The Impacts of Covid 19

Hana Getu Gessess

501099010

CIND 820: Big Data Capstone Project

Dr. Tamer Abdou

Table of Contents

[Abstract 3](#_Toc409783205)

Abstract

The Covid-19 pandemic has posed unprecedented challenges globally, affecting millions of people and straining healthcare systems. This capstone project aims to provide insight into the pandemic by addressing a few questions to understand the trends of COVID-19 cases globally and looking at the number of confirmed cases, deaths, and recovered questions. The chosen theme revolves around leveraging a comprehensive dataset, available at https://www.kaggle.com/datasets/imdevskp/corona-virus-report/data, to derive meaningful conclusions.

The primary problem in this project is to address and look at the trends of the COVID-19 pandemic. The research questions delve into aspects such as the geographical spread of the virus, the confirmed number of cases and deaths, and the impact of various interventions. These research questions aim to enhance the understanding of the dynamics of the pandemic and help make informed decisions.

The dataset used for this project is sourced from Kaggle and encompasses a rich collection of COVID-19-related data. It includes information on confirmed cases, deaths, recoveries, testing rates, and various demographic factors. The dataset's depth and breadth make it an ideal resource for conducting a comprehensive analysis of the pandemic and addressing the research questions at hand.

Data analysis techniques will be employed to tackle the stated problem. These include classification algorithms to identify patterns and trends in the data, clustering to group regions with similar characteristics, and text mining to extract valuable insights from textual information. Model evaluation techniques will be applied to ensure the robustness and reliability of the results.

The tools proposed for this analysis include popular data science libraries and frameworks such as Pandas, NumPy, sci-kit-learn, and TensorFlow. These tools will facilitate data preprocessing, exploratory data analysis, and the implementation of machine learning algorithms. Visualizations generated using Matplotlib and Seaborn will aid in presenting the findings in an easily interpretable manner.

In conclusion, this capstone project aims to contribute valuable insights into the ongoing COVID-19 pandemic. By addressing these critical questions, leveraging a comprehensive dataset, and employing sophisticated data analysis techniques, the project seeks to enhance our understanding of the pandemic's dynamics and inform effective strategies for mitigating its impact.

References

P., D. K. (2020, August 7). *Covid-19 dataset*. Kaggle. https://www.kaggle.com/datasets/imdevskp/corona-virus-report/data