

Predictive Analytics to Forecast the next upcoming Pandemics

*<https://www.researchgate.net/publication/347453332>

Shivanarayana Dhulipala
Department of Computing
Bournemouth University
Bournemouth, United Kingdom
s5229871@bournemouth.ac.uk

Abstract—Predictive analytics uses data mining techniques, statistics from past research, and AI to analyse data and predict the future that may or not occur. This proposed research paper uses predictive analysis to analyse past pandemics and forecast on the next one to come. The predictions and information found in research sources can identify the pandemic's coming risks and opportunities. The history of pandemics will be discussed in this paper to identify linked factors and relationships among pandemics that already occurred [1]. In this case, we will also use both structured and unstructured data to analyse the pandemics. Predictive data changes the behaviours of the researcher. They become anxious about the analysis results, and to some point, it impacts them to know how to behave if the outcome comes as predicted or vice versa. The researcher is likely to benefit a lot from the predictive analysis if they are proactive and forward-looking to the results. The results of predictive analysis are involved, and they later yield recommendations towards the proposed prediction.

Keywords—Data mining; Statistics; AI; Pandemics; Predictive analytics; Structured data; Unstructured data.

I. INTRODUCTION

Several organisations approve predictive analytics on the collection, analysis, and prediction of future happenings. It supports various processes including data mining, machine learning, AI, statistics and modelling. With the use of predictive analytics, the researcher discovers patterns among data and how they are related. Among these, data and information are current and historical data of the happenings that are being predicted. After detecting the proposed prediction trends, the researcher can utilise the data they acquire. The forecast carries with the possibility of a risk or an opportunity [2]. Predictive analysis has several advantages, including making the future look accurate and reliable to the researcher. Predictive analysis can help the researcher make correct decisions and save money based on the prediction. The researcher also knows whether to employ weak or strong precautionary measures to predict the risks that occur during the forecast. The researcher can prepare for the opportunities that arise during the prediction, and they can alert other people on the oncoming prediction and, in this case, a pandemic. Pandemics can be controlled in

so many ways if, during the first breakout, scientists identify a way to prevent the spread. With predictive analytics, scientists can be alerted and commence research to curb the pandemic or reduce the risk's effects. They can research and develop ways to live with the pandemic without increasing the spread before the pandemic occurs. The use of predictive analysis on forecasting the next upcoming pandemics can be used as an advantage to the people worldwide and help them be ready for the risk and opportunities that come with it.

II. LITERATURE REVIEW

With the increase in medical faculties' development globally, we have improved in handling significant diseases, which have rendered the world or some parts of the world helpless to cope with the conditions. For instance, in the year 1918, the world experienced a deadly influenza pandemic, which was a severe disease that saw the world's population decrease by a considerable margin of roughly 50 million people due to its harmful causes. However, with the advancement of technology, the virus would have been more tackled than the year it happened. Creating a model known as a prediction model can be a significant step in decision-making. Another step that would help predict pandemic diseases in different parts of the world works on big data analytics. Massive data storage compartments or repositories containing the geospatial and wellness information give significant numerical findings on surveillance and give important leeways into the disease-causing factors. Utilising these substantial information sources and assessments to analyse the factors causing these pandemic diseases can help create reliable prevention and disease control measures for these diseases. The spatial data usually has the volume and velocity required for substantial data sources and has extra information regarding the disease's location. Modern information sources for medicine-related issues, e.g., medicine claims, location pinned tweets, phone call history records, have penetrated infectious infections epidemiology research as sources of information to reflect the essential transmissible infections surveillance. Another critical factor is the processing of electronic medical records to increase the outcomes of

the predictive analytics; this is because readmission to the hospital is expensive but at the same time can have hugely been prevented. To accomplish these, scientists have utilised predictive analysis to create samples that can reduce patients' hospital readmission's extensive economic and social effects. In recent years, the African continent with several countries experienced a wave of Ebola diseases, which deprived the lands of a massive number of individuals who succumbed to the disease. A year later, after the containment of the virus, it was observed that the countries compliance to the worldwide precautions helped lower the risk of a person's body to the disease-causing virus, which is usually transmitted via blood and also the fluids found in the body. The change in people's behaviour helped lower the virus to the point of containment of the virus. The spread of the virus was predicted through a novel. Epidemic trend analysis, predictions are paramount for helping the health system reduce the population's adverse effects. Public health control and Artificial intelligence are used to curb the results and calculate the number of people infected by the virus to help predict vaccines' statistics to be developed. In the past, the predictions were created based on technology's underdevelopment and the medical system.

The earlier forecasts of the pandemics in history were conducted by predictive analytics without the advanced help of artificial intelligence and big data. Artificial intelligence and data mining have helped the researcher alert the people to help them be ready for pandemic opportunities and risks. The health care sector will also develop vaccines and medicine that will help in reducing deaths and infections. It had a significant impact on the health system globally, economically, and the economic system. The behavioural change was enhanced by the universal precautions set by health organisations. An example of such heavily hit countries includes Sierra Leone, an immense number of 14000 individuals who succumbed to the disease. Worse still is the fact that the corpse of the people who succumb to the disease is highly infectious, and thus any exposure to the people would alleviate the spread of the disease at a rate double the standard rate. The deceased's families and the community had to compromise their traditions to enable a proper burial to prevent any infected body contact. These standards have also incorporated in the fight against coronavirus/covid 19, which is currently affecting the whole world. The government has to ensure dignified areas, which set to bury the people, and these places were fumigated to prevent any spread of the virus. These are among the several measures taken to curb the pandemic diseases which affect the world. To ensure all this is done accurately, the predictive analysis is used to collect data, compile results, and allocate the requirements to the different sectors, which call for attention.

III. HISTORY OF PANDEMICS

The outbreak of infectious diseases has held several effects on society throughout history. These effects have gone ahead to shape the social, economic, political aspects of humans. The impact has been felt over the centuries, even after the pandemic has ended. Pandemics have shaped several aspects

of human life and how they behave. It has led to people choosing prevention over cure by immunisation, coming up with prevention measures and epidemiology, and scientists developing modern medicines to help in the current situation and reduce the future effects [3]. The pandemics have influenced the way people react and respond. Globally, people understand the impact of the outbreak of these diseases; therefore, the ability to deal with them through prevention, control, and treatment has improved [4]. During these pandemics, history is altered, and the civilisation of humanity is halted. Dating back to 5000 years ago, a pandemic stroked through the northeast of China. Everyone was affected regardless of their age. The pandemic happened so fast that a proper burial for the infected was not carried out, and the corpse was stuffed in a building and burnt. Another pandemic hit the Sparta people around 430BC. The scientists to this age still make their prediction on what precisely the disease was. Healthy people were said to start feeling hot in their heads, inflammation of the eyes, and pain in their throats or tongue. The pandemic lasted five years, and the cause was said to be the crowding of Athens during the war between Athens and Sparta. This pandemic is said to have killed approximately 100,000 people. In the 16th century, the American people experienced several Eurasian diseases, which were first introduced by European explorers. These diseases included smallpox, and research shows that they almost swept out a whole population in the western hemisphere. The pandemic paralysed the Aztec and the Inca civilisation, and the Spanish took over the territories of both cities since the armies were affected by the pandemic. In the late '80s, a pandemic that started in Russia killed over one million people globally. It has spread easily through modern industrialisation, and new means of transport [5]. Between 1957-1958, an Asian flu that started in china killed 1.1million people globally. The flu was a combination of avian flu viruses, and it killed more people in the United States. HIV/AIDS was a global pandemic before medication was developed in the '90s. It came from an infected chimpanzee in the West of Africa and spread throughout the world, claiming the lives of 35million people before a way of controlling it. The current Covid-19 pandemic that started in Wuhan, China, has killed 1.26million people globally. Scientists are learning how to prevent, control, or treat it. Scientists indicate that the virus originated from bats. The pandemic has changed the world economy and people's lifestyles.

IV. METHODOLOGY

The time series forecasting model will be the most suitable approach as it majorly focuses on analyzing the past observations of any given random variable to aid in the development of a model that could best capture various existing relationships and their patterns, after which it is utilized in the prediction of future variables or instead values of the given random variable. This approach is specifically useful in two significant ways: when the available data generating process has limited knowledge or lacks an explanatory model that is effective enough in predicting variables to other explana-

tory variables. The time series forecasting model has various forecasting models that include Auto Regressive Integrated Moving Average (ARIMA), The Holt-Winters additive model, the Facebook's prophet, the deeper and the N-Beats model, in which this study will maximize on the Auto-Regressive Integrated Moving Average model. The Auto-Regressive Integrated Moving Average model is among the well-known and widely used family of time series due to its statistical properties, effective implementation of the well-known box methodology during its model training process, and expertise in the performance of exponential models for smoothing. This type of model often assumes a correlation that is linear in between the values offered however the time series model and their attempt in exploiting the various linear dependencies to extract local patterns while at the same time effectively removing noise from the data that is considered to be high frequency. As a result of this approach, the interpret ability of high level is offered. Based on the model's assumptions, there is a clear comprehension between the independent and dependent variables. It is thereby enabling a more straightforward explanation, eventually allowing the researcher to comprehend the existing relationship between the current state as a function of endogenous variables and any influencing inputs outside the state of this series that might have exogenous variables. Auto-Regressive Integrated Moving Average models can also perform automated model selection, thus maximizing the resulting accuracy of their prediction.

V. PROS

With the use of predictive analytics, scientists can predict an oncoming pandemic and alert the people. They can also develop modern medicine and preventive measures to reduce the risk and maximise the prediction opportunities' effects. The researchers can acquire more knowledge about current and past pandemics, which can help develop future pandemics technologies and techniques. The experience of a pandemic that is coming is crucial since the affected people prioritise their activities and methods of reducing the risk. Various people can be prepared for a pandemic, including adoption, taking no action, which can be dangerous and have expertise.

VI. CONS

There are several cons of using predictive analytics include using data that is incomplete. Missing information or values could limit usability. The section of data that is not available could be the one that is to stop the effects of the prediction or reduce them to a sustainable amount. If one uses data from a survey, there is a likelihood of having inaccurate data because people don't always give correct answers. Data from different sources, if mixed up, may vary in format and quality depending based on the data. Such data requires a lot of processing before sending it to the cumulative database [1].

VII. CONCLUSION

Due to the frequency and pandemics patterns over the recent years, there is a probability of a pandemic hitting again after

the covid-19 pandemic. By using predictive analytics, some ways are relationships between the various pandemics. Predictive analytics can be used to help prepare people before the pandemics hit hard. It can also help scientists develop a modernized way of controlling, preventing, or treating the oncoming pandemics. People can also learn from past pandemics and how they affected social, economic, and political behaviours. Researchers and scientists can predict approximately when the next pandemic will occur using the acquired information. With the help of big data and predictive analytics, citizens of the world are becoming more aware and smarter on what to do in case of an outbreak. Non-profit organizations and universities use predictive analytics to discover the most common cause of pandemics and their hotspots. Researchers are coming up with new ways where scientists can get information on the pathogens in one database. An example of such a database is Enhanced Infectious Disease Database (EID2). This database helps bring together all the researched information about pandemics. Scientists can anticipate and prepare a model to predict the next upcoming pandemics [6]. Some diseases recur over the years, and scientists have become more aware of reducing their effects. Most of the pandemics are first discovered in animals and then transmitted to humans. Predictive analysis helps identify the various animals previously determined to spread infectious diseases to humans.

ACKNOWLEDGEMENT

The author wishes to thank Dr Philips Davies and Dr Festus Adedoyin in the Department of Computing, Bournemouth University, UK for their continuous help and support.

REFERENCES

- [1] Bari, A., Chaouchi, M. and Jung, T., 2020. The Limitations Of The Data In Predictive Analytics - Dummies. [online] dummies. Available at: <https://www.dummies.com/programming/big-data/data-science/the-limitations-of-the-data-in-predictive-analytics/>; [Accessed 9 November 2020].
- [2] Edwards, J., 2020. Predictive Analytics: Transforming Data Into Future Insights. [online] CIO. Available at: <https://www.cio.com/article/3273114/what-is-predictive-analytics-transforming-data-into-future-insights.html>; [Accessed 9 November 2020].
- [3] Huremovic, D., 2020. Huremovic D, editor. Psychiatry of Pandemics: a Mental Health Response to Infection OutbreakSpringer International Publishing 2019; 185 pages; ISBN978-3-030-15346-5 (e-book), ISBN978-3-030-15345-8 (softcover). Croatian Medical Journal, 61(3), pp.306-306..
- [4] K.Saunders-Hastings, P. and Krewski, D., 2016. Reviewing the History of Pandemic Influenza: Understanding Patterns of Emergence and Transmission. Pathogens, 5(4), p.66.
- [5] Jarus, O., 2020. 20 Of The Worst Epidemics And Pandemics In History. [online] livescience.com. Available at: <https://www.livescience.com/worst-epidemics-and-pandemics-in-history.html>; [Accessed 9 November 2020].
- [6] Locklear, M., 2020. With Big Data And Predictive Analytics, Scientists Are Getting Smarter About Outbreaks. [online] Discover Magazine. Available at: <https://www.discovermagazine.com/technology/with-big-data-and-predictive-analytics-scientists-are-getting-smarter-about>; [Accessed 9 November 2020].
- [7] Olalekan Adebimpe, W., 2016. Knowledge, Attitude, and Practice of Use of Safety Precautions Among Health Care Workers in a Nigerian Tertiary Hospital, 1 Year After the Ebola Virus Disease Epidemic. Annals of Global Health, 82(5), p.897.

- [8] Siegel, Eric. Predictive Analytics: the Power to Predict Who Will Click, Buy, Lie, or Die. Wiley, 2016.
- [9] Shmueli, Galit, and O. Koppius. "Predictive Analytics in Information Systems Research." SSRN Electronic Journal, 2010, doi:10.2139/ssrn.1606674.
- [10] Mahalle, Parikshit, et al. "Forecasting Models for Coronavirus (COVID-19): A Survey of the State-of-the-Art." 2020, doi:10.36227/techrxiv.12101547.v1.