A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Santhosh Shanmugasundaram |
| *Student Number* | Sba24100 |
| *Module Title* | Programming for AI |
| *Assessment Title* | CA1 |
| *Assessment Due Date* | 15th December 2024 |
| *Date of Submission* | 15th December 2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Contents

[1. Introduction 1](#_Toc185173509)

[2. Exploratory Data Analysis 2](#_Toc185173510)

[2.1 Understand the Dataset 2](#_Toc185173511)

[2.2 Calculate Correlation Matrix Using NumPy 2](#_Toc185173512)

[3. Dense Neural Network Model 2](#_Toc185173513)

[4. Hyperparameter Tuning 3](#_Toc185173514)

[4.1 Connect 3](#_Toc185173515)

[4.2 Database 3](#_Toc185173516)

[4.3 Insert 3](#_Toc185173517)

[5. Classification Report 3](#_Toc185173518)

[6. References 3](#_Toc185173519)

[7. GitHub Link 4](#_Toc185173520)

# Introduction

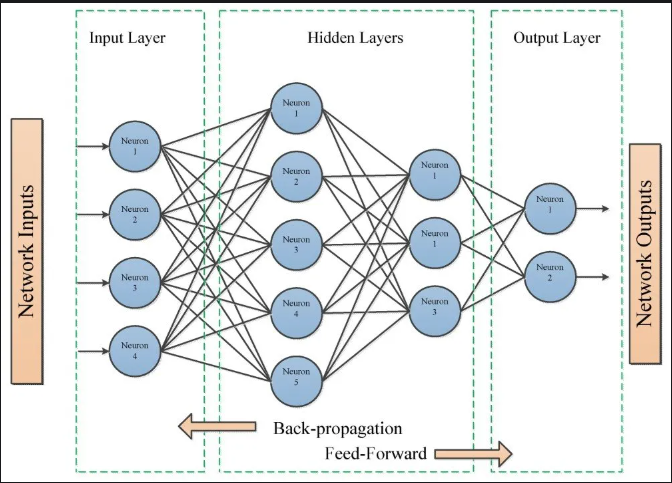
Artificial neural network is a machine learning model which is inspired by the biological neural networks of human brain on how information is processed. In a dense neural network system, there will be one input layer which receives data, one output layer that returns results and there can be 1 or many hidden layers between input and output layer. (Melanie)

Each layer is made up of neurons/nodes and each neuron/node in one layer is connected to the neurons in the previous and/or next layer. Each neurons combine inputs from a dataset with a weight factor and bias factor, to increase or decrease their value. In doing so, neural network system performs calculations.

In addition to neurons, layers, weights and bias, an activation function is performed at each node and a loss function at output nodes. Some of the most used activation functions are RELU, SIGMOID, SOFTMAX and etc. Some of the most used loss functions are ADAM, SGD.

During training of the model, forward propagation and backward propagation activities are performed which modifies the weight and bias factor consistently till model is ready. During testing/prediction phase, based on the weight and bias factors discovered earlier, only forward propagation activity is performed.

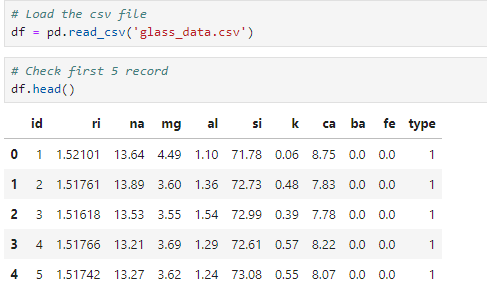
(Corpnce)



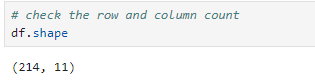
# Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a critical step in the machine learning process. EDA provides a deep understanding of the data, enabling data scientists to make informed decisions and select the right algorithms for their machine learning projects. It involves examining datasets to uncover patterns, spot anomalies, and test hypotheses before moving on to model building.   
(Tariq A)

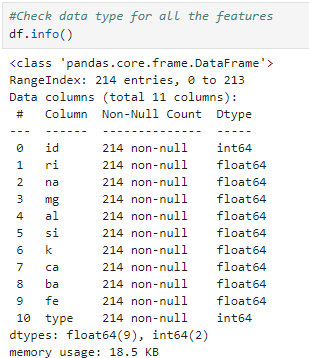
## Understand and Adjust the Dataset

Import the dataset using panda function (READ\_CSV), followed by displaying the top 5 rows of the glass dataset using dataframe function ‘head’.   


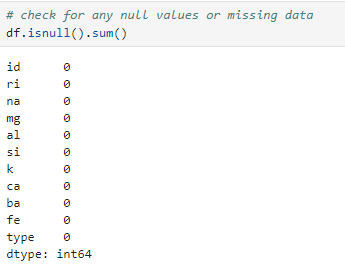
The shape of the dataset can be displayed using function ‘shape’. This helps in identifying the number of rows and columns.



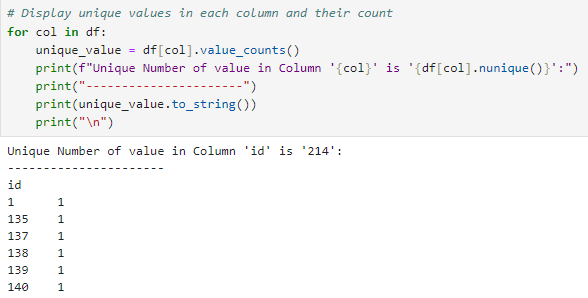
The datatypes of the columns can be checked using ‘info’ function. This tells what type of data is stored in each column.



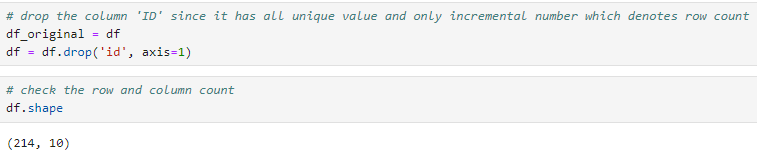
In order to check for any missing values in each column, we can use the function ‘isnull’ and to get the count of missing value, we can use function ‘sum’.



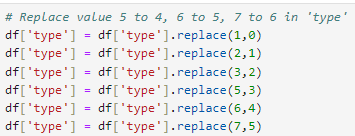
Unique values of specify column can be found using ‘unique’ function of pandas. In a very large dataset, this helps in identifying unique values in each column which is very important to list the possible values.



Checking the unique values in each column, the column ‘id’ contains all unique value in incremental order which acts similar to row number. Column with all unique values is no significance to model training, thereby, this column can be deleted which is used using ‘DROP’ function.



Checking the unique value of the dependent/output column, we noticed that the values start from 1 till 7 with no record for value 4. The loss function ‘SPARSE\_CATEGORICAL\_CROSSENTROPY’ expects the output column to have continuous numbers starting 0. Therefore, the value in the column ‘TYPE’ is replaced to start from 0 with continuous numbers till 5. The same is achieved by using the ‘REPLACE’ function. The result by model prediction needs to be converted back to the original numbers.



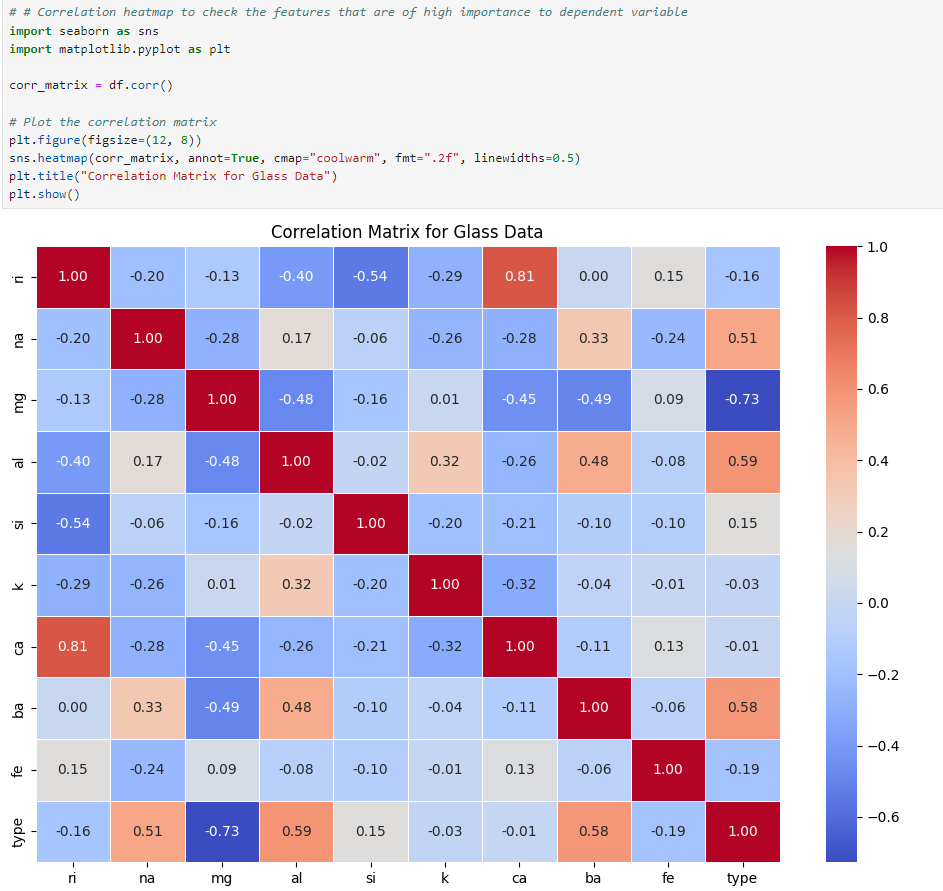
It is important to make sure that there is no duplicated rows in the dataset. In order to verify, we can use function ‘DUPLICATED’ and duplicate rows can be deleted using function ‘DROP\_DUPLICATES’



## Visualize and Examine the Relationship Between Variables

Correlation matrix helps in identifying the relationship between 2 features. Correlation value of 0 indicates no relationship between 2 features. Correlation value of 1 indicates both are strongly proportionate relationship. This means that value for both features increase or decreases at a constant ratio. Correlation value of -1 indicates that both are disproportionate to each other.

Two important libraries namely, SEABORN and MATPLOTLIB is used derive the correlation matrix.



For the given dataset, the dependent variable ‘TYPE’ has positive correlation with variables namely ‘al’, ‘ba’, and ‘na’ and strong negative correlation to ‘mg’.

# Dense Neural Network Model

**Scope**

.

# Hyperparameter Tuning

data.

## Connect

First

## Database

## Insert

Once

# Classification Report

length.

# References

(Melanie)

Melanie. “Dense Neural Networks: Understanding Their Structure and Function.” *Data Science Courses | DataScientest*, 5 Mar. 2024, datascientest.com/en/dense-neural-networks-understanding-their-structure-and-function.

Available at:

<https://datascientest.com/en/dense-neural-networks-understanding-their-structure-and-function>

(Corpnce)

Corpnce. “Corpnce.” *Corpnce*, 22 Dec. 2023, www.corpnce.com/5541-2/.

Available at:

https://www.corpnce.com/5541-2/

(Tariq A)

Tariq A. “Exploratory Data Analysis (EDA) Is a Critical Step in the Machine Learning Process. It Involves Examining Datasets to Uncover Patterns, Spot Anomalies, and Test Hypotheses before Moving on to Model Building.” *Linkedin.com*, 21 Aug. 2024, www.linkedin.com/pulse/comprehensive-guide-exploratory-data-analysis-eda-machine-tariq-ali-lnwmc/. Accessed 15 Dec. 2024.

Available at:

https://www.linkedin.com/pulse/comprehensive-guide-exploratory-data-analysis-eda-machine-tariq-ali-lnwmc/

# GitHub Link

<https://github.com/santhosh-sba24100/CA1-Python_Programming>