In this Assignment, I have made a jupyter notebook which is able to download the credit card dataset from the github link provided and then I cleaned the dataset and balanced the dataset using imblearn library. I also divided the dataset into an 80-20 train test split.

Then I applied five different sampling techniques on the dataset like simple random sampling, systematic sampling, stratified sampling, cluster sampling and convenience sampling.

Then I trained five different ML models (Logistic Regression, Decision Tree Classifier, K Nearest Neighbours, Random Forest Classifier and Support Vector Classifier (SVC)) and compared each model using different sampling techniques. I finally chose the most accurate model which was Random forest using convenience sampling.

These are the accuracies that I obtained :

|  | Simple Random Sampling | Systematic Sampling | Stratified Sampling | Cluster Sampling | Convenience Sampling |
| --- | --- | --- | --- | --- | --- |
| Logistic Regression | 81.81 | 94.11 | 88.96 | 96.68 | 93.42 |
| Decision Tree Classifier | 97.40 | 96.07 | 95.45 | 97.15 | 98.68 |
| K Nearest Neighbours | 77.92 | 91.50 | 91.55 | 96.20 | 94.73 |
| Random Forest Classifier | 98.70 | 99.34 | 98.70 | 99.52 | 100 |
| Support Vector Classifier | 89.61 | 95.42 | 95.45 | 99.05 | 98.68 |