ML Capstone Project for Forecasting Covid-19 Situation KSA

The covid-19 pandemic has been started in Saudi Arabia in March 2020, which causes to close the schools, universities, and even the markets and shops in Saudi Arabia. The COVID-19 is a virous comes from large family of viruses that are known to infectious disease that causes illness in the respiratory system in humans. It can multiply rapidly as it is spreading from person to person. It was first identified in December 2019 in China. It has influenced day life. This pandemic has influenced numerous of peoples, it causes a large number of people infected by covid-19 which makes them either sick or death consequence the spread of this disease. In this research, use effective models that solve problems in order to predict confirmed cases, recovered cases, and deaths in KSA. Further, look to some ways how can decrease their chances of getting the virus.

Vision 2030

The COVID-19 pandemic has disrupted daily services due to community-level mitigation measures taken by many countries. Global efforts have focused heavily on social distancing and, in many cases, completely shutting down cities and the country as the only solutions to containing the epidemic. These mitigation measures necessitated the use of technology to preserve jobs in all aspects of life. When using data analysis, we can predict what will happen in the future and thus Saudi Arabia has also contributed to the global data gathering of MERS-Covid-19 information.

During the current COVID-19 pandemic, the Kingdom of Saudi Arabia has been proactive in implementing the measures to contain the disease and working to meet the needs and demands of the community in a very short time.

Bearing in mind the importance of rapid and timely sharing of digital data for policy actions, which has also been emphasized by the World Health Organization (WHO).

The Kingdom of Saudi Arabia has achieved Vision 2030 in the health transformation, which is e-health through the creation of applications that serve the community, which in turn helped facilitate access to vaccines and early detection of cases by facilitating and simplifying the work of examining Covid-19Predicting cases before and after vaccinations, so we used Machine Learning.

The Dataset of Saudi Arabia Coronavirus Disease (COVID-19)

First, we chose the dataset that can help in prediction how is the pandemic daily situation in Saudi Arabia started from March 2020 from King Abdullah Petroleum Studies and Research Center (KAPSRC). The dataset called Saudi Arabia Coronavirus disease (COVID-19) situation which contain 531,350 records and seven columns are presenting the dates, the cases type, and total cases per day in each city for all the regions in Saudi Arabia. The types of cases in the dataset are the Cases which is the new people who are affected by COVID-19, the Active which is present all people who are affected by COVID-19 and they need a medical supervision, the Recoveries, and the Mortalities. The dataset is presented in the following figure.

	Daily / Cumulative	Indicator	Date	Event	City	Region	Cases (pe
1	Daily	Cases	February 1, 2022		Al Asyah	Al Qaseem	7
2	Daily	Cases	February 1, 2022		Khubash	Najran	5
3	Daily	Cases	February 1, 2022		'Uqlat As Suqur	Al Qaseem	1
4	Daily	Cases	February 1, 2022		Makkah Al Mukarramah	Makkah Al Mukarramah	122
5	Daily	Cases	February 1, 2022		Ar Riyad	Ar Riyad	1,408
6	Daily	Cases	February 1, 2022		Bishah	Aseer	7
7	Daily	Cases	February 1, 2022		Al Mandaq	Al Bahah	5
8	Daily	Cases	February 1, 2022		Al Ha'It	Hail	5
9	Daily	Cases	February 1, 2022		Al Bijadiyyah	Ar Riyad	5
10	Daily	Cases	February 1, 2022		'Uyun Al Jiwa'	Al Qaseem	4
11	Daily	Cases	February 1, 2022		Al Qaysumah	Eastern Region	16
12	Daily	Cases	February 1, 2022		Al Bada'I'	Al Qaseem	12
13	Daily	Cases	February 1, 2022		Al Ghat	Ar Riyad	2

Figure 1: The Dataset

After choosing the dataset the Data Ninjas starts reading it in the Jupyter Notebook by using the Python language to work on the programing. They start with cleaning the dataset from the nulls in the Event column and inserting two new columns are Year and Month from the Date column to assist in visualization the dataset. The Data Ninjas team create two plots to present the data, they are present the following.

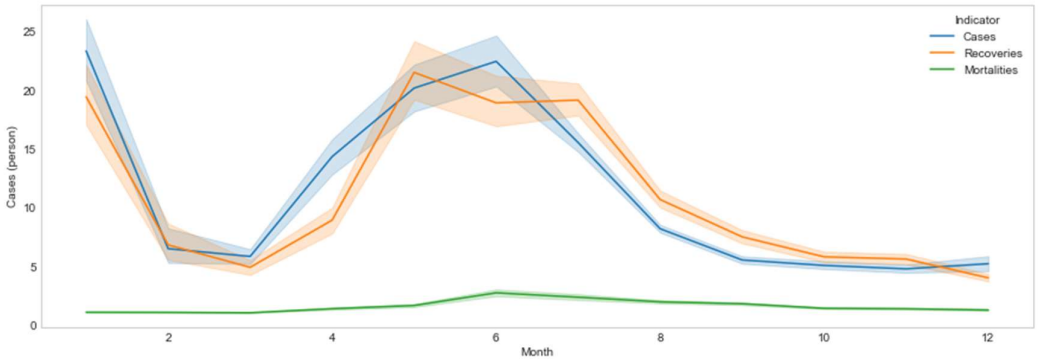

Figure 2: Data Visualization 1

The plot presents the increase and decrease of the cases types in each month based on the dataset.

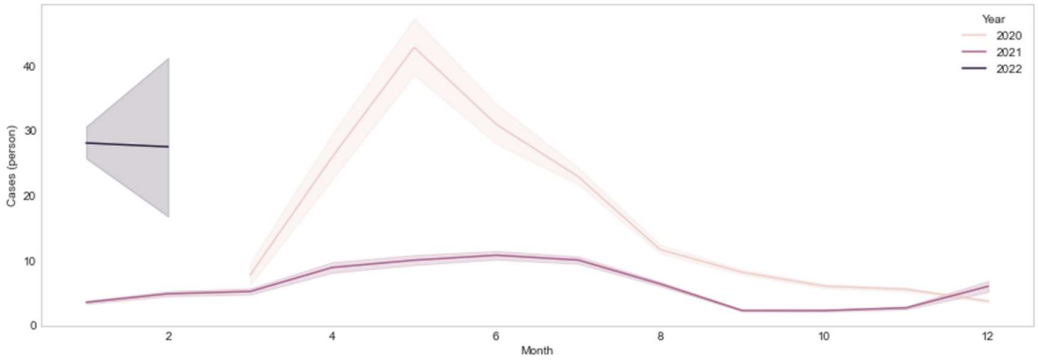

Figure 3: Data Visualization 2

The second plot presents the cases situations in general for each year started from 2020 to 2022.

Later the Data Ninjas filtered the regions by deleting the totals from the region column because the totals consider as a repetition for all the real results of the

regions. Then, they split the Daily OR Cumulative column into two data frames for each, after that they filter the Daily data frame for three regions are Al-Riyadh Region, Eastern Region, and Makkah Al Mukarramah to become a three different data frames.

The Data Preprocessing Techniques

The preprocessing techniques are starts after the cleaning and EDA the data. At the beginning they start with the Al-Riyadh Region data frame by rolling it to be able to handle the outliners. After that, they split the Al-Riyadh Region data frame into train and test. Where the train is about 80 percent of the data frame and the test is about 20 percent of the data frame. The process of splitting the Al-Riyadh Region data frame into train and test, it has been prepared again for the Eastern Region, and Makkah Al Mukarramah.

Machine Learning Model

The Data Ninjas has chosen the Prophet model. Where it is introduced and designed by Taylor and Letham. The Prophet model for the analysis and forecast of time-series data which is available in Python and R14. It adopts a Bayesian-based curve fitting method to smooth and forecast time-series data. To implement the model, they need to create a model and define it after that make fitting to the data frame, the train, and teat. Then later they have to identify the parameters of the model. The forecasting came after identifying model's parameters. In forecasting the Data Ninjas have to specify the prediction period. The last step of implementing the Prophet model is to do the model evaluation based on mean absolute error.

The Result of Implementing the Prophet Model to the regions:

1. Al-Riyadh Region

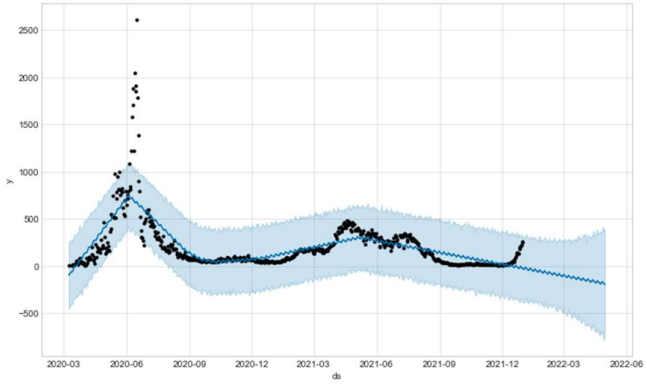

Figure 4: The Forecasting Result for Al-Riyadh Region

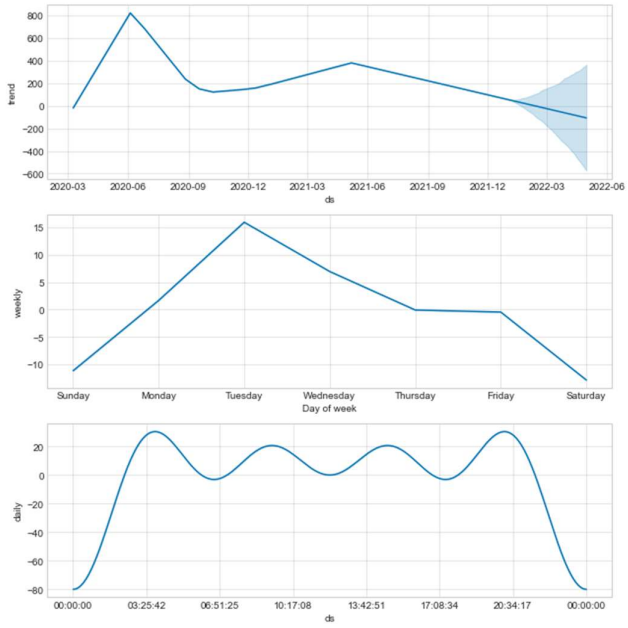

Figure 5: The Components of Al-Ryadh Region

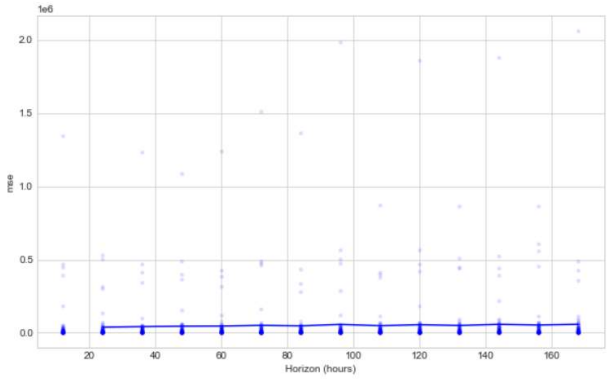

Figure 6: The Model Evaluation Result

2. Eastern Region

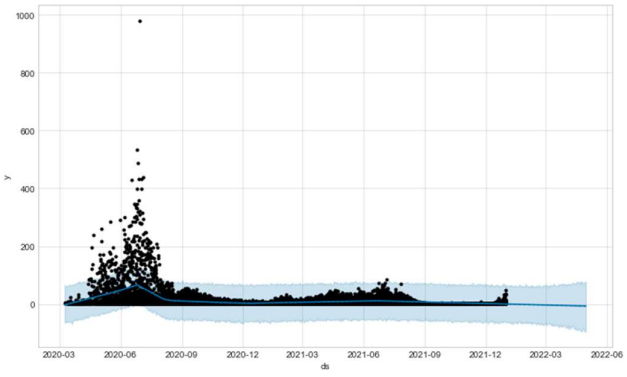

Figure 7: The Forecasting Result for Eastern Region

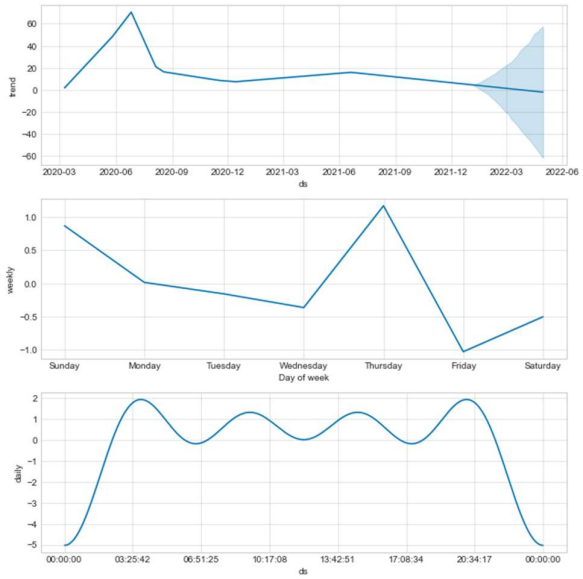

Figure 8: The Component of Eastern Region

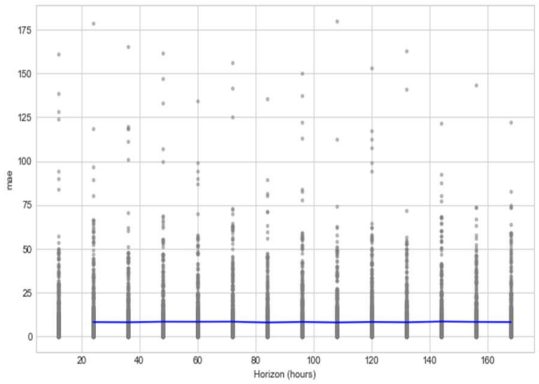

Figure 9: The Model Evaluation Result Eastern Region

3. Makkah Al Mukarramah

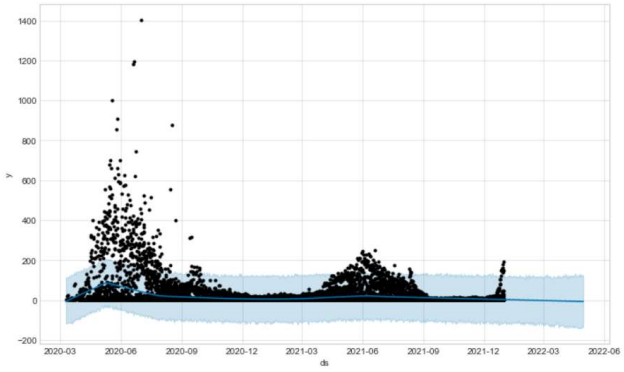

Figure 10: The Forecasting Result for Makkah Al-Mukarramah

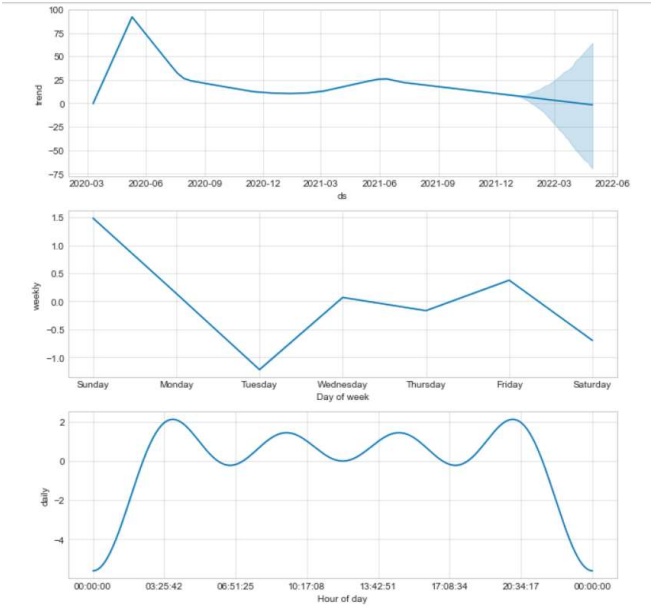

Figure 11: The Components of Makkah Al Mukarramah

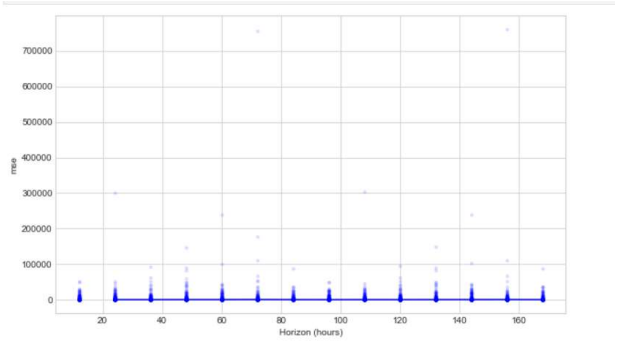

Figure 12: The Model Evaluation Result Makkah AlMukarrmah

The Dashboard

We created dashboard by using dash plotly to make a visualization for cases of Covid-19 over 3 years in each region and we make 3 plots for forecasting data that we extracted from applying the model in 3 deferent regions.

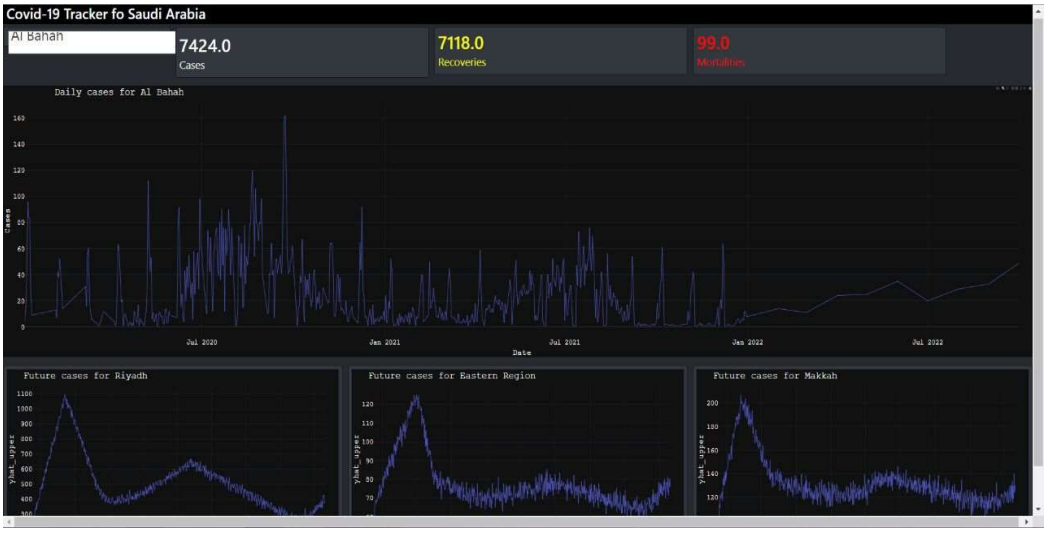

Figure 13: Dashboard

In conclusion, we can say that the forecasting model for Covid 19 cases helps the Kingdom of Saudi Arabia in developing appropriate plans to reduce the number of Covid 19 cases. The team Data Ninjas are planning to implement the model to all the other regions Saudi Arabia.

Rescores

1. *Saudi Arabia Coronavirus disease (COVID-19) situation*
https://datasource.kapsarc.org/explore/dataset/saudi-arabia-coronavirus-disease-covid-19-situation/

2. *Digital Response During the COVID-19 Pandemic in Saudi Arabia*
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7473704/

3. *Trend Analysis and Forecast of daily reported incidence of hand*
Trend analysis and forecast of daily reported incidence of hand, foot and mouth disease in Hubei, China by Prophet model | Scientific Reports (nature.com)