COVID-19 on the Market

Raymond Hong
College of Engineering
San Jose State University
San Jose, USA
raym.hong@gmail.com

Zoe Lie
College of Engineering
San Jose State University
San Jose, USA
zoelie215@gmail.com

Michaela Molina
College of Engineering
San Jose State University
San Jose, USA
michaela.molina@sjsu.edu

Yathukulan Thayabaran
College of Engineering
San Jose State University
San Jose, USA
ythayabaran@gmail.com

***Abstract*—This dissertation describes the project where the economic effects of the Novel Coronavirus (COVID-19) were explored and useful predictions were made using Machine Learning. The predictions were displayed to potential users through a website. User interactivity was included. Predictions were limited to the U.S. economy.**

***Index Terms*—COVID-19, Macroeconomics, Machine Learning, IBM Watson Machine Studio**

I. Introduction

The effects of the current COVID-19 pandemic are felt globally. The exponential spread of the pandemic during the first few months is the highest in recent history. Figure 1 shows the severity of the U.S. cases [2].

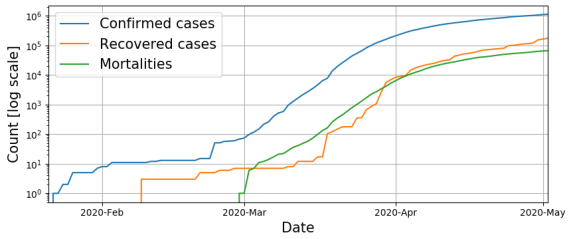

Fig. 1. U.S. COVID-19 spread. Note how the morality rate closely follows the Infection rate trend.

As with all pandemics, the effect on an individual's health and mortality are the first order effects. However, the second, third, and the consequential order effects, like the resultant economic disruption, the loss of productivity, and the psychological effects on the populace, are harder to measure and quantify; thus are more difficult to predict. This project attempts the difficult task of predicting COVID-19's effect on the U.S economy: the effect being twofold. Firstly, the pandemic and its resultant government mandated lock-down caused a fall in the productivity of the average business and thus reduced the supply of goods and services on the market. This also results in a cycle where the demand for supporting materials and business services also plummets. Secondly, the faith in the market and the economy erodes as the pandemic spreads and adds a psychological element that contributes to the downturn of stocks and market optimism reducing consumer spending. A good indicator is consumption. Figure 2 shows the sharp decline in retail sales since the pandemic started.

Fig. 2. Bar Graph depicting the decline in retail sales.

The pandemic affects both consumers and businesses. The goal to assist those affected by the repercussions of shelter-in-place is to allow them to visualize and see what is coming their way. Currently, a family business can only see the non-existent traffic coming through their doors. Our platform gives these people the ability to predict and draw business decisions based on concrete data and user input.

II. Related Work

There has been recent work that is currently ongoing to predict the spread of COVID-19. In India, where this paper is describing the setting of the spread, uses Deep learning to predict the spread [1]. Using data that dates between January and April it uses Long Short Term Memory Neural Networks to predict it. With a prediction spread of 90 days, it was able to predict with good accuracy the number of confirmed cases. The error margin was fairly strong. What this proves that machine learning models can give health officials a better forecast on what is to come, and proceed with necessary measures to contain the spread.

The topic of stock market prediction is also popular in research. In "Short-term prediction for opening price of stock market based on self-adapting variant PSO-Elman neural network" [3] researchers continue this well-trod path and use the Elman network for predicting opening stock prices. Elman networks belong to the family of Neural Networks, which are in short terms representations of finite state machines. However, the researchers used a self-adapting variant called the PSO-Elman neural network to account for certain deficiencies in the Elman network. Similar to our project, the researchers used as data the short-term stock market prices from recent history. The researchers concluded that when there are no significant

plunges in the market the predictions provide solid advice for realistic usage. This provides evidence that our project is feasible and applicable in real-world situations.

III. Architecture

The architecture requires several platforms. The reiterate, this platform is a website that displays important economic data related to the pandemic, it also provides predictions such as stock prices, and unemployment. This stack consists of technologies such as IBM Watson, Django Web Framework, and PythonAnywhere. The full Architecture diagram can be seen in Figure 3.

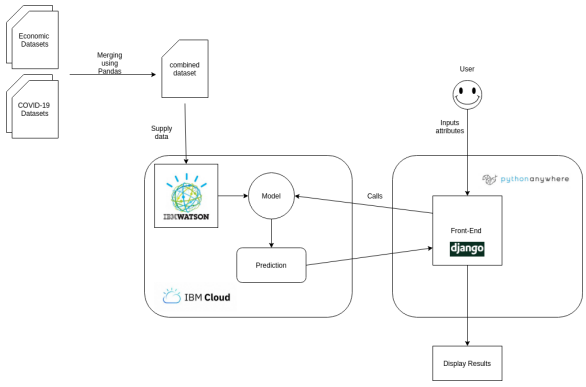

Fig. 3. Diagram of Architecture

The dataset was constructed out of many different datasets relating to economic factors that could be affected by the pandemic. Once a dataset is made, it is uploaded into IBM Watson's machine learning platform. The platform then makes a model that has callable functions that can be called from the Django front-end. The front-end receives user input about what their intentions are, and either shows visualized data, or submits input data for predictions. Once IBM Watson responds, the results are showed on the front-end. In addition, the website is hosted and deployed on PythonAnywhere.

IV. Core Implementation

A. Dataset

The dataset was chosen to represent various parts of the U.S. economy. Both leading indicators (like stock market status) and lagging indicators (like gasoline prices) were chosen to better create the machine learning model. Given the recency of the pandemic, the use of data with monthly frequency was omitted, and only weekly and daily data are used. Some data were missing certain points (e.g: weekends for the stock markets) and were extrapolated from existing data. The dataset contains data for predictions of the stock market as a general health indicator for businesses, unemployment rate for individuals, and petroleum prices for infrastructure, among multiple other variables. All data were gathered from primary sources, mostly from the U.S. government institutions. The data manipulation and validity checking was automated using the Python Pandas framework. The datasets were plotted against the main cause to establish visual correlation (and thus probable causation) before they were added to the input of the Machine Learning Model. An example would be the Seasonally Adjusted Unemployment rate which is currently at its worst since the great depression (See figure 4).

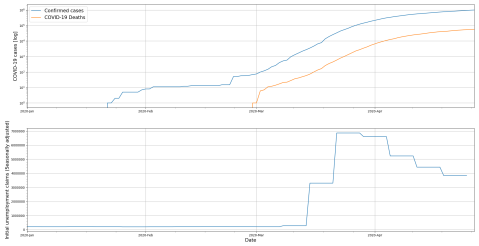

Fig. 4. Seasonally adjusted initial unemployment claims vs coronovirus spread in the U.S.

B. IBM Watson

IBM Watson is a Software-as-a-Service (SaaS) where developers can use its platform to create machine learning models without having to code anything in machine learning libraries, such as Tensorflow. It recently developed a new service, called "Auto-AI", that automatically analyzes the data fed into the platform, and suggests the best type of classifier to you.

After feeding the application the constructed dataset, we obtain a regression model that we can use. Each model is able to predict one column, so in order to implement many prediction services, we must create multiple models. In this case, we have created a model to predict SP500, Dow Jones, and Unemployment rates. Figure 5 shows the top 5 models for the Dow Jones prediction model.

Rank	↑	Name	Algorithm	RMSE (Optimized)	Enhancements
★ 1		Pipeline 8	XGB Regressor	290.690	HPO-1 FE HPO-2
2		Pipeline 7	XGB Regressor	322.434	HPO-1 FE
3		Pipeline 6	XGB Regressor	331.546	HPO-1
4		Pipeline 1	Extra Trees Regressor	344.475	None
5		Pipeline 2	Extra Trees Regressor	344.475	HPO-1

Fig. 5. Top five regression models

C. Front-End

The website was built using Django, Python, HTML, and CSS. The website dashboard is shown in Figure 6. The dashboard shows a summary of statistics including the number of tests conducted, the number of confirmed cases, the number of deaths, and the number of people in the ICU in either the United States as a whole or just California. The dashboard also displays a variety of line graphs and bar charts for data visualization, including graphs for the 10-year breakeven inflation rate, an overview of COVID-19 statistics, and an overview of stock market prices.

In the side menu of the website, the user can choose to "Get New Predictions" of the SP 500, Dow Jones, and

unemployment by selecting one of the options, shown in Figure 7. After selecting an option, the user can then upload a csv file with their own data, or enter up to three rows of data manually as shown in Figure 8. After submitting the request, the user will be shown a graph showing predictions that uses their input for the option that they choose.

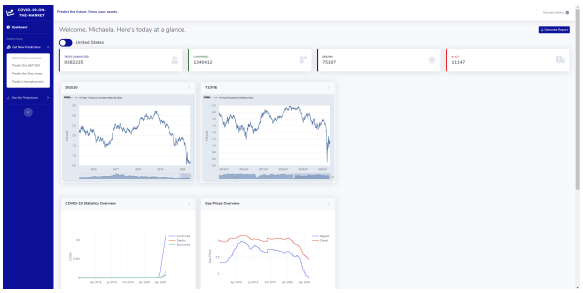

Fig. 6. Website Dashboard

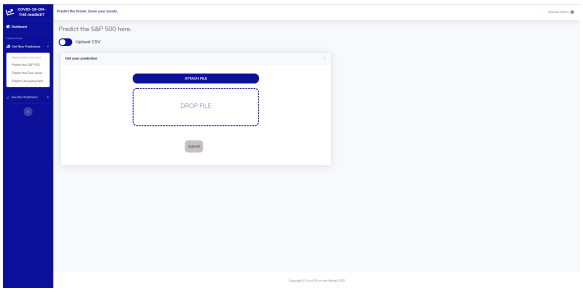

Fig. 7. User Predictions Screen on Website

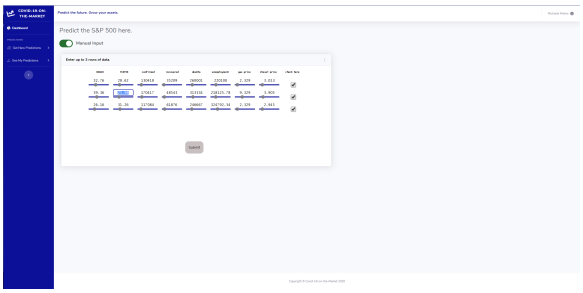

Fig. 8. User Predictions with Manual Input on Website

D. PythonAnywhere Cloud Service

Instead of using other cloud services such as Amazon Web Services (AWS) or Google Cloud Platform (GCP), we decided to use PythonAnywhere to host and run our Django app. PythonAnywhere is an web hosting service based in Python. We decided to use PythonAnywhere because our website mainly uses Python and Django, so PythonAnywhere was a logical choice for hosting our website.

V. Conclusion

The IBM Watson machine learning model binds the macroeconomics of the pandemic and its effects to the national economy with the microeconomic predictions to the user (businesses and individuals alike). Given the scope of the predictions, a dataset with thousands of factors, often accessible to data aggregators and governments, would further refine the predictions of this product. Possible improves to this project is possible once more information/data becomes available. One of the challenges of creating this dataset was not having enough information to work with. We believe with more data, our model would be more accurate. We would also strive to create more models that are relevant to certain businesses, like retail sales for example. All-in-all, these tools are especially useful for dangerous times like this pandemic, and we can rely on them to create innovative solutions in the future.

References

[1] A. Tomar, N. Gupta, “Prediction for spread of COVID-19 in India and effectiveness of preventive measures” Science of the Total Environment, vol. 728, April 2020.

[2] time-series-19-covid-combined,” DataHub, 11-May-2020. [Online]. Available: https://datahub.io/core/covid-19readme. [Accessed: 11-May-2020].

[3] Z. Zhang, Y. Shen, G. Zhang, Y. Song and Y. Zhu, ”Short-term prediction for opening price of stock market based on self-adapting variant PSO-Elman neural network,” 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 225-228, doi: 10.1109/ICSESS.2017.8342901.