A logo of a company

Description automatically generated A close-up of a logo

Description automatically generated

**Joint Tech Internship Community Program**

**Assignment1**

**Dead Line for Submission:** 26th July 2024

For Questions, write to [internship@genaicm.com](mailto:internship@genaicm.com)

**Step1)** You don’t need to code anything. The first assignment is a documentation work. You need to spend some time to understand following list of terminologies/concept in Machine Learning, by yourself. You can use Books, Google, Youtube or anything that you feel convenient to help you understand them. After you’ve fully understood, you need to write one example test data (in table format, with very few records 5 or 6) and walk through the test data to clarify each terminology.

List of Terminologies:

* Feature
* Label
* Prediction
* Outlier
* Test Data
* Training Data
* Model
* Validation Data
* Hyperparameter
* Epoch
* Loss Function
* Learning Rate
* Overfitting
* Underfitting
* Regularization
* Cross-Validation
* Feature Engineering
* Dimensionality Reduction
* Bias
* Variance

**Example Assignment Format:**

Here is a sample test data of used car’s resale price history.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Make Year** | **Brand** | **Variant** | **Mileage** | **Fuel** | **Transmission** | **Resale Price (INR)** |
| 2015 | BMW | 520D | 80000 | Diesel | Automatic | 2500000 |
| 2016 | Audi | A6 | 92000 | Petrol | Automatic | 1900000 |
| 2018 | Mercedes Benz | E200 | 61000 | Petrol | Automatic | 3400000 |
| 2014 | Skoda | Superb | 95000 | Petrol | Automatic | 600000 |
| 2020 | Benz | E200 | 35000 | Petrol | Automatic | 12000000 |

In the above Example Training Data Set,

* **Feature**: Individual measurable properties (e.g., Make Year, Brand , Variant etc) used as inputs to the model.
* **Label**: The output variable that the model aims to predict (e.g., Resale Price).
* **Outlier**: A data point that deviates significantly from the rest of the data (e.g., if there was a car with resale price at 12000000in this dataset).
* **Training Data:**…..?

**…**…

**…….**

**….. (all remaining Terminologies)**

**Put everything in Microsoft Word and save.**

**Step 2)** Create a Github Project (if you don’t have Github account, ask your friends how to get there)

Upload the Microsoft Word Document into the Github

**Step 3)** Once uploaded, click on following link to respond that your Assignment 1 is completed [**https://forms.gle/QTiuUux7Kxvo3gQAA**](https://forms.gle/QTiuUux7Kxvo3gQAA)

**Step 4)** Once you’re done with the assignment, time to showcase what you have learnt! Request you to post the highlights of this assignment in LinkedIn and tag **@Generative AI Consortium**

**Note:**

1. The Assignmen1 submission is mandatory to complete the internship
2. Plagiarism will be rejected.
3. Do not use AI tools to generate answers, we need to you understand and do this by youso you can do the Assignment 2 later by yourself. Ask question if you have doubts.