Project Milestone 2

Kiran Chowdary Komati

College of Science and Technology, Bellevue University

DSC680-T301: Applied Data Science (2231-1)

Dr. Catherine Williams

January 14, 2023

Proposal:

Topic: The topic I have selected for my second project is "Covid-19 Image Classification".

Business Problem:

We all are aware that covid-19 has taken over the world from the year 2020. Most people who fall sick recover under 2 weeks with minimal symptoms and without any special treatment. But there were cases where they needed medical treatment due to severe symptoms. The diagnosis is done primarily by nasal / throat swab tests. This project explores the alternative way by analyzing the x-rays of the patients diagnosed with Covid-19.

Data sets:

I'm using the dataset https://www.kaggle.com/datasets/pranavraikokte/covid19-image-dataset I found in Kaggle. This has a total of 454 images organized in to 3 folders representing 3 classes out of which 137 images are of Covid-19 and 317 in total with the viral pneumonia and the normal chest X-rays grouped into training and test folders.

Methods:

The analysis methods for this topic may change as the project progress. I'll first perform some EDA to understand the data. I'm planning to use VGG16 initially and may add other models such as CNN to compare the accuracy against the logistic regression model.

Ethical considerations:

As this project is related to the health of individuals and as it contains sensitive information, I made sure to not use any PII information such as individual names. Also, the data needs to be presented in accurate form without any modifications or misrepresentations. Also, steps must be taken to avoid any bias towards any gender.

Challenges/Issues: Potential challenges and issues must be identified as a best practice when exploring data science projects. I need to validate if the data is incomplete and must take measures to handle it ethically. Also, I'm planning to use google-colab as I have memory/processor restrictions in my personal laptop. Despite these potential challenges, I'm looking forward to this interesting data science project, "Covid-19 Image Classification"

References:

https://www.kaggle.com/datasets/pranavraikokte/covid19-image-dataset?select=Covid19-dataset