COVID-19 Vaccination and Death Rates*

"People are dying that have never died before." - Donald J. Trump

Santiago Bermudez†
Computer Science Department
California State University, Sacramento
Sacramento, California, U.S.
santiagoabermudez@csus.edu

Eric Truong
Computer Science Department
California State University, Sacramento
Sacramento, California, U.S.
etruong2@csus.edu

Amad Shah
Computer Science Department
California State University, Sacramento
Sacramento, California, U.S.
amadshah@csus.edu

ABSTRACT

With COVID-19 shocking the whole world with its growth and deadliness, we were unprepared. With people's beliefs, be it religion or doubtful towards the governemnt, they were unkeen to seek vaccination when available. We want to predict the amount of deaths in a country given details about it, including vaccination rate. To approach the problem we used two models to predict the amount of deaths country given vaccination rate, case rate, population, etc. Our models were fairly close to the actual rate of deaths.

1 INTRODUCTION

Given how COVID has taken us by storm, we were fast to act. It is easy to obtain data and use it to predict future events and avoid risk. This is important because we want to save as many people as possible. We want this information to be seen by the public eye to sway their ideals on vaccinations.

We first got the dataset for CovidDeaths and Vaccination rate for each country. We then transformed our data and built models to predict the death amount for each country. We show that countries with lower vaccination rates experience higher total deaths amount compared to countries that have higher vaccination rates.

First, we will go over the **Problem Formulation** seen in section 2. This will show how we will predict the number of deaths in a country. It will go into more detail in which inputs we used to predict the output, total deaths. We would first have to take the data provided and process and split them to be ready for our FCNN and CNN models.

Next we will go over the **System/Algorithm Design** of our model. We used FCNN and CNN models to train and predict outputs. With these two models we are predicting the total deaths within each country. We will tune the parameters of each model and compare the different models with each other. With each FCNN models, we changed activations, neuron counts, and optimizers to see how they would affect each model. With the CNN model, we chose to do a CNN model without transfer learning and changed the parameters such as activation and neuron counts. To get our best model we tuned our parameters to see which was giving us the best results.

Then came our results. For the **Experimental Evaluation,** we compared the results of each the FCNN and the CNN models to see how they varied with different parameters. We were looking to compare our models predicted total death amount and the actual total death in a country. First we would compare each activation test and see which one had the best results. After activation tests, we looked at the models with varying neuron counts. Next, we checked the results of each optimizer. It seemed only one of our activation tests were near close to the actual number of deaths. We would take the best model out of each tune and use that for the next tests. After comparing the FCNN models between each other, we would do the same comparisons for the CNN model. We compare each model with different tuning and take away the best model out of each. Finally, we compared our best FCNN model and our best CNN model to see which one of those performed best at predicting the total death amount. We then printed out the total deaths in each country given their vaccination rate. We found that countries with higher vaccination rates have lower amount of deaths in their country.

2 PROBLEM FORMULATION

Our current plan is to use the data we have to try and predict the total number of deaths per country. To see if we can predict whether vaccination and testing help to reduce death. We would also like to see if any other factors play a significant role, like age or a nation's GDP.

Our goal is to use the following inputs to predict total_deaths:

*From CovidVaccinations.csv:
- location
- date

- total_tests
- positive_rate
- testspercase
- total_vaccinations
- people_vaccinated
- totalvaccinationsper_hundred
- stringency_index
- population_density
- median_age
- gdppercapita
- diabetes_prevalence
- female_smokers
- male_smokers
- hospitalbedsper_thousand

*From CovidDeaths.csv:
- location
- date
- population
- total_cases
- totalcasesper_million
- totaldeathsper_million
- reproduction_rate
- Total_tests

Our output (*from CovidDeaths.csv):
- total_deaths

3 SYSTEM/ALGORITHM DESIGN

3.1 System Architecture

For our FCNN model we will train it by taking the input to predict the output, total deaths. Since we are predicting deaths we will be using regression model. We are able to directly used the processed data to train our models.

For our CNN model, we simply take our inputs and convert them into a sort of image format for our model to process and analyze for regression.

3.2 FCNN

For our Fully Connected Neural Network, we will be using different activations, optimizers, and different neuron counts in each layer. We will keep it at two layers of neurons for our model. When testing the different activations for our model, we kept neuron count and optimizer the same. We tested “sigmoid”, “relu”, and “tanh” activations while keeping neuron count in the first layer at 50, 25 in the second layer and with our optimizer as “adam.” We learned that “relu” activation gave us the best results so we went with that to test neuron count. For testing neuron count, we kept the same activation and same optimizer. We tested first with our control, 50 and 25. Then we tested 100 and 50 in each layer. Next we cut those down and tested with 20 and 10 in layer 1 and 2. For the different optimizer tests, we tested the “sgd” optimizer. It did not perform even close the our “adam” optimizer that we have been using.

3.3 CNN

There are three types of layers that would make up the CNN. These are the convolutional layers, pooling layers, and fully-connected (FC) layers. For starters, we have our convolutional layer, which is used to extract features from input images, or in our case, data that has been converted into a sort of image format. We start by declaring a sequential model format, then we add the input layer and use Conv 1D. We then flatten our convolutional layer before passing it to the fully connected Dense layer. Now all we need to do is to compile the model by defining the loss function and the optimizer, and then we'll be ready to train it. We also use early stopping and fit. We run our model, tune some of the hyperparameters and that is it.

4 EXPERIMENTAL EVALUATION

4.1 Methodology

For this project, the data that we used was a Covid-19 dataset from Kaggle. This dataset consisted of two CSC files, one for Covid deaths and the other for Covid vaccinations. For data preprocessing, we would first check for and get rid of duplicate rows for each CSV file separately and deal with missing values by first getting rid of unnecessary columns in our dataframes. Afters some basic preprocessing, we would merge both data frames on common identifying attributes that both data sets share. In our case, we would merge the datasets on ‘location’, ‘date’, and ‘total tests’.

In terms of experimental setting, we did not actually go out and conduct any experiments anywhere physically. We just used a dataset that was provided to us on Kaggle, which is a collection of statistics from all around the world. We just worked with this data on Google Colab with at least two different models.

When it comes to comparing different methods, because our problem is based on regression, we would use the RMSE score and lift charts to compare the effectiveness of the methods that we utilized.

For our project, the methods that we used are Fully-Connected Neural Networks and Convolutional Neural Networks without transfer learning. These were the methods that we utilized and compared with some parameter tuning.

4.2 **Results**

For our Fully-Connected Neural Network model, below are the RMSE scores and lift charts that we got from that model with the results that we got from parameter tuning as well:

Basic FCNN model:

Activation Tests:

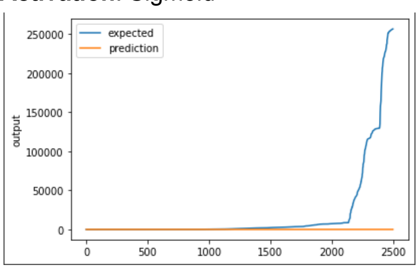
(Model 1 Activation - Sigmoid)

Neuron Count (Layer 1): 50

Neuron Count (Layer 2): 25

Optimizer: Adam

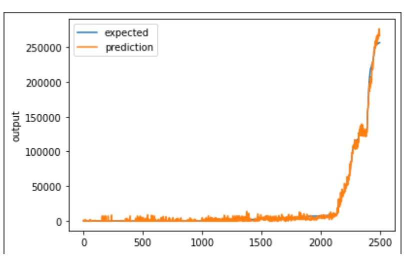
Activation: Sigmoid



Final Score (RMSE): 3334.77

(Model 2 Activation - Relu)

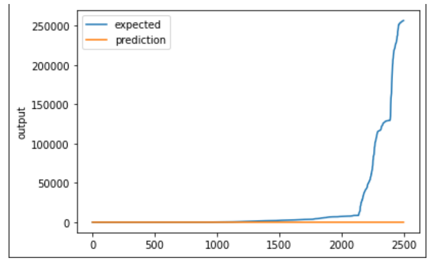
Activation: Relu



Final Score (RMSE): 55888.75

(Model 3 Activation - Tanh)

Activation: Tanh



Final Score (RMSE): 55888.74

Neuron Count:

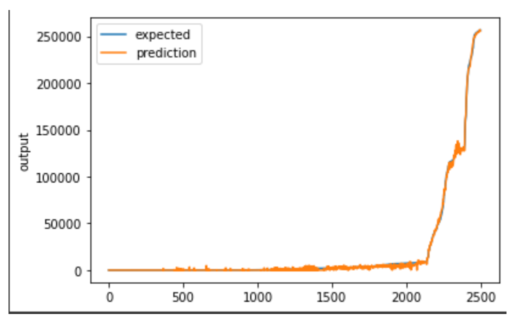
(Model 1 Neuron Count - 100 - 50)

Neuron Count (Layer 1): 100

Neuron Count (Layer 2): 50

Optimizer: Adam

Activation: Relu

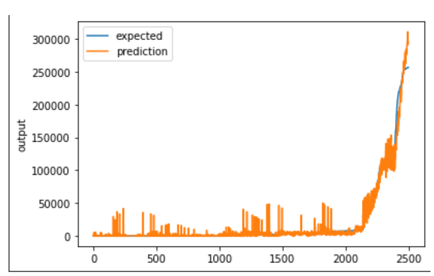


Final Score (RMSE): 1512.92

(Model 2 Neuron Count - 20 -10)

Neuron Count (Layer 1): 20

Neuron Count (Layer 2): 10



Final Score (RMSE): 8671.95

Optimizers:

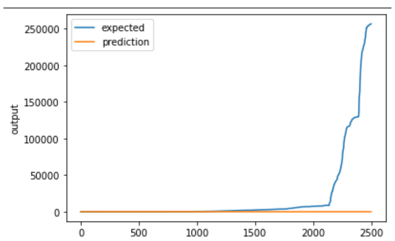
(Model 1 Optimizer - sgd)

Neuron Count (Layer 1): 50

Neuron Count (Layer 2): 25

Optimizer: sgd

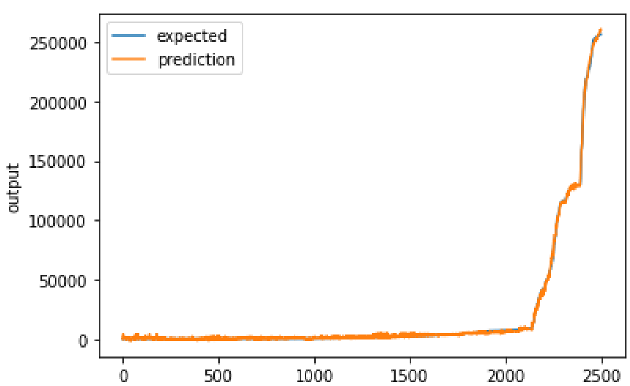
Activation: Relu



Final Score (RMSE): 55889.11

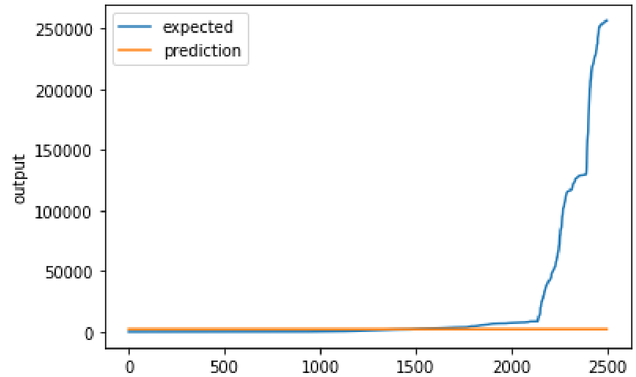
For our Convolutional Neural Network model, below are the RMSE scores and lift charts that we got from that model with the results that we got from parameter tuning as well:

Basic CNN model:



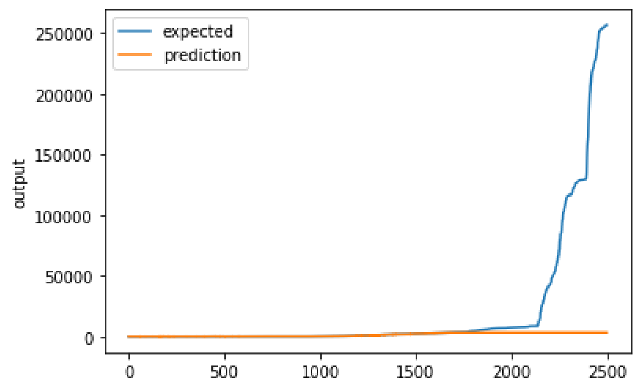
Final score (RMSE): 1114.9791259765625

CNN model with Sigmoid activation:



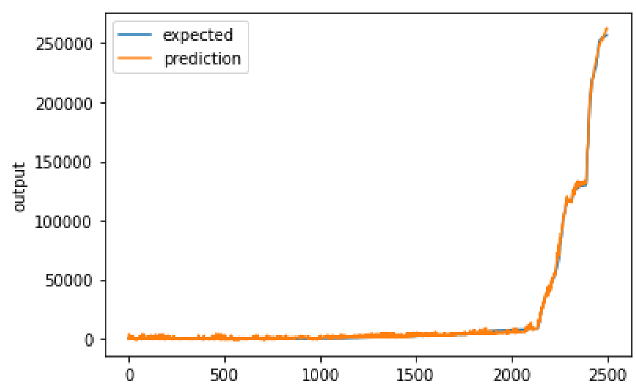
Final score (RMSE): 55141.9609375

CNN model with Tangent activation:



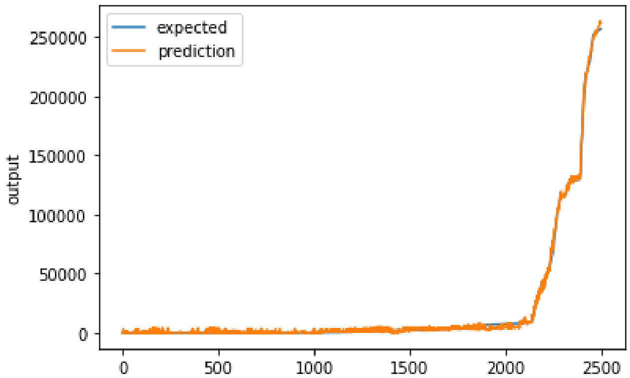
Final score (RMSE): 54783.0078125

CNN model with bigger neuron count:



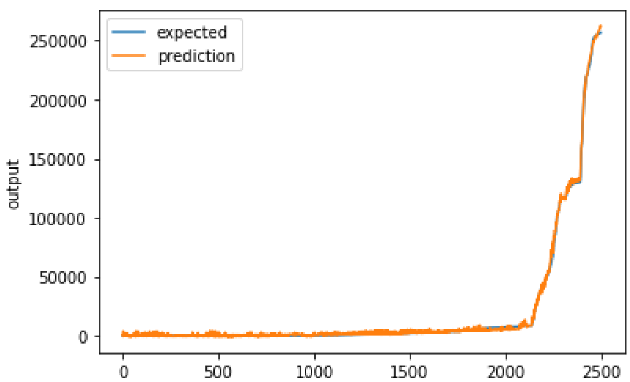
Final score (RMSE): 1471.493408203125

CNN model with smaller neuron count:



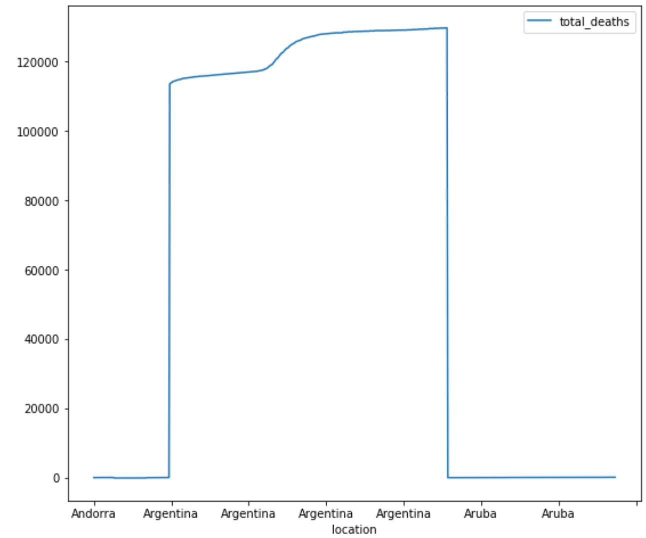
Final score (RMSE): 1654.90673828125

Below is the lift chart and RMSE score of the best model we have for CNN:
CNN model with bigger neuron count:



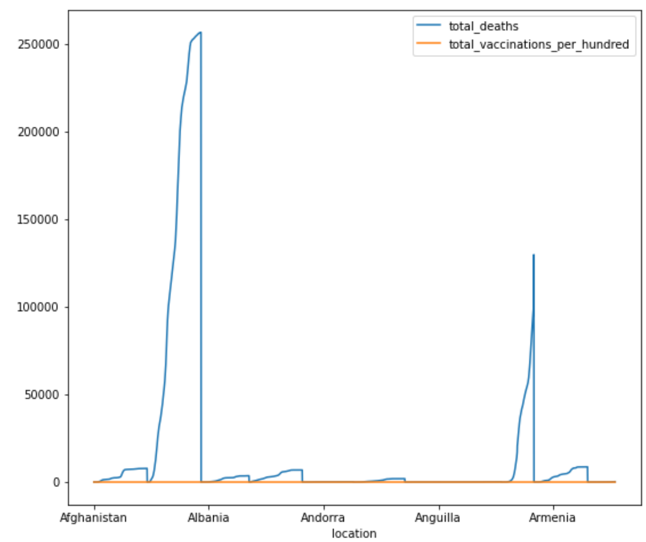
Final score (RMSE): 1471.493408203125

Below are some other graphs we got from using our models to predict total deaths:
For our countries in periods of high vaccination here:



If you look at the graph above , you can see that there are fewer deaths and the deaths are more spread out.

For our vaccinated countries in periods of low vaccination:



If you look at the graph above, you can see that the death counts are higher and that there are more spikes.

5 CONCLUSION

After creating our models and training them, the best model we found was a basic CNN model with ReLU activation, an Adam optimizer, and input that is one row or one line. Using this model, we then went on to split our dataset into something easier to graph and would then graph our results for total deaths based on vaccination rates. For periods of high vaccination, we have seen that the death count is lower and more spread out, which is a good thing because it means that medical resources are less likely to be stretched thin for that period of time. During periods of lack of vaccination, we have seen spikes in death count, which are harder to control and manage.

6 WORK DIVISION

Our work division is all mostly based on the final project proposal that we submitted. For data preprocessing, Amad would do the first third of the tasks, followed by Santiago for the second third of the tasks, and Eric for the final third of the preprocessing tasks. For the FCNN model, this would be done by Amad with aid form Santiago, and likewise, Eric would work on the CNN model with some aid from Santiago. Once the coding was done, we would split the final project report into thirds and do some of the last few tasks together, such as work division, learning experience, checking overall quality, and the powerpoint slides.

7 LEARNING EXPERIENCE

From this project, we have learned that getting vaccinated does in fact help, even if it does not make the problem go away entirely. While the results we got from graphing did confuse us a bit at first, it did turn out to confirm exactly what we are expecting. Generally with better self-protective practices, such as vaccination, we see a trend where fewer people fall ill and perish. We also see a trend where the deaths are more spread out, which means that people are getting impacted at a slower rate, which is preferable because it means that our medical resources and health care workers are not being stretched too thin or strained. We also found that our best model for predicting the results so far was the CNN model and that is it! Thank you!