Analysing Covid-19 Total , Active, Recovered cases and Deaths in Pakistan using Time Series Decomposition

Zulfiqar Tariq
Department of Computer Science
Riphah International University
Lahore, Pakistan
zulfiqartariqburmi@gmail.com

Adan Khan Niazi
Department of Computer Science
Riphah International University
Lahore, Pakistan
adankhanniazi4@gmail.com

Khadija Ilyas
Department of Computer Science
Riphah International University
Lahore, Pakistan
ikhadija645@gmail.com

Maira Kamran
Department of Computer Science
Superior University
Lahore, Pakistan
lanamarii7@gmail.com

Marium Malik
Department of Computer Science
Superior University
Lahore, Pakistan
maryammalick2@gmail.com

Muhammad Tahir Zaman
Department of Computer Science
Superior University
Lahore, Pakistan
tahirzee8@gmail.com

***Abstract*—Since the first case of covid-19 was reported in Wuhan china in December 2019 the disease has been declared as global pandemic in March by WHO in March 2020. The spread of this disease is resulting in high mortality rates in Pakistan. In this study time series decomposition technique has been used to analyze the four systematic components such as trends, levels, seasonality, and noise of series including deaths, recovered cases, active cases, and confirmed cases. Finally, data has been analyzed by assuming data to be both addictive and multiplicative. The decomposed time series components reflect an exponential increase in corona active cases.**

Keywords: Social media, text mining, sentiment analysis, business intelligence, competitive intelligence, competitive analytics

I. INTRODUCTION

In December 2019 china started to report a novel form of corona virus named SARS CoV-2 (stands for Severe Acute Respiratory Syndrome Coronavirus 2) which was later name covid-19 in Feb 2020 and was declared as a global pandemic in march 2020 [1]. The first official case of corona was reported to WHO (world health organization) on 31 December 2019. Initially 44 patients with pneumonia of unknown etiology were reported. [2]. Covid-19 is caused by the SARS C0V-2 which is highly contagious disease which effects human lungs and it can be fatal for people having weak immune systems [6] and for people who are already having a certain medical condition like diabetes, tuberculosis.

The first case of corona in Pakistan was confirmed on 26 Feb 2020 in Karachi, Sindh province by the ministry of health [3]. Since then the number of cases reported have been increasing. Outbreak of corona in Pakistan was due to several factors which mainly includes international travel from various countries including Iran and UK. The high numbers of cases in neighboring countries like China and Iran played a major role in increasing the cases. The major of initial confirmed cases had a travel history of Iran and UK. Till 31st May 2020 the total number of confirmed cases were 72,460 [4] .

Time series is collection of data points measure over specific time intervals. It can be define in a mathematical form as follows

$$F(x)= y \quad x=0,1,2,3,4,5$$

Time series can be discrete or continuous in nature. In discrete the values are measured at discrete point of time like daily temperature, daily sales while in continuous the values are measure at every point to time.

Time series comprises of four integral components which are trend, seasonality, cycle, and noise.

A time series can have a positive or a negative trend over a period. If the graph is increasing over a period, we say that the series has an increasing and if the graph in is decreasing over a period of time we say that trend in negative. In a population time series generally, the trend is positive because population of the world is increasing over time and similarly the trend is negative in series related to number of deaths due to a specific disease like polio.

Seasonality is the variation during a regular period. It means that you can predict them based on the past i.e. sale of bottle water increases during the summer because there is high consumption and drops during the winter, similarly sales of jackets increase in winter due to cold weather and drop in summer. The seasonality should not be associated with seasons, seasonality can be found in weekly or daily data i.e. weekly sales of food at a restaurant can have seasonality as sales tend to increase on weekends due to holidays.

Cycle is the variation in time series but its is not periodic and we cannot predict it. This component is usually caused by external forces acting on time series for example wars floods and famines.

Time series are observed in many fields like business, science, commerce, forex, and other industries. In retail a time series can be observed in terms of daily weekly and quarterly sales.

Time series have different model for prediction which are not limited to moving average, single exponential smoothing (SES), double exponential smoothing (DES), triple exponential smoothing (TES) and Decomposition model.

In moving average an average of datapoints is calculated over a period of time and based on the average future values are calculate and predictions are made. Moving average has many forms like cumulative moving average, weighted moving average and exponential moving average. All the models can be effective in respective cases.

II. LITERATURE REVIEW

A. Covid-19

Various Artificial intelligence techniques have been used to predict different aspects of pandemics in the past. There are various factors of a pandemic that can be studied and analyzed like mortality rate in a specific pandemic, how quickly it will spread, how many people can be affected. Different techniques of AI can help to us to solve these problems.

Mohammad Pourhomayoun and Mahdi Shakibi came up with a predictive model based on AI algorithms to predict the mortality risk in patients having COVID-19.Their final set includes demographic features such as age, sex, province, country, age, travel history, general medical information such as comorbidities (diabetes, cardiovascular disease, and also patient symptoms such as chest pain, chills, colds, conjunctivitis, cough, diarrhea, discomfort, dizziness, dry cough, dyspnea, emesis, expectoration, eye irritation, fatigue, gasp, headache, lesions on chest radiographs, little sputum, malaise, muscle pain, myalgia, obnubilation, pneumonia, myelofibrosis, respiratory symptoms, rhinorrhea, somnolence, sputum, transient fatigue, weakness. They have used different machine learning algorithms such as Support Vector Machine (SVM), Neural Networks, Random Forest, Decision Tree, Logistic Regression, and K-Nearest Neighbor (KNN). According to their research ANNS gave the best performance. [1].

V. Prema and K. Uma Rao have devised a decomposition and TES(Triple exponential smoothing) model for predicting wind speeds in which they have tested on past data over a period of one year. The two models were compared with other forecasting models which are ARIMA, BPNN and persistence. All the models were tested with three 1 month,2 month and 4-month duration. They finally conclude that decomposition model gives better results compared to conventional forecasting methods [8].

Pasapitch Chujai, Nittaya Kerdprasop, and Kittisak Kerdprasop used ARIMA(auto regressive integrated moving average) and ARMA (auto regressive moving average) models for forecasting household electric consumption, therefore concluding that ARIMA model is best for finding the most suitable forecasting period in monthly and quarterly[9].

A research carried out by members of Cairo University, Faculty of Computers and Artificial Intelligence, Cairo, Egypt presented various methods to predict daily forecast of COVID-19 [5],their study presents a comparison of day level forecasting models that helps to predict daily spread and compares models moving average, weighted moving average and single exponentials smoothing. The model successfully predicts the confirmed cases on based of three error measuring techniques which include mean absolute deviation (MAD), mean square error (MSE), root mean square error (RMSE) and mean absolute percentage error (MAPE). Different AI based simulation models have been suggested to control and help reduce the impact of Covid-19.

III. RESEARCH QUESTIONS

This study focusses on decomposition of time series to analyze each individual component and try to answer following questions.

What is the trend in number of deaths, recovered, active and confirmed cases?
Is there any seasonality in deaths, recovered, active and confirmed cases?

IV. METHODS

A. Dataset

In this paper the data has collected from official website of government of Pakistan. The dataset comprises of daily cumulative confirmed cases, deaths, active and recovered cases for the months of April, May, and June in the form of CSV. For decomposition of time series, the data should not contain null or 0 value for that the deaths for initials dates were set to 1. Data is preprocessed using python library pandas.

B. Time Series

Out of all the model this paper focus on decomposition model of time series. The decomposition model breaks down the time series systematic components trend, seasonality, level, and noise, and helps to give insights about the time series. Time series decomposition has been used to predict and extract insights in data.

C. Additive Model of Time Series

The additive components are added as follows

$$f(x) = (\text{trend} + \text{level} + \text{seasonality})\ \text{systematic components} + (\text{noise})\ \text{non-systematic component} \quad (1)$$

A time series is said to follow an addictive model when changes over time a consistent and made by the same amount.

D. Multiplicative Model of Time series

The multiplicative components are multiplied as follows

$$f(x) = (\text{trend} * \text{level} * \text{seasonality})\ \text{systematic components} * (\text{noise})\ \text{non-systematic component}$$

The time series in said to follow multiplicative model when it nonlinear such as quadratic or exponential and tend to increase or decrease over time.

E. Decomposition

To answer our research question, I decomposed the time series in systematic components to find out the level's trends and seasonality. The goal was to decompose each

timer series including deaths, recovered cases, active cases, and confirmed cases. The first step was to decompose death time series. The death time series was decomposed using both additive and multiplicative model using python seasonal decompose library. Three other time series were also decomposed the method mentioned prior. The results were plotted for each time series including recovered cases, active cases, and confirmed cases. The plots were also drawn for original observations. Figure 1 shows observed deaths plotted against daily recorded values. By observation we can see that the graph is exponentially increasing which indicates that the death time series will fit multiplicative model very well. Figure 2, Figure 3, and Figure 4 shows active, confirmed and daily recovered cases, respectively.

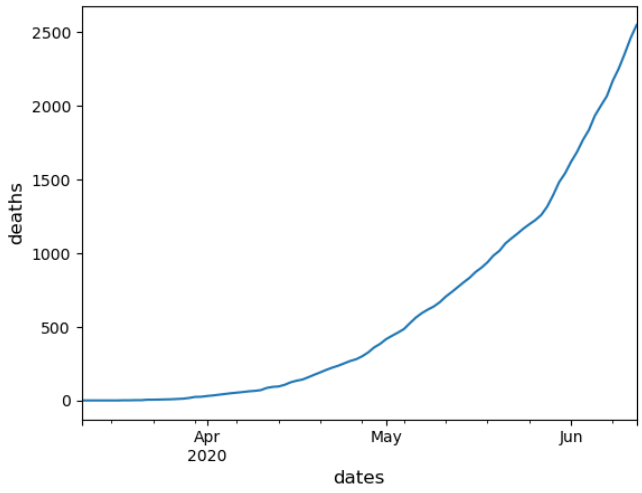

Figure 1 Daily observed Deaths

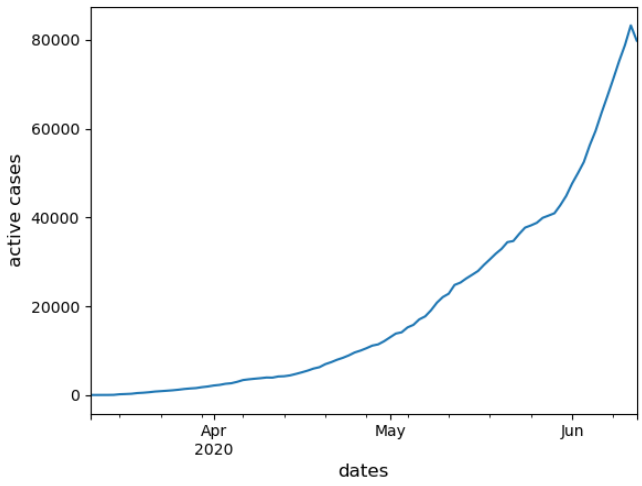

Figure 2 Daily active cases

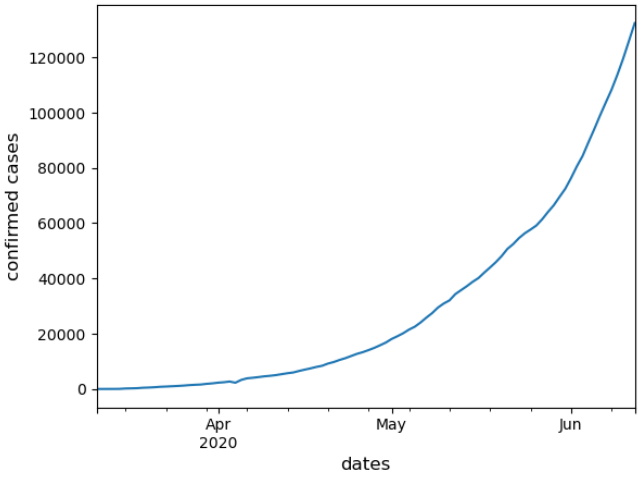

Figure 3 Daily confirmed Cases

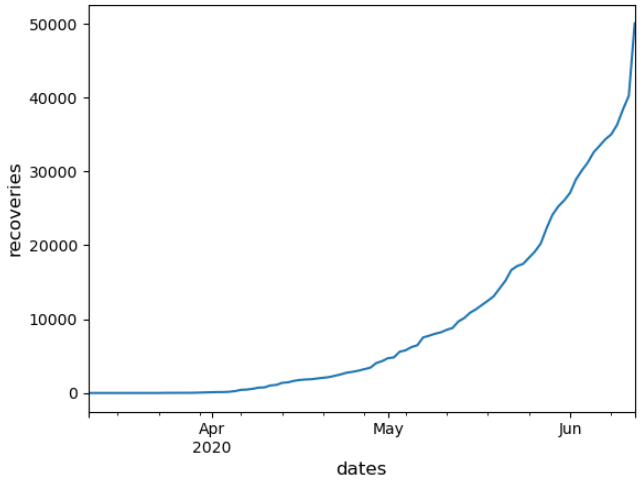

Figure 4 Daily recovered cases

V. RESULTS

All the observed time series graphs show that they are increasing exponentially. Figure 5 shows the additive death time series. There is an exponentially increasing trend on the death time series and there exists a seasonality in the death time series

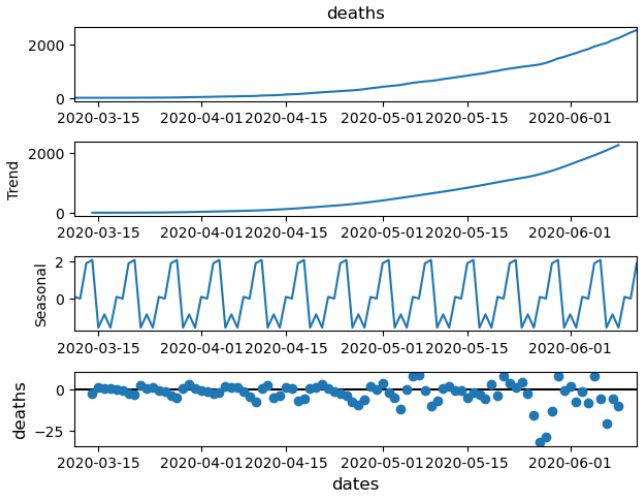

Figure 5. Decomposed additive deaths

Figure 6 shows the multiplicative decomposition of death time series and the results are almost the same in nature the trend is exponentially increasing, and seasonality is also present in the time series.

Figure 7 shows the additive decomposition of active cases time series and figure 8 shows multiplicative decomposition. There active case has an exponentially increasing trend and seasonality exists in this time series. The trend is exponential in all the models and times series and seasonality exists in all the time series.

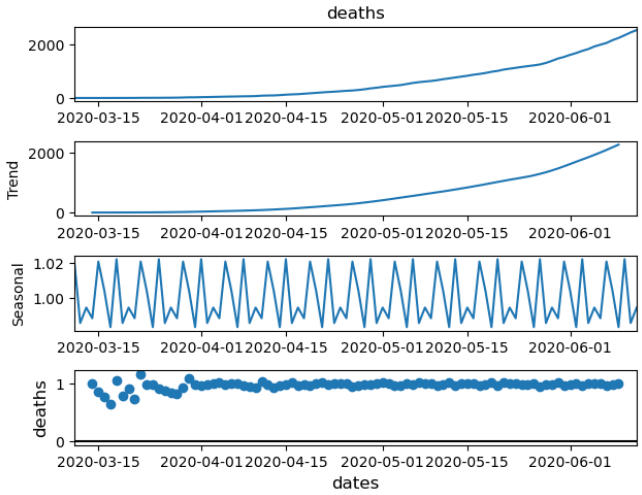

Figure 6. Decomposed multiplicative deaths

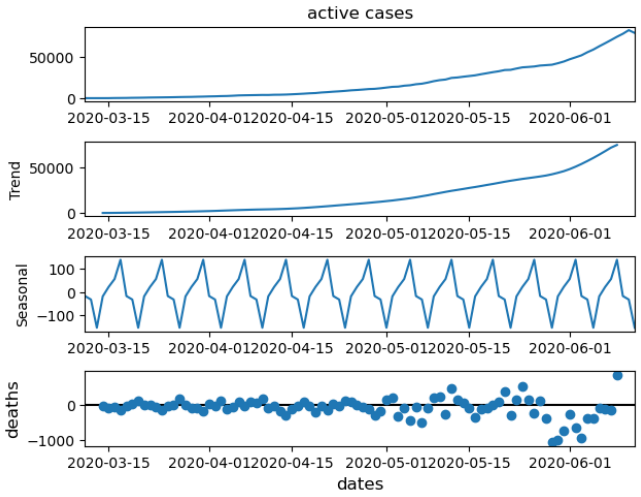

Figure 7 Decomposed additive active cases

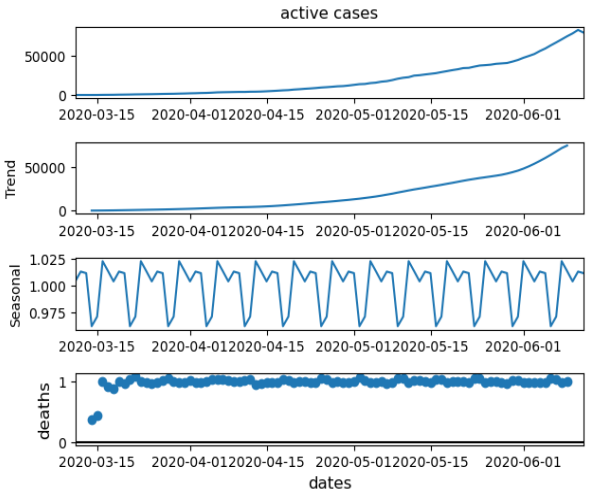

Figure 8 Decomposed multiplicative active cases

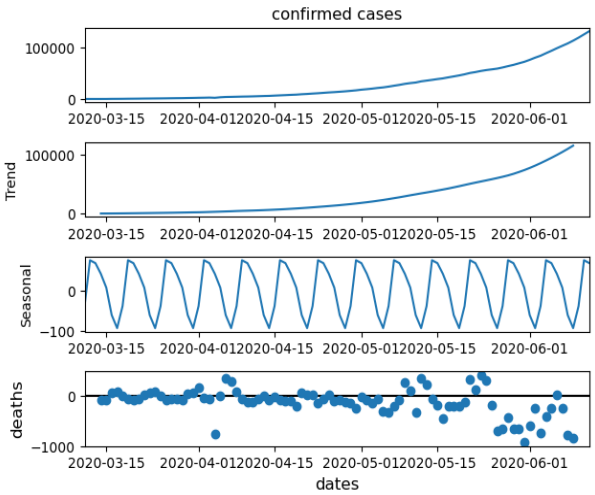

Figure 9 Decomposed additive confirmed cases

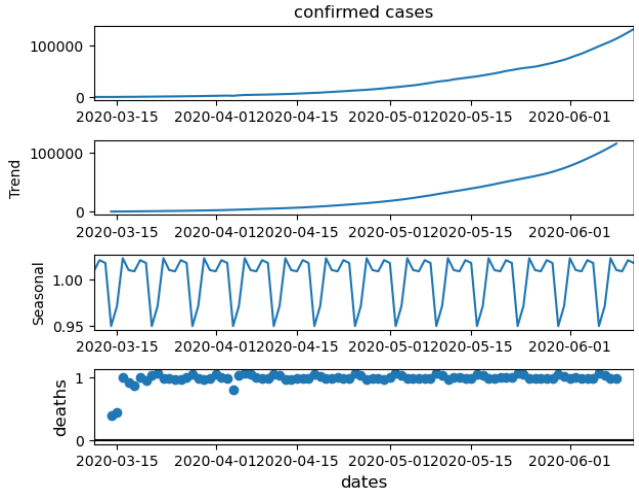

Figure 10 Decomposed multiplicative confirmed cases

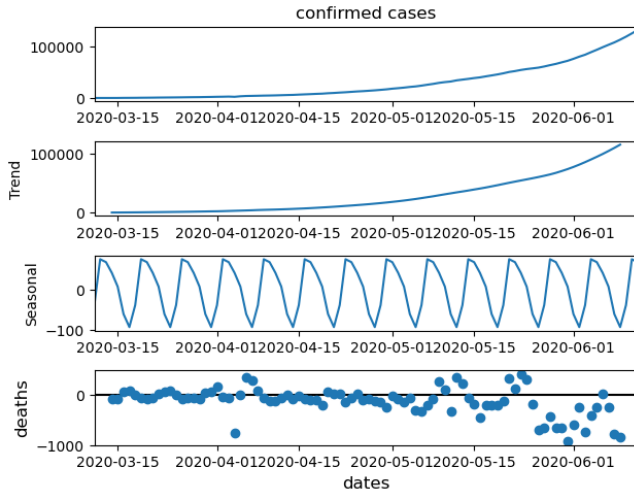

Figure 11 Decomposed Additive Confirmed cases

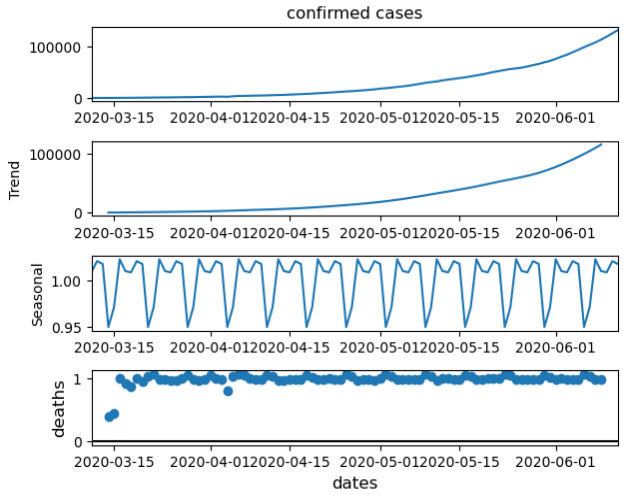

Figure 12 Decomposed multiplicative confirmed cases

VI. Limitations

Since COVID-19 is a novel disease so there is not ample data available. This study has used government of Pakistan website which was the most authentic data source available. The data available was lacking features. Only four type of data was available, considering additional features like number of tests might have showed different trends.

VII. Conclusion

The purpose of this study was just to decompose the time series. In this study I was able to decompose and analysis its key components. The further study can by using time series forecasting models like SRIMA and Moving averages. Also forecast performance measures such as Mean Forecast Error (MFE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) can be used to predict errors in prediction.

References

[1] Mohammad Pourhomayoun and Mahdi Shakibi “Predicting Mortality Risk in Patients with COVID-19 Using Artificial Intelligence to Help Medical Decision-Making” medRxiv.org, The preprint server for Health Sciences

[2] “Pneumonia of unknown cause – China” World Health Organization website https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/.

[3] A. Waris , U. K. Ata , M. Ali1 , A. Asmat and A. Baset “COVID-19 outbreak: current scenario of Pakistan”. Science direct New Microbes and New Infections Volume 35, May 2020, 100681

[4] K. Haytham H. Elmousalami and Aboul Ella Hassanien “Day Level Forecasting for Coronavirus Disease (COVID-19) Spread: Analysis, Modeling and Recommendations” arXiv preprint arXiv:2003.07778

[5] Farzad Taghizadeh-Hesarya and Hassan Akbar “The powerful immune system against powerful COVID-19: A hypothesis” Preprints, 2020 (2020), Article 2020040101, 10.20944/preprints202004.0101.v1

[6] Christine S.M. Currie, John W. Fowler and Kathy Kotiadis “How simulation modelling can help reduce the impact of COVID-19” Journal of Simulation. 2020 doi: 10.1080/17477778.2020.1751570

[7] Ratnadip Adhikari , R. K. Agrawal “An Introductory Study on Time Series Modeling and Forecasting” arXiv preprint arXiv:1302.6613, 2013

[8] V. Prema and K. Uma Rao “Time series decomposition model for accurate wind speed forecast” Renewables Wind Water & Solar, 2015. 2(1): p. 1-11.

[9] Pasapitch Chujai, Nittaya Kerdprasop, and Kittisak Kerdprasop “Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models” IMECS 2013;2013, [I].

[10] Andrew T. Jebb1, Louis Tay1, Wei Wang and Qiming Huang“Time series analysis for psychological research: examining and forecasting change” Frontiers Psychol., vol. 6, p. 727, Jun. 2015.