

Economic Impacts of COVID-19 Globally

Steven son Koninti Mark, MSc Big Data Analytics, L00162007

Abstract—This article addresses the economic effects of the COVID-19 crisis across markets and across nations. It addresses the economic mechanisms by which economic growth can be affected. With asymmetrical findings across countries and sectors. It also seeks a preliminary approximation of the possible global economic impacts of COVID-19 in diverse circumstances. The COVID-19 pandemic is an unexpected blow to the global economy, and its possible scarring consequences are also impossible to anticipate. About every country in the world has been affected by the coronavirus pandemic. Its proliferation left global economies and industries counting losses, while policymakers are struggling with new lock-up policies to combat the spread of the virus. Despite the introduction of new vaccines, many also question what recovery would look like. We will develop machine learning models using this dataset to help us predict economic losses across countries. Two separate Supervised Machine Learning Models were considered for this analysis, namely Logistic Regression and Random Forest. We analysed the data set in Logistic Regression by assessing its probabilistic importance while we carried iterations for the global impact of COVID-19 in Random Forest. We used the Apache Spark Analytical Engine to build machine learning models on the Databricks platform.

Index Terms—Coronavirus, Covid-19, pandemic, countries, global economy, economic crisis, Apache Spark, Random Forest (RF), machine learning, supervised learning, Databricks, logistic regression.

I. INTRODUCTION

The globe panicked to a threatening fact on 11 March 2020 when the World Health Organization (WHO) proclaimed the novel coronavirus (COVID-19) a pandemic [1]. The cases originated in Wuhan, China and transmitted exponentially to Japan, South Korea, Europe, and the United States when they approached global proportions. Towards the official announcement of a pandemic, substantial economic signs from various sources, weeks ago, suggested that the globe was leaning towards an unparalleled peak in our lifetime, if not in human history. The widespread of the COVID-19 pandemic has had a significant impact on the global economy. People all over the world are moving together to make a concerted effort to deal with this economic shock. Around the same time, we are seeing a global effort to address the economic crisis and to turn it into a spark for future growth. The Covid-19 outbreak is a global crisis 'like nothing else' including concurrent destruction of both supply and demand in the integrated global economy. Not only has the virus been a public health concern, but it has also impacted the world economy. Substantial economic impacts have also arisen across the world due to decreased economic output, sector shutdowns, import and export interruptions and extinction of the tourism industry. COVID-19 could be a wake-up sign for united nations to step up coordination on disease resilience and provide the appropriate funding for global concerted action. On the supply side, diseases limit labour supply and output, while lockdowns, market suspensions and social distances often interrupt supply. On the demand hand, layoffs, and lack of profits (from complications, lockdown, and wage stagnation) and deteriorated economic circumstances limit residential usage and spending by enterprises. Ample research has been provided on the potential socioeconomic impacts of epidemics, [2] however the global community has struggled to engage

enough in prevention and preparedness programs to minimize the effects of major epidemics. The primary issue in the world today is to save human lives. Lockdown and psychological distancing are the only methods to monitor the spread of contagious diseases, specifically practiced by several countries all over the world as a lockdown. However, regardless of all this, the weakening economy cannot be discounted, as a result of which several world-class companies are on the verge of sinking. For example, lockdowns have limited numerous industries, such as traveling to contain viruses, as a result of a sudden global downturn in this industry. While some countries are now starting to step steadily towards relaxing lockout controls, the consequences of the pandemic have already devastated several people's livelihoods and harmed opportunities for potential development. The main elements of globalization have already failed to work adequately or have entirely vanished. The COVID-19 pandemic has expanded with unparalleled momentum and has affected millions of people to shut down their economies, while countries have implemented strict travel restrictions to prevent the dissemination of the virus. The economic damage is now evident, with the health and human burden increasing and the world's greatest economic impact in decades. It is said that it will fall by 3 per cent, while the worst affected for developing countries is 4 per cent on average, while others are over 6.5 percent [3]. Unique strategies must be implemented in order to restore the economic and social opportunities which will dominate the next decade. To this end, we study the economic impacts of COVID-19 internationally by using machine learning models and Apache Spark as a data processing platform for large open-source data sets. It is designed to provide the processing power, scalability and programmability required for big data, especially data streaming, graph data and machine learning applications.

II. PROPOSED WORK

In the current project, we intend to come up with a design that can forecast countries that will have a major effect on the economy on the basis of dataset fed to it. the Covid-19 pandemic ramifications for the world economy we must examine for each nation the effects of the COVID-19 pandemic. We use a group version of Databricks as a research method to explore these indigenous elements. It offers a cloud-based online network for big data studies and gives cluster and micro cluster managers access to a notebook environment. We analysed and found the top countries with the largest influence due to COVID-19 while developing a model for prediction.

III. FRAMEWORK AND SOFTWARE

The data for the impact of Covid-19 Pandemic on the global economy is available online from the Kaggle website. The dataset is composed of nation, code, total caseloads, total mortality, stringency index, GDP, human development index and population. This collection of data includes at least 210 countries, which have an effect on the global economy of each region after the COVID-19 pandemic. The dataset is chosen to demonstrate how the economy has affected the countries worldwide. For the analysis of these results, we used the Databricks Group Version as an analytical tool. It offers an online cloud-based portal for the study of large data as well as access to the notebook ecosystem and cluster and micro-cluster managers. There are many big data frameworks available on Data Bricks from which we used the Apache Spark platform. Apache Spark is not only an unified engine for solving various data problems instead of learning and managing a number of separate methods, but also a general purpose system that shortens the path from lab-based informative analysis to operational analysis in production data applications and frameworks [4]. It is an analytical engine that uses its machine learning algorithms to construct an accurate predictive model for the economic effects of COVID-19. The understudy data collection is allocated to the data processing cluster and parallel processing areas. The SPARK CONTEXT is attached independently to the worker nodes, forming the main processing unit for the computation of the dataset within the spark system. These worker nodes live in executors and are linked to the background of cluster managers. [5] As data from the website was then organized in a highly structured format, we used the logistic regression method, which is a supervised learning method which allows the used framework to recognize and classify input data-based problems or situations. The Random Forest Algorithm was provided the same data set as an input for decision makers. We have to do this by choosing 10 decision-making processes. We used PySpark SQL library in Apache's Sparke both for data analysis and for visualization development. After data pre-processing, the MLlib library was used to hold predictions. Many big data structures are provided for data bricks from which we have used the Apache Spark framework. It is an empiric engine that uses its machine learning models to assess the impact of the Covid-19 Pandemic on the global economy. The data collection understudy is divided into parallel data

processing and cluster portions. In order to calculate the data collection within the Spark scheme, the Spark Context is separately linked to the operating nodes. This worker nodes are connected by the executor to the history of the Cluster Managers.

IV. DATA PROCESSING

A. Data description

The dataset Economic impacts of Covid-19 is taken from the online website Kaggle and the description is as follows: Format: CSV file separated by commas Number of Records: 50419 The dataset includes country name, code, total cases, total deaths, stringency index, GDP per capita, population and human development index. This dataset includes at least 210 countries, which are summarized below on the impact of the COVID-19 pandemic on the global economy of each region. The dataset is chosen to demonstrate how the economy has had a global influence in various countries.

Attributes	Description
ISO Code	Country Code
Location	Name of the country
Date	Days of the cases registered
Total Cases	Number of cases
Total deaths	Number of deaths
Stringency index	Measure based on responsive indicators
Population	Total population of country
GDP per capita	GDP of the country
Human development index	Measure of average achievement

Fig 1: Dataset Description

B. Data Pre-processing

Data Pre-processing is the step that transforms or encodes the data to make it easy to parse. Data Pre-processing In other words, the data characteristics can now be easily represented using the algorithm. A series of data objects may be considered as data objects, also often called records, points, vectors, patterns, instances, incidents, measurements, observations, or entities. The dataset contains the information about the total number of cases each day and the total number of deaths, GDP per capita, stringency index and human development index which shows the how the cases and deaths everyday impacting the GDP of each country which gives us how the economy of each country will affect due to the pandemic. There were several null values in the dataset. We used data fill commands from python to fill in these missing values.

```
1 Covid_data = Covid_data.na.drop()

Covid_data: pyspark.sql.dataframe.DataFrame = [iso_code: string, location: string ... 7 more fields]

Command took 0.18 seconds -- by stevensn303@gmail.com at 3/7/2021, 7:04:23 AM on BDA
```

Fig 2:Data Pre-Processing

V. DATA MODELLING

A. Concept and techniques

1) *Predictive Modelling*: Predictable modelling is a statistical technique designed to predict future events or consequences, also known as predictive analytics, by observing trends that can anticipate future outcomes. Predictive analysis models have their advantages and disadvantages and are better used by particular applications. The two subsections which the algorithm is used to train the dataset and predict desired outcomes they are regression and classification, while regression helps to predict continuous quantities, classification predicts discrete class labels. There are two subclasses in pattern classification, supervised and unsupervised learning. A model based on computational, machine-learning methods and the example data pack is built in this process. After the implementation, the test dataset is checked to verify the validity of the model by a portion of the main data collection, and the model is thought to suit if accurate. The establishment of predicted predictive properties relates to the a priori determination of the characteristics and expected consistency and accuracy of the forecasts [6]. The model, once fitted, will predict accurately the new data as machine input. The multi-model solution has been chosen for a problem in many applications.

VI. METHODS OF PREDICTIVE MODELLING

A. Logistic Regression

Logistic regression is the effective regression analysis for the dichotomous component. Logistic regression is a statistical analysis, as are other analyses of regression. Logistic regression is used in the description of data to illustrate the relationship between one binary variable dependent and one or more independent nominal, ordinal, interval, or ratio-level variables. Machine learning can be completed for certain basic binary classification problems, and the implementation of machine learning in this respect is quite advanced, so that machine learning already has significant scientific importance. Among them, logistic regression has been commonly used as a general form of data processing for binary classification and prediction purposes.[7] This algorithm uses the dataset and predicts the GDP of each country and how it will affect

the economy.

```
Cmd 34
>
1 evaluator = MulticlassClassificationEvaluator(
2 labelCol='gdp_per_capita_temp', predictionCol='prediction', metricName='accuracy')
3 accuracy = evaluator.evaluate(predictions)
4 print('Accuracy = %g' % (accuracy))

(1) Spark Jobs
Accuracy = 0.0817052

Command took 1.44 seconds -- by stevensn303@gmail.com at 3/7/2021, 7:04:52 AM on BDA
```

Fig 3: Logistic Regression

B. Random Forest

Random Forest is a scalable, easy-to-use machine learning algorithm that delivers a fantastic result much of the time, even without hyperparameter tuning. It is also one of the most popular algorithms due to its simplicity and variety. A significant benefit of the random forest is that it can be used for both classification and regression problems, which make up the bulk of modern machine learning systems. Let us look at the classification of random forest, since classification also constitutes the machine learning component. The Decision Tree estimates the significance for the next results as applied for regression analysis, based on the previous iteration.[8] This algorithm collects and compile the N number of values from the data sets in a decision-tree and estimates the value for the upcoming result on the basis of the preceding iterations when applied to the analysis.

```
1 # Select (prediction, true label) and compute test error
2 evaluator = MulticlassClassificationEvaluator(
3 labelCol='gdp_per_capita_temp', predictionCol='prediction', metricName='accuracy')
4 accuracy = evaluator.evaluate(predictions)
5 print('Accuracy = %g' % (accuracy))

(1) Spark Jobs
Accuracy = 0.179881

Command took 1.97 seconds -- by stevensn303@gmail.com at 3/7/2021, 7:04:46 AM on BDA
```

Fig 4: Random Forest

VII. TABLE

New attacks are the other obvious form of adversarial i/p which can fully disregard classifiers. Attacks wouldn't occur too much, but it is always necessary to find a way

to cope to them because they are catastrophic. Latest attacks are the obvious method of adversarial variables which can be completely ignored by the algorithms. There are not many new attacks, but it's always essential to find ways to deal with them, as they are often quite dangerous. There are also many possible probable reasons for fresh assaults, in our research, these above two categories of events are likely to trigger them.

TABLE

Model	Accuracy
Random Forest	81.17%
Logistic Regression	17.98%

Fig 4: Accuracy and comparison

VIII. VISUALIZATION

Following visualisation describe briefly about the economic impact of COVID-19 globally from the data set under study.

A. Countries with sum of total cases

After analysing the data, we can see the largest number of covid cases in Russia, at 53 percentage, while Fiji is the least number of cases at less than 1percent. This allows one to analyse the largest number of cases per dataset in Russia.

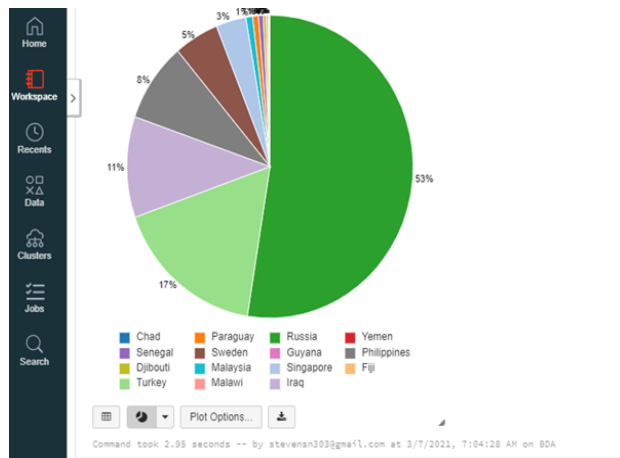


Fig 4: Total cases comparison

B. 2) Countries with total number of deaths per day

After analysing the data, we can see the highest number of covid deaths in Italy, at 294, while china is the least number of deaths at 271. This allows one to analyse the largest number of deaths per dataset in Italy.

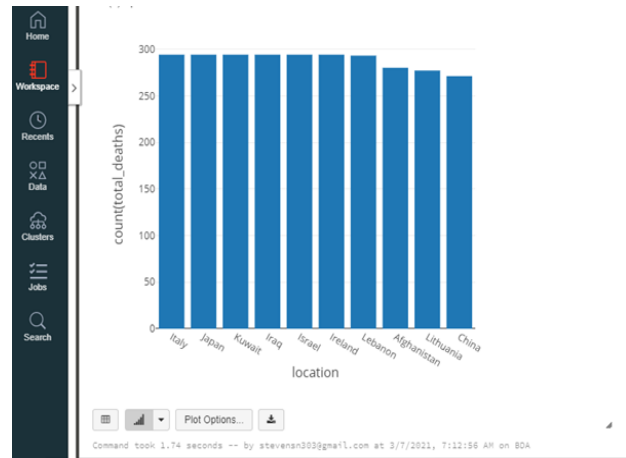


Fig 4: Total deaths comparison

C. GDP per capita

After the analysis of data, we can notice that most of the countries are having the same GDP whereas china has slightly lower as per the dataset.

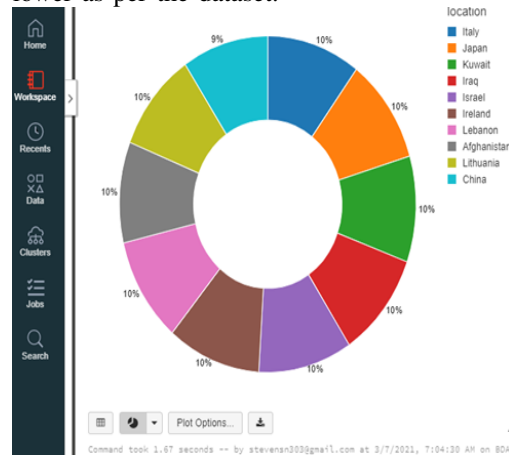


Fig 4: Total deaths comparison

1) Human development index based on different locations:

On analysing the dataset, we can notice the human development index is same for most of the countries whereas china has slightly lower when compared with the other countries.

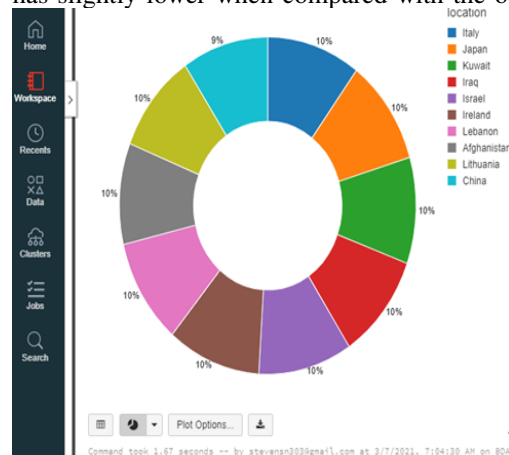


Fig 4: Human development index

IX. CONCLUSION AND FUTUREWORK

Given that the propagation of the virus is expected to continue to disrupt economic activity and adverse impacts on the manufacturing and service industries, particularly in developing countries, we expect financial markets to remain volatile. It remains to be asked if the ongoing recession would have a permanent effect on the world economy, or largely a short-term economic and financial impact. In this case, it is obviously possible to inflict serious economic and financial burdens on international and global economies with communicable diseases such as COVID-19. For predicting the Impact of Covid-19 on global economy, we have found random forest more useful as it gave the accurate results when compared to the logistic regression. As seen from the visualizations, the countries which are most affecting and the countries which are having a greater number of deaths and the countries with a greater number of COVID-19 cases and different countries human development index we can predict the countries that are going to be more affected due to COVID-19.

X. REFERENCES

- [1] Sohrabi C., Alsafi Z., O'Neill N., Khan M., Kerwan A., Al-Jabir A., Iosifidis C., Agha R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19) *Int. J. Surg.* 2020.
- [2] Yamey G, Schäferhoff M, Aars OK, Bloom B, Carroll D, Chawla M, et al. Financing of international collective action for epidemic and pandemic preparedness. *Lancet Global Health.* (2017) 5:e742–4. doi: 10.1016/S2214-109X(17)30203-6
- [3] Maliszewska M, Mattoo A, Van Der Mensbrugghe D. The potential impact of COVID-19 on GDP and trade: A preliminary assessment. *World Bank Policy Research Working Paper*, 2020,9211
- [4] Ryza, S., Laserson, U., Owen, S., Wills, J.: *Advanced Analytics with Spark: Patterns for Learning from Data at Scale.* O'Reilly (2015).
- [5] Stoica, I. and Zaharia, M. (2016). "Introducing Databricks Community Edition: Apache Spark for All", Company Blog, Databricks, viewed on Feb 18th, 2020 at <https://databricks.com/blog/2016/02/17/introducing-databricks-community-edition-apache-spark-for-all.html>
- [6] Refsgaard, J. C., J. P. van der Sluijs, A. L. Højberg, and P. A. Vanrolleghem. 2007. Uncertainty in the environmental modelling process: a framework and guidance. *Environmental Modelling and Software* 22:1543–1556.
- [7] S. Rongheng, *Applied Mathematical Statistics*, CA:Science Press, 2014.
- [8] Nigam, R. and Govinda, K. (2017). "Cloud Based Flight Delay Prediction using Logistic Regression", *Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017)*, IEEE, p. 662-667.