GROUP ASSIGNMENT

BEJ42803 ARTIFICIAL INTELLIGENCE

MACHINE LEARNING TECHNIQUES FOR PATTERN RECOGNITION APPLICATIONS

A. Problem Statement

The problem statement is based on the Shinkansen Bullet Train in Japan, and passengers' experience with that mode of travel. This machine learning exercise aims to determine the relative importance of each parameter with regards to their contribution to the passengers' overall travel experience.

B. Goal

The objective of this problem is to understand which parameters play an important role in swaying passenger feedback towards a positive scale. The goal of the problem is to **predict whether a passenger was satisfied or not** with his/her experience of travelling on Shinkansen Bullet train.

C. Dataset

The dataset contains a random sample of individuals who traveled on this train. The dataset consists of information related to passengers and attributes related to Shinkansen train. and the post-service experience. Each passenger was explicitly asked whether they were satisfied with their overall travel experience or not, and that is captured in the data of the survey report under the variable labeled **'Overall_Experience'**.

INSTRUCTIONS

1. **Load** the dataset using Pandas in Jupyter Notebook.
2. **Understand the data** (check for each of the following in both the train & test dataset)
   a. Check the head and tail of the data.
   b. Use the info() and describe() functions for more information
   c. Look for the presence of null values in the dataset.
3. **Preprocessing**
   a. Remove the unsignificant parameters (ex: ID)
   b. Encode the categorical object variables in both the train & test set
   c. Separate the dataset into Training dan Testing data.

4. **Perform model building** using any algorithm you feel will suit this problem:
   a. Fit the model on the training data.
   b. Make predictions on the testing data.
   c. Store the predicted values in an array.
5. Now **go back to Step 4** (this is an iterative process)
   a. Check to see if scaling the data helps with performance.
   b. Check to see if other models help with performance.
   c. Try removing unnecessary variables & check to see if that helps with performance (if necessary)
6. **Show the output of the model** which includes the overall accuracy, precision, recall and f1-score. Provide supported figures (eg: models, layers, validation accuracy graph, validation loss graph, confusion matric table, etc)
7. **Analyze and discuss** your result.
8. **Write a short repo**rt which consists of all the **details required from 1-7. Cite and provide the references** at the end of your report.

B. ASSESSMENT (TOTAL 20%)

1. ***Final Report (10%) – Week 14 (8th July 2023)***
   - Title
   - Introduction
   - Problem Statement
   - Methodology
   - Dataset Descriptions
   - Results
   - Analysis and Discussion
   - Conclusion
   - References
   - Not more than 10 pages

2. ***Output File (5%)-Week 14(8th July 2023)***
   - Save output file in .html
   - Write necessary comments and descriptions of output

3. ***Rubrics Evaluations***

Final Report 10%

Task	Mark distributions					
CLO3 (LOD5-PLO5-P4) Final Report 5%	0	[1 - Very poor]	[2 - Poor]	[3 - Average]	[4 - Good]	[5 - Excellent]
	Task is not performed	Less evidence to assume works is perform due to limited analysis and inappropriate project design flow	Acceptable evidence to assume works is perform due to limited analysis and inappropriate project design flow	Acceptable evidence to assume works is perform with sufficient analysis and proper project design flow	Clear evidence to assume works is perform with sufficient analysis and proper project design flow	Very clear evidence to assume works is perform with sufficient analysis and proper project design flow
CLO3 (LOD5-PLO5-P4) Selection of Tools 5%	0	[1 - Very poor]	[2 - Poor]	[3 - Average]	[4 - Good]	[5 - Excellent]
	Task is not performed	Inappropriate methods, no justification and not suitable for the selected AI applications	Less acceptable methods, not relevant justification and less suitable for the selected AI applications	Acceptable methods, relevant justification and suitable for the selected AI applications	Appropriate methods, valid justification and suitable for the selected AI applications	Highly appropriate methods, valid justification and suitable for the selected AI applications

Output File (.html) 5%

Task	Mark distributions					
CLO4 (LOD15-PLO8-A3) Outcome 3%	0	[1 - Very poor]	[1.5 - Poor]	[2.0 - Average]	[2.5 - Good]	[3 - Excellent]
	No outcome and unable to performed task ethically	Shows inappropriate outcome, as well as not being efficient and demonstrate inappropriate ethical work.	Shows limited outcome, as well as being less efficient and demonstrate acceptable ethical work.	Shows outcome, as well as being efficient and demonstrate ethical work.	Shows good outcome, as well as being efficient and demonstrate ethical work.	Always shows excellent outcome, as well as being efficient and demonstrate good ethical work.
CLO4 (LOD15-PLO8-A3) Group task management 2%	0	[0.5 - Poor]	[1.0 - Average]	[1.5 - Good]	[2 - Excellent]	
	No work integrity and bad behaviour in carrying out the task ethically	Less appropriate work integrity and inconsistent behaviour in carrying out the task ethically	Acceptable present work integrity and good behaviour in carrying out the task ethically	Able to present work integrity and good behaviour in carrying out the task ethically	Highly present work integrity and good behaviour in carrying out the task ethically	