

Assignment - Sampling

General Instructions – Must Read

- **Submission Guidelines**
 - Upload the assignment to a GitHub repository and submit the repository link through the Google Form.
 - **Zero marks** will be awarded if the code or results are found to be plagiarized.

Program 1

Objective

The objective of this assignment is to understand the importance of sampling techniques in handling imbalanced datasets and to analyze how different sampling strategies affect the performance of various machine learning models.

Problem Statement

You are given a highly imbalanced credit card dataset. In real-world applications, such imbalance can significantly affect model performance. Your task is to balance the dataset, apply different sampling techniques, and evaluate how these techniques influence the accuracy of multiple machine learning models.

Program need to perform the following task:

1. **Dataset**
Download the dataset from the following GitHub link:
 https://github.com/AnjulaMehto/Sampling_Assignment/blob/main/Creditcard_data.csv
2. Convert this data-set into balanced class data-set.
3. Create five samples
4. Apply five different sampling techniques (Sampling1, Sampling2, Sampling3, Sampling4, Sampling5) on five different ML models (M1, M2, M3, M4 and M5)

	Sampling1	Sampling2	Sampling3	Sampling4	Sampling5
M1	50.10	52.24	63.18	69.23	70.12
M2	59.25	65.27	68.72	28.36	30.25
M3	90.45	72.41	32.17	42.58	41.85
M4	78.25	56.24	47.23	33.44	40.12
M5	81.25	12.85	57.36	32.25	52.74

5. Determine which sampling technique gives higher accuracy on which model.
6. The put the solution on the “Github Repository” with discussion and then submit the “Github” link using the submission link.