BITI 3133 : NEURAL NETWORK
SEM 2 2022/2023
FINAL REPORT

TITLE : THYROID DISEASE DETECTION USING MULTICLASS MULTILAYER NEURAL NETWORK

LECTURER : DR. BURHANUDDIN MOHD ABOOBAIDER

NAME	MATRIC NUMBER	GROUP
DEVESHWER A/L NAGANTHIRAN	B032110343	S1G1
TARRSHAN TAMILARASU	B032110400	S1G1
SUREND A/L ESVARAN	B032110339	S1G1
TUAN SYED FAIQ EIZAHAN BIN TUAN AZMAN	B032110229	S1G1

Table of Contents

1. INTRODUCTION

The prevalence of thyroid problems and their major effects on general health make them common worldwide. Traditional thyroid detection techniques rely on expert medical analysis of numerous diagnostic procedures and judgmental interpretations. It is crucial to create automatic and precise detection systems because this procedure can be time-consuming and prone to mistakes. This project's goal is to employ a multiclass multilayer neural network to determine a patient's likelihood of having hypothyroidism based on their medical history and symptoms.

This case study's main goal is to develop and put into use a neural network-based thyroid detection system. The following objectives are pursued by the system:

a. Enhance accuracy of thyroid diagnosis by utilizing neural network skills, which are excellent at pattern recognition and classification activities.
b. Reduce the time needed for diagnosis by developing an automated system that can quickly process diagnostic data and deliver accurate results.
c. Assist medical professionals: Give them a tool that will help them make better judgements and increase the effectiveness of thyroid screening and diagnosis as a whole.

2. DETAIL DESCRIPTION ABOUT SAMPLE DATA

- Number of attributes: 21 (15 attributes are binary, 6 attributes are continuous)
- Number of classes: 3

1 : normal (not hypothyroid),

2 : hyperfunction

3 : subnormal functioning

Number of learning examples: 3772

Number of testing examples: 3428

NO of ATTRIBUTE	NAME	DESCRIPTION
1	AGE	- Continuous - Changed to decimal point to make it easier for coding
2	SEX	- Binary - 0 for male and 1 for female
3	ON_THYROXINE	- Binary - 0 if no problems with thyroxine gland - 1 if there are any problems
4	QUERY_ON_THYROXINE	- Binary - 0 for no and 1 for yes
5	ON_ANTITHYROID_MEDICATION	- Binary - 0 if the person doesn't take antithyroid medicine - 1 if yes
6	SICK	- Binary - 0 if healthy - 1 if sick
7	PREGNANT	- Binary - 0 if not pregnant - 1 if pregnant
8	THYROID SURGERY	- Binary - 0 if patient have not gone through thyroid surgery - 1 if yes
9	I131 TREATMENT	- Binary - 0 if not under treatment - 1 if yes
10	QUERY_HYPOTHYROID	- Binary - 0 if no possibilities on hypothyroid - 1 if possible
11	QUERY_HYPERTHYROID	- Binary - 0 if no possibilities on hyperthyroid - 1 if possible

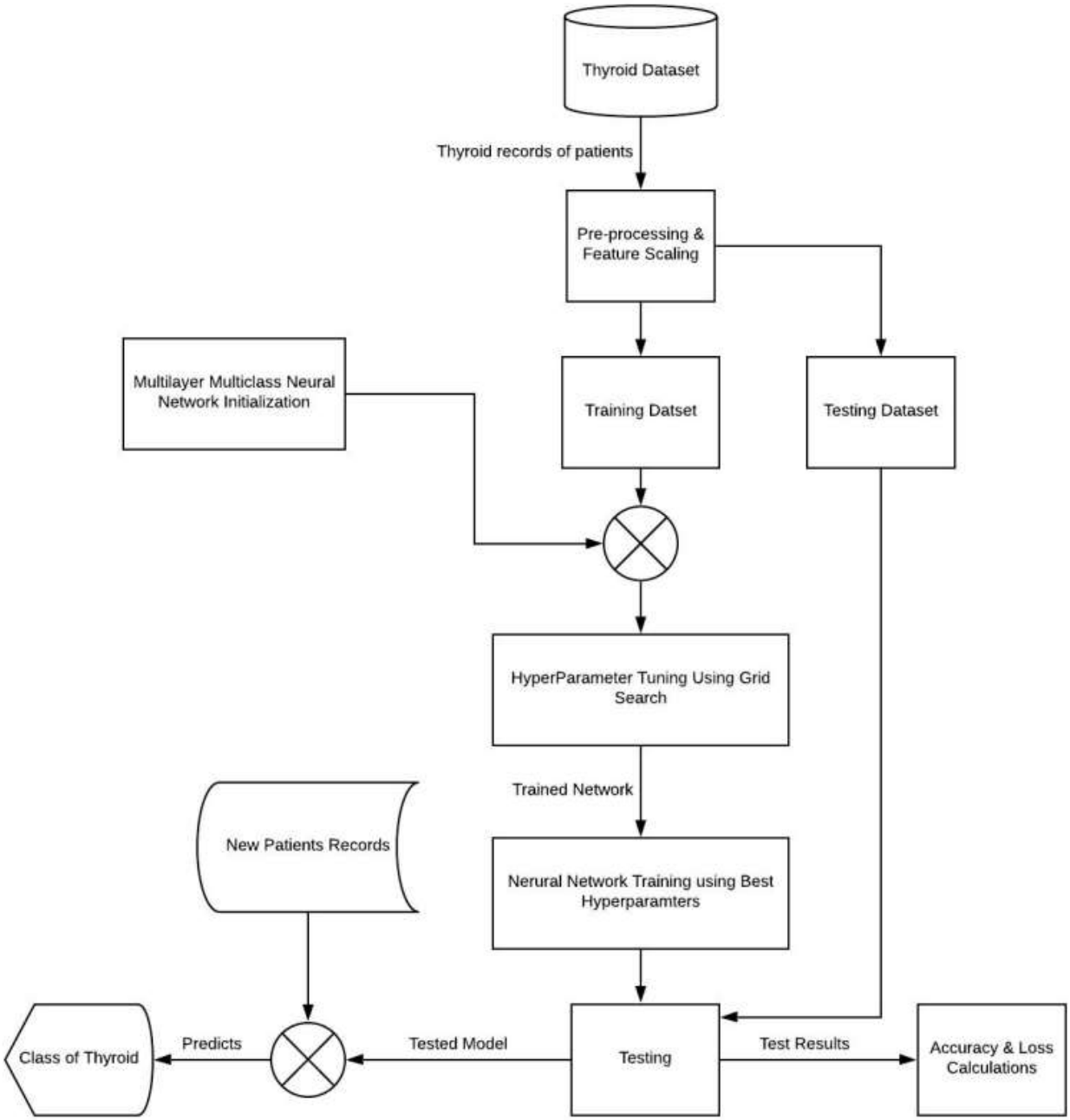
12	LITHIUM	- Binary - 0 if no usage / exposure to lithium - 1 if traces of lithium are found
13	GOITRE	- Binary - 0 if no goitre - 1 if goitre is found
14	TUMOUR	- Binary - 0 if no tumour - 1 if tumour is found
15	HYPOPITUITARY	- Binary - 0 if no hypopituitary - 1 if diagnosed with hypopituitary
16	PSYCH	- Binary - 0 if no psychiatrist is needed - 1 for the help of psychiatrist
17	TSH	- Continuous
18	T3	- Continuous
19	TT4	- Continuous
20	T4U	- Continuous
21	FTI	- Continuous

The dataset can be downloaded from the link given below :

https://archive.ics.uci.edu/dataset/102/thyroid+disease

A zip folder will be downloaded which contains all datasets related to thyroid. We chose the dataset with the most instances as larger dataset can train the model accurately. Other files that are related to these datasets can be found in the folder too.

3. FLOW CHART AND LEARNING PROCESS



4. STEP BY STEP LEARNING AND ANALYSIS

a) **Importing Libraries**:
- The necessary libraries, including NumPy, Matplotlib, Pandas, Keras, TensorFlow, and scikit-learn, are imported.

b) **Loading and Preprocessing Data**:
- Pandas is used to load the dataset from the Ann_train.csv file.
- The dataset is used to extract the input characteristics (X) and the target variable (y).
- Using the pd.get_dummies() method, the target variable (y) is one-shot encoded.
- The dataset is divided into training and testing sets using scikit-learn's train_test_split() method.
- Another dataset from the Ann_test.csv file is imported for further testing.

c) **Model Creation**:
- Keras is used to generate a neural network model.
- Using the Sequential() method, the model is initialized as a sequential model.
- The Adam optimizer, categorical cross-entropy loss function, and accuracy measure are used to construct the model.

d) **Model Architecture**:
- Dense() is used to add the input layer, which has 'relu' activation and input dimension 21.
- With 'relu' activation and 6 units each, two hidden levels are added.
- The output layer is enhanced with softmax' activation and three units (corresponding to the three classes).

e) **Model Training**:
- The fit() method is used to train the built model on the training dataset (X_train and y_train).
- The batch size and epoch count are chosen by the optimal hyperparameters found through grid search.
- The neural network's weights and biases are updated during the training phase to minimize the loss function and enhance accuracy.

f) **Model Evaluation**:
   - Using the predict() method, the trained model is utilized to generate predictions on the testing dataset (X_test).
   - Using a threshold of 0.5, the predictions are transformed to binary values (0 or 1).
   - Finding the index of the greatest value in each prediction row yields the expected class labels.
   - The confusion matrix is computed by combining anticipated and actual class labels (y_test).
   - The accuracy_score() method from scikit-learn is used to calculate the model's accuracy on the test set.

g) **Prediction on New Data**:
   - Three test records are created.
   - The records are fed into the trained model using the predict() method.
   - Using a threshold of 0.5, the predictions are transformed to binary values (0 or 1).
   - Finding the index of the greatest value in each prediction row yields the expected class labels.
   - Each test record has its expected and projected output indicated.

5. HOW ANALYSIS CAN BE CONDUCTED

The numerous thyroid disease groups or categories may be handled by this kind of neural network, which also has the ability to include numerous hidden layers to extract intricate patterns from the input data. The following steps are commonly included in the analysis process when using a neural network:

a. **Data collection and preparation**: A dataset comprising pertinent data on thyroid illnesses, such as patient symptoms, medical history, and test results, would have been gathered and created.

b. **Training**: Using the prepared dataset, the dataset would be trained. In order for the network to understand the underlying patterns and correlations, it must be fed input data and target outputs.

c. **Testing and Validation:** Following training, the effectiveness would have been assessed using a different set of data. This procedure aids in evaluating the network's classification of thyroid disorders for accuracy and dependability.

d. **Iteration and Improvement**: The dataset design may have been improved, parameters might have been changed, or other strategies might have been investigated based on the testing and validation findings.

6. DECISION/BUSINESS INTELLIGENCE MODEL

A framework or approach used to evaluate data and offer insights that support sensible business decisions is known as a decision/business intelligence model. It entails gathering pertinent information from a variety of sources, processing it into the appropriate format, and then evaluating it using methods like data mining, statistical analysis, and machine learning. The analyzed data is then presented in a simple and understandable way using visualization tools. Insights from the research are used to inform strategic and tactical decisions made inside the company, such as locating areas for expansion, streamlining procedures, raising client happiness, or reducing risks. Depending on the requirements of the business and the specifics of the sector, several methodologies, tools, and models may be used.

7. CONCLUSION

A Multiclass Multilayer Neural Network was developed and was successfully trained and tested on the Thyroid disease datasets, which yielded an accuracy of 98.51% . Neural Network was able to perform the multiclass classification of the three classes of the Thyroid disease.As the dataset contained around 92% records with class 3(normal subjects) it was very important to for the accuracy of the neural network to be greater than 92% and as the accuracy achieved is 98.51%, thus it would be safe to conclude that the neural network has achieved a respectable accuracy in predicting the Thyroid disease.

8. REFERENCES

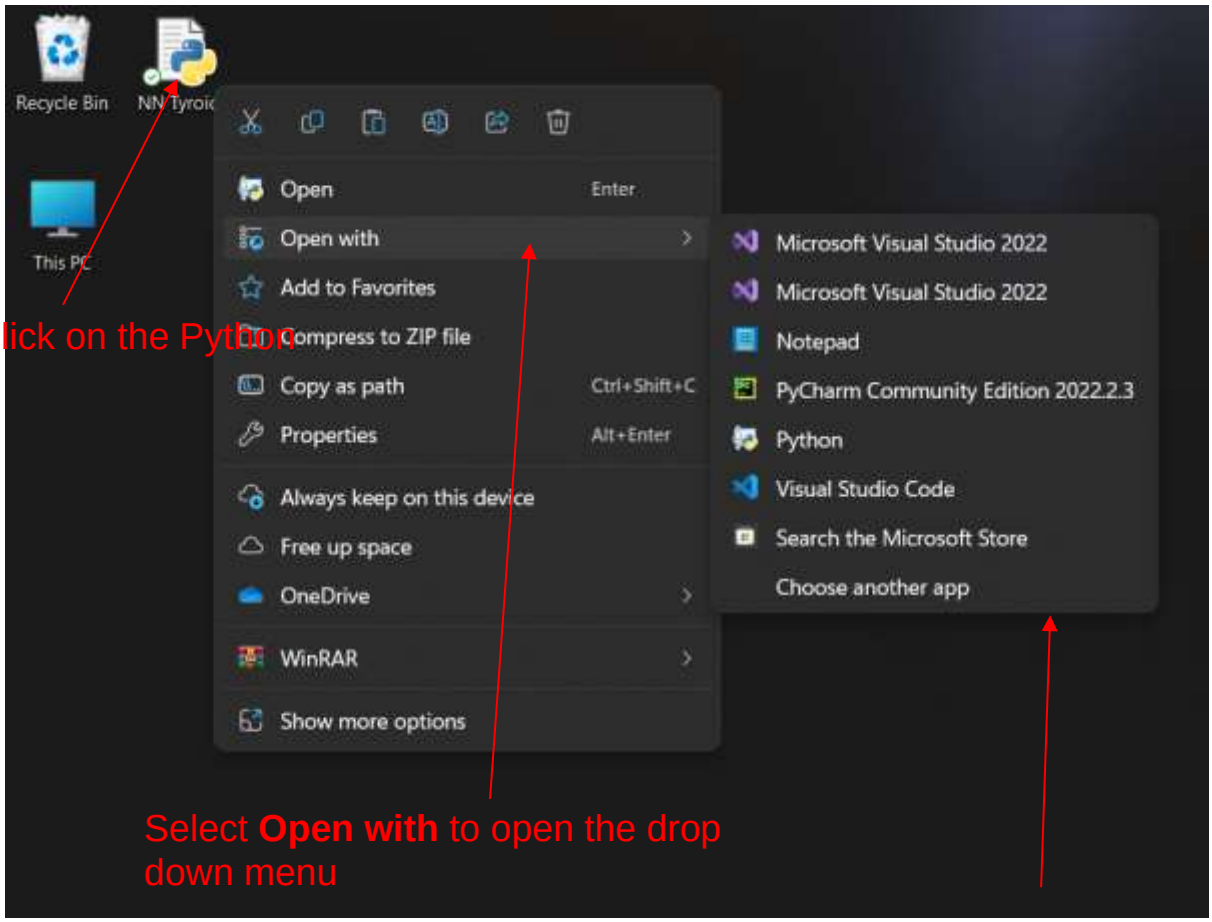
1. https://archive.ics.uci.edu/dataset/102/thyroid+disease
.
2. https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html

3. https://www.jeremyjordan.me/hyperparameter-tuning/

4. https://en.wikipedia.org/wiki/Neural_network

9. APPENDIX

User Manual

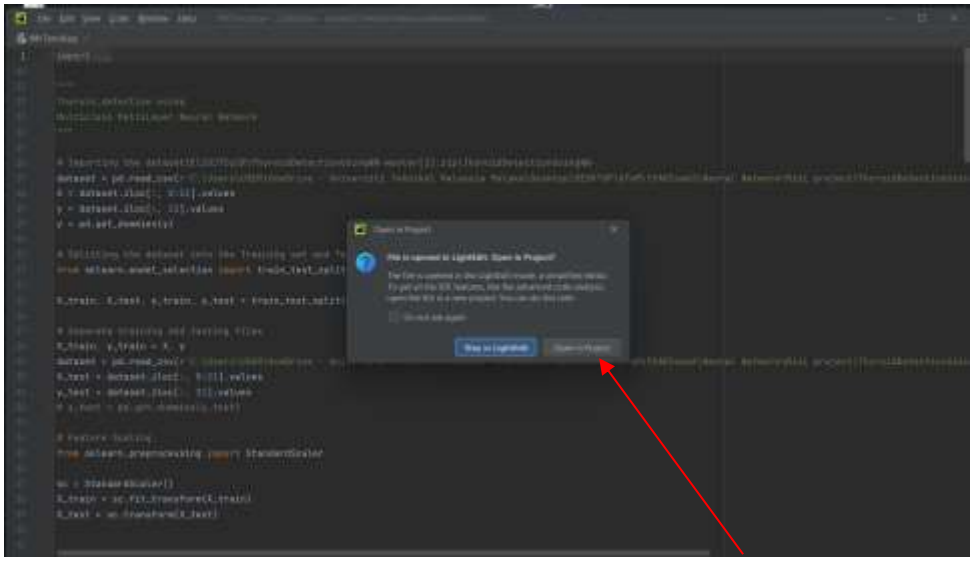


Right-Click on the Python File

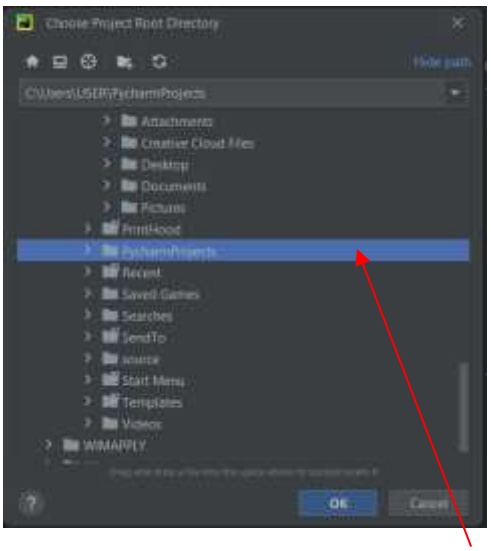
Select **Open with** to open the drop down menu

Select an Application that can run the Python file

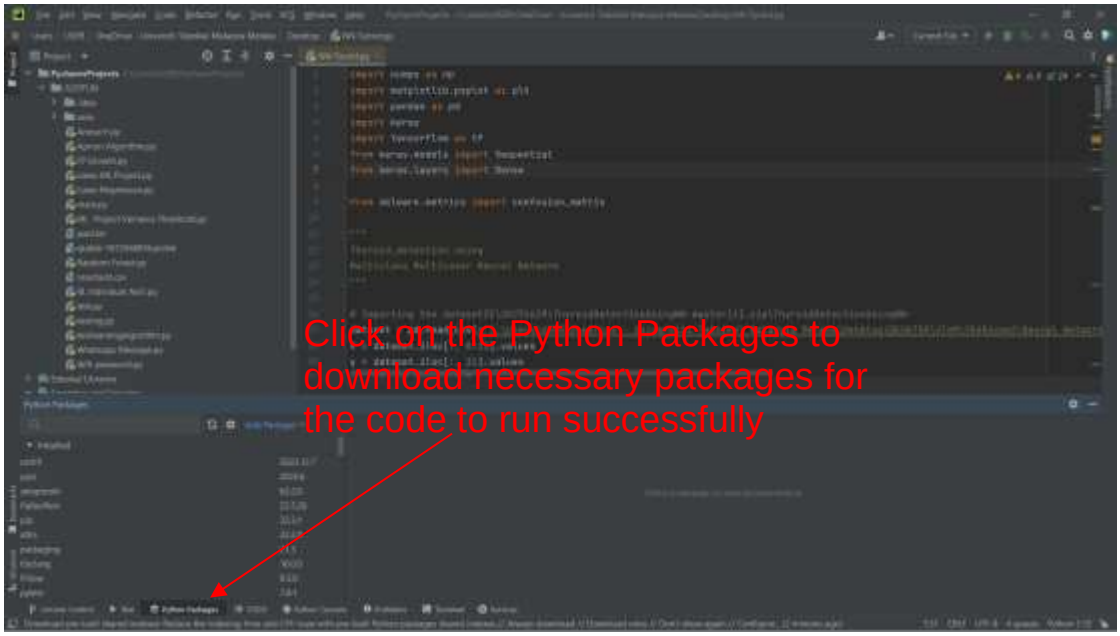
We Selected PYCHARM Application



Click **Open in Project**

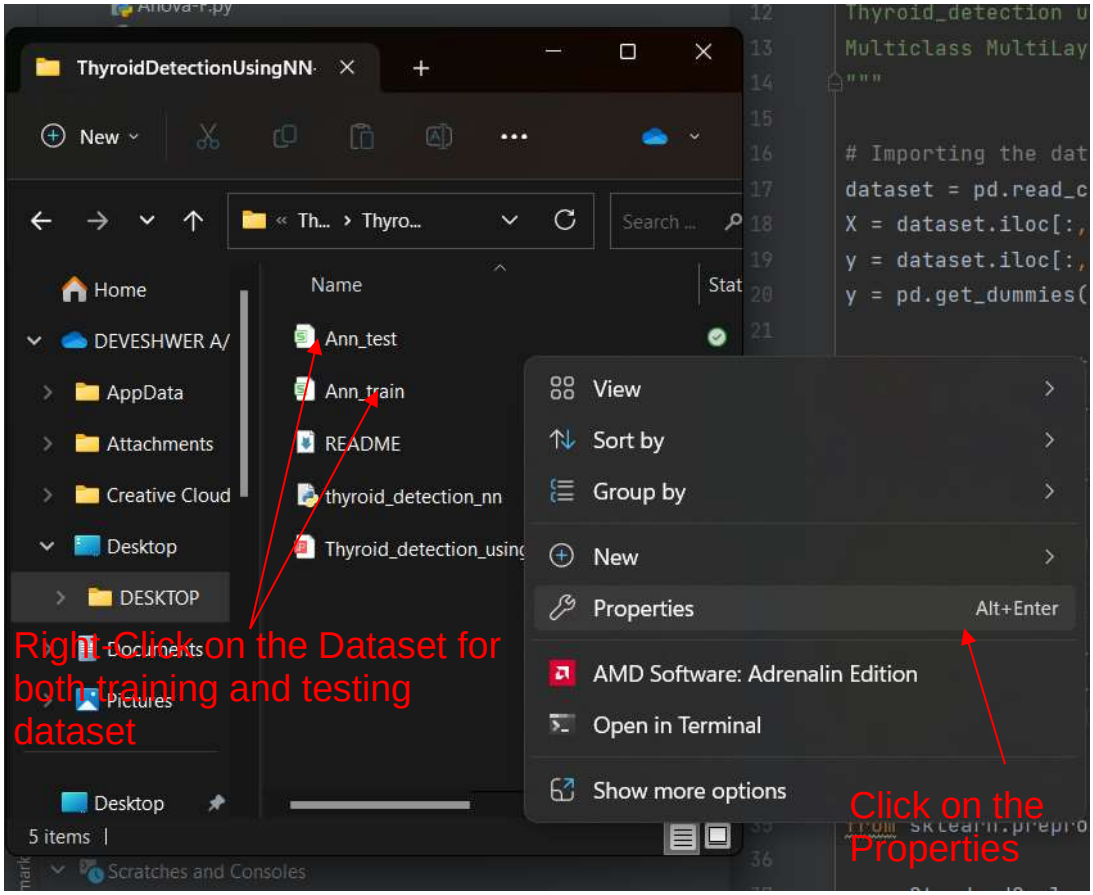


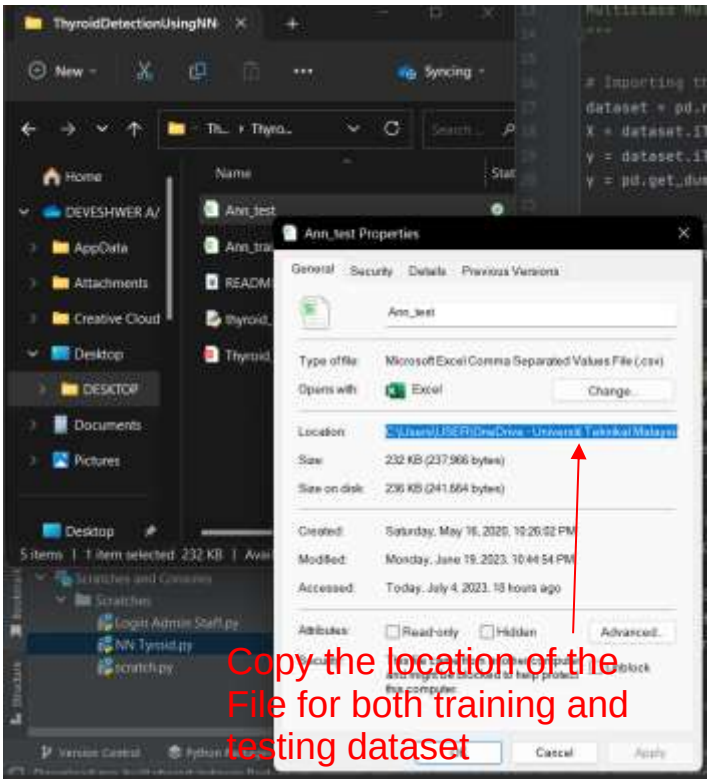
Select the Location to save the File

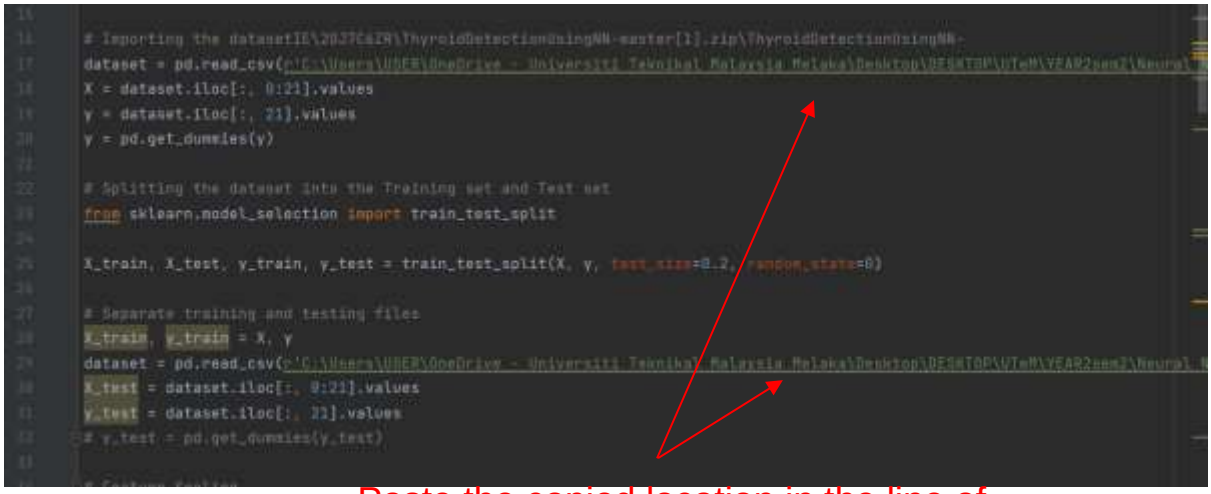


The necessary packages that should be downloaded are as follows:-

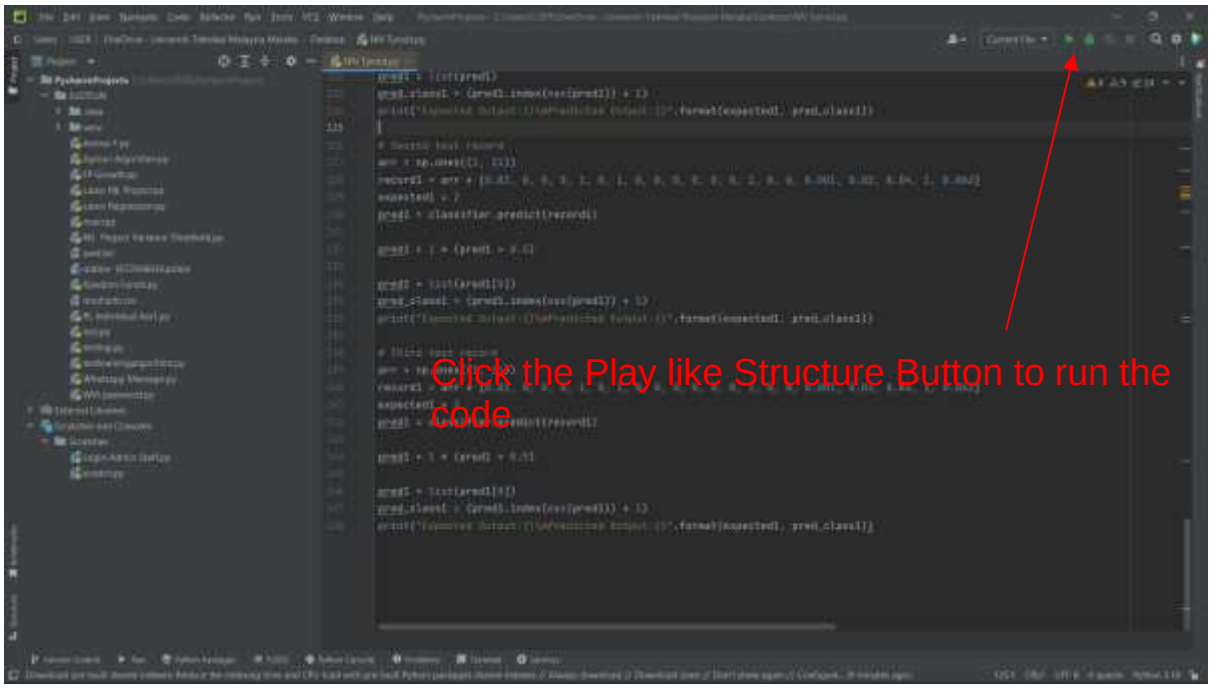
- Numpy
- Matplotlib
- Pandas
- Keras
- Tensorflow
- Sklearn


ThyroidDetectionUsingNN
New
Th... › Thyro...
Search ...
Home
DEVESHWER A/
AppData
Attachments
Creative Cloud
Desktop
DESKTOP
Documents
Pictures
Desktop
5 items
Scratches and Consoles
Name
Stat
Ann_test
Ann_train
README
thyroid_detection_nn
Thyroid_detection_using
View
Sort by
Group by
New
Properties
Alt+Enter
AMD Software: Adrenalin Edition
Open in Terminal
Show more options
Right Click on the Dataset for both training and testing dataset
Click on the Properties


Ann_test Properties
General
Security
Details
Previous Versions
Ann_test
Type of file:
Microsoft Excel Comma Separated Values File (.csv)
Opens with:
Excel
Change...
Location:
Size:
232 KB (237,966 bytes)
Size on disk:
236 KB (241,664 bytes)
Created:
Modified:
Accessed:
Today, July 4, 2023, 18 hours ago
Attributes:
Read-only
Hidden
Advanced...
Cancel
Apply
Copy the location of the File for both training and testing dataset

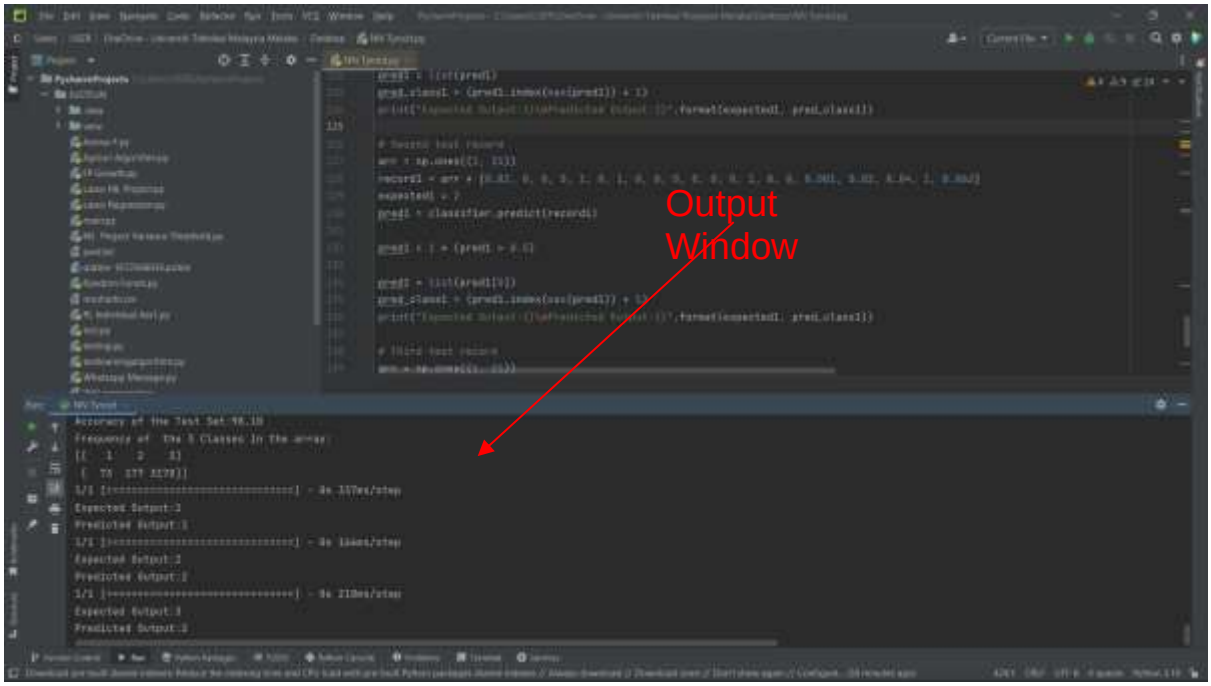


For an example, the location is **C:\Users\USER\Downloads\Ann_train.csv**



The Output will be shown as follows:-

IMPORTANT the OUTPUT may take a while



Sample Of Raw Data

- **Testing Data**

A	B	C	D	E	F	K	L	M	N	O	P	Q	R	S	T	U	V
AGE	SEX	ON_THYRO	QUERY_ON_T	ON_ANTITH	SICK	QUERY_HYPERTH	LITHIUM	GOITRE	TUMOUR	HYPOPITUITA	PSYCH	TSH	T3	TT4	T4U	FTI	Class
0.29	0	0	0	0	0	0	0	0	0	0	0	0.0061	0.028	0.111	0.131	0.085	2
0.32	0	0	0	0	0	0	0	0	0	0	0	0.0013	0.019	0.084	0.078	0.107	3
0.35	0	0	0	0	0	0	0	0	0	0	0	0	0.031	0.239	0.1	0.239	3
0.21	0	0	0	0	0	0	0	0	0	0	0	0.001	0.018	0.087	0.088	0.099	3
0.22	0	0	0	0	1	0	0	0	0	0	0	0.0004	0.022	0.134	0.135	0.099	3
0.22	0	0	0	0	0	0	0	0	0	0	0	0.0016	0.02	0.123	0.113	0.109	3
0.39	0	0	0	0	0	1	0	0	0	0	0	0.0016	0.036	0.133	0.144	0.093	3
0.77	1	0	0	0	0	0	0	0	0	0	0	0.0008	0.02	0.08	0.096	0.0832	3
0.23	0	0	0	0	0	0	0	0	0	0	0	0.0003	0.014	0.113	0.096	0.1175	3
0.23	0	0	0	0	0	0	0	0	0	0	0	0.0026	0.011	0.104	0.104	0.099	3
0.45	1	0	0	0	0	0	0	0	0	0	0	0.061	0.006	0.023	0.087	0.026	1
0.65	0	1	0	0	0	0	0	0	0	0	0	3E-05	0.023	0.154	0.09	0.17	3
0.53	0	0	0	0	0	1	0	0	0	0	0	4E-05	0.022	0.115	0.093	0.124	3
0.79	0	0	0	0	0	1	0	0	0	0	0	0.0042	0.011	0.078	0.097	0.081	3

- **Training Data**

	AGE	SEX	ON_THYRO	QUERY_ON_T	ON_ANTITH	SICK	PREGNAN	THYROID_S	I131_TREA	QUERY_HY	QUERY_HY	LITHIUM	GOITRE	TUMOUR	HYPOPITU	PSYCH	TSH	T3	TT4	T4U	FTI	Class
2	0.73	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0.0006	0.015	0.12	0.082	0.146	3
3	0.24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00025	0.03	0.143	0.133	0.108	3
4	0.47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0019	0.024	0.102	0.131	0.078	3
5	0.64	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0009	0.017	0.077	0.09	0.085	3
6	0.23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00025	0.026	0.139	0.09	0.153	3
7	0.69	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00025	0.016	0.086	0.07	0.123	3
8	0.85	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00025	0.023	0.128	0.104	0.121	3
9	0.48	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00208	0.02	0.086	0.078	0.11	3
10	0.67	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0013	0.024	0.087	0.109	0.08	3
11	0.76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0001	0.029	0.124	0.128	0.097	3
12	0.62	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0.011	0.008	0.073	0.074	0.098	2
13	0.18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.0001	0.023	0.098	0.085	0.115	3
14	0.59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0008	0.023	0.094	0.099	0.09475	3
15	0.49	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0006	0.023	0.113	0.102	0.111	3