Data Curation Lab

Sampling

Tasks to do:

1. Read the attached CSV file which contains the customer data for a bank's loan approval, where the target is whether the loan was approved or not.

Columns:

- `ID`: Unique identifier
- `Age`: Age of the applicant
- `Income`: Monthly income in INR
- `Credit_Score`: Creditworthiness on a scale of 300 to 900
- `Loan_Approved`: `1` for approved, `0` for not approved (imbalance: 70% not approved)

2. Use random sampling of 10 samples with replacement from the dataset
3. Use random sampling of 5 samples without replacement from the dataset
4. Use systematic sampling of every 4th element
5. Use clustered sampling where sample 2 data from the following categories
    a. High cibil score (750 and above)
    b. Medium cibil score (650 to 749)
    c. Low cibil score (less than 650)
6. From the dataset , it is clear the approval vs not approval of the loan follows the 70:30 ratio, which is an imbalance. Solve the imbalance using
    a. Random oversampling
    b. Random undersampling
    c. Synthetic Minority Oversampling Technique (SMOTE)

Submission Guidelines is similar to that of previous experiment

Feel free to ask any questions .