Mark Angelo S. Panaguiton
BSCS-3A

Thesis I

Chpater 1. Introduction

Background of the study

Coronaviruses are a family of viruses that can cause illnesses such as the common cold, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). In 2019, a new coronavirus was identified as the cause of a disease outbreak that originated in China. The virus is known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease it causes is called coronavirus disease 2019 (COVID-19). In March 2020, the World Health Organization (WHO) declared the outbreak a pandemic. Public health groups, including the U.S. Centers for Disease Control and Prevention (CDC) and , are monitoring the pandemic and posting updates on their websites. These groups have also issued recommendations for preventing and treating the virus that causes **(Mayo Clinic, 2020)**.

As the world suffers from Covid-19, there are many individuals and organizations helping the world to mitigate this illness. One of these organizations involved in mitigating this illness are those people who practices analytics. In this study, machine learning models or algorithms will be used to predict the possibility of having a covid 19 based in the symptoms of a suspected individual.

In machine learning, a classifier is an algorithm that automatically sorts or categorizes data into one or more "classes." Targets, labels, and categories are all terms used to describe classes **(Dutta, 2022).** Example of classification models are the following: Support vector machine (SVM), Percepton, Logistic Regression, Naive Bayes, K-Nearest Neighbours, and Random Forest. Each model has their own accuracy rate, and the very core of this study is to evaluate classification models and use the model which has the greatest accuracy rate, in predicting the possiblity of having covid 19.

Statement of the problem

Covid 19 is a pandemic declared by the World Health Organization or commonly known as WHO. As this pandemic affects the whole world, many organizations takes the initiatives in mitigating this illness, however due to huge number of cases, the organizations are facing delays in diagnosing a suspected covid-19 patient. This pandemic continues to challenge medical systems worldwide in many aspects, including sharp increases in demands for hospital beds and critical shortages in medical equipment, while many healthcare workers have themselves been infected. Thus, the capacity for immediate clinical decisions and effective usage of healthcare resources is crucial. The most validated diagnosis test for COVID-19, using reverse transcriptase polymerase chain reaction (RT-PCR), has long been in shortage in developing countries. This contributes to increased infection rates and delays critical preventive measures **(Zoabi et al., 2021)**.

Objectives

In this study, the aim is to develop a machine learning classification model to predict the possibility of having covid 19 based on symptoms. There are many classification models that have high accuracy rate , however these models will undergo evaluation to get the model which has the greatest accuracy rate. In addtion, to help organizations in administering preventive measures to lessen the infection rates of covid-19.

Scope and Limitations

The scope of this study are the individuals who have diagnosed as covid-19 patients, wherein a dataset that contains the different symptoms of these patients will be used in order to procure the machine learning classfication moodel. The possible limitations of this study are the following: Time contraints and a valid dataset.

Significance of the study

The importance of this study is to give support to organizations in diagnosing possible covid-19 patients. Thus, will result to a much faster diagnosing rates and will prevent delays of preventive measures administered by health care professionals.