National College of Ireland

Configuration Manual

MSc Research Project
MSc in Data Analytics

Irfan Pervez
Student ID: 23174854

School of Computing
National College of Ireland

Supervisor: Hamilton Niculescu

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Irfan Pervez
Student ID:	23174854
Programme:	MSc in Data Analytics **Year:** Jan - 2024
Module:	MSc Research Project
Lecturer:	Hamilton Niculescu
Submission Due Date:	12/12/2024
Project Title:	A Generative AI Framework for Data Augmentation Employing Generative Adversarial Networks (GANs) to Predict Parkinson's Disease
Word Count:	1748 **Page Count:** 17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

Signature:	Irfan Pervez
Date:	10-12-2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Irfan Pervez
Student ID: 23174854

Environment Setup –

Software Requirements
Operating System : Windows 10/11 (64-bit)

Hardware Requirements
Processor: Dual-core CPU (e.g., Intel Core i3, AMD Ryzen 3).
Recommended: Quad-core CPU (e.g., Intel Core i5/i7, AMD Ryzen 5/7) for faster data processing and model training.

Memory: Minimum: 8 GB RAM.
Recommended: 16 GB or more for handling large datasets and running both frontend and backend concurrently.

Storage
Minimum: 20 GB of free disk space.
Recommended: SSD storage for improved performance during model training and data generation.

Installation Setup - Installation Steps for Running the SyntheticDataQ.ai(Generative AI) Application on Your System.

Step 1 : Install Microsoft Visual Studio code app for windows - https://code.visualstudio.com/docs/setup/windows highlighted below in Figure 1. in yellow .

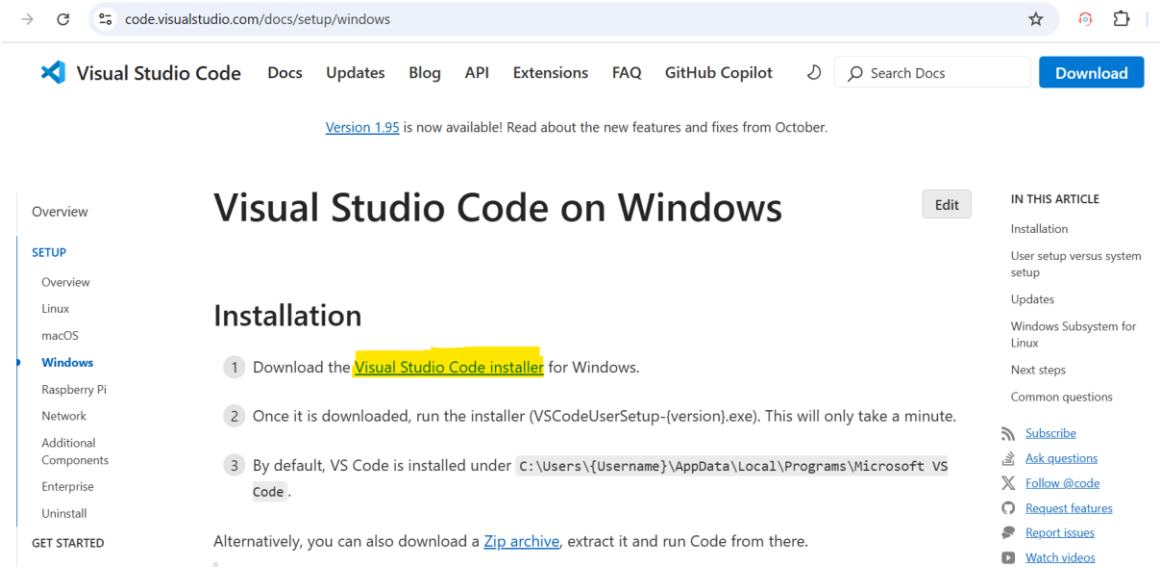

Figure 1. Visual studio code installation webpage (Source: Microsoft, 2023)

Step 2 : Open Visual studio code and Install below python libraries shown below in Figure 2.

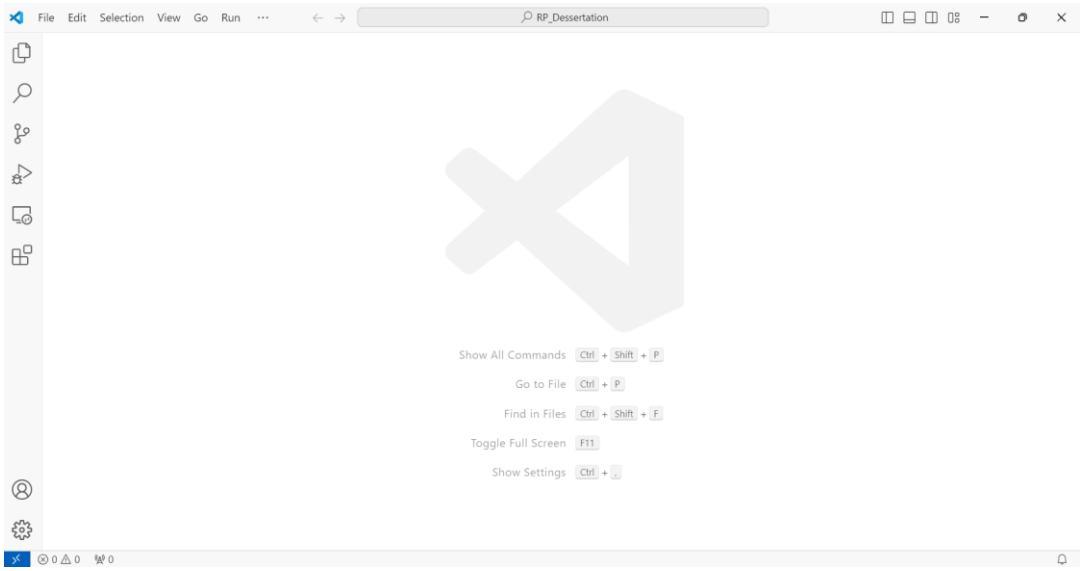

Figure 2. Visual studio developer toolkit (Source: Microsoft, 2023)

Step 3 : Go to Terminal -> New Terminal and open a new Terminal as shown below in Figure 3. Run the below commands to install all the required python libraries.

- pip install flask
- pip install flask-cors
- pip install pandas
- pip install numpy
- pip install torch
- pip install sklearn

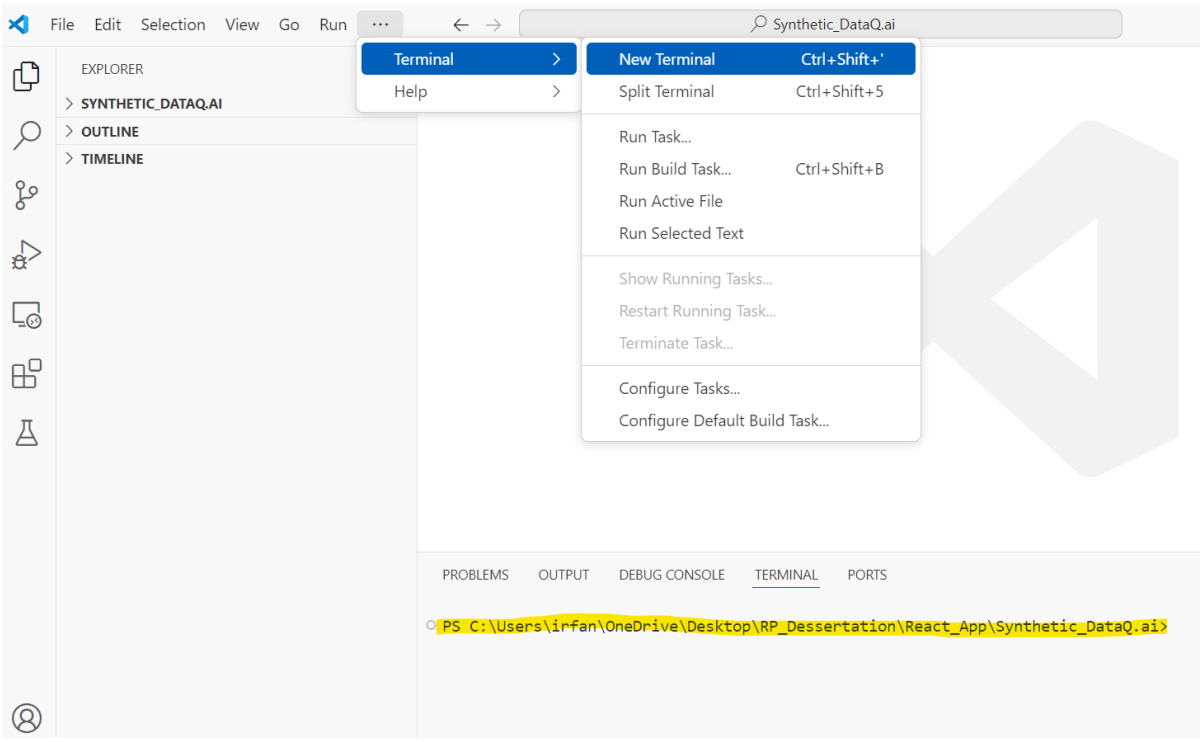

Figure 3. New Terminal in Visual Studio code (Source: Microsoft, 2023)

Step 3: Download the SyntheticDataQ.ai source code from Moodle with file name (**Synthetic_DataQ.ai)**. Extract the code and upload the source code to workspace using open workspace File -> Open workspace from file shown in figure 4.

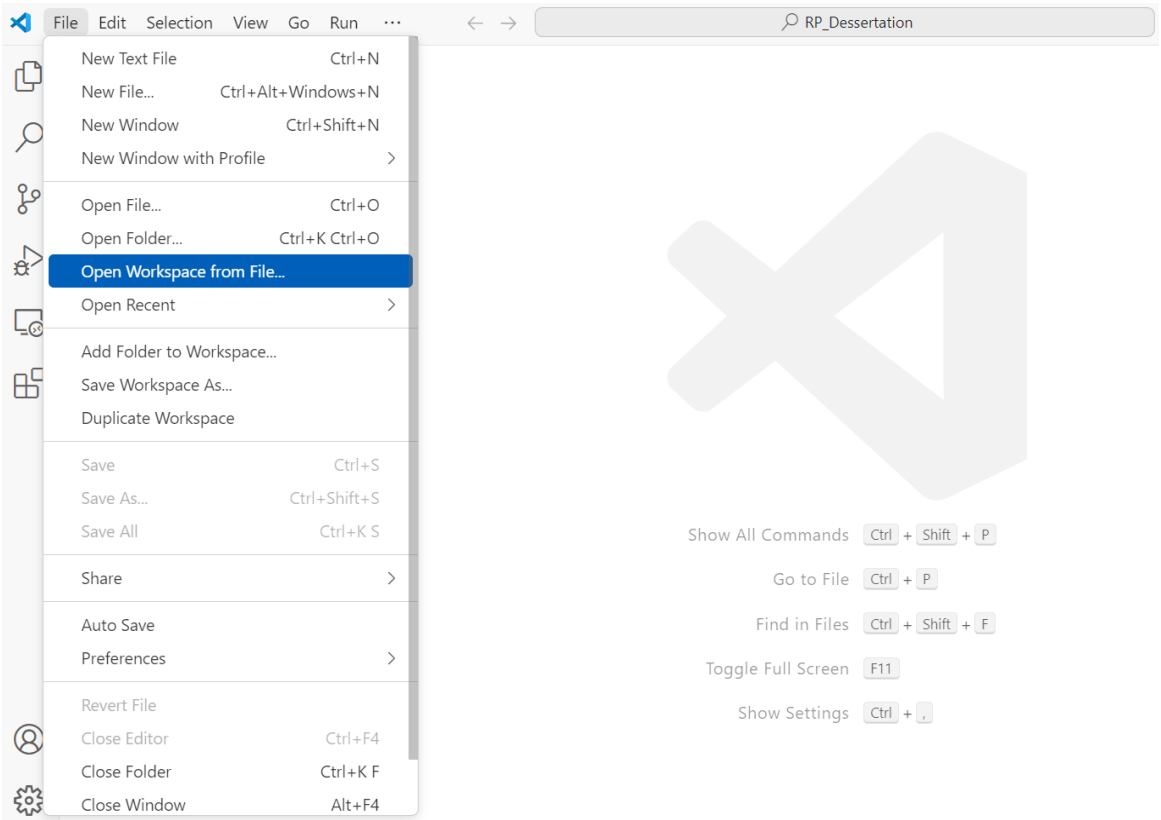

Figure 4. Visual studio code option to start import (Source: Microsoft, 2023)

Step 4: Select the source code for SyntheticDataQ.ai you have downloaded and upload it in workspace shown in figure 5.

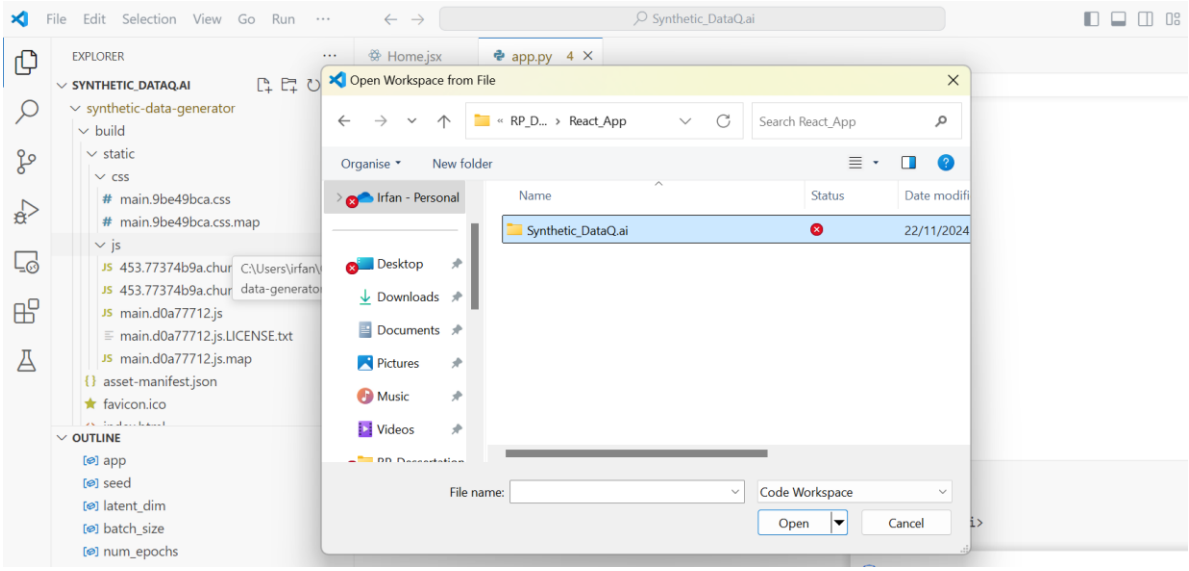

Figure 5. Visual studio code to select the project (SyntheticDataQ.ai) source code (Source: Self-created)

Step 5: Open a new terminal to start the React based user interface shown in Figure 6.

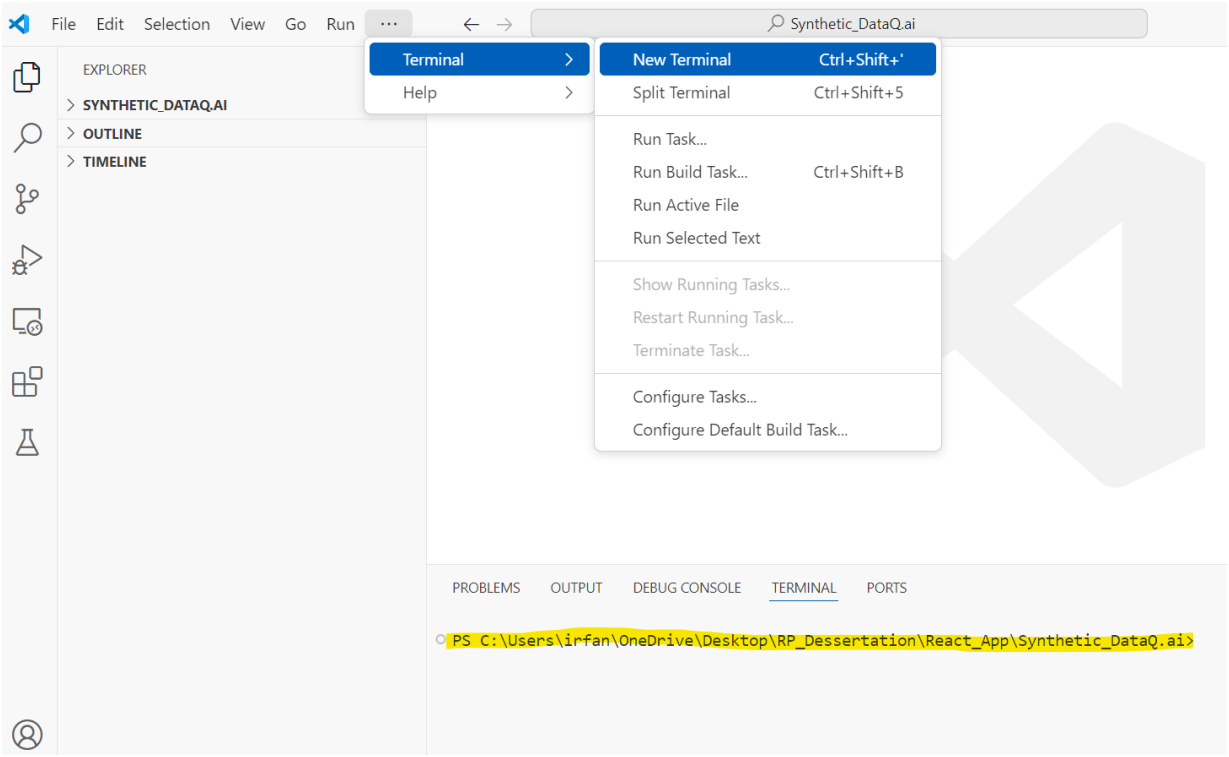

Figure 6. Opening a new terminal (Source: Microsoft, 2023)

Step 6: Go to src folder inside Synthetic_DataQ -> synthetic-data-generator folder ->src highlighted in yellow in Figure 7.

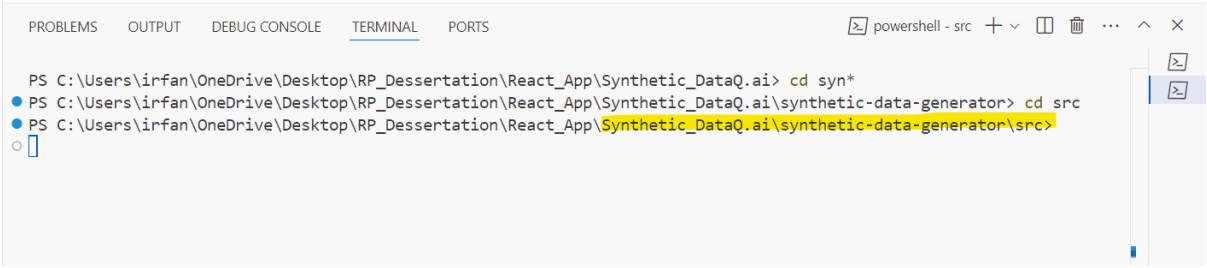

Figure 7 . Terminal src folder (Source: Microsoft, 2023)

Step 7: Execute **npm start** command in terminal to start the React web application. You will see the below Highlighted message "webpack compiled successfully " shown in Figure 8.

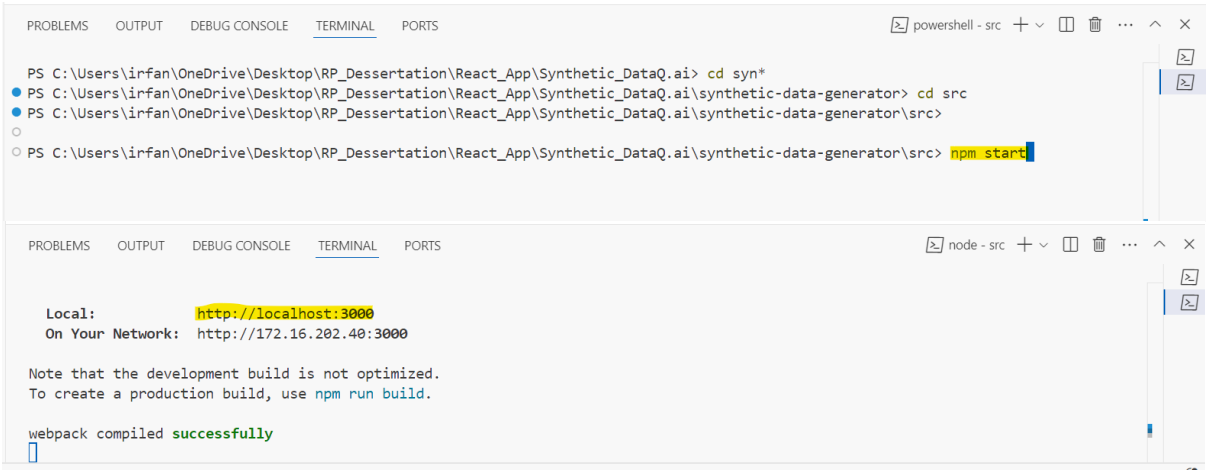

Figure 8. npm start command and SyntheticDataQ.ai app is up on http://localhost:3000/ URL (Source: Microsoft, 2023)

Step 8 : Google browser will automatically open with SyntheticDataQ.ai . Please use this link to access the SyntheticDataQ.ai UI http://localhost:3000/ if it's not opening automatically.

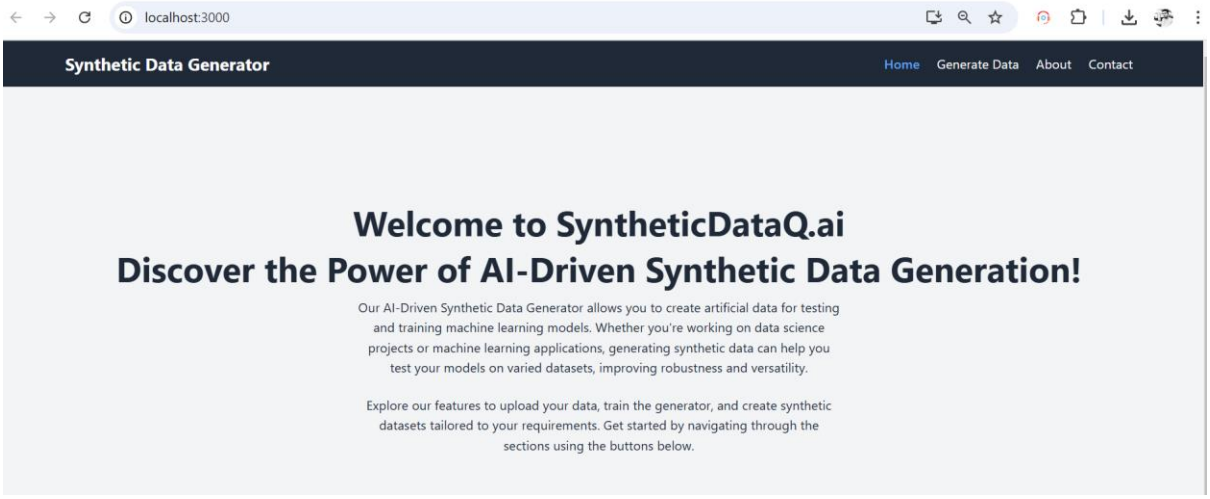

Figure 9. Home screen of SyntheticDataQ.ai (Source: Self-created)

Step 9 : Go to Microsoft Visual Studio app and open a new terminal again to run Backend flask server to run **app.py** python file which contains Generative Adversarial Networks (GANs) AI model. Go to backend folder to run GAN model shown below. Go to following path highlighted in yellow in Figure 10. Synthetic_DataQ.ai -> synthetic-data-generator -> src -> backend

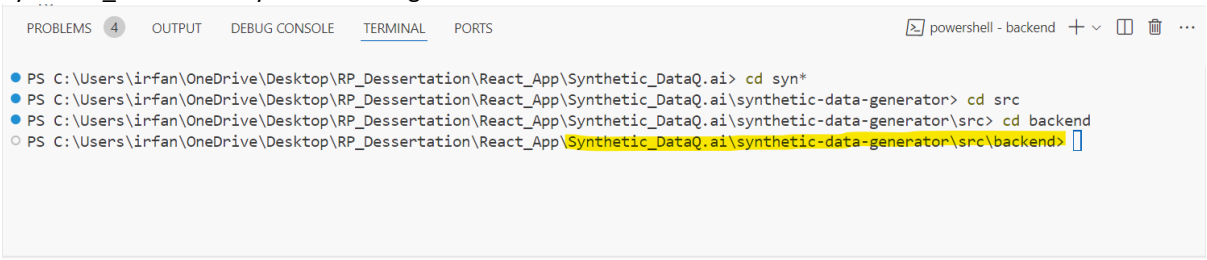

Figure 10. Path to the backend where Generative Adversarial Networks (GANs) model deployed

Step 10: Execute the **Python app.py** in the terminal highlighted in yellow in Figure 11. contains the GAN model code which is deployed in Flask server.

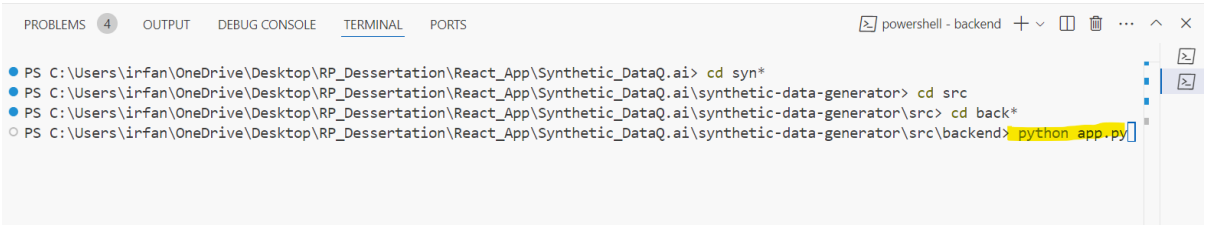

Figure 11. Run the app.py on Flask server

Step 11 : Click on generate data tab in Right corner highlighted below in Figure 12 in Green.

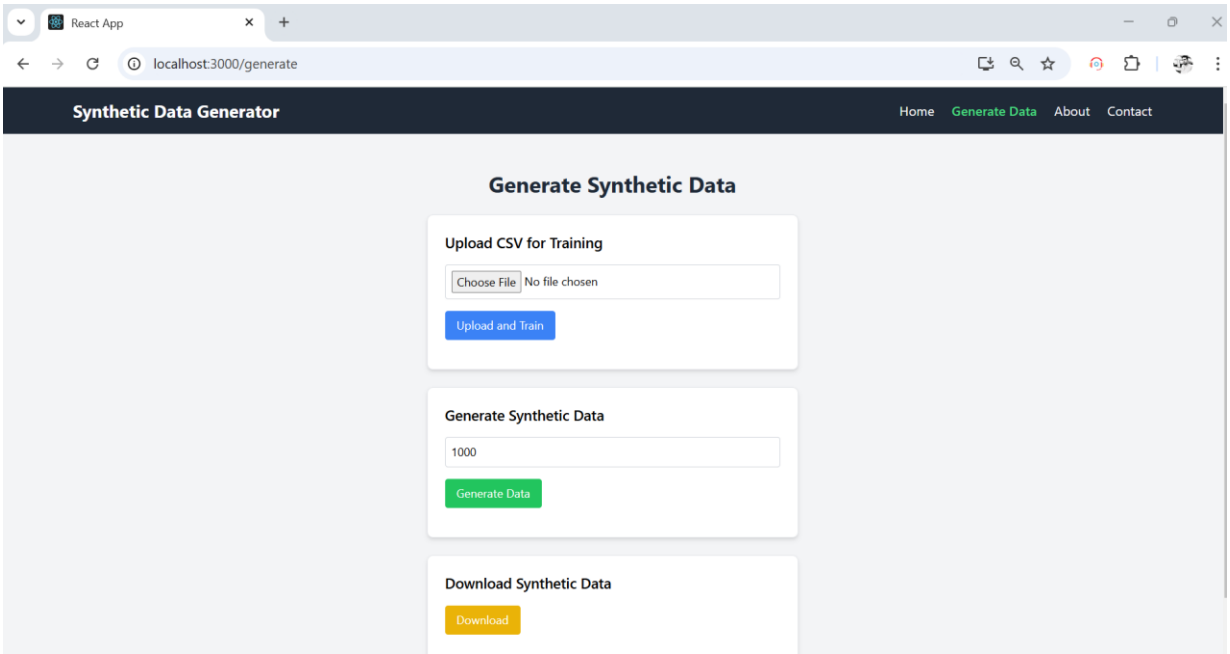

Figure 12. Generate Data screen of of SyntheticDataQ.ai (Source: Self-created)

Step 12 : Click on chose file and select the Parkinsons Dataset file parkinsons.csv available in the **Synthetic_DataQ.ai**. Again, click on upload and Train the model. You will see the message **Model Training Complete** highlighted in yellow shown in Figure 13.

-Add 5000 in Generate Synthetic data textbox and click on Generate Data button. You will see the message "Synthetic data generated successfully!" highlighted in yellow in Figure 13.

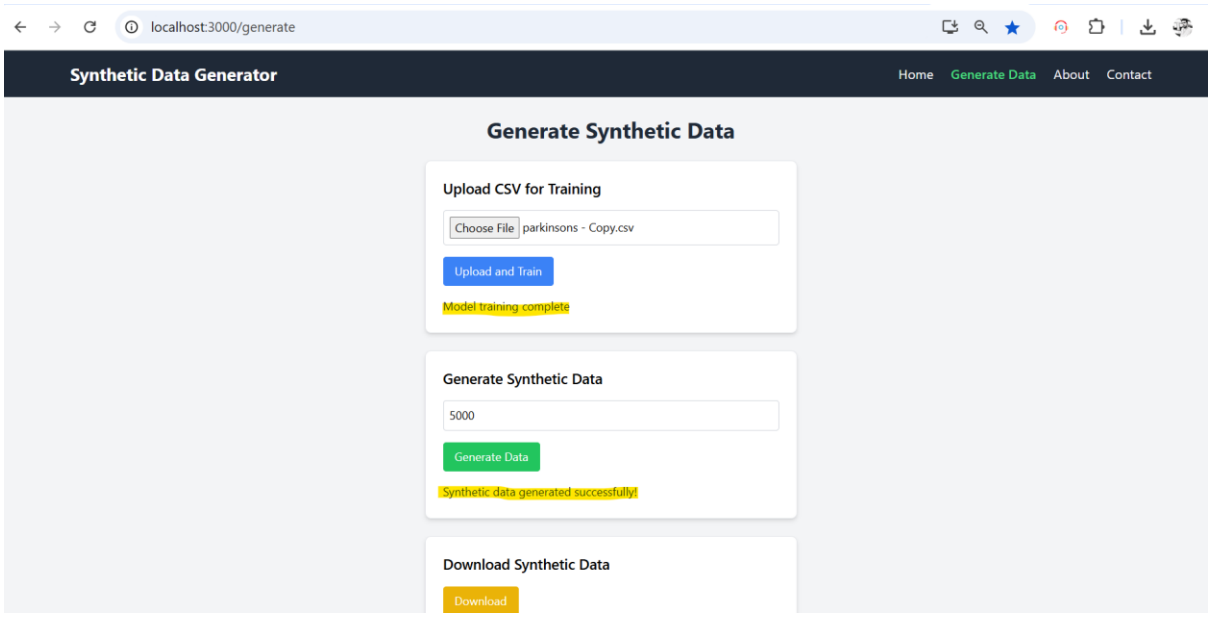

Figure 13. Generate Data process of SyntheticDataQ.ai(Source: Self-created)

Step 13: Click on Download button and Download Synthetic data. You can see the file downloaded in the download folder **synthetic_data.csv**.

- Version 1 (Synthetic Data-5000 records)

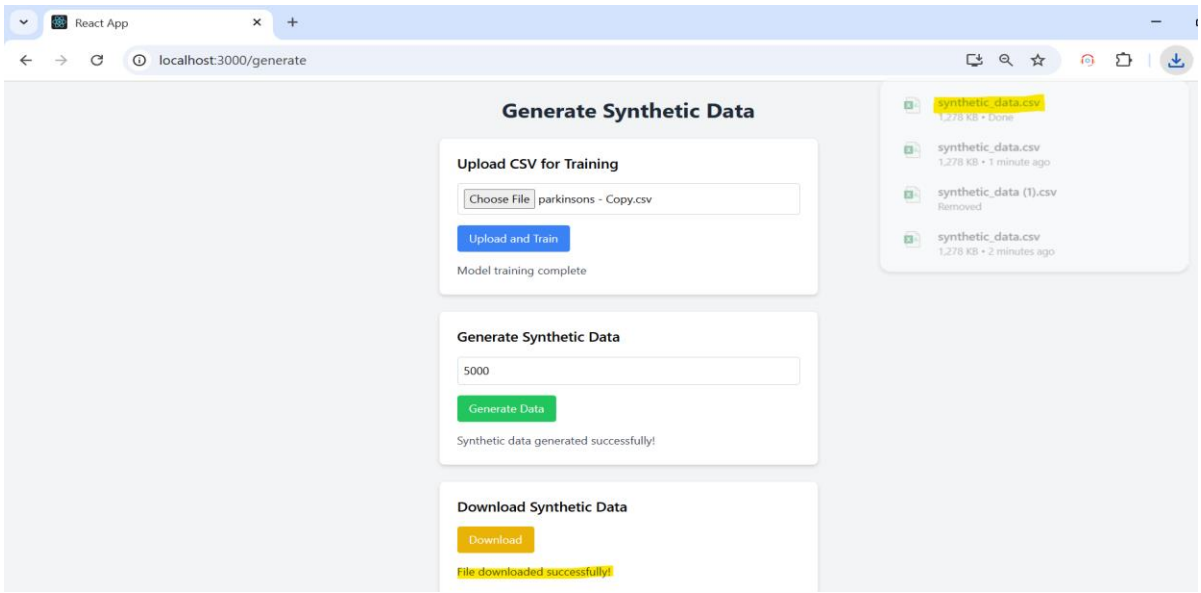

Figure 14. Download synthetic data size 5000 records from SyntheticDataQ.ai(Source: Self-created)

Step 14: Synthetic Data generated as **synthetic_data.csv file** and can be downloaded from the download folder. Repeat the steps Step 13 to generate the synthetic data for 10,000 and 15,000 voice recordings of Parkinson's diseases.

- Version 2 (Synthetic Data-10000 records) shown in Figure 15.
- Version 3 (Synthetic Data-15000 records) shown in Figure 16.

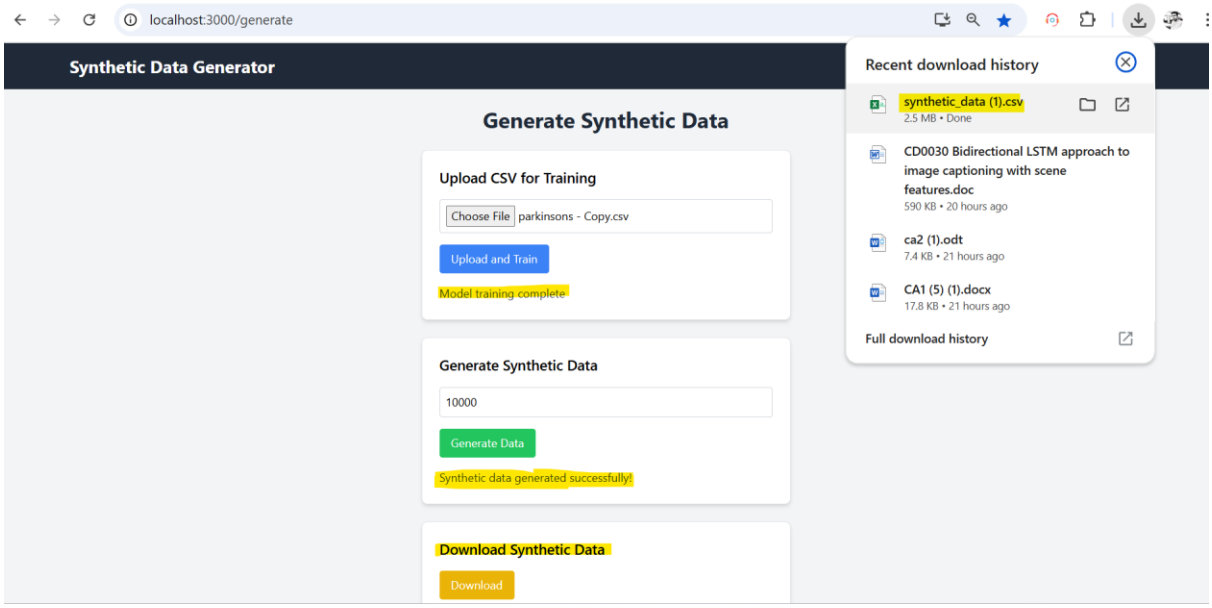

Figure 15. Download synthetic data size 10000 records from SyntheticDataQ.ai (Source: Self-created)

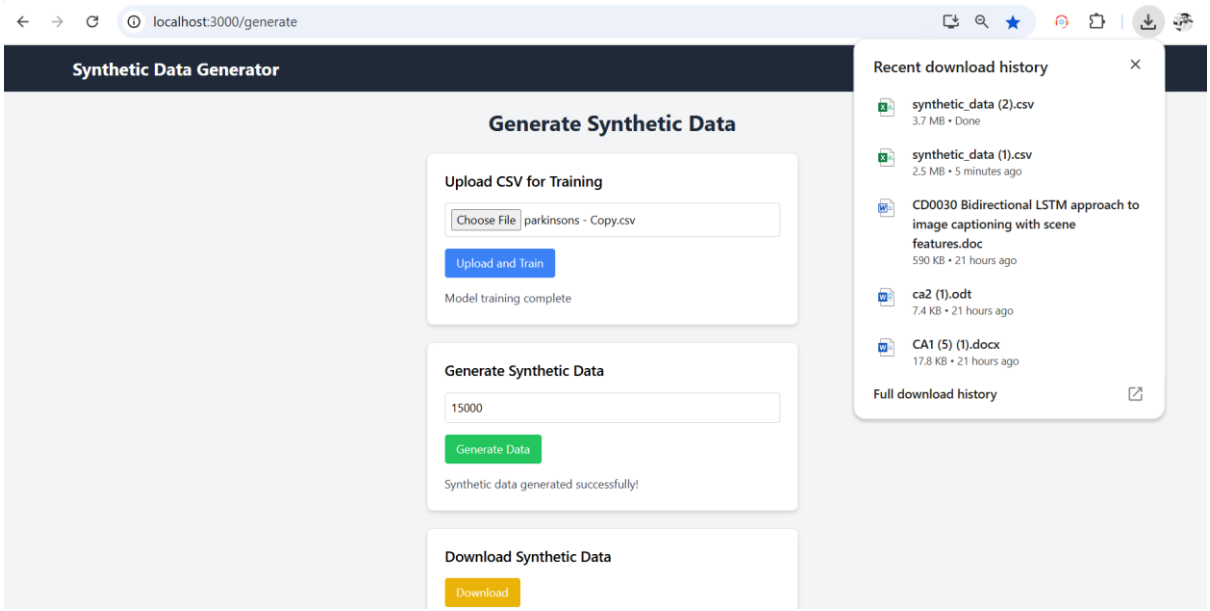

Figure 16. Download synthetic data size 15000 records from SyntheticDataQ.ai (Source: Self-created)

Evaluation setup- Steps to Evaluate All Three Versions of Synthetic Data Using Machine Learning and Deep Learning Models.

Step 16: Download and Install Anaconda
https://www.anaconda.com/download/success shown below in Figure 17.

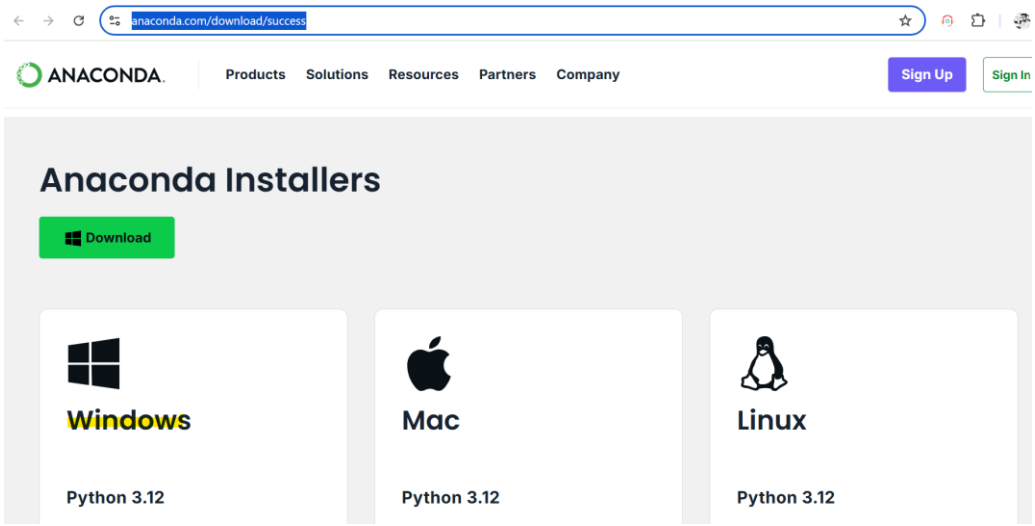

Figure 17. Anaconda Installer Webpage. Anaconda Inc. (2023).

Step 17 : Install the below python libraries as we needed these libs for Explore the Dataset, Preprocessing Steps, Model Building and Model Evaluation.

Open Anaconda Prompt from command prompt and install below libraries shown below Figure 18.

- pip install pandas
- pip install numpy
- pip install matplotlib
- pip install tensorflow
- pip install plotly
- pip install seaborn
- pip install sklearn

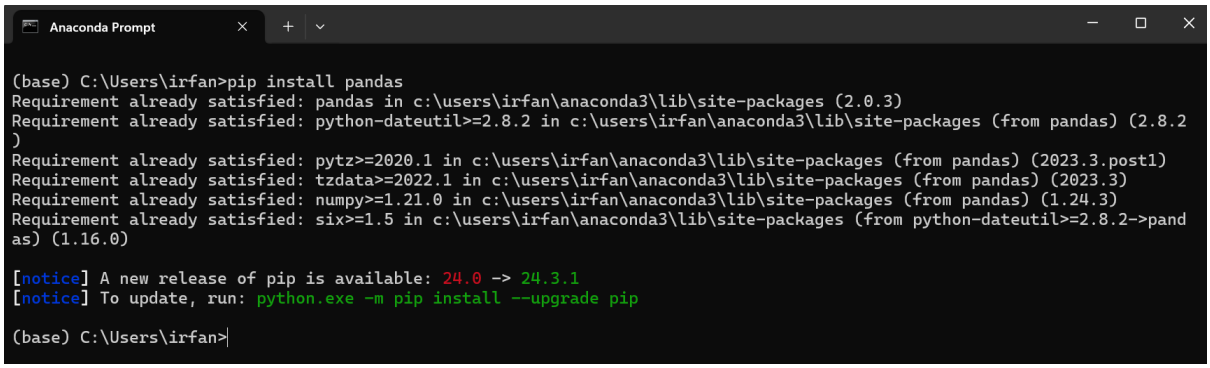

Figure 18. Anaconda prompt to install python libs

Step 18 : Download the Synthetic_Data_Evaluation_Artifacts source code folder from **Moodle** (folder name **Synthetic_Data_Evaluation_Artifacts**) to run the evaluation of synthetic data.

Step 19: Load the python files and Jupyter Lab and open the **detecting_Parkinson's_disease_Experiment_1.ipynb** python file to analyze the Version 1 (Synthetic Data-5000 records) from the artifacts . Copy the **synthetic_data.csv** generated by **Synthetic dataQ.ai application** as per Step 14.

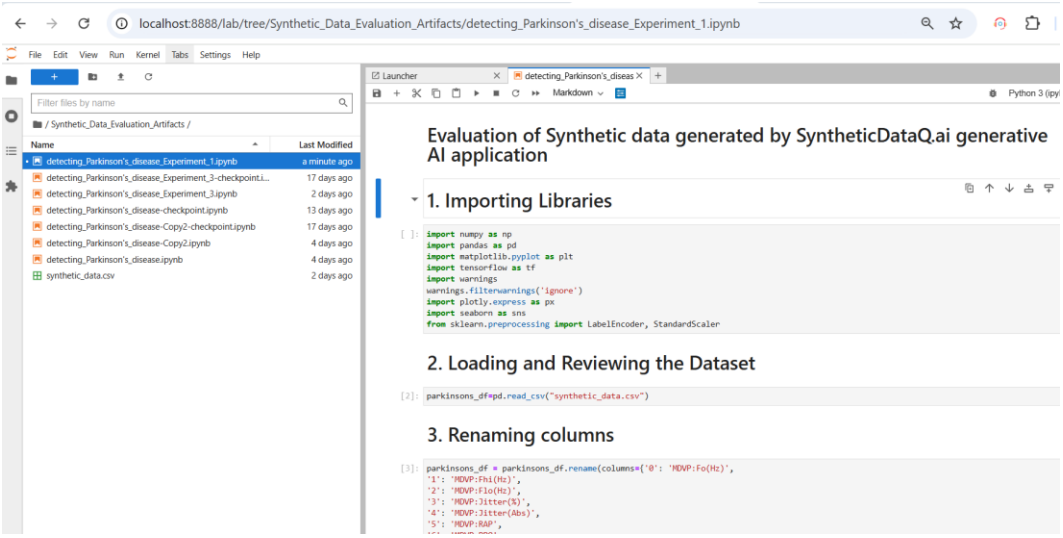

Figure 20. detecting_Parkinson's_disease_Experiment_1.ipynb python file to analyze the Version 1 (Synthetic Data-5000 records)

Step 20 : Click on **Restart Kernal and Run All cells** to run all the execute to follow below steps.

- Explore the Dataset.
- Conduct Extensive Exploratory Data Analysis (EDA):

4. Exploratory Data Analysis (EDA)

Feature Distributions

```
[34]: features = ['MDVP:Fo(Hz)', 'MDVP:Fhi(Hz)', 'MDVP:Flo(Hz)', 'MDVP:Jitter(%)']
parkinsons_df[features].hist(bins=20, figsize=(10, 8))
plt.suptitle('Distribution of Voice Measurements', fontsize=16)
plt.show()
```

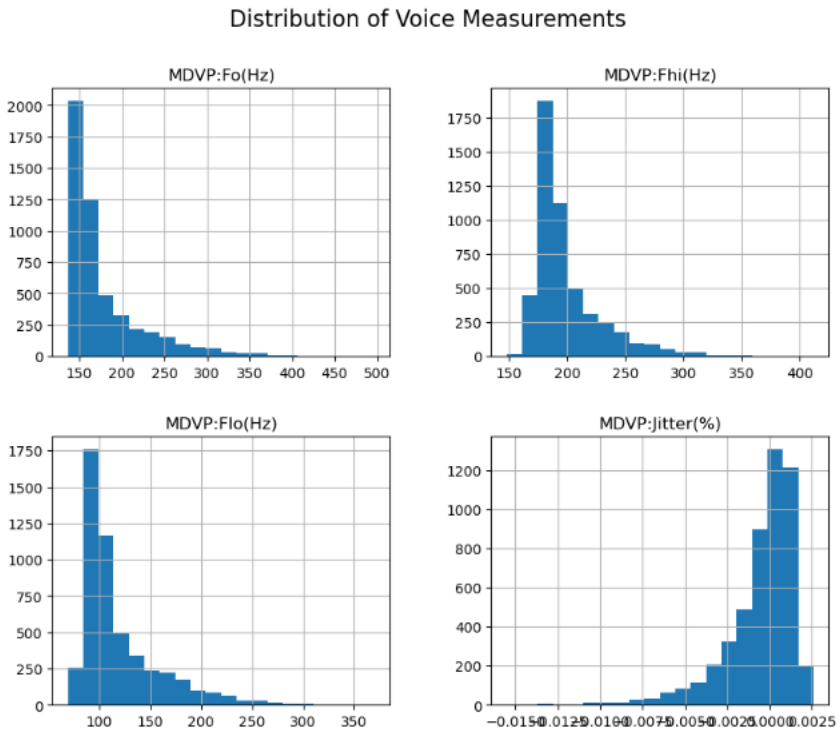

Figure 21. Exploratory Data Analysis (EDA) (Source: Self-created)

- Preprocessing Steps.
- Model Building.
- Evaluation Model Performance shown in Figure 22

Random Forest Classifier

```
[55]: rf_clf = RandomForestClassifier(random_state=42)
rf_clf.fit(X_train, y_train)
```

```
[55]: RandomForestClassifier(random_state=42)
```

```
[56]: rf_pred = rf_clf.predict(X_test)
```

```
[57]: rf_accuracy = accuracy_score(y_test, rf_pred)
print("Random Forest Accuracy:", rf_accuracy)
```

```
Random Forest Accuracy: 0.991
```

```
[58]: from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn import metrics
from sklearn.metrics import mean_squared_error
```

```
[59]: print(classification_report(y_test, rf_pred))
```

```
              precision    recall  f1-score   support

           0       0.97      1.00      0.98       267
           1       1.00      0.99      0.99       733

    accuracy                           0.99      1000
   macro avg       0.98      0.99      0.99      1000
weighted avg       0.99      0.99      0.99      1000
```

Figure 22. Randon forest classifier performance (Source: Self-created)

Step 21 : Follow the Steps 18 and 19 to Run the other versions of Synthetic data **Version 2 (Synthetic Data-10000 records)** and **Version 3 (Synthetic Data-15000 records)** and capture the results .

-Load the python files and Jupyter Lab and open the **detecting_Parkinson's_disease_Experiment_2.ipynb** python file to analyze the Version 2 (Synthetic Data-10000 records) from the artifacts .

- Load the python files and Jupyter Lab and open the **detecting_Parkinson's_disease_Experiment_3.ipynb** python file to analyze the Version 3 (Synthetic Data-15000 records) from the artifacts .

FAQs –

Q: Any presentation available to go through to setup the Artifact code ?
A: Presentation URL to setup the Artifact - https://studentncirl-my.sharepoint.com/personal/x23174854_student_ncirl_ie/_layouts/15/stream.aspx?id=%2Fpersonal%2Fx23174854%5Fstudent%5Fncirl%5Fie%2FDocuments%2FRecordings%2FMSc%20Project%2D20241210%5F184611%2DMeeting%20Recording%2Emp4&referrer=StreamWebApp%2EWeb&referrerScenario=AddressBarCopied%2Eview%2Eb283af4c%2D91eb%2D4e2e%2D9e49%2D7b778190de96

Q: What are the operating systems are support SyntheticDataQ.ai?

A: The application is designed for Windows 10/11 (64-bit)

Q: Do I need add GPU to run this application?

A: A GPU is recommend for faster training of the GAN model, but the application can also run on a CPU with less performance.

Q: What if a Python package failing while installations?

A: Check and ensure if your Python version is 3.7 or higher and run the command:
bash
Copy code
pip install --upgrade pip

Q: How do I install Node.js?

A: go to the Node.js download page and download the new LTS version and follow the installation prompts.

Q: I got permission denied the issue while running commands in the terminal. What should I do?

A: Run the terminal as an administrator (Windows)

Q: How do I start the React based frontend?

A: Navigate to the project folder in Visual Studio Code, open the terminal, and run:
bash
Copy code
npm start

Q: How do I start the Flask-based backend?

A: Open a new terminal in Visual Studio Code, navigate to the backend folder, and run:
bash

Copy code
python app.py

References

Anaconda Inc. (2023). *Anaconda Installer Webpage*. Available at: https://www.anaconda.com/download/success

Anaconda Inc. (2023). *Installing Python Libraries Using Anaconda Prompt*. Available at: https://docs.anaconda.com/anaconda/user-guide/tasks/install-packages/

Microsoft (2023). *Explorer Panel in Visual Studio Code*. Available at: https://code.visualstudio.com/docs/getstarted/userinterface

Microsoft (2023). *Navigate to src Folder in Visual Studio Code*. Available at: https://code.visualstudio.com/docs/editor/codebasics

Microsoft (2023). *Opening a New Terminal in Visual Studio Code*. Available at: https://code.visualstudio.com/docs/editor/integrated-terminal

Microsoft (2023). *Open Workspace in Visual Studio Code*. Available at: https://code.visualstudio.com/docs/editor/workspaces

Microsoft (2023). *Run app.py on Flask Server Using Terminal*. Available at: https://flask.palletsprojects.com/en/2.3.x/

Microsoft (2023). *Set up Visual Studio Code on Windows*. Available at: https://code.visualstudio.com/docs/setup/windows

Microsoft (2023). *Using Terminal in Visual Studio Code*. Available at: https://code.visualstudio.com/docs/editor/integrated-terminal

Microsoft (2023). *Visual Studio Developer Toolkit*. Available at: https://code.visualstudio.com/docs

React.js (2023). *React Application Running on localhost:3000*. Available at: https://react.dev

React.js (2023). *SyntheticDataQ.ai User Interface Home Screen*. Available at: http://localhost:3000/