National College of Ireland

Comparative Study of Machine Learning Algorithms to Predict COVID-19 Cases

MSc Research Project
Data Analytics

Noel Joseph
Student ID: x18184499

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Noel Joseph
Student ID:	x18184499
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Noel Cosgrave
Submission Due Date:	17/12/2020
Project Title:	Comparative Study of Machine Learning Algorithms to Predict COVID-19 Cases
Word Count:	4660
Page Count:	18

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	17th December 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Comparative Study of Machine Learning Algorithms to Predict COVID-19 Cases

Noel Joseph
x18184499

Abstract

Covid-19 which is caused by SARS-CoV2 an infectious epidemic was declared as a pandemic by the WORLD Health Organization(WHO) in the month of march. The virus has affected almost the whole world.which leads lot of researchers all over the world to study, whats about to come? how bad is it going to affect the society? what measures or precautions which can be taken to prevent or restrict the spread etc. In this study we will forecast and predict the number of confirmed cases and deaths of the world using the prophet model. Which gave a MAPE value of 8.151203314201725 for the confirmed cases prediction and a MAPE value of 2.1322274771230485 for the deaths prediction. This paper also predict the number of confirmed cases of Ireland using the neural network techniques like Multilayer Perceptron(MLP) which gave a MAPE value of 0.10325138342004382 and also Long short-term memory(LSTM-RNN) which gave an MAPE value of 0.12573520838004137 all of this predictions that were performed gave a good prediction accuracy with MLP being the best.

Keywords: Covid-19, Machine-learning, Neural Networks, Prophet model.

1 Introduction

A new virus, later labeled as COVID-19, triggered an outbreak in Wuhan, Hubei, in early December 2019. The province, China and other areas of China and many other countries in the world have expanded further Roda et al. (2020). The WHO had then declared a public health emergency for the outbreak to be an international problem. Which kept on increasing throughout all the months which came ahead the deaths were also increasing from hundreds to thousands and then to millions. This disease effects the respiratory system and is highly infectious disease the transmission is from one human to another and even anything they touch like through surfaces the virus can live on surfaces.

COVID-19 is susceptible to ultraviolet rays and heat it is assumed that at 27 degree this virus can be inactivated and also things like ethanol peroxyacetic acid things that contain chlorine disinfectant leaving chlorhexidine. Various virus like SARS and the MERS that effect the human health with infections like cold were able to affect a lot of people in the history Wieczorek et al. (2020).

The government and the agencies in the effected countries have being reporting the impact of the outbreaks, it is important for modelers to estimate the magnitude of the epidemic in terms of the total number of infected, the total number of reported cases,

the total number of casualties and the basic number of reproductions, and to forecast the epidemic's timing, peak time and total duration. Such data can help to make informed decisions for public health organizations Roda et al. (2020). Various estimations were made but there was too little data to work with at that given time which was the major drawback in applying the models Understanding how the available data (confirmed cases) correlates with model predictions is a key problem that can clarify the uncertainty in model predictions. Confirmed cases are persons with symptoms that have been contacted and checked by a hospital and whose COVID-19 infection has been confirmed by DNA or imaging tests Roda et al. (2020).

There are various question that rise with the Covid19 spread like, what will be the number of the people affected tomorrow or in the near future? i.e. a week a month and so on. What measures can be taken to control the spread? When will it reach its peak? Which models can be implemented to get the answers for such questions? Erraissi et al. (2020) .

The main aim of this paper is to try o predict the number of confirmed cases and deaths worldwide as well as in Ireland with the help of the data provided University of Johns Hopkins, Baltimore, MD, United States Dong et al. (2020) applying various techniques like Facebook's Prophet model for forecasting the future trajectory and then the (James Usher 1new).and then applying models like LSTM and MLP to predict the number of confirmed cases Car et al. (2020) and Pano-Azucena et al. (2018).

1.1 Research Question

In this study we try to answer the question:

RQ: How well can machine learning models help in accurately predicting deaths and cases for COVID-19?

2 Related Work

In this section we will discuss the papers which were researched that had done similar work on the topic as well as things which are related to the topic the difficulties they faced and the results,

2.1 Analysis and prediction

This part will talk about the previous research done on the topics or models which are related to the work presented in this paper and discus about their findings and thoughts.

There are various difficulties that are being faced in predicting covid-19 the author Roda et al. (2020) in this paper states that the data of the confirm cases is non-identifiable in the models due to which there are large variations. They have said that the main reason for a wide range of prediction models was that there was limited amount of data available. The calibration of the model was based on the Bayesian Inference and the selection of the model was based on the Akaike Information Criterion(AIC). They have improved the Markov chain monte calo and presented a better version of it. They have used two different frameworks the SIR and the SEIR and have applied a model selection algorithm to decide which one of them is better they found out that the SEIR model when used had a higher amount of data loss so they used the SIR model then as it gave better results.

They have also stated the effects and the benefits of the lockdown in lowing the spread of the epidemic and how the people returning to work lead to the second wave.

A forecasting model for Covid-19 with the help of the Neural network is been developed by the author in this paper Wieczorek et al. (2020) has given information about the disease and the factors that are effecting the disease with the history and background of similar kind. The limitation stated that was taken under consideration was that a simple mathematical problem can determine the new potential cases. They have tried using both the RNN and the ANN neural network and the Ann was better than the RNN both in time and accuracy. They have run various algorithms to optimize and NAdam gave the best result. The major disadvantage was the training time of the model or any sudden changes could effect the model predictions. In future the author plans to increase the precision and improve the network efficiency even if there was limited information available with the help of various statical models.

To predict the deaths worldwide due to covid 19 Chaurasia and Pal (2020) has made use of Arima model for timeseries and regression model. They have analyzed the correlation between all the attributes and the mortality. they used the ARIMA to differentiate the time series and make them stationary and then they have combined both of the model expressions they found that the correlation between the confirmed and the recovered was highly positive. They have said that in future the mortality rate is going to decrease based on the ARIMA graph and have concluded by saying that there is a possibility to reduce the death worldwide

To find out the number of people that have been infected by the covid-19 Erraissi et al. (2020) proposed a model that was based on Artificial Intelligence and Machine learning. In the model that the author used There were a number of predefined instructions which generates the output in the authors case it was the number of people the number of cases of the covid-19. They have used the Spark ML API in order to perform the prediction. There were various algorithm used in this like the Decision tree, SVM, Random forest. They conclude by saying that they have a possibility of getting a better prediction with the help of the method proposed

Timeseries model for the prediction of the trend and to short term prediction model has been shown in this paper by Tran (2008). They have created a time series model on the basis of mathematical formula. The results indicate that the time which is required to suspect the population has a significant impact on the peak. He also says that if there is an increase in the mortality it will also cause a peak in the number of the confirmed cases. They three different types of they have used the ARIMAX model for the timeseries analysis the limitations was that they yet don't know anything about the seasonal fluctuations. The Arima x performed had done the best among all the other 6 models one of the major limitations found in most is the lack of information in the dataset.

Various deep learning models have been used by Arora et al. (2020) to predict the covid-19 positive cases In India. The methodology used in this paper was based on the RNN (LSTM) like the bidirectional LSTM, convolutional LSTM and deep LSTM to predict the amount of the positive cases in India. The linear data moving average is been taken to fill the missing values of the data for better accuracy and feasibility for the predictions the bi-directional LSTM had the lowest amount of errors against all others with a highly accurate short term predictions and the convolutional LSTM gave the worst result of all.

SEIR model as previously seen by the author **?** which was not that efficient, In

this paper the author Yang et al. (2020) has proposed a modified SEIR model and has used an Ai approach which was proposed in the year 2003 to predict the same kind of epidemic. RNN's LSTM model was used in this research and was optimized with the help of the Adam Optimizer. Both the models performed well in the predictions but the SEIR model was a bit better that the LSTM model SEIR model was also helpful in reducing the epidemic size.

In this research study by James and Pierpaolo (2020) the author develops two ML models using the Facebook's Prophet model in the form of (ARIMA) financial time series and an additive regression financial time series to investigate the hypothesis that a Conservative victory's poll prediction could be confirmed by predicted pound increase. Efficiency of the models used for forecasting was examined on the MAPE and the MSE value. They also stated that ARIMA and the prophet model were verry effective in forecasting and the arima gave the lowest MAPE and the MAE value.

A research article by Car et al. (2020) they have performed a multilayer perceptron on a timeseries data it is a similar kind of data which has been used in this paper but they have transformed the data to a regression form of dataset and used it in training a MLP -ANN. They aimed to obtain a model which is capable of calculating the amount of patients worldwide of the time unit. They used different hyperparameters in MLP were determined with the help of grid search a total number of 48384 were trained which contained the infected, recovered and deceased the evaluated the model based on the r2 value and cross validated it with the help of the k-fold algorithm. They used the Relu activation which has been used in the model in this paper. The high robustness of the deceased reveals Patient model, strong verified robustness, and poor robustness for the recovered model.

According to Pano-Azucena et al. (2018) he shows the utility of the multilayer perceptron (MLP) in the ANN family to predict chaotic time series that have been obtained from chaotic systems based which were on saturated nonlinear function series and from the Rossler method in the research paper they have shown the implementation of hardware in the in the systems for the prediction which was verified with the help of FPGA they have provided the RMSE which shows the how stable the electronic system is. The paper shows how MLP is being applied in the prediction of the timeseries and how useful ANN's are they were able to achieve a verry good performance when it came to the prediction.

2.2 Working with real time data

As the part of the real time data in this paper the author Swapnarekha et al. (2020) has analyzed various on-going methods like Deep learning and Machine Learning for predicting the covid-19. Various mathematical and also statistical models have been used on the covid dataset. Machine learning techniques like SVM, regression, k means, have been applied and also deep learning techniques like CNN, LSTM, residual neural network have been performed. They were successfully used in the prediction each with having certain limitations.

Another research study by author Moorthy et al. (2020) they have used real time data of COVID-19 to perform forecasts on various models like the SIER model in areas like china and around which were based on the daily observations of the number of cases deaths and recovered. Their findings showed that they had assumed the outbreak would reach the peak and then will start to drop. Their literature work had found to be more

of the negative articles rather than positive ones. Thy also said that the SIER model can be obtained more accurately after the successful containment of the outbreak.

A web based dashboard was created by Dong et al. (2020) to track covid-19 on a real time basis the repository was made available on GitHub for the public health authorities and the general people so that they could track covid on a real time bases. The data gets updated twice a day one in the morning and once at night first it was done manually then it became semiautomated information from various other trusted sources was taken to make an aggregate and the dataset is always growing and updating.

A real time tracking with the help of self enrolment based on health conditions systems was an application created by the author Menni et al. (2020) there were a huge number of participants/People that have used this app the generated results about various symptoms like loss of smell and taste from which 65 percent of the people were tested positive and the negative test result ere 21 percent. the model was combining the systems to predict the result. The limitation faced was that the Physiological tests of olfactory and the gustatory activity or nucleotide-based function may not be replaced by SARS-CoV-2 research. False negatives as well as false positives which could be included in the dataset as the questions asked were in a different way.

In another research study the author Marian et al. (2020) have developed a modified transmission model of SIR metapopulation and Analytically parameterized it by additionally setting the values of parameter depending on Or by believing their plausible beliefs, the literature. They contrasted the results of their In twenty selected nations with the number of cases diagnosed, those that provide Reliable figures, however they varied greatly in the terms of intensity and speed of implementation. Interventions of Non-Drugs. Their results showed a infectious period of 4.6 days and also stated that the probability to develop a very serious health condition was verry low and also said that the virus might have been misestimated.

3 Methodology

Data mining is defined as the method of extracting valuable information from the vast volume of data to obtain certain insights. The three most common methodologies for data mining are KDD, CRISPDM, and SEMMA. In terms of the number of steps they have each technique varies from one another. [1]

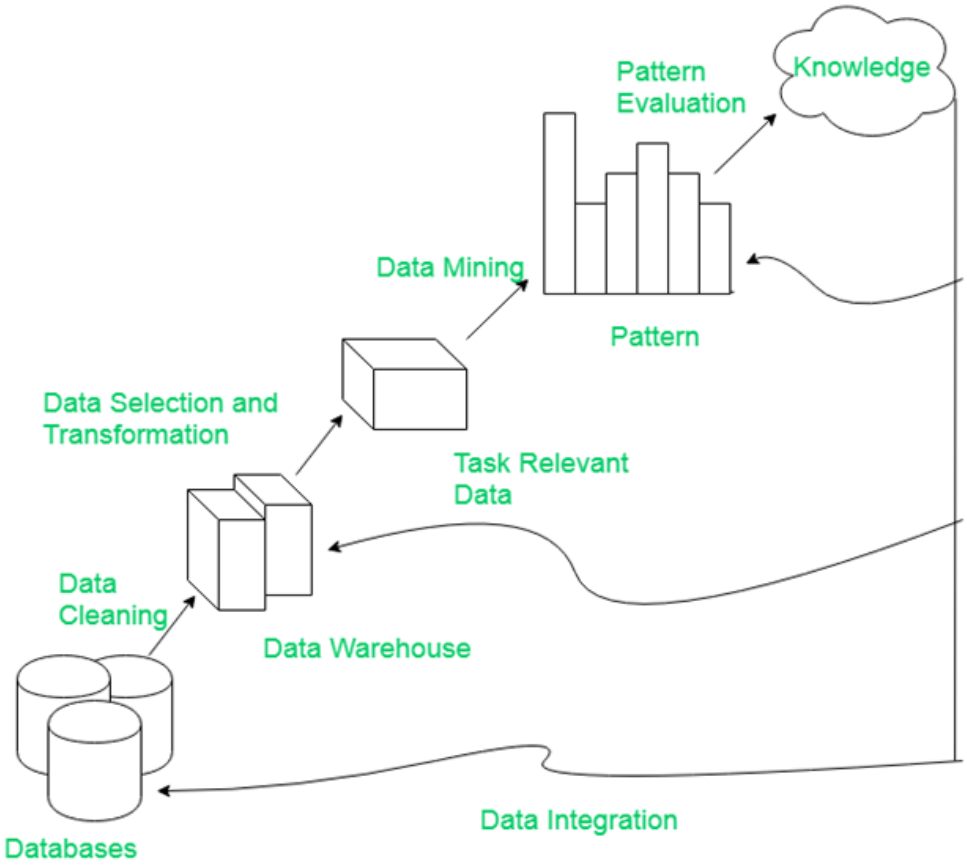

Figure 1: KDD Methodology

- Data Cleaning: In this stage we check for any values which are missing in the data also check for noise or uneven data or removing the data which is irrelevant.
- Data Integration: Merging the data if collected from various sources in this paper we have collected data from three different files of the same repository and merged it into one.
- Data Selection: In this process we select which data is relevant and required this can be done using models this is one of the most important step.
- Data transformation: here we transform the data and make it relevant for the mining process. Data which is relevant to the modeling and visualisations is kept.
- Data Mining: Various models are applied to the data to get the desired output/results.

[1] Fig `https://media.geeksforgeeks.org/wp-content/uploads/KDD_process.png`

- Pattern Evaluation: The models applied are then evaluated on the basis on their performance and output and evaluating the results and summarizing them .
- Knowledge representation: Representation of the acquired knowledge with the help of visualisation .

3.1 Process Architecture

The objective of this study is to accurately predict the total number of covid-19 confirmed cases by using Machine learning

- The uncleaned data is being loaded into the notebook.
- EDA has being performed for the better understanding of the data things like transforming and extracting the data.
- Applying different algorithms of machine learning on the data.
- The algorithms then help in forecasting and making predictions.
- Finally the performance of the models that have been applied are evaluated and the best one is then chosen.

3.2 Data Collection

The raw data here was taken from the repository by the university of Johns Hopkins(CSSE)[4] and also Applied Physics Lab (APL) of the university which presented the data of covid-19 it consists of data all over the world collected from various sources. The data is being updated on the daily basis and stored into .csv files. The data is available and can be used publicly for research and educational purposes. there were three different csv files used in this paper which consisted of data from a number of countries and provinces about their deaths number of confirmed cases and the number of recovered people on a daily bases. the data consisted of columns like Province/State, Country/Region the latitude the longitude, dates , Confirmed, Recovered, Dead further data Transformation was performed and new columns were added for better understanding of the data.

3.3 Models

The algorithms applied in this research study are as follows

3.3.1 Prophet

the prophet model is a model which has been proposed by Facebook. It performers really well with time-series data which comprises of daily, weekly and hourly data which should at least consist data of a few months for better predictions [new 1]. A reasonable number of missed findings or large outliers, historical pattern shifts, such as product launches or logging changes and patterns that are non-linear growth curves, where a trend reaches a natural limit or saturates, are major holidays that occur at irregular intervals that are known in advance [new1].The Prophet tries to make it easy to use, the model is sufficiently advanced to provide useful observations and findings.

3.3.2 MLP

Multilayer perceptron is an ANN which comprises of neurons which are arranged in layers. MLP is made up of three different layers first is the input layer the second is the output layer and the third is the hidden layer. The reason for choosing MLP as one of the techniques used in this study was the simplicity of such techniques being implemented. In contrast to more complicated approaches, MLP is often known to provide high-quality models, thus keeping the training time relatively low. MLP is based on the measurement of neuron values in the current layer as an activated total of weighted neuron outputs in the previous layer linked to the neuron [new 2].

3.3.3 LSTM

LSTM is a RNN which is widely used for prediction where there is timeseries data. It also consists of three different gate the input the output and the foregate gate. The past data remains in the hidden state [9]. The activation function used here is the ReLu as it is the most common and also is really good at performance. The LSTM is known to give really accurate results when it comes to timeseries prediction. This approach here was used in the prediction of covid-19 cases.

3.4 Performance Evaluation

The models were evaluated based on their MAPE value and the MAPE accuracy the Prophet model was evaluated on the basis of the R2 score as well. Mean Absolute Percentage Error is used in the comparison of the performance between all the three models applied. MAPE is one of the common measures which is used to measure in forecasting the lower the values the better the prediction [new 1].

4 Implementation

4.1 EDA

As a part of the Exploratory Data Analysis there were various transformation performed like new columns were introduced with the help of the existing information the data was transformed with the help of the melt function. To show what the data comprises of various visualisations were done. Two separate data frames were created one for the whole world and another just for Ireland.

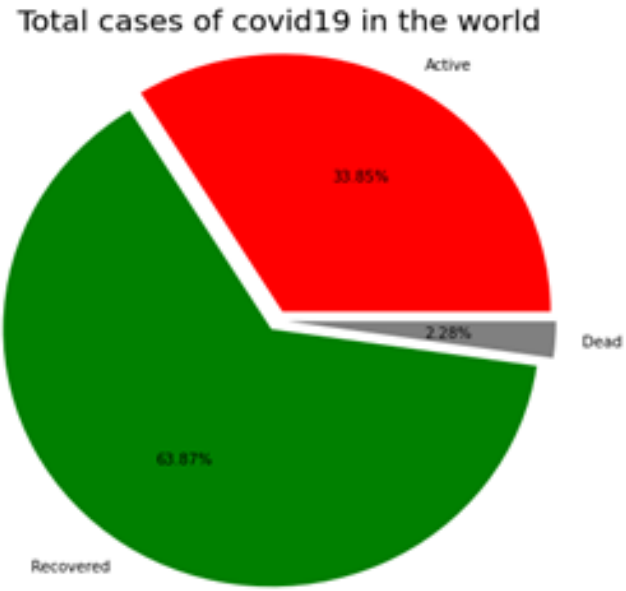

Figure 2: Pie diagram of the world covid-19 data
This pie diagram shows the total cases of Covid-19 in the world the percentage of Deaths, Active and Recovered.

According to the figure 2 we can tell that the number of people which have recovered(63.87 percent) from the disease(33.85 percent) is higher than the number of active cases and that the death rate is really low at 2.28 percent

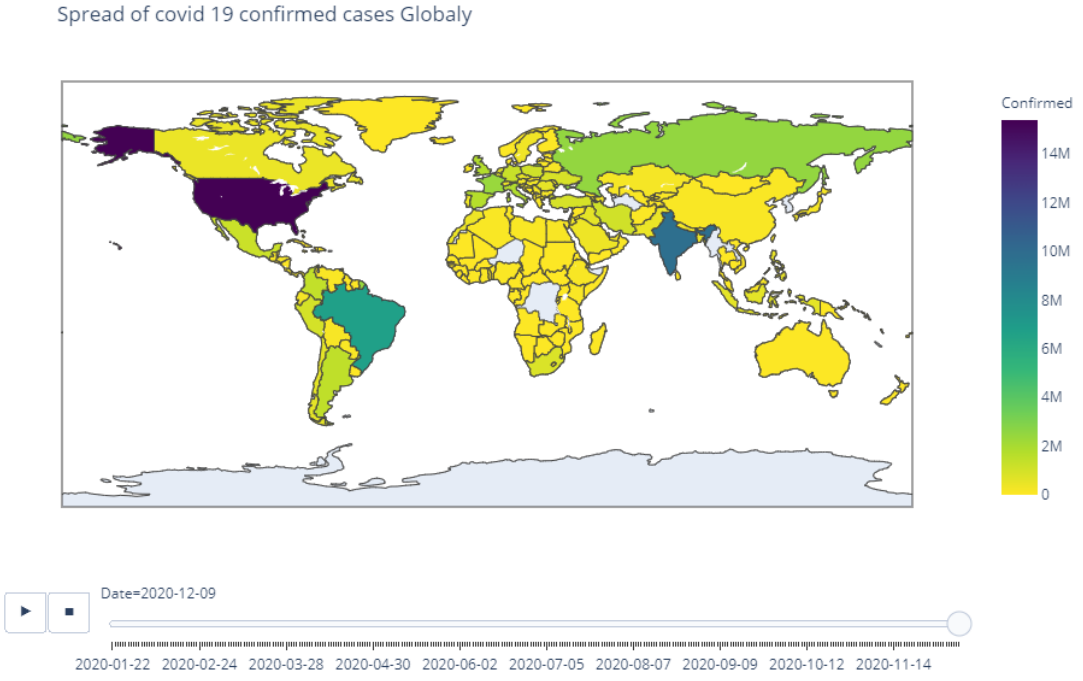

Figure 3: Choropleth map which shows the number of confirmed cases Globally

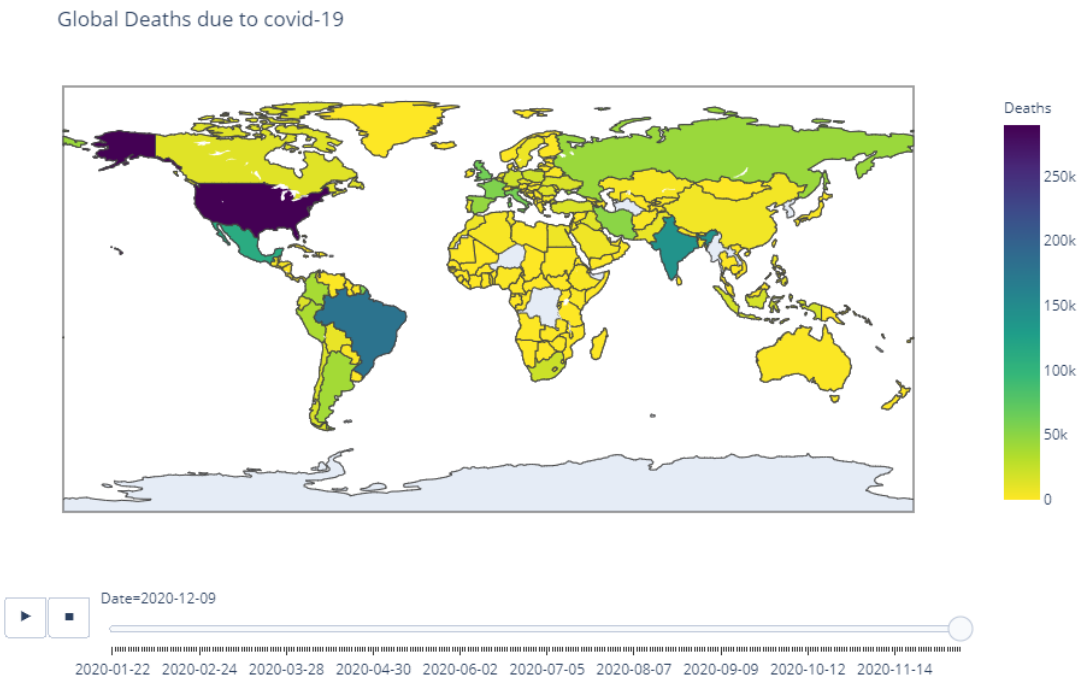

Figure 4: Choropleth map which shows the number of Deaths Globally

The figure 3 shows the confirmed cases of the countries and the figure 4 shows the deaths according to the countries. It is a time-series graph with a scrolling option according to the dates it also it ids colour scaled lighter colour being the lowest affected and darker colour being the most affected.

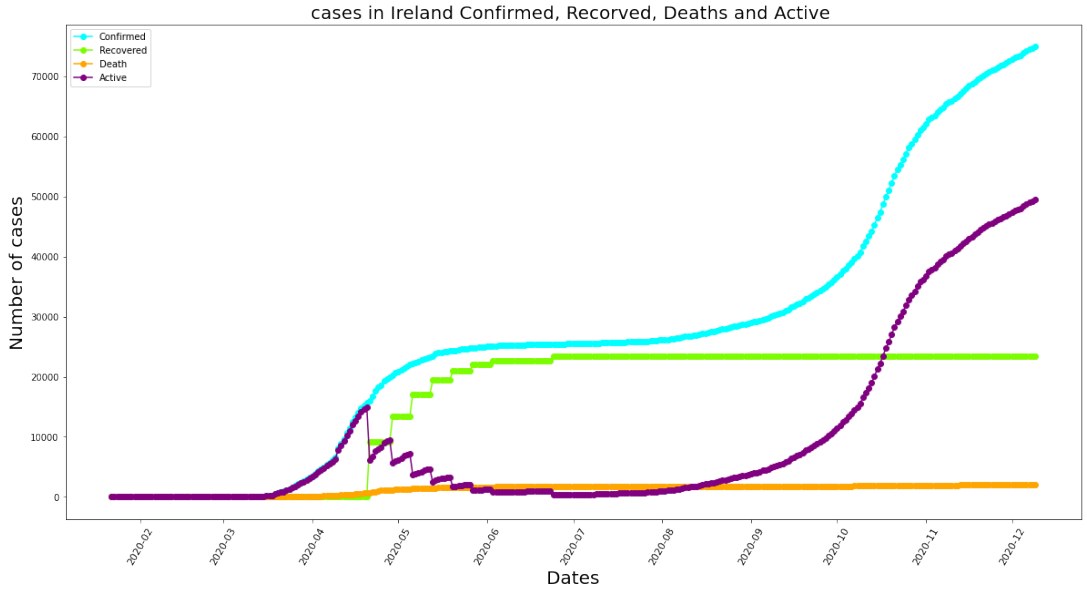

Figure 5: Ireland's Covid-19 Cases Data

In this figure 5 Ireland's data of covid cases which consists of the confirmed, recovered, dead and active are represented according to their dates by this graph we can say that as

the number of cases rises the number of active cases also rises which means the recovery rate is low as shown in the graph the recovered is gone flat after time there is very limited recovery after june.

4.2 Model Implementation

The implementation of the models were done on Jupyter notebook environment by Anaconda. Python was used to implement the models various libraries were imported and packages were installed.In this section the models which were implemented will be discussed. Packages like the statsmodels, sklearn, fbprophet, datetime, keras.

- Facebook's Prophet: To implement the prophet model the fbprophet package has been imported. A new data frame has been created which contains the date and values to implement the model the number of days to predict that is the period is given another data frame has been created which contains the future values the future values .predict() function is used and the output of the future data frame has been passed as the input which gives the future values.

- MLP: In this paper we perform Multilayer Perceptron with one hidden layer model. after splitting the data into test and train we scale the data with the help of MinMaxScaler() which has been imported from the package sklearn we create an instance with the use of TimeseriesGenerator() in which we set the batch size and the length which is the number of steps we impoet the model Sequential() and the different layers. early stopping has being used and actual loss and validation loss have been plotted finally a new data frame has been created which shows the predicted the actual and the additional forecast for the future days.

- LSTM: For the LSTM model the keras package has been used. LSTM and the other libreries have been imported. Pandas was used in the date time conversion and timeseries generator was imported from keras. The number of days n-input and the number of days to predict n-steps were set. The validation set was created and then used in the timeseries generator from tenserflow early stopping was imported and applied to stop the training if there was no improvement in the training.

5 Evaluation

5.1 Evaluating the Prophet Model

This model has been evaluated on the basis of the r2 score and the MAPE value for both the number of confirmed cases and the number of deaths FOR "Confirmed" MAPE = 8.151203314201725 and the r2 score = 0.7410487570227484 FOR "Dead" MAPE = 2.1322274771230485 and the r2 score = 0.9667686498788048

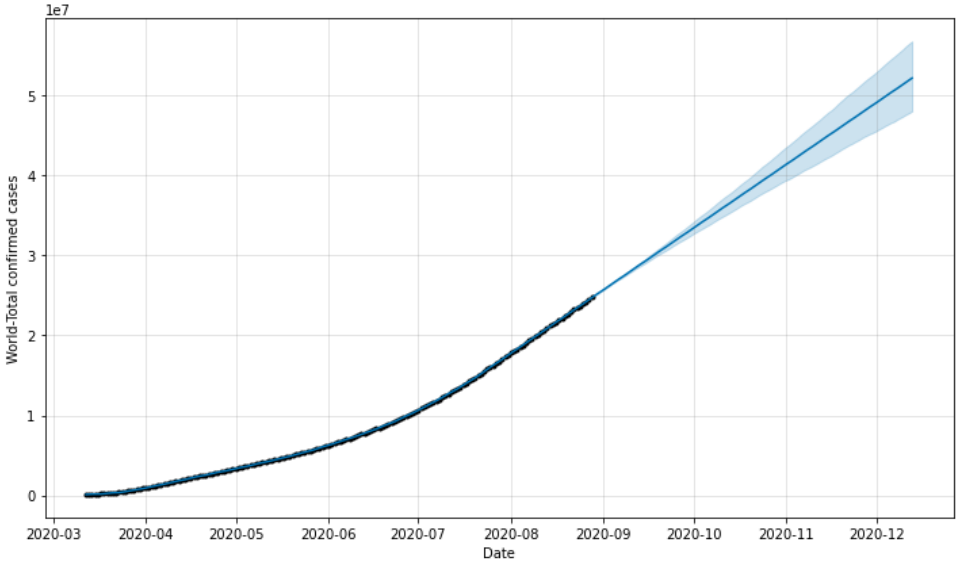

Figure 6: Confirmed Cases Prophet prediction

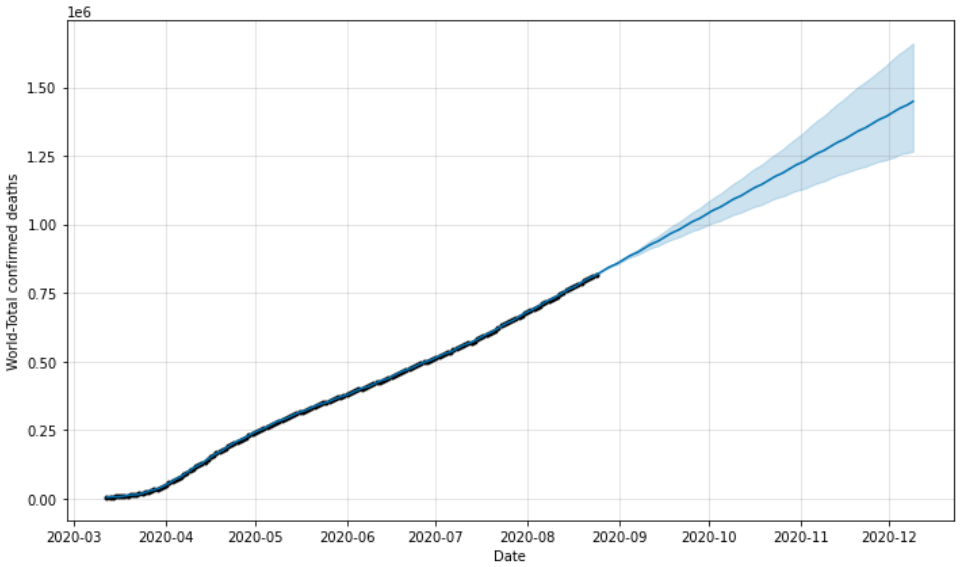

Figure 7: Confirmed Deaths Prophet prediction

5.2 Evaluating the Multilayer Perceptron Model

The MLP model was evaluated based on the mean absolute percentage error that is the MAPE value and the accuracy of the mape= 0.10325138342004382 which means that the model is just 0.1 percent of error and is 99 percent accurate.

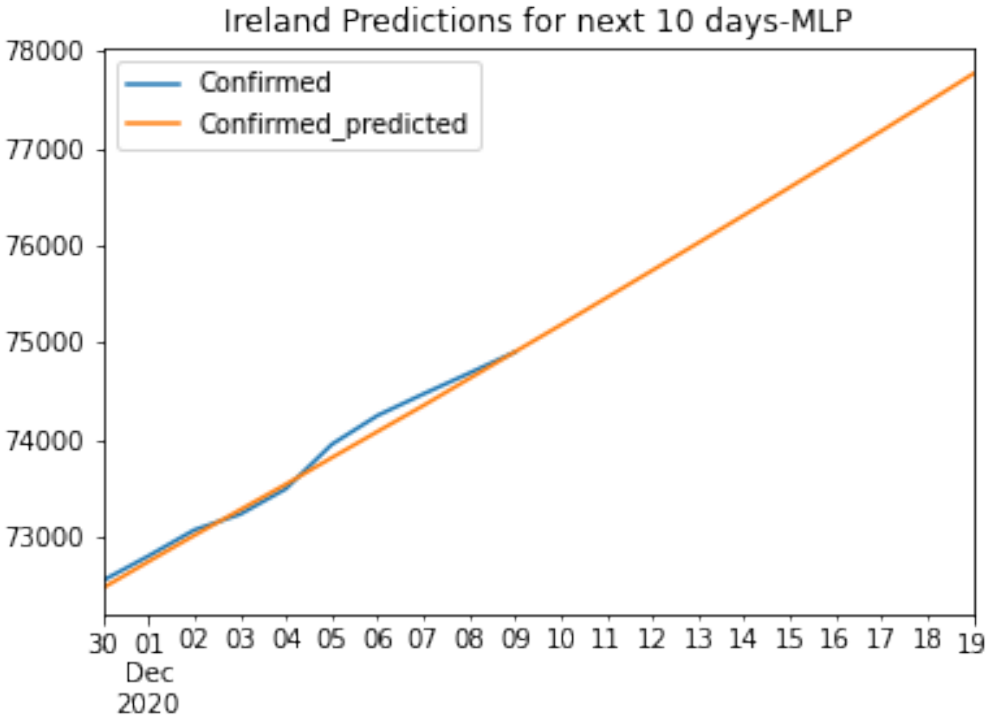

Figure 8: Confirmed Cases of the previous and next 10 days MLP

Comparing the actual values with the confirmed predicted values and also showing the prediction for the future days.

Dates	Confirmed	Confirmed predicted
2020-12-05	73948	73810.982498
2020-12-06	74246	74080.781527
2020-12-07	74468	74348.548171
2020-12-08	74682	74626.175324
2020-12-09	74900	74902.174847
2020-12-10	NaN	75179.319740
2020-12-11	NaN	75460.374389
2020-12-12	NaN	75742.428006
2020-12-13	NaN	76025.471981
2020-12-14	NaN	76311.073659

Table 1 MLP Predictions

5.3 Evaluating the LSTM Model

The LSTM model was evaluated based on the Mape value and the accuracy to predict the model MAPE= 0.12573520838004137 percent and the model accuracy =99.9987426479162 percent.

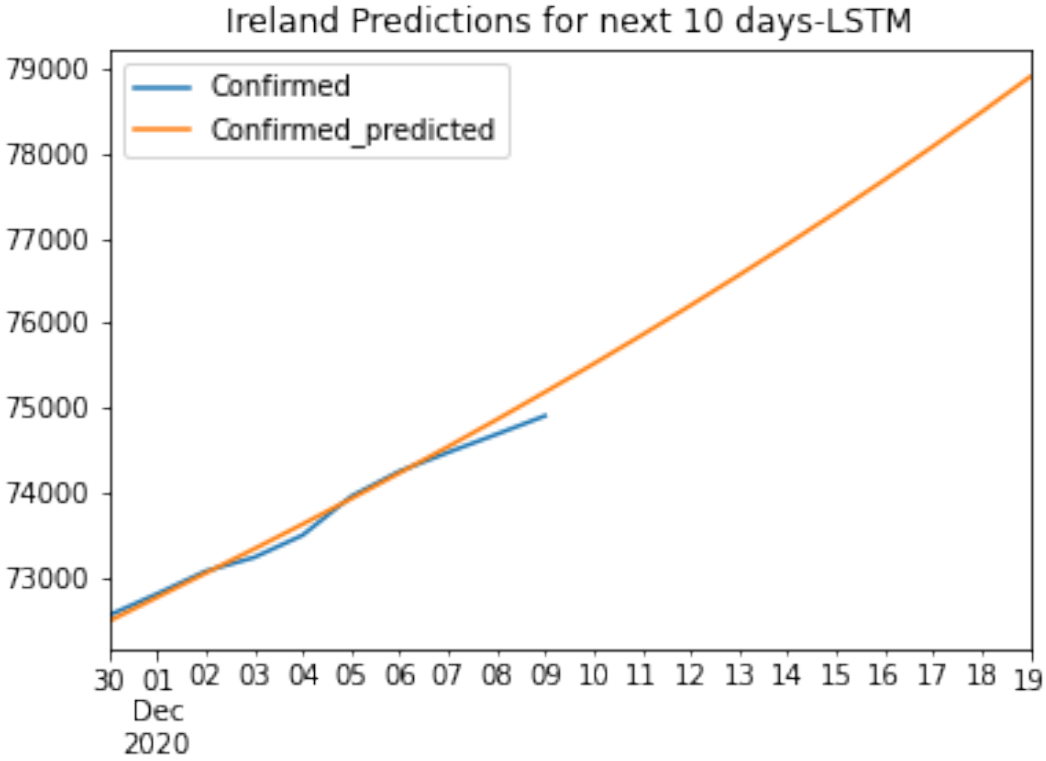

Figure 9: Confirmed Cases Prophet prediction

Dates	Confirmed	Confirmed predicted
2020-12-05	73948	73920.679585
2020-12-06	74246	74226.071176
2020-12-07	74468	74535.630877
2020-12-08	74682	74854.258802
2020-12-09	74900	75180.852640
2020-12-10	NaN	75512.837464
2020-12-11	NaN	75854.321101
2020-12-12	NaN	76203.460628
2020-12-13	NaN	76560.798586
2020-12-14	NaN	76926.886131

Table 2 LSTM Predictions

5.4 Discussion

The above experiments gave us predictions both for the world as well as Ireland specific it was a process that had various trial and errors as the dataset used is a live dataset and has been updating every single day it was challenging to fit the model. As there was an easy access to the data thanks to the Johns Hopkins University which as it was verry important for the experiment to have a good dataset. As well all know that this pandemic is going to be for a long time which was announced by the world health organization as well. Therefore, it was important to forecast and have a prior knowledge and letting people know how it would affect and how bad the impact is going to be. As the health department and the government are taking certain measures which would impact the forecasts.

There were various factors which were which could have been implemented to increase or help in forecasting like the few months in the future and so on but they couldn't be included dude to lesser availability of resources and time. Things like the health care centers the availability of the beds in the area or country people under treatment, people tested positive which are under home quarantine, etc. This paper will help to forecast the number of cases in Ireland and also the world.

	MLP	LSTM
MAPE	0.10325138342004382	0.12573520838004137
Accuracy	99.9989674861658	99.9987426479162

Table 3 Comparison between models on Ireland data

In the above given Table 3 we can see that the models have been evaluated on the bases of the MAPE value and model accuracy. Both the models MLP and LSTM performed well in predicting the number of cases of Ireland as well as forecasting, the lower the error the better the model.

Prophet	Confirmed	deaths
MAPE	8.151203314201725	2.1322274771230485
r2 score	0.7410487570227484	0.9667686498788048

Table 4 Prophet prediction for the world data

As you see in the table 4 is an evaluation of the prophet model we have evaluated this model on the basis of the MAPE and the r2 score for the confirmed as well as the deaths prediction. MAPE value is less than 10 so it is said to be a good prediction and also the deaths had a lower mape then the conformed it means the model was more accurate in predicting the deaths than the confirmed cases. also we can see the r2 score which shows that the prophet model was 74 percent accurate in predicting the confirmed cases and 96 percent accurate in predicting the deaths.

6 Conclusion and Future Work

In this research we have implemented various machine learning models for time-series forecasting to predict the future and understand what is to come, As covid-19 has affected the lives of mostly everyone in the world it is better to be prepared from this research of various models like prophet model for the world prediction and MLP and LSTM for the prediction of cases in Ireland we can say that all the models performed well in the prediction with MLP being the best model having the MAPE value of 0.10325138342004382.

As the vaccine distribution starts and people start getting the vaccine there can be various other factors that can be considered as there was many research done on covid-19 based on classification models. In this study we can try to predict future values or the take a totally different approach things like effects of lockdown, healthcare availability etc can be worked with.

7 Acknowledgment

I would like to thank Prof. Noel Cosgrave for his continuous support and encouragement thought this research work.

References

Arora, P., Kumar, H. and Panigrahi, B. K. (2020). Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of india, *Chaos, Solitons & Fractals* **139**: 110017.
URL: *https://doi.org/10.1016/j.chaos.2020.110017*

Car, Z., Baressi Šegota, S., Anđelić, N., Lorencin, I. and Mrzljak, V. (2020). Modeling the spread of covid-19 infection using a multilayer perceptron., *Computational Mathematical Methods in Medicine* pp. 1 – 10.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=a9hAN=143493576site=eds-livescope=sitecustid=ncirlib*

Chaurasia, V. and Pal, S. (2020). COVID-19 pandemic: ARIMA and regression model-based worldwide death cases predictions, *SN Computer Science* **1**(5).
URL: *https://doi.org/10.1007/s42979-020-00298-6*

Dong, E., Du, H. and Gardner, L. (2020). An interactive web-based dashboard to track covid-19 in real time., *The Lancet. Infectious diseases* **20**(5): 533 – 534.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=cmedmAN=32087114site=eds-livescope=sitecustid=ncirlib*

Erraissi, A., Azouazi, M., Belangour, A. and Banane, M. (2020). Machine learning model to predict the number of cases contaminated by COVID-19.
URL: *https://doi.org/10.21203/rs.3.rs-23330/v1*

James, U. and Pierpaolo, D. (2020). Brexit election : Forecasting a conservative party victory through the pound using arima and facebook's prophet., *Web Intelligence, Mining and Semantics* pp. 123 – 128.

URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip, cookie,shibdb=edscmaAN=edscma.3405967site=eds-livescope=sitecustid=ncirlib*

Marian, S., Pawel, S. and Marlena, S. (2020). From the index case to global spread: the global mobility based modelling of the covid-19 pandemic implies higher infection rate and lower detection ratio than current estimates., *PeerJ* **8**: e9548.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip, cookie,shibdb=edsdojAN=edsdoj.21022aae988444f98111c825ccfe7e2dsite=eds-livescope=sitecustid=ncirlib*

Menni, C., Valdes, A. M., Freidin, M. B., Sudre, C. H., Nguyen, L. H., Drew, D. A., Ganesh, S., Varsavsky, T., Cardoso, M. J., Moustafa, J. S. E.-S., Visconti, A., Hysi, P., Bowyer, R. C. E., Mangino, M., Falchi, M., Wolf, J., Ourselin, S., Chan, A. T., Steves, C. J. and Spector, T. D. (2020). Real-time tracking of self-reported symptoms to predict potential COVID-19, *Nature Medicine* **26**(7): 1037–1040.
URL: *https://doi.org/10.1038/s41591-020-0916-2*

Moorthy, V., Restrepo, A. M. H., Preziosi, M.-P. and Swaminathan, S. (2020). Data sharing for novel coronavirus (COVID-19), *Bulletin of the World Health Organization* **98**(3): 150–150.
URL: *https://doi.org/10.2471/blt.20.251561*

Pano-Azucena, A. D., Tlelo-Cuautle, E. and Tan, S. (2018). Electronic system for chaotic time series prediction associated to human disease., *2018 IEEE International Conference on Healthcare Informatics (ICHI), Healthcare Informatics (ICHI), 2018 IEEE International Conference on, ICHI* pp. 323 – 327.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip, cookie,shibdb=edseeeAN=edseee.8419377site=eds-livescope=sitecustid=ncirlib*

Roda, W. C., Varughese, M. B., Han, D. and Li, M. Y. (2020). Why is it difficult to accurately predict the covid-19 epidemic?., *Infectious Disease Modelling* **5**: 271 – 281.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie, shibdb=edselpAN=S2468042720300075site=eds-livescope=sitecustid=ncirlib*

Swapnarekha, H., Behera, H. S., Nayak, J. and Naik, B. (2020). Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review, *Chaos, Solitons & Fractals* **138**: 109947.
URL: *https://doi.org/10.1016/j.chaos.2020.109947*

Tran, M. (2008). Barack obama to be america's first black president.
URL: *https://www.theguardian.com/world/2008/nov/05/uselections20083*

Wieczorek, M., Siłka, J. and Woźniak, M. (2020). Neural network powered covid-19 spread forecasting model., *Chaos, Solitons and Fractals: the interdisciplinary journal of Nonlinear Science, and Nonequilibrium and Complex Phenomena* **140**.
URL: *http://search.ebscohost.com/login.aspx?direct=trueAuthType=ip, cookie,shibdb=edselpAN=S0960077920305993site=eds-livescope=sitecustid=ncirlib*

Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., Xie, Y., Liu, B., Wang, Z., Zhang, S., Wang, Y., Zhong, N. and He, J. (2020).

Modified SEIR and AI prediction of the epidemics trend of COVID-19 in china under public health interventions, *Journal of Thoracic Disease* **12**(3): 165–174.
URL: *https://doi.org/10.21037/jtd.2020.02.64*