88%
National Revenue	95%	100%	67%	87%

Sensitivity for all sectors was 100%. Nevertheless, specificity was 0%, 0%, and 67% for employment level, worker earnings, and national revenue, respectively. However, applying k-fold cross validation where k=5, suddenly, the accuracy showed different results. We get the mean value of 5 folds, 82% for employment level, 88% for worker earnings, and 87% for national revenue, respectively, as displayed in Table 1.

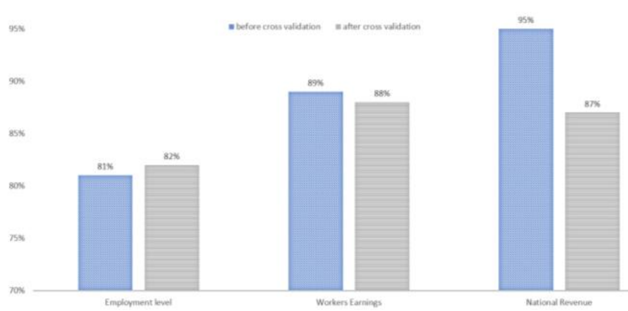

Fig. 3 Results from MLP classifier before and after cross validation (CV)

Figure 3 demonstrates accuracy before and after cross-validation of all three sectors. For employment level, accuracy increased after cross-validation, but accuracy decreased for worker earnings and national revenue.

4-1-2 Random Forest classifier

By using the Random Forest classifier, we got an accuracy of 81% for employment level, 89% for worker earnings, and 85% for national revenue shown in Table 2. We have used n-jobs equal to -1 for faster processing, logging output as verbose for continuous feedback, n-estimator is 100 and used criterion as *gini* for the Gini impurity.

Table 2: Results from Random Forest classifier

Section	Accuracy	Sensitivity	Specificity	Accuracy(cv)
Employment level	81%	100%	0%	98%
Worker Earnings	89%	100%	0%	99%
National Revenue	85%	100%	67%	98%

Sensitivity for all sectors was 100%. Nevertheless, specificity was 0%, 0%, and 67% for employment level, worker earnings, and national revenue, respectively. However, applying k-fold cross validation where k=5, suddenly, the accuracy showed different results. We get the mean value of 5 folds, 98% for employment level, 99% for worker earnings, and 98% for national revenue, respectively, as displayed in Table 2.
Figure 4 exhibits accuracy before and after cross-validation of all three sectors. For all three sectors, accuracy increased after cross-validation.

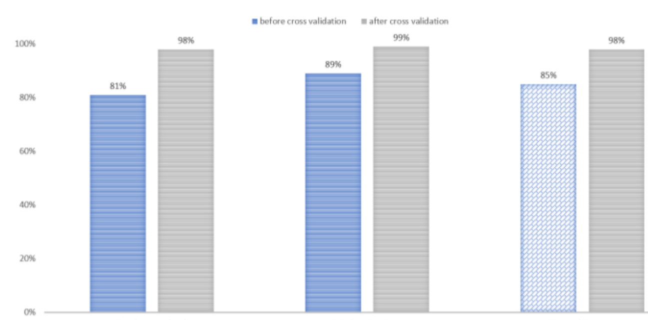

Fig. 4 Results from Random Forest classifier before and after cross validation (CV)

4-2- Result of Regressors

We used two different regressors, long short-term memory and random forest, to determine how much loss a country might face for a pandemic effect. This section will discuss experimental results found from them described in section 4.2. The formulas for calculating Table 1 values for LSTM and Random Forest regressors are equations no. 23,24, 25 and 26.

$$Accuracy = \frac{Correctly\ predicted\ class*100\%}{Total\ testing\ class} \quad (23)$$

$$Mean\ absolute\ error = (\sum_{i=1}^{n} y_i - k_i)/n \quad (24)$$

In the equation 24, y_i=predicted value k_i=true value and n=total number of data points

$$Mean\ Squared\ error = (\sum_{i=1}^{n} (y_i - y_s)^2)/n \quad (25)$$

In the equation 25, y_s= predicted value y_i= true value and n = total number of data points

$$Root\ Mean\ Squared\ error = \sqrt{(\sum_{i=1}^{n} (y_i - y_s)^2)/n} \quad (26)$$

In the equation 26, y_s= predicted value y_i = true value and n = total number of data points.

4-2-1 Long term short-term memory

From Table 3, we got an accuracy of 90.04% for employment level, 90.33% for worker earnings, and 92.62% for national revenue using the LSTM. We get 2.63% mean absolute error, 0.12% mean squared error, and 3.46% root mean squared error for national revenue. Again, 4.87% mean absolute error, 0.24% mean squared error, and 4.89% root means squared error for employment level. Finally, 4.68% mean absolute error, 0.22% mean squared error, and 4.72% root means squared error for worker earnings.

Table 3: Results from LSTM regression

Section	*Accuracy*	*MAE*	*MSE*	*RMSE*
Employment level	90.04%	4.87%	0.24%	4.89%
Worker Earnings	90.33%	4.68%	0.22%	4.72%
National Revenue	92.62%	2.63%	0.12%	3.46%

The visual representation of actual and predicted data using LSTM regression are in Figures 5, 6 and 7.

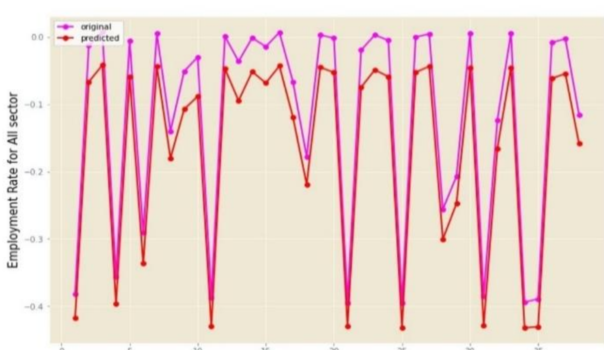

Fig. 5 Comparison of real and modeled (LSTM) data of employment rate of all sectors.

Figure 5 illustrates actual (original labeled) and predicted (predicted labeled) data we found using LSTM for employment rate or level. This figure, the x-axis, shows the number of days, and the y-axis shows the employment rate for all sectors. The predicted rates are 90.04% accurate.

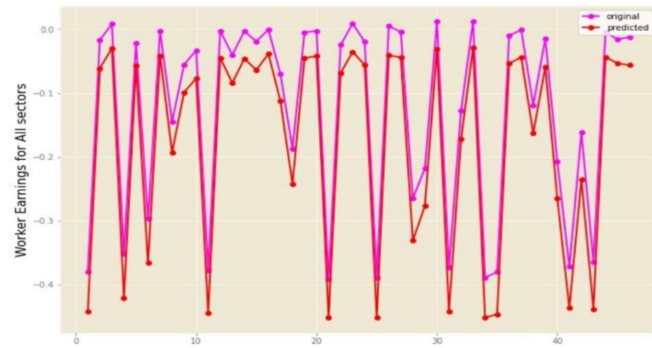

Fig. 6 Comparison of real and modeled (LSTM) data of worker earnings of all sectors

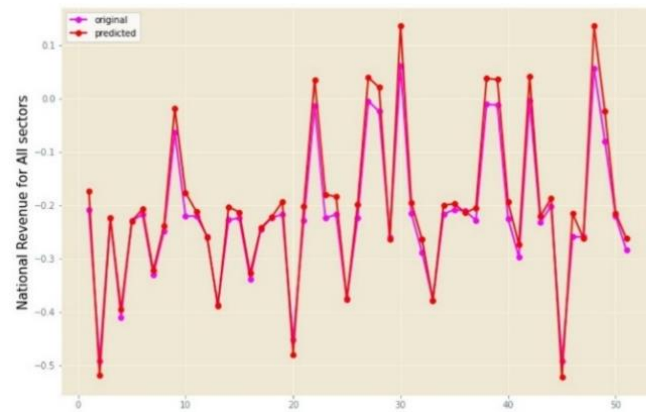

Fig. 7 Comparison of real and modeled (LSTM) data of national revenue all sectors.

Figure 6 represents actual (original labeled) and predicted (predicted labeled) data we found using LSTM for worker earnings. This figure, the x-axis, shows the number of days, and the y-axis shows worker earnings for all sectors. The predicted earnings are 90.33% accurate.

Figure 7 shows actual (original labeled) and predicted (predicted labeled) data we found using LSTM for national revenue. This figure, the x-axis, shows the number of days, and the y-axis shows the national revenue for all sectors. The predicted rates are 92.62% accurate.

4-2-2 Random Forest regression

From Table 4, we get an accuracy of 93.16% for workers' earnings, 93.18% for employment level, and 95.53% for national revenue using the random forest regression. We get 3.63% mean absolute error, 0.16% mean squared error, and 4.05% root means squared error for employment level. Again, 3.65% mean absolute error, 0.15% mean squared error, and 3.97% root means squared error for worker earnings. Finally, 2.02% mean absolute error, 0.07% mean squared error, and 2.69% root mean squared error for national revenue. From Table 4, it is clear that accuracy for the employment level, worker earnings, and national revenue is consistent, and the error percentage is low.

Table 4: Results from Random Forest regression

Section	*Accuracy*	*MAE*	*MSE*	*RMSE*
Employment level	93.18%	3.63%	0.16%	4.05%
Worker Earnings	93.16%	3.65%	0.15%	3.97%
National Revenue	95.53%	2.02%	0.07%	2.69%

The visual representation of actual and predicted data using random forest regression are in Figure 8, 9 and 10.

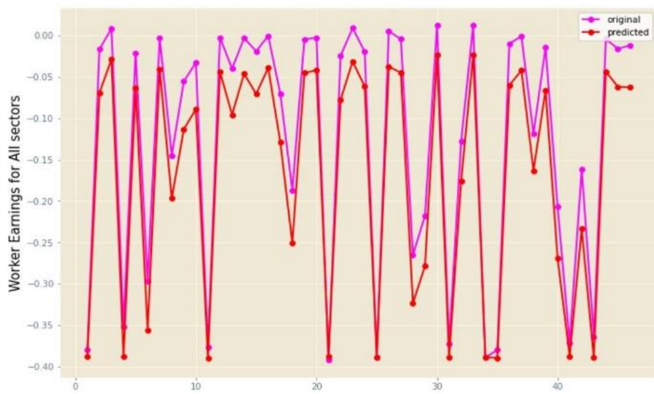

Fig. 8 Comparison of real and modeled (Random Forest Regression) data of worker earnings of all sectors.

Figure 8 illustrates actual (original labeled) and predicted (predicted labeled) data we found using random forest regression for employment rate or level. This figure, the x-axis, shows the number of days, and the y-axis shows the employment rate for all sectors. The predicted rates are 93.18% accurate.

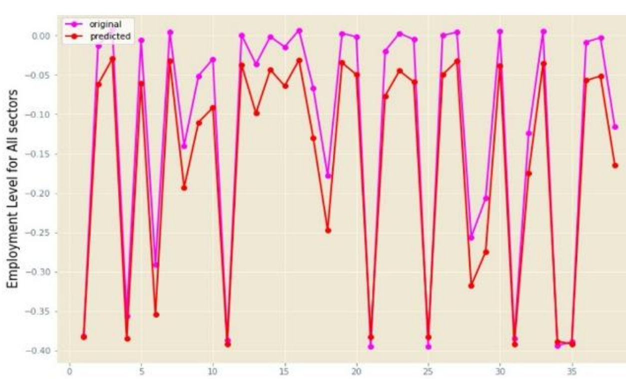

Fig. 9 Comparison of real and modeled (Random Forest Regression) data of employment rate of all sectors.

Figure 9 represents actual (original labeled) and predicted (predicted labeled) data we found using random forest regression for worker earnings. This figure, the x-axis, shows the number of days, and the y-axis shows worker earnings for all sectors. The predicted earnings are 93.18% accurate.

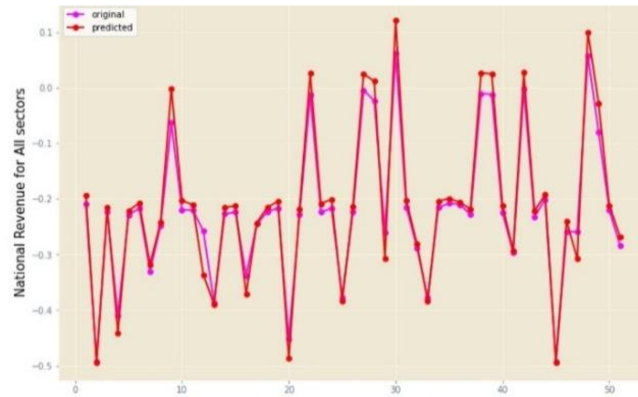

Fig. 10 Comparison of real and modeled (Random Forest Regression) data of national revenue all sectors.

Figure 10 shows actual (original labeled) and predicted (predicted labeled) data we found using random forest regression for national revenue. This figure, the x-axis, shows the number of days, and the y-axis shows the national revenue for all sectors. The predicted rates are 95.53% accurate.

4-3- Discussion

The goal of this model is to provide approximately financial downfall due to the pandemic by analyzing eight months (239 days) data of different economic sectors. The models demonstrated that it is possible to achieve a quality model using economic indicators as inputs through multi-layer perceptron (MLP), long short-term memory (LSTM) RNN, and Random Forest classifier and Regressor. The revenue, employment level, and worker earnings models use the same learning rate. When more data is available, a new detection can be produced with the latest data set by the neural network and random forest algorithms. The findings demonstrate the potential to use these algorithms in the future to model almost the same phenomenon. We have developed this model solution on all available data, which is incredibly constrained. Although the volume of data is minimal, in all three situations, we have superior accuracy.

In the MLP classifier, we got 81% accuracy for employment level, 89% for worker earnings, and 95% for national revenue. Although applying k-fold, employment level and worker earnings model accuracy increased, national revenue model accuracy decreased from 95% to 88%. On the other hand, LSTM revenue got 92.62%, LSTM employment rate got 90.04%, and LSTM earnings got 90.33%, which is appreciable with this limited amount of data. Including one year of training data and validation data to improve these models to decrease the data limitation problem. Besides, to improve this model for higher accuracy in the future by enclosing bidirectional LSTM, recurrent neural networks, and other appropriate algorithms in the current model.

For the global pandemic situation like COVID-19, we used different economic sectors reports as inputs. We correctly predicted the economic recession outbreak, including revenue, employment rate, and earnings data set. Like the LSTM, MLP, we use random forest classifier and regression to learn national revenue, employment level, and workers' earnings. This model has done an excellent job of predicting the most recent data. The model gives us a positive indicator for foreseeing the economic recession. Though we have used limited data for this detection, we got excellent accuracy in all the cases.

The advantage of the random forest classifier is its tall precision for multi-class classification, which is of the highest need [12]. In addition, as the random forest classifier builds different decision trees, and the ultimate result is assessed depending on the voting of these trees, the issue of overfitting happening in a single choice tree approach is killed. The random selection of feature vectors and random choice of features during learning makes the Random Forest classifier and Regressor solid and productive for any dataset [12]. Forecast analysis from [35] appears to have an accuracy of up to 90% in foreseeing classes with an ensemble approach. Using the random forest classifier model, we got an accuracy of 81% for employment level, 89% for worker earnings, and 85% for national revenue.

On the other hand, using the Random Forest regression model, we got an accuracy of 95.53% for national revenue, 93.18% for employment level, and 93.16% for worker earnings. After applying k-fold cross validation where k=5, suddenly the accuracy level showed different results for random forest classifier. Taking the mean value of 5 folds, we get 98.94% for employment level, 99.04% for worker earnings, and 98.43% for national revenue, prepared with a limited amount of data. Random Forest Classifiers and regression are helpful tools for economists and practitioners dealing with forecasting economic recession detection.

We did provide a warning model system for increasing awareness of an upcoming shock event. Economic crisis detection is essential for practitioners and policymakers since it provides an in-depth understanding of economic linkage breakdowns after a crisis. We proposed a system where we selected some crucial indicators that can be used to predict the economic crisis. As we know that economic conditions are continually changing, and during a pandemic, it changes drastically. So economic downfall detection will remain an open research issue with many situations and challenges to address. Continuously train the data-set with a more diverse set of machine learning algorithms and deep learning architectures that will benefit this research. The use of a developed and more robust technique will enhance the detection of the forecasting economy. Finally, a large amount of data-set will allow us to predict more accurately by filtering out noise embedded in the time series data.

Furthermore, cross-validation and comparison of training and testing data sets are essential. Since it helps one truly understand the models' meaning and efficiency and thus improve the outcome's reliability. In this prospective research aspirants, the importance of these latest technologies needs to be discussed. Before a country faces an economic recession, it is crucial to identify which sector to emphasize to minimize this unexpected scenario. Analyzing the data and detecting the economic impact to help save a country's significant capital using machine learning.

5- Conclusion

Several studies demonstrate that pandemic outbreaks caused considerable economic changes. In this present day, Covid-19 makes a tremendous economic impact on the global economy in terms of revenue, employment level, workers' earnings, and many more. The situation worsens due to the covid-19 pandemic, so prediction of the economic factor is necessary for taking early steps. So, this paper uses machine-learning algorithms to identify and predict the recession. This paper detected the recession in three different sectors of the USA that occurred due to this pandemic and got higher accuracy. When more data is available, the performance will be much higher. Including other algorithms and more data, we can forecast sector-wise economic recessions. We can achieve a model that can detect and forecast recessions that might occur in the future and help stakeholders take decisions as early as possible. To save a country's economy before it breaks apart, this model could play a vital role.

Acknowledgments

This research is funded by Woosong University Academic Research in 2023.

References

[1] M. K. Goyal and A. K. Gupta, “Integrated risk of pandemic: Covid-19 impacts, resilience and recommendations,” Berlin: Springer, 2020.

[2] D. Hanna and Y. Huang, “The impact of sars on asian economies,” Asian Economic Papers, vol. 3, no. 1, 2004, pp. 102–112.

[3] W.Y. Lin, Y.H. Hu, and C.F. Tsai, "Machine learning in financial crisis prediction: a survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 4, 2011, pp. 421– 436.

[4] M. R. Keogh-Brown and R. D. Smith, "The economic impact of sars: how does the reality match the predictions?," Health policy, vol. 88, no. 1, 2008, pp. 110–120.

[5] R. D. Smith, M. R. Keogh-Brown, and T. Barnett, "Estimating the economic impact of pandemic influenza: an application of the computable general equilibrium model to the uk," Social science and medicine, vol. 73, no. 2, 2011, pp. 235–244.

[6] M. Fern´andez-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," The journal of machine learning research, vol. 15, no. 1, 2014, pp. 3133–3181.

[7] S. Balcaen and H. Ooghe, "35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems," The British Accounting Review, vol. 38, no. 1, 2006, pp. 63–93.

[8] P. R. Kumar and V. Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques–a review," European journal of operational research, vol. 180, no. 1, 2007, pp. 1–28.

[9] S. J. Fong, G. Li, N. Dey, R. G. Crespo, and E. Herrera-Viedma, "Composite monte carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction," Applied Soft Computing, vol. 93, 2020, pp. 106282.

[10] K. Bluwstein, M. Buckmann, A. Joseph, M. Kang, S. Kapadia, and O. Simsek, "Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach," 2020.

[11] R. Nyman, and P. Ormerod, "Predicting economic recessions using machine learning algorithms," arXiv preprint arXiv:1701.01428, 2017.

[12] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, 2001, pp. 5–32.

[13] U. Gr¨omping, "Variable importance assessment in regression: linear regression versus random forest," The American Statistician, vol. 63, no. 4, 2009, pp. 308–319.

[14] Z. Car, S. B. Šegota, N. Anđelić, I. Lorencin, and V.Mrzljak, "Modeling the spread of COVID-19 infection using a multilayer perceptron", Computational and mathematical methods in medicine, Hindawi, vol. 2020, 2020.

[15] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy", Journal of accounting research, vol. 18, no. 1, 1980, pp. 109–131.

[16] R. C. West, "A factor-analytic approach to bank condition", Journal of Banking and Finance, vol. 9, no. 2, 1985, pp. 253–266.

[17] A. F. Atiya, "Bankruptcy prediction for credit risk using neural networks: A survey and new results," IEEE Transactions on neural networks, vol. 12, no. 4, 2001, pp. 929–935.

[18] J. H. Min and Y.-C. Lee, "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters," Expert systems with applications, vol. 28, no. 4, 2005, pp. 603–614.

[19] K.-S. Shin, T. S. Lee, and H.-j. Kim, "An application of support vector machines in bankruptcy prediction model," Expert systems with applications, vol. 28, no. 1, 2005, pp. 127–135.

[20] S. Sarkar, and R. S. Sriram, "Bayesian models for early warning of bank failures", Management Science, vol. 47, no. 11, 2001, pp. 1457–1475.

[21] E. Fedorova, E. Gilenko, and S. Dovzhenko, "Bankruptcy prediction for russian companies: Application of combined classifiers", Expert Systems with Applications, vol. 40, no. 18, 2013, pp. 7285–7293.

[22] J. Abell´an and C. J. Mantas, "Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring," Expert Systems with Applications, vol. 41, no. 8, 2014, pp. 3825–3830.

[23] R. Chetty, J. N. Friedman, N. Hendren, M. Stepner, and T. O. I. Team, "The economic impacts of covid-19: Evidence from a new public database built using private sector data," National Bureau of Economic Research, Working Paper 27431, June 2020.

[24] R. M. Simon, J. Subramanian, M.-C. Li, and S. Menezes, "Using crossvalidation to evaluate predictive accuracy of survival risk classifiers based on high-dimensional data," Briefings in bioinformatics, vol. 12, no. 3, 2011, pp. 203–214.

[25] Haykin S., "Neural networks and learning machines," 3/E: Pearson Education India, 2010.

[26] I. Syarif, A. Prugel-Bennett, and G. Wills, "Svm parameter optimization using grid search and genetic algorithm to improve classification performance," Telkomnika, vol. 14, no. 4, 2016, pp. 1502.

[27] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," Computer methods and programs in biomedicine, vol. 108, no. 1, 2012, pp. 10-19.

[28] V. Y. Kulkarni and P. K. Sinha, "Pruning of random forest classifiers: A survey and future directions," in 2012 International Conference on Data Science and Engineering, IEEE, 2012, pp. 64–68.

[29] Y. E. Cakra and B. D. Trisedya, "Stock price prediction using linear regression based on sentiment analysis," in 2015 international conference on advanced computer science and information systems, IEEE, 2015, pp. 147–154.

[30] B. Wang, L. Gao, and Z. Juan, "Travel mode detection using gps data and socioeconomic attributes based on a random forest classifier," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 5, 2017, pp. 1547–1558.

[31] F. A. Gers, J. Schmidhuber, and F. Cummins, "Neural Nets WIRN Vietri-99," Continual prediction using LSTM with forget gates: Springer, 1999.

[32] K. Suri and R. Gupta, "Transfer learning for semg-based hand gesture classification using deep learning in a master-slave architecture," in 3rd International Conference on Contemporary Computing and Informatics (IC3I), IEEE, 2018, pp. 178–183.

[33] S. Polyzos, A. Samitas, and A. E. Spyridou, "Tourism demand and the covid-19 pandemic: An lstm approach," Tourism Recreation Research, vol. 46, no. 2, 2021, pp. 175–187.

[34] L. Breiman, "Bagging predictors," Machine learning, vol. 24, no. 2, 1996, pp. 123–140.

[35] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for

motor-imagery BCI system," Biomedical Signal Processing and Control, vol. 31, 2017, pp. 398–406.

[36] Md. Monirul Islam, Mohammod Abul Kashem, Jia Uddin, "Fish Survival Prediction in an Aquatic Environment Using Random Forest Model," IAES International Journal of Artificial Intelligence (IJ-AI), Indonesia, 2021, vol. 10, no. 3, pp. 614-622.

[37] J. Ferdoush, B. N. Mahmud, A. Chakrabarty, J. Uddin, "A Short-Term Hybrid Forecasting Model for Time Series Electrical-Load Data using Random Forest and Bidirectional Long Short-Term Memory" International Journal of Electrical and Computer Engineering, Indonesia, 2020, vol. 11, no. 1, pp. 763-771.

[38] R. Islam, J. Uddin, J.M. Kim, "An Acoustic Emission Sensor based Fault Diagnosis of Induction Motors using Gabor filter and Multiclass SVM", Journal of Ad-hoc and Sensors Wireless Networks, Old city publisher, 2016, vol. 34, no. 1, pp. 273-287.