Trends in Undergraduate Research (2021) 4(1): e8-17
https://doi.org/10.33736/tur.3201.2021

Visualisation System of COVID-19 Data in Malaysia

REHMAN ULLAH KHAN*, NOR SYAZA SYAMIMI, CLADIA SIMBUT ANAK MAMBANG, IVY ANAK THOMAS, & TZI NI WEE

Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
*Corresponding author: rehmanphdar@gmail.com

ABSTRACT

Pandemics are highly unlikely events, therefore, we need a system to understand the statistics about the pandamic. Machine learning algorithms can analyse the data and then we can plan for handling the pandamic. To date, many people are suffering because of the lack of reliable information system. The problem is that there is no integrated system to use the data and plan for pandemic management to minimise social panic. This study aims to provide a system, using COVID-19 data as a sample to visualise and analyse cases, deaths, discharged ICU cases updates in Malaysia as a whole state wise of COVID-19 daily statistics. The results provide visualisation and case comparison among states in Malaysia to easily and quickly understand the situation. This will help and assist the management in decision-making.

Keywords: COVID-19, Decision support system, Disaster management, Panic, Pandemic

INTRODUCTION

The Coronavirus outbreak which began in Wuhan, China, in December, has expanded to every corner of the globe. Millions of people around the world have been sickened and died. With the pandemic affecting worldwide, it is crucial to look at the numbers to counter the outbreak. World Health Organization states that Coronavirus (COVID-19) is contagious caused by coronavirus (World Health Organization, 2020). This viral disease is caused by a recently discovered coronavirus mutation; SARS-CoV-2; a disease that causes respiratory infections in humans. This virus was brought to our attention through cases reported from Wuhan, China in December 2019. The earliest case of COVID-19 in Malaysia was reported on 24 January 2020.

COVID-19 virus is spread by droplets in a short distance and settles on surfaces when an infected person coughs or sneezes without covering their mouth and nose (Ministry of Health, 2021). There are effective ways of avoiding catching the virus, such as washing your hands regularly with soap and water, using alcohol-based hand sanitiser, avoid touching your face and eyes, maintaining a social distance of six feet apart, wearing two-ply face masks and staying up-to-date with COVID-19 virus news.

Alternative infection prevention and control measures can be applied in the healthcare environment such as wearing personal protective equipment to help prevent the spread of infectious diseases. People must meet proper standards of healthcare, the equipment must be worn correctly in the appropriate context (Ministry of Health, 2021). A study by MacIntyre and Hasanain (2020), wearing a mask may protect people from becoming infected and prevent transmission of infection from infected people because a surgical mask can filter three micrometres droplets. Hence, wearing a mask is important because new research on face coverings shows that the risk of infection to the wearer is decreased by 65 percent (Chu *et al*., 2020). Besides, during this COVID-19 crisis, it is necessary to curb this pandemic.

The need to build up a system to analyse and visualise the COVID-19 crisis as a step to counter the pandemic is crucial for the nation. The government can take further action and deliberate on timely planning to control the pandemic by implementing a plan for targeted high-risk states, high-risk individuals and various health

backgrounds. By using the programming tool, python, via an open-source web application Jupyter Notebook, we aim to visualise and analyse the cases, deaths, discharged cases, and current ICU status within a particular time in Malaysia or the specifications of each state with the COVID-19 outbreak daily statistics. Besides, the system also focuses on the total number cases in each state and visualises it at a specific time. Finally, the system also seeks to compare cases between states in Malaysia. With the following specific tasks, the system will target to complete each aspect to get the best results from the system.

The next section relates to studies that provide literature reviews in the field. Materials and methods explain the methodology of how the system was designed and developed. The last section is about the results and conclusion.

Related Studies

Time series prediction

Machine learning (ML) has recently emerged as one of the key computing technologies and is increasingly applied in day-to-day life and various industrial domains (Deparday, Gevaert, Molinario, Soden, & Balog-Way, 2019). ML is an artificial intelligence application that uses algorithms that work on characteristics of available data to make further predictions. In the era of emerging technologies such as unmanned aerial vehicles, internet of things, and satellite-based technology, the network is becoming more autonomous. Such systems require several local decisions to be made, such as bandwidth selection, data rate selection, power control, and user association to a base station. We can use ML algorithms to address these issues and lower human intervention in uncertain and stochastic environments.

Researchers are using different machine learning algorithms to detect or predict COVID-19 cases. One technique is time series, which uses information from past data, values, and patterns to predict future activity. Time series is a set or series of data points ordered in time whereby the independent variable would be the time, and the forecast for the future would be the goal for the time series. It is often modelled via stochastic process, Y(t), whereby in a forecasting setting, the method Y(t+h) would be used with what information is available during that particular time and setting (t) would be applied (Kerrigan *et al*., 2019). The difference between time series data and cross-sectional data is the fact that time-series data is collected from various points in time. In contrast, cross-sectional data would collect data at a single point in time. Some well-known examples of forecasting models used in time series would be ARIMA, TBATS, Prophet, LSTM, ANFIS, MNETAR, and GARCH. The application of time series can be found in various contexts such as daily weather temperature, allocation of resources, business planning, and stock price forecasting (Erica, 2021). Univariate is called when a time series data contains records of a single variable. When the dataset has more than one variable, it is multivariate data. Time series can be either continuous or discrete. The observations are calculated at every instance of time in an ongoing time series. However, a discrete-time series contains observations measured at distinct points of time (Adhikari & Agrawal, 2013).

LSTM network is a special kind of recurrent neural network (RNN) that could learn long-term dependencies proposed by Hochreiter and Schmidhuber (1997). LSTM is explicitly designed to avoid the long-term dependency problem. The architecture of this network is shown in Figure 1 (Franklin, 2018).

Bouhamed (2020) proposed deep learning nested sequence prediction models with LSTM to monitor the infection and recovery process of the Covid-19 cases continuously. This research used COVID data from 79 countries. This model is capable of controlling the Covid-19 pandemic by making the right decisions. Yudistira (2020) compared the LSTM model with the precedent model of RNN. This study involved 100 countries data cases from 22 Jan 2020 until 1 May 2020. LSTM was concluded as a promising tool to predict the COVID-19 pandemics by learning from big data and can potentially predict future outbreaks.