AIML Online

Project Milestone – Unsupervised Learning

This is a comprehensive project for all the topics you have learned so far from all the modules. This project will give a good grip on most of Unsupervised Learning techniques you have learned in this module.

Criteria	Pts	Week	Topic
Part 1.1 a - Read 'Car name.csv' as a DataFrame and assign it to a variable.	1 pts	Python	Pandas functions
Part 1.1 b - Read 'Car-Attributes.json as a DataFrame and assign it to a variable.	1 pts	Python	Pandas functions
Part 1.1 c - Merge both the DataFrames together to form a single DataFrame	2 pts	Python	Pandas functions
Part 1.1 d - Print 5 point summary of the numerical features and share insights.	1 pts	EDA	Understand the distribution of your data
Part 1.2 a - Check and print feature-wise percentage of missing values present in the data and impute with the best suitable approach.	2 pts	Data Pre-processing	Missing value treatment
Part 1.2 b - Check for duplicate values in the data and impute with the best suitable approach.	1 pts	Data Pre-processing	Checking duplicate rows in the data
Part 1.2 c - Plot a pairplot for all features.	1 pts	EDA	Multivariate Analysis
Part 1.2 d - Visualize a scatterplot for 'wt' and 'disp'. Datapoints should be distinguishable by 'cyl'.	1 pts	EDA	Bivariate Analysis
Part 1.2 e - Share insights for Q2.d.	1 pts	EDA	Your observations
Part 1.2 f - Visualize a scatterplot for 'wt' and 'mpg'. Datapoints should be distinguishable by 'cyl'.	1 pts	EDA	Bivariate Analysis
Part 1.2 g- Share insights for Q2.f.	1 pts	EDA	Your observations
Part 1.2 h - Check for unexpected values in all the features and datapoints with such values.	2 pts	Data Pre-processing	Anomaly or unexpected value detection
Part 1.3 a - Apply K-Means clustering for 2 to 10 clusters.	3 pts	Week-1	K-Means Clustering
Part 1.3 b - Plot a visual and find elbow point.	2 pts	Week-1	K-Means Clustering
Part 1.3 c - On the above visual, highlight which are the possible Elbow points.	1 pts	Week-1	K-Means Clustering, Elbow plot
Part 1.3 d - Train a K-means clustering model once again on the optimal number of clusters.	3 pts	Week-1	K-Means Clustering

Criteria	Pts	Week	Topic
Part 1.3 e - Add a new feature in the DataFrame which will have labels based upon cluster value	2 pts	Week-1	Pandas functions, K-Means Clustering
Part 1.3 f -Plot a visual and color the datapoints based upon clusters.	2 pts	Week-1	Visualizing built Clusters on the data
Part 1.3 g - Pass a new DataPoint and predict which cluster it belongs to.	2 pts	Week-1	K-Means Clustering
Part 2. 1 a - Read 'vehicle.csv' and save as DataFrame.	1 pts	Python	Pandas functions
Part 2.1 b- Check percentage of missing values and impute with correct approach.	1 pts	Data Pre-processing	Missing value treatment
Part 2.1 c- Visualize a Pie-chart and print percentage of values for variable 'class'.	2 pts	EDA	Univariate Analysis of the column 'class'
Part 2.1 d - Check for duplicate rows in the data and impute with correct approach.	1 pts	Data Pre-processing	Checking duplicate rows in the data
Part 2.2 a - Split data into X and Y.	1 pts	Modelling	Separating dependent and Independent variables
Part 2.2 b - Standardize the Data.	1 pts	Data Pre-processing	Z score scaling/ Standardizing
Part 2.3 a - Train a base Classification model using SVM.	1 pts	Revision-Supervised Learning	SVM model building
Part 2.3 b - Print Classification metrics for train data.	1 pts	Revision-Supervised Learning	Performance metrics evaluation
Part 2.3 c - Apply PCA on the data with 10 components.	3 pts	Week 2	Principal Components Analysis
Part 2.3 d - Visualize Cumulative Variance Explained with Number of Components.	2 pts	Week 2	Principal Components Analysis
Part 2.3 e - Draw a horizontal line on the above plot to highlight the threshold of 90%.	1 pts	Week 2	Principal Components Analysis
Part 2.3 f - Apply PCA on the data. This time Select Minimum Components with 90% or above variance explained.	2 pts	Week 2	Principal Components Analysis
Part 2.3 g - Train SVM model on components selected from above step.	1 pts	Revision-Supervised Learning	PCA, SVM
Part 2.3 h - Print Classification metrics for train data of above model and share insights.	2 pts	Revision-Supervised Learning	Performance metrics evaluation
Part 2.4 a - Train another SVM on the components out of PCA. Tune the parameters to improve performance.	2 pts	Revision-Supervised Learning	PCA, SVM, Hyper parameters tuning
Part 2.4 b- Share best Parameters observed from above step.	1 pts	Revision-Supervised Learning	Hyper parameter tuning

Criteria	Pts	Week	Topic
Part 2.4 c- Print Classification metrics for train data of above model and share relative improvement in performance in all the models along with insights.	2 pts	Revision-Supervised Learning	Performance metrics evaluation
Part 2.5 a - Explain pre-requisite/assumptions of PCA.	2 pts	Week 2	Need of PCA for your data
Part 2.5 b - Explain advantages and limitations of PCA.	3 pts	Week 2	PCA

Wishing the best for every step in this journey!

****************************Happy Learning****************************