IFEANYI ANTHONY OKPALA - 202203449

UNDERSTANDING ARTIFICIAL INTELLIGENCE

771763-2022/2023

ASSIGNMENT: COMPONENT 1

ABSTRACT

In this report we will examine the breakdown structure of various agents and how they perform in their environment.

1.0 INTRODUCTION

A decision or action can be made by an agent based on its own evaluation of the surrounding circumstances and the data it has access to. In the case of computer agents, there is more emphasis about their performance in an environment *[Russell and Norvig, 2016].*

2.0 METHODOLOGY

This shows the breakdown structure of the agent, including pictorial representation using microsoft paint to show the agent's behaviour in an environment and some pseudocode demonstration for the agent function.

2.1 THE TASK ENVIRONMENT

Table 1. This is the list of the task environment for each agent.

AGENT TYPE	THE TASK ENVIRONMENT
Motion Detection Agent	Improve Office Productivity
Solar Power Agent	Reduce CO2 Emission Usage
Human Agent	Detect Hazard in an Environment
Motion Detection Security Agent	Prevent Physical Theft Prevention
Magnetic Resonance Imaging (MRI) Agent	Display Irregularities in the Brain
Computed Tomography (CT) Scan Agent	Display Cancer Tumors in the body

2.2 THE PEAS DESCRIPTION

Table 2. The PEAS (Performance Measures Environment Actuators Sensors)

AGENT TYPE	PERFORMANCE MEASURES	ENVIRONMENT	ACTUATORS	SENSORS
Motion Detection	The accuracy of the motion prediction.	The office space and people	Lightening and Heating systems.	Infrared and Ultrasonic

Solar Power	The efficiency of the energy from sunlight.	Sunlight and trees.	Batteries and Inverters.	Solar Panels.
Human	The accuracy in detecting hazard.	People and equipment.	Human Actions	Human Eyes Human Noise
Security Motion Detection	The accuracy of the motion prediction.	People and stationary objects.	Security light and Alarm bells.	Infrared and Ultrasonic
Brain Surgery MRI	Clear 3D image of a patient facial organs	patient head, patient facial organs	3D Image display and Sound	Fiber Optic Temperature sensors
(CT) Scan for Cancer Detection	Clear 3D image of a patient body organs	Patient body and organs	3D Image display and Sound	Image sensor

2.3 PERCEPTS OF THE AGENT

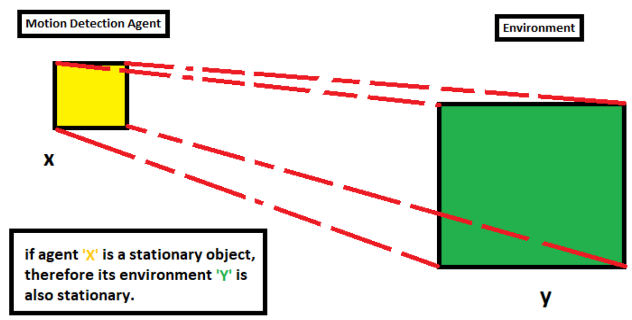
Table 3. The percept refers to the input provided to the system.

AGENT TYPE	PERCEPTS
Office motion detection agent	Human motion in the environment
Solar power agent	Sunlight intensity
Human agent	Hazard in the environment
Security motion detection agent	Human motion in the environment and time of motion
Brain surgery MRI agent	Images in the environment
CT scan for cancer detection agent	Images in the environment

2.4 AGENT FUNCTION

Table 4. Some pictorial representation of sensors behavior in an environment and a code demonstration of the agent function.

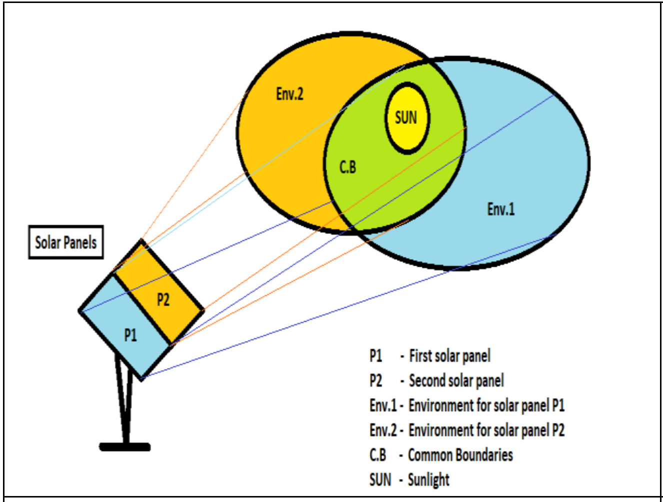
<table>
<tr><th>SENSOR AND ENVIRONMENT</th><th>PSEUDOCODE DEMONSTRATION</th></tr>
<tr><td>

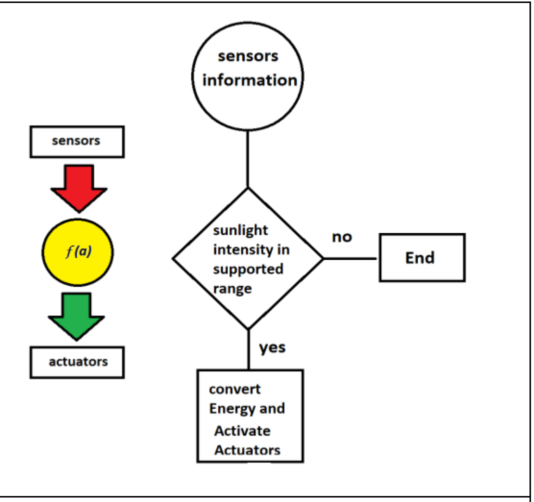


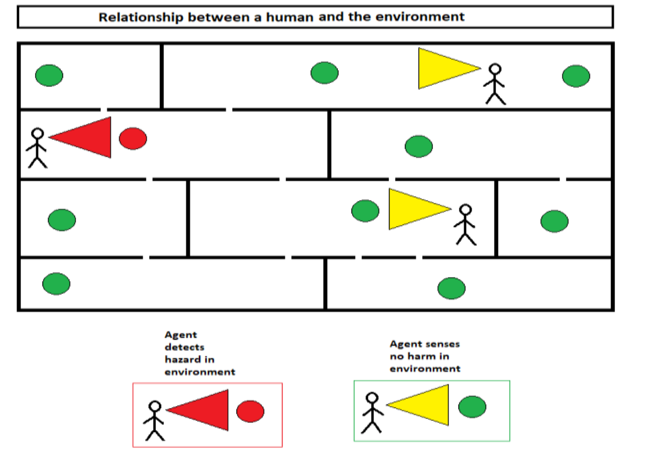
</td><td>

```
if sensor_detect_movement == True:
    RESET timer
    if actuator == OFF:
        ON actuator
elif timer == 0 and actuator == ON:
    OFF actuator
```

</td></tr>
</table>



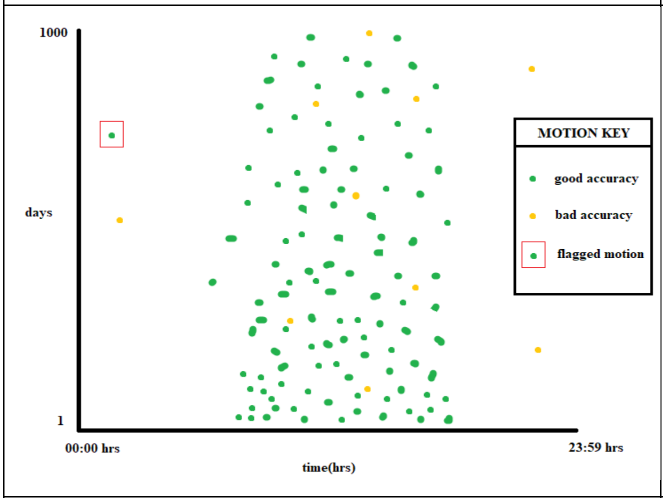


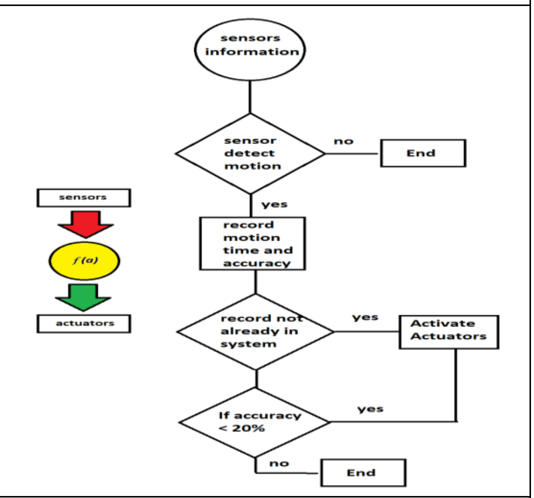


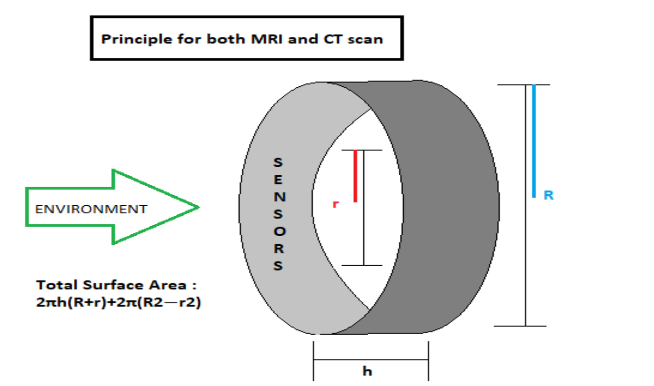

```
Def human ():

    If hazard == True:
        Activate Actuators

human ()
```








```
Def MRI_CT ():
    issue = []
    if sensor_detect_issue == True:
        issue.append(image_display)
    return issue # To be sent to actuators

Def MRI_CT ()
```

2.5 EXTERNAL STIMULI

Table 5. External stimulus (actions) could affect (trigger) the behaviour of the agents.

AGENT TYPE	EXTERNAL STIMULI
Motion Detection Agent for Office Productivity	Obstruction and Rapid Distance
Solar Power Agent for Climate Change Solutions	Weather condition Trapped dirt on the solar panel
Human Agent for Hazardous Environment	Poor sight to see clearly and Emotional stimuli
Motion Detection Security Agent for Physical Theft Prevention	Camera placement and Movement in the environment
Magnetic Resonance Imaging (MRI) for Brain Surgery	Metal objects, Patient movement, Patient size
Computed Tomography (CT) Scan for Cancer Detection	Bad electronic component (IC), Patient size, Patient movement

2.6 PROPERTIES OF THE TASK ENVIRONMENT

Table 6. The properties of the task environment of the agents

TASK ENVIRONMENT	OBSERVABLE	AGENTS	DETERMINISTIC	EPISODIC	STATIC	DISCRETE
Motion Detector	Fully observable	Single agent	Deterministic	Episodic	Dynamic	Discrete
Solar Power System	Fully observable	Single agent	Non-deterministic	Episodic	Dynamic	Continuous
Human	Partially observable	Multi-agent	Non-deterministic	Episodic	Dynamic	Discrete
Security Motion Detector	Fully observable	Single agent	Deterministic	Sequential	Dynamic	Discrete
Brain MRI	Fully observable	Single agent	Deterministic	Episodic	Static	Discrete
CT Scan Cancer Detector	Fully observable	Single agent	Deterministic	Episodic	Static	Discrete

2.7 AGENT STRUCTURE

Table 7. Agent program category

AGENT TYPE	AGENT PROGRAM
Motion Detection Agent for Office Productivity	SIMPLE REFLEX AGENT
Solar Power Agent for Climate Change Solutions	SIMPLE REFLEX AGENT
Human Agent for Hazardous Environment	LEARNING AGENT
Motion Detection Security Agent for Physical Theft Prevention	MODEL-BASED FLEX AGENT
Magnetic Resonance Imaging Agents for Brain Surgery	SIMPLE REFLEX AGENT
Computed Tomography (CT) Scan for Cancer Detection	SIMPLE REFLEX AGENT

2.8 RESULT/JUSTIFICATION

The performance measures for humans and motion detection for both office productivity and security depend on the accuracy to ensure correct actions by the actuators while MRI and CT depend on clarity for easy operation of the patient. The sensor receives percepts from the environment and gives output to the agent function, while the agent function does the mathematical computation that activates the actuators.

For both motion detection agents, the sensor is interested in the percept involving human motion to feed the agent function for an accurate output. In the case of human and solar power system agent, we are looking at the hazard and sun intensity respectively as they are the requirement needed for the agent function after detected by a sensor. The MRI and CT agent percepts are images of the environment detected by the sensors for the agent function to take further action.

Motion detectors, solar power system, MRI and CT scan agent are all fully observable because they can see the whole environment at every instant, and they are static. These set of agents are also single agent as they all operate alone in the environment. Human on the other hand is not static and they could be multiple agents in the environment. For this purpose, human and solar power system are non-deterministic because they are independent of the environment previous state while the rest of the agent are. Due to there action from previous step, all are episodic except security motion detector being sequential. MRI and CT environment does not change unlike others. All are discrete except solar power system due to their finite or infinite steps.

3.0 CONCLUSION

We were able to describe what constituted in our agents such as the percepts, task environment, PEAS, environment, and properties of the task environment.

ASSIGNMENT: COMPONENT 2

ABSTRACT

In order to prevent serious environmental implications in the future, CO2 emission must be tackled. In this report, our aim is to see how to predict the CO2 emission of vehicles in a dataset and find steps to classify the vehicles based on the categorical variables.

1.0 INTRODUCTION

Carbon dioxide (CO2) is produced through the production and burning of fossil fuels like coal, oil, and natural gas as well as during wildfires and other natural processes *[NASA, 2017]*. The amount of CO2 emitted by a vehicle depends on several factors which we will examine using various models.

2.0 METHODOLOGY

2.1 STEPS REQUIRED TO TRAIN A MODEL

There are various steps required for us to start training testing and validating models as shown below:

- **Data cleaning and preparation** is the process of fixing or removing incorrect data within a dataset.
- **Data visualization and analysis** gives a true representation of our dataset. Using Exploratory Data Analysis (EDA), we can detect variable performance which is necessary for feature selection.
- **Feature selection** is the process of selecting a subset of relevant features for use in model construction.
- **Model selection** is the process of building a machine learning model best fit for a task.

2.2 REGRESSION MODEL SELECTION PROCESS AND ANALYSIS

It is important to know that not all numerical variables are continuous. After selecting my numerical continuous variables, I decided to scale the data before imploring a model.

Using Multiple linear regression model to predict CO2 emission, the root mean squared error RMSE was **23.5**. The lower the RMSE the more accurate the model. Comparing RMSE for different regression models using cross validation we were able to see that the decision tree regressor performance was the best as shown in figure 1 below.

	Model	cv_score 1	cv_score 2	cv_score 3	cv_score 4	cv_score 5	avg_score
0	Linear_R	23.184607	24.847968	24.909708	23.021667	23.427049	23.878200
1	Dtree	12.387146	12.203859	8.318004	5.946689	8.611148	9.493369
2	SVR	26.258449	27.391193	28.665844	25.727745	25.377219	26.684090

Figure 1. *Cross validation RMSE for regression models before EDA and Feature selection.*

After undergoing Exploratory Data Analysis, I decided to drop either the Fuel_Consumption_COMB (L/100 km) or Fuel_Consumption_City (L/100 km) since they had a perfect correlation as shown in figure 2.

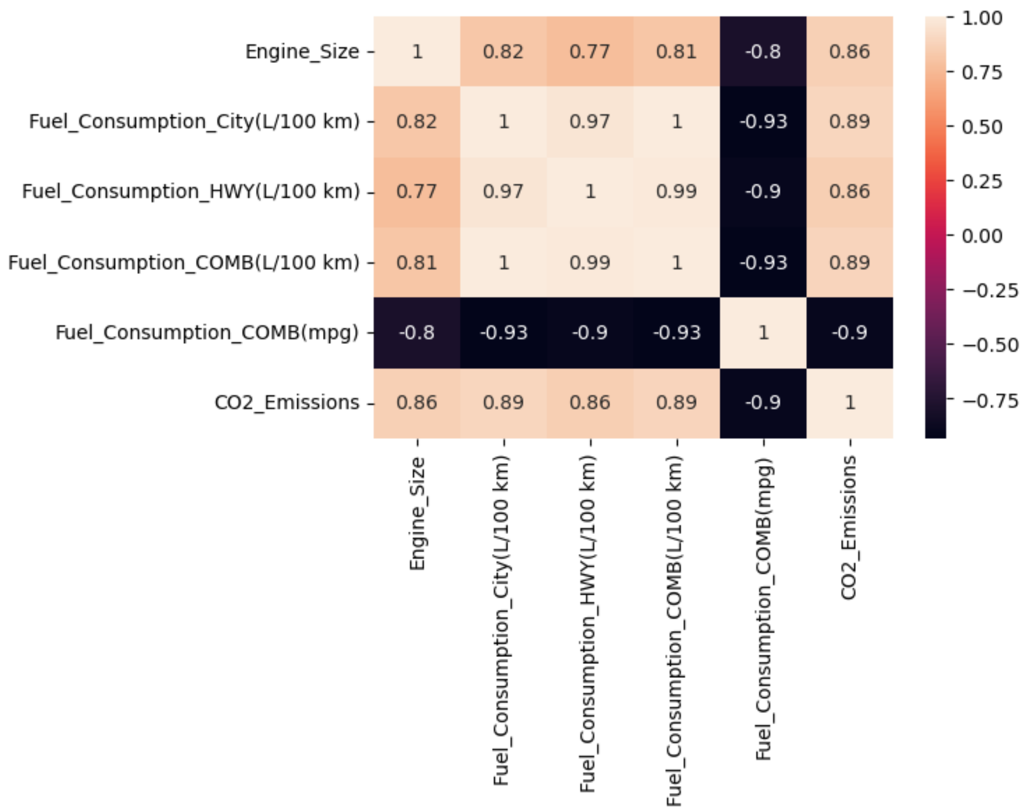


Figure 2. *Exploratory data analysis (EDA) using heatmap.*

A new subset was created to repeat the procedure after dropping Fuel_Consumption_City (L/100 km), the result is shown in figure 3 below.

	Model	cv_score 1	cv_score 2	cv_score 3	cv_score 4	cv_score 5	avg_score
0	Linear_R	23.181949	24.844155	25.004827	23.058002	23.524561	23.922699
1	Dtree	10.706445	12.904897	6.432219	6.379179	8.124584	8.909465
2	SVR	25.736334	26.763897	28.040629	25.527637	24.933863	26.200472

Figure 3. *Cross validation RMSE for regression models after EDA and Feature selection.*

From the above comparison, the decision tree regressor is the best model as clearly shown in the visualization in figure 4 below.

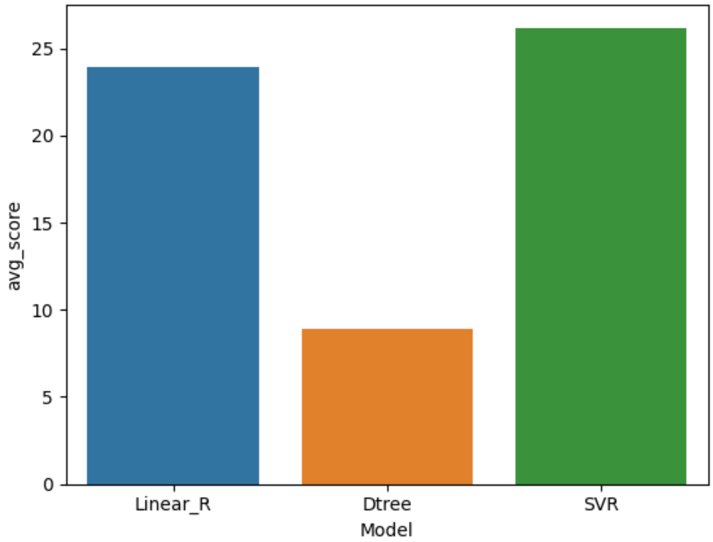


Figure 4. *Bar plot showing the regression models performance.*

In terms of the performance, we can see that the RMSE is **slightly more accurate from 9.49 to 8.91** for the decision tree regressor after using a subset of the input data. This performance was further improved to **7.92** using Hyperparameter.

2.3 CHANGE IN CO2 EMISSION FROM YEAR 2010 TO YEAR 2014

To know the change in CO2 Emission from the year 2010 to 2014, we will need to plot both data for the model year and CO2 Emissions. From the average CO2 Emissions per year between 2010 to 2014 as shown in figure 5, there is a decrease in CO2_Emissions from 279 to around 255, respectively.

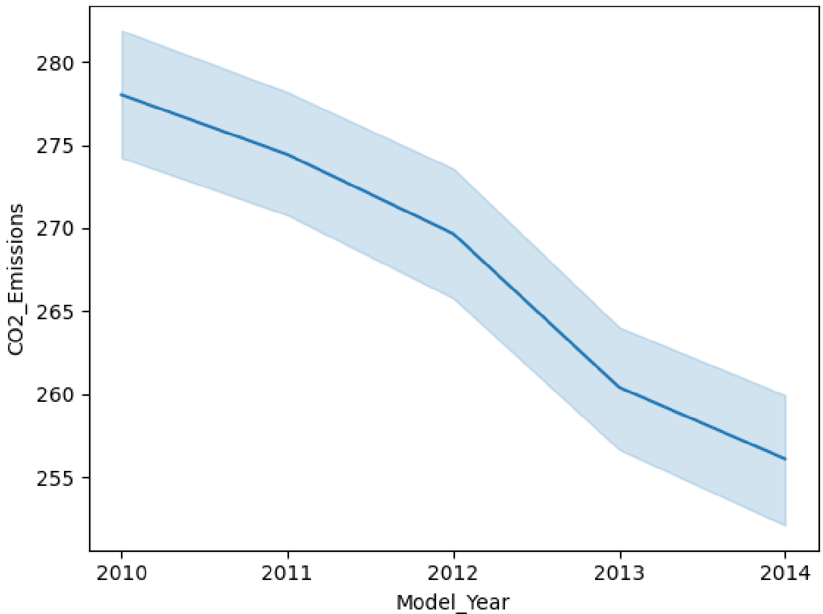


Figure 5. *Average CO2 Emissions per year between 2010 to 2014.*

2.4 CATEGORICAL VARIABLE COMPARISON BY CLASSIFYING THE DATASET

For us to determine the variable that performs best in classifying the dataset, we will first choice a model by using cross validation on a single categorical variable e.g., ‘Fuel’. Synthetic Minority Oversampling Technique was first applied to our categorical variable before training and cross validation. After applying cross validation, it is observed that Random Forest Classifier (RF) performed best as shown in figure 6 below.

model	cv score1	cv score1	cv score1	cv score1	cv score1
LR	0.760997	0.749817	0.785033	0.776229	0.827586
KNN	0.8761	0.856933	0.868672	0.870873	0.866471
Dtree	**0.934751**	**0.942773**	**0.953778**	**0.937638**	**0.949376**
SVC	0.359971	0.336757	0.338225	0.362436	0.376376
RF	0.942815	0.946442	0.954512	0.949376	0.955246

Figure 6. *Cross validation accuracy for classification models*

Selecting RF as the model, there was a need to create a function that accepts two categorical data X_train and y_train which returns both test data and predicted data. From figure 7 below we can see the Confusion Matrix and accuracy of the categorical data. The **‘Fuel’ variable seems to have the best classification accuracy of 98%** and it takes less computational power due to the low number of categories.

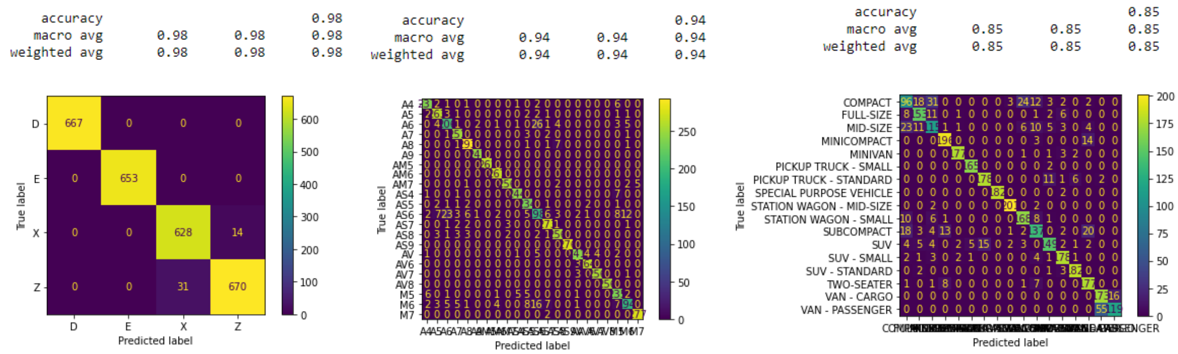


Figure 7. *Shows the confusion matrix and classification accuracy of the categorical data.*

2.5 OVERFIT CHECK

Over fitting check was done in both our regession and classification model . **Cross Validation Check** was the way I checked for over fitting. The model is probably overfitting if it performs noticeably better on training data than on the validation data.

2.6 PERFORMANCE MEASURE(S)

- For Regression, **root mean squared error** was prefered because it mantains the original data unit making it easy to interpret.
- For Classification, **Accuracy** was perfered because it is easy to understand and we have a balanced dataset.

2.7 MODEL DEPLOYABILITY

The models are deployable for both regression and classification. In both cases, the selection of the models was done using cross validation and the results were accurate and reliable. The regression RMSE was 7.92 and classification had an accuracy of 98%.

2.8 CATEGORICAL VARIABLE DESCRIPTION USING INTERNAL AND EXTERNAL EVALUATION METRICS

For internal evaluation metrics, it is using data that was used to train the model. For the evaluation, I made use of both Davies-Bouldin score and Silhouette score to determine the number of clusters. Elbow method was also used to visually confirm the optimal number of clusters needed.

	K	silhouette_score	davies_bouldin_score	avg_score
0	2	0.567549	0.608284	0.222318
1	3	0.543206	0.558200	0.239988
2	4	0.527002	0.546258	0.239123
3	5	0.504279	0.582376	0.210599
4	6	0.518497	0.563652	0.226245
5	7	0.513542	0.556114	0.227954

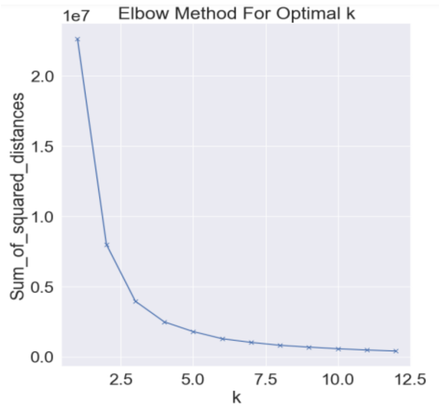


Figure 8. *Internal Evaluation Outcomes.*

For External evaluation metrics, it is using data that was not used to train the model. V-measure score was used as an external evaluation metric for a clustering algorithm as shown in figure 9. The highest score is the Model.

object variable	rand_score	v_measure_score
Model_Year	0.768387	0.013544
Make	0.912525	0.276378
Model	**0.945891**	**0.512122**
Vehicle_Class	0.872974	0.231854
Transmission	0.84806	0.176471
Fuel	0.594292	0.200315

Figure 9. *External evaluation metrics using V-measure score outcomes and rand_score.*

3.0 CONCLUSION

We were able to predict the CO2 emission of vehicles in a dataset and find steps to classify the vehicles based on the categorical variables.

REFERENCES

NHS Choices (2021). *CT scan*. [online] NHS. Available at: https://www.nhs.uk/conditions/ct-scan/.

Russell, S.J. and Norvig, P. (2016). *Artificial intelligence: a modern approach*. Upper Saddle River: Pearson.

NASA (2017). Carbon Dioxide Concentration | NASA Global Climate Change. [online] Climate Change: Vital Signs of the Planet. Available at: https://climate.nasa.gov/vital-signs/carbon-dioxide/.