A FORECASTING TOOLKIT FOR EPIDEMIC SPREADING

D.Dandeniya

(199312M)

MSc in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

June 2023

A FORECASTING TOOLKIT FOR EPIDEMIC SPREADING

D.Dandeniya

(199312M)

This dissertation submitted in partial fulfillment of the requirements for the degree MSc in Computer Science specializing in Software Architecture

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

June 2023

DECLARATION

I declare that this is my own work, and this Project Report does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other University or institute of higher learning and to the best of my knowledge and belief. It does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic, or another medium. I retain the right to use this content in whole or part in future works.

.... 2023/06/29 ....

D.Dandeniya Date

I certify that the declaration above given by the candidate is true according to the best of my knowledge and that this project report can be accepted for evaluation of the Final Research.

.... 29/06/2023 ……………………...

Prof. Indika Perera Date

ABSTRACT

The study introduces a novel approach to predict the presence or absence of COVID-19 without the use of laboratory tests, kits, or equipment. It uses machine learning algorithms. Instead, the method relies on the symptoms experienced by a person to make predictions.

To achieve the best possible performance, the study applied seven supervised machine learning methodologies, including Naive Bayes, Logistic Regression, Random Forest, KNN, Gradient Boosting Classifier, Decision Tree, and Support Vector Machines. The algorithms were tested on the COVID 19 Symptoms and Presence Dataset in Kaggle. Then to improve their performance hyperparameter optimization was used.

The study found that the Gradient Boosting Classifier was the most effective algorithm, achieving an accuracy of 97.4%. The proposed method has the capacity to accurately discover the presence or absence of COVID-19, without requiring any devices or laboratory tests. This suggests that the method may offer a convenient and efficient way to quickly identify COVID-19 cases without relying on traditional laboratory-based testing methods.

The research suggests that machine learning algorithms can be useful tools for disease detection, even in the absence of laboratory tests. The proposed approach can help overcome the challenges of limited access to laboratory tests and kits, making disease detection more accessible and efficient.

ACKNOWLEDGEMENTS

My profound gratitude to my supervisor, Dr. Indika Perera, Head of the Department of Computer Science and Engineering, Faculty of Engineering, University of Moratuwa for his expertise and support, which have been instrumental in achieving this milestone.

I would like to extend my sincere gratitude to all the lecturers, my family and friends for their unwavering motivation and continuous support during this significant period of my life. Their encouragement, understanding, and presence have been invaluable in keeping me motivated and focused on my goals. I am deeply grateful for their unwavering support, which has been a driving force behind my success.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES