Prediction of impact on GDP and Unemployment due to COVID19

Jashwanth Reddy Baggari, Jyothi Vasamsetty, Nikhil Velakurthy

February 2023

1 Introduction

Covid 19 is spreading worldwide. Scientists have debated the outbreak's cause. The World Health Organization estimates that over a quarter million people had died from the illness by May 27, 2020. Within months, this virus had infected millions of individuals worldwide. Many facets of the covid problem exist in one country. Countries around the world have taken steps like staying indoors, social isolation, hand washing, travel restrictions, lockdowns, and more to stop the spread of the disease. Lockdowns, for example, are extreme, have unparalleled effects on daily life, and have major economic consequences. The recent global lockdowns have had a major influence on global GDP, making precise forecasting of COVID-19-related aspects even more important.

The lockout stretched the supply chain and made streamlining crucial products opaque. The informal sector and hourly workers are especially at danger. Uncertainty plagues many US food farmers. Hotels and airlines are laying off workers and lowering salaries. We evaluate the spread under the economy's effect using machine learning algorithms with linear regression models. Predicting future trends by analyzing past or present data. Such analysis and forecasting involve defining the job, acquiring relevant data from many sources, assessing the data, undertaking statistical analysis, constructing a data model, deploying the collected data using numerous approaches, and monitoring the model.

Forecasting has been one of the most successful statistical methods for detecting and analyzing patterns and predicting future events, allowing for early and mitigating action. In this paper, we use a linear regression model and clustering to examine the data on COVID-19 cases and the pandemic's economic impact. We also provide many

data visualizations. We evaluate, clean, load, characterize, partition, apply the model, and visualize the data in this procedure.

2   Literature Survey

Our study project used these three papers and publications. It aids economic studies on COVID-19.

In the first study, they seek a correlation between socio-economic factors and spread in each affected country. The key contribution is projecting spread using categorization models based on over 1000 socioeconomic parameters and 190 nations. Given constructed classifiers, an importance analysis is suggested to identify the most important classification indication

The second paper examined COVID-19's effects on elementary, secondary, and tertiary sectors. The COVID 19 outbreak reaction is unthinkable. A quarter of the world's population is under lockdown as governments declare emergencies, gather cash, and ask residents to change their behaviour to stop the epidemic. With a new recession and economic crisis, health, industry, government, and community policy making is crucial. A comprehensive social-economic development strategy that promotes businesses with stable and sustainable business models through infrastructure and sector-by-sector programmes.

The third paper examines the influence of COVID-19 on unemployment rates in 13 European nations from January 2019 to August 2020 using a panel data technique. The report also examines the causes that contributed to the pandemic's surge in unemployment rates. According to the survey, the closure of non-essential enterprises such as restaurants and retail establishments has been a key contributor to the rise in unemployment rates. The study also indicates that government interventions, such as vacation plans and unemployment compensation, have helped reduce the impact of the epidemic on unemployment rates.

In our paper, we used two data sets covid related unemployment and country population statistics to assess the pandemic's economic impact. We utilized Linear Regression and Clustering to determine how population drop and unemployment during COVID-19 affected global GDP.

3 Project Background

Worldwide markets have been significantly impacted by the Covid-19 pandemic, with GDP growth decreasing and unemployment rates rising in several nations. The pandemic has impacted numerous economic sectors in diverse ways, including as the closure of enterprises and the adoption of social segregation policies, which have changed consumer behavior and decreased economic activity.

Understanding the short-term and long-term effects of the epidemic on the global economy is crucial as it continues to develop. The accurate planning of governments' and organizations' reactions to the crisis depends on accurate GDP and unemployment rate predictions.To anticipate future economic results, this model will make use of historical data on GDP and unemployment rates, as well as information on Covid-19 cases and relevant governmental actions.

The prediction model can help decision-makers evaluate the efficacy of various policy actions to lessen Covid-19's negative economic effects. Also, it can give firms information about probable upcoming economic conditions, allowing them to modify their activities accordingly. In general, the model can offer useful information to a variety of stakeholders, assisting them in making decisions during these unsure times.

3.1   Tools :

- Python: A high-level, all-purpose programming language is Python. Code readability is prioritized in its design approach, which makes large use of indentation. Python uses garbage collection and has dynamic typing. It supports a variety of paradigms for programming, including functional, object-oriented, and structured programming.

- Jupyter Notebook : The Researchers have conducted research project, which once had an IPython Notebook project of its own, gave rise to Jupyter Notebooks. The primary programming languages it supports are Julia, Python, and R, hence the name Jupyter. There are presently more than 100 additional kernels offered, however Jupyter comes with the IPython kernel, which supports Python programming.

- Numpy : For accessing, manipulating, and acting on data in vectors, matrices, and higher-dimensional arrays, array programming provides a strong, condensed, and expressive syntax. The main Python library for array programming is called NumPy. It is an essential component in research analysis pipelines in a number of fields, including engineering, finance, physics, chemistry, astronomy, geoscience, biology, psychology,

and materials science. For instance, in astrophysics, NumPy was a critical part of the software stack used in the first imaging of a black hole and the discovery of gravitational waves.

Working with huge datasets is a common task of machine learning algorithms, and NumPy offers rapid and successful array operations that can greatly accelerate learning. Also, a lot of data preprocessing, which involves transforming and standardizing data before feeding it to a machine learning algorithm, uses NumPy.

- Matplotlib : The most prominent and established Python plotting libraries used in machine learning is Matplotlib. Using various visualisations, machine learning helps to grasp the vast amount of data.

  Data visualization is a core part of machine learning because it enables practitioners and academics to examine and comprehend their data, spot patterns and trends, and explain their conclusions to others. With Matplotlib, users can easily customize and fine-tune their plots to suit their unique needs. Matplotlib offers a flexible and powerful interface for developing a wide range of visualizations.

  Many elements of machine learning, such as input data, output predictions, model performance measures, and more, can be visualized using Matplotlib. Matplotlib, for instance, can be used to make line graphs that display changes in a model's performance over time or scatter plots that show the relationship between two input variables.

- Seaborn : A matplotlib-based Python data visualization library is called Seaborn. It provides an elegant drawing tool for creating eye-catching and educational statistical visuals.

  Heatmaps, violin plots, box plots, and other types of statistical visualization are all available with Seaborn. These visualizations are made to make it easier for users to study and grasp complicated statistics as well as to spot trends and interactions between different variables.

  For the visualization of multivariate data, or data with numerous variables, Seaborn offers a number of tools. For instance, Seaborn can be used to make heatmaps that display the correlation between various variables or scatter plots with multiple dimensions.

3.2 Infrastructure :

We are going to require strong infrastructure that can support the data processing and analysis processes in order to develop the predictive model. These are some essential infrastructure elements that we will require:

- Data Storage: In order to store the huge datasets we'll be employing, we'll need a dependable and scalable data storage system. A cloud-based

storage solution like Amazon S3 or Google Cloud Storage may fall under this category.

- Data Processing: We will require a system to handle the data processing and format conversion into an analysis-ready state. Both a batch processing system like Apache Hadoop and a stream processing system like Apache Kafka might be used to do this.

- Data Analysis: To evaluate the data and create the prediction model, we will need a system. A machine learning framework like TensorFlow or PyTorch could be used for this.

- Deployment: We will require a method of deployment once the model has been trained so that it can produce predictions in real-time. A cloud-based deployment tool like AWS Lambda/SageMaker or Google Cloud Functions might be used to do this.

- Monitoring: Last but not least, we'll need a monitoring system to keep tabs on the model's performance and spot any problems early. Prometheus or Grafana are two monitoring tools that could be used for this.
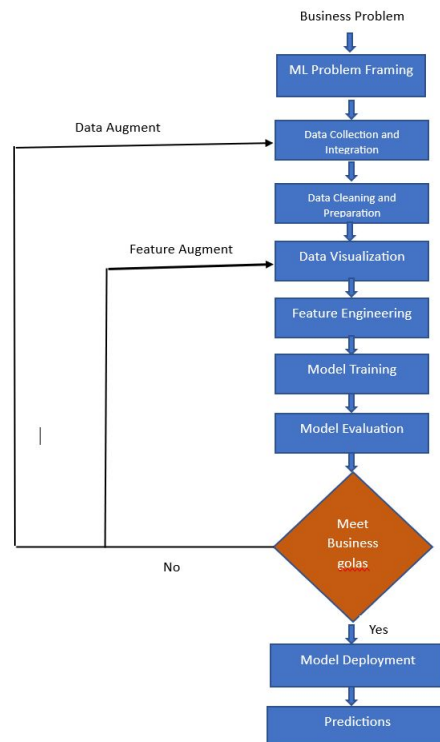
3.3   Algorithms:

- Linear Regression : The link between two or more variables can be modeled using this straightforward but efficient algorithm. It might be applied to forecast how Covid-19 cases and associated policy actions will affect GDP and unemployment rates.

- Random Forest : An ensemble technique mixes different decision trees to increase the precision of predictions. By taking into account a number of variables associated to Covid-19 instances and relevant policy actions, it could be utilized to produce projections of GDP and unemployment rates that are more accurate.

- XGBoost : A class of group machine learning methods known as gradient boosting can be applied to classification or regression predictive modeling issues.Decision tree models are used to build clusters. In order to repair the prediction mistakes caused by earlier models, trees are added one at a time to the ensemble and fitted. Boosting is a term used to describe this kind of batch machine learning model.

- Decision Tree : A decision tree creates tree-like models for classification or regression. It incrementally develops an associated decision tree while segmenting a dataset into smaller and smaller sections. The outcome is a tree containing leaf nodes and decision nodes.

- Support Vector Machines (SVM) : By identifying the ideal decision border between data points, this approach can be applied to both classification and regression issues. It might be used to find the model that fits the data the best and forecasts how Covid-19 would affect GDP and unemployment rates.

3.4   Cloud :

Users may create, train, and deploy machine learning models at scale using Amazon SageMaker, a fully managed machine learning platform. For data scientists and developers, it offers a range of tools and services that make it simpler for them to create, train, and deploy machine learning models on the cloud.

- Data Preparation : Data preparation, cleansing, and transformation are made easier by the data wrangling service offered by Amazon SageMaker, known as SageMaker Data Wrangler.

- Model Building : Machine learning model construction is simplified by SageMaker's array of pre-built algorithms and frameworks, including XGBoost, TensorFlow, and PyTorch. Also, users can use SageMaker with their own algorithms.

- Training and Tuning : SageMaker offers distributed training features that enable users to rapidly train models on huge datasets. Additionally, it offers a service for automatically adjusting models, which improves hyperparameters for better model performance.

- Deployment : Machine learning models may be quickly and easily deployed to production with SageMaker. It supports many deployment scenarios like batch inference, real-time inference, and edge inference and offers a fully managed, scalable infrastructure for serving models.
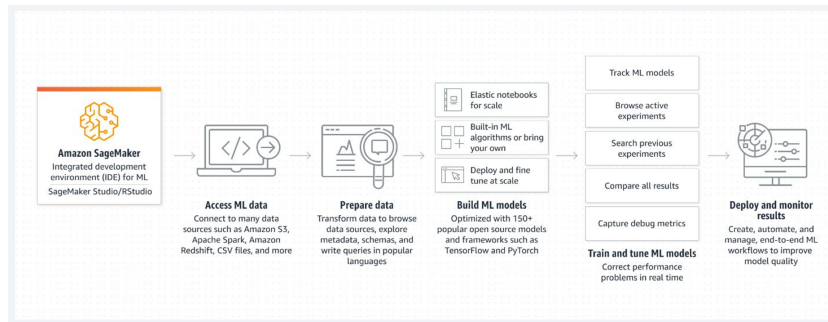
4 References:

- Senthil Kumar, Mohan , John A , Ahed Abugabah, Adimoolam M, Shubham Kumar Singh, Ali Kashif Bashir, Louis Sanzogni. An approach to forecast impact of Covid-19 using supervised machine learning model,2021

- R. Kumari et al., "Analysis and predictions of spread, recovery, and death caused by COVID-19 in India," in Big Data Mining and Analytics, vol. 4, no. 2, pp. 65-75, June 2021, doi: 10.26599/BDMA.2020.9020013.

- Su, C.W., Dai, K., Ullah, S. and Andlib, Z., 2022. COVID-19 pandemic and unemployment dynamics in European economies. Economic Research-Ekonomska Istraživanja, 35(1), pp.1752-1764.