**Project 4 Group Number:4**

Title of the project : Classifying the MNIST Dataset with Questions

Team members: Aaron Billow and Adam Mayer

The aim of our project is to uncover patterns in images provided to us from the MNIST Dataset to be able to correctly identify the type of scan taken.

We shall be looking into the relationship our testing model creates and the accuracy with which it predicts the image scanned.

1) Find a problem worth solving, analyzing, or visualizing.

The problem is the accuracy at which machines can determine which human body part has been scanned with a CT or MRI automatically.

If this was correctly fixed and adjusted for, the findings/code could go on to be used to assist with the identification of medical problems for patients with great accuracy and speed. This would benefit the medical community.

2) You must use Scikit-learn and/or another machine learning library.

We have decided to use Scikit-learn for predictions.

Nevertheless tensorflow.keras may also be used.

3) You must use at least two of the following: Python Pandas, Python Matplotlib, HTML/CSS/Bootstrap, JavaScript Plotly, JavaScript Leaflet, Tableau, SQL Database, MongoDB Database, Google Cloud SQL, Amazon AWS.

From the following list we shall use at least: Python Pandas and Python Matplotlib.

4) The data set that shall be used is the MNIST Medical, containing 58954 Images.

Data sources:

<https://www.kaggle.com/datasets/andrewmvd/medical-mnist>

### The author who provides us with this dataset:

[@misc](https://www.kaggle.com/misc){Medical MNIST Classification, author = {apolanco3225}, title = {Medical MNIST Classification}, year = {2017}, publisher = {GitHub}, journal = {GitHub repository}, howpublished = {\url{[https://github.com/apolanco3225/Medical-MNIST-Classification}}](https://github.com/apolanco3225/Medical-MNIST-Classification%7D%7D) }

Scans provided are from one of 6 following classes: Abdomen CT, Breast MRI, CXR, Chest CT,

Hand, Head CT.

Task Breakdown: TBD on 22/12/23- due to Aaron’s power outage on the 21/12/23.

Data Collection: provided by Kaggle

Loading data images in and determine what size they are and the number of dimensions.

Pre process data; Including (not necessarily in order) Scale the images all to a specific size and number of dimensions. Ensure that the data images are normalized as well and that we have both a train and a testing data set. Make sure the number of classes is set as well as knowing the number of images in each class.

Coding: make sure the code is clear and that any function that shall be reused numerous times gets defined. Use the accuracy scores, predictions and confusion matrix’s to make sense of the data.

Visualize: Data must be presented clearly and simply. The goal here is to tell the story of how well how model can predict the image rather than displaying images themselves. Graphs and scatterplots could be used to attempt to tell said story.

Data Analysis: Reviewing the predictions from the code and making an evaluation on the model. Giving thoughts on how the model can be improved, possible limitations and things we may like to see get tested or released in the future for us to be able to test.

Presentation: Both group members shall talk and prepare accordingly for the speech. The data shall be introduced, and the author shall be credited. Very important to make sure that we display a reference section both in the presentation and in the git hub.

Looking at using HTML/CCSS/ Flask to assist in producing the final presentation.

Ensure the website has a user interface section, ideally an upload image section. This then triggers the model that we have created to predict which of the 6 classes, identified above in the Scan list section, the image belongs to.

Communicating and organizing:

Adam is responsible for the planning organising of the group meetings and shall accommodate to the needs of Aaron whilst he is away over Christmas.

3x meetings shall be held prior to the return of class to ensure that communication is clear between both parties. The reason for the break down is to assist both team members in understanding and keeping track of where they are up to within the project and its time line for completion.

Working log, 1`meeting was held on the 22nd, the 2nd meeting was done via messages and kept short due to the Christmas break and planned holidays. The final meeting was held prior to the class starting on the 2nd of January.

Edited Plans to the project.

Plans changed on the 5th of January

The data set that we used would be the same as would most of the code. Kaggle was where most of these notebooks where ran and steamlit will be the application used to present and run our model.’

As of the 8th of the First.

Make sure we can call efficient net module to finish the code lines

Upload the end modules to streamlit

Divide who shall present what during the presentation

Make sure we know how to upload another plot/table into streamlit

On the 9/

Finish importing any plots/tables into streamlit

Finish formatting streamlit

Make sure the GitHub Repository and ReadMe file are up to date.