

NAME-DIKSHA RANI

ROLL NUMBER-102016116

SUB GROUP-CS7

sampling refers to the process of selecting a subset of data from a larger dataset. Sampling is often used in machine learning and statistics to create training and testing sets, or to balance imbalanced datasets. There are several sampling techniques that can be used, including:

1. Simple random sampling: randomly selecting data points from the dataset without any bias.
2. Stratified sampling: dividing the dataset into subgroups (or strata) and sampling from each subgroup to ensure that the sample is representative of the population.
3. Systematic sampling: In systematic sampling, a sample is chosen by selecting every "kth" individual from a population of "N" individuals, where "k" is a constant that is determined by the total number of individuals in the population and the desired sample size.
4. Cluster sampling: In cluster sampling, the population is divided into clusters (or groups) based on some natural grouping factor (such as geographic location or occupation).

STEPS FOR ACCOMPLISHING THE TASKS ARE:-

1. We will first download the dataset from the given link  
“[https://github.com/AnjulaMehto/Sampling\\_Assignment/blob/main/Creditcard\\_data.csv](https://github.com/AnjulaMehto/Sampling_Assignment/blob/main/Creditcard_data.csv)”
2. Since the given dataset is imbalanced as maximum of the class are 0's, we can use one of the techniques to balance the dataset. For example, we can use the oversampling technique such as Synthetic Minority Over-sampling Technique (SMOTE) or Adaptive Synthetic Sampling (ADASYN) to balance the dataset. we can use libraries such as imbalanced-learn or SMOTE-variants to perform these techniques.
3. To create five samples, we will use the sample size detection formula, to determine the sample size. Then we can use random sampling techniques such as simple random sampling, stratified sampling, or systematic sampling to create the samples.
4. we can choose five different sampling techniques such as simple random sampling, stratified sampling, systematic sampling, cluster sampling, or multistage sampling, based on the problem and the data. Then we can apply these techniques on five different ML models such as logistic regression, decision trees, random forests, support vector machines, or neural networks. To apply these techniques, we can use libraries such as scikit-learn or TensorFlow.
5. We can split the data into training and testing sets, fit the models on the training set, and evaluate the performance on the testing set using appropriate metrics such as accuracy, precision, recall, F1-score, or ROC-AUC.

6. To determine which sampling technique gives a higher accuracy for each model, we can modify the code by keeping track of the highest accuracy and the corresponding sampling technique for each model.