Practical 8
19BCE248
2CS501

Aim: Handling Imbalanced Dataset using various techniques.

Dataset:
This dataset consisted of 4024 rows and 16 columns which is a classification-based problem where output depicts whether a person is dead or alive based on the input parameters given.

Preprocessing:

- First of all categorical data are converted into a numerical value.
- Then removing the label column from the original dataset and labelling it as y.
- They are classified into training and testing sets.
- Then various imbalance handling techniques are applied to overcome the problem.

Model Training with various algorithms and corresponding accuracy obtained are shown below in the tabular form:

1.) Without any handling imbalance approach

Model	Accuracy
LogisticRegression	0.8993788819875776
KNN	0.8968944099378882
SVC	0.8894409937888199
DecisionTree	0.8335403726708075

2.) BorderLine Smote:

Model	Accuracy
LogisticRegression	0.7705278592375366
KNN	0.8724340175953079
SVC	0.7631964809384164
DecisionTree	0.8702346041055719

It might look strange that the accuracy decreased at a significant rate but the reason is in the above training their was a problem of model overfitting which leads to bad results when tested in a purely new type of instance in real-world scenarios.

```
0    3408
1     616
Name: Status, dtype: int64
```

Here clearly we can see the dominance of class 0 making the dataset imbalanced.

3.) SVMSmote

Model	Accuracy
LogisticRegression	0.7690615835777126
KNN	0.8629032258064516
SVC	0.7741935483870968
DecisionTree	0.8533724340175953

4.) KMeans SMOTE

Model	Accuracy
LogisticRegression	0.7712609970674487
KNN	0.8878299120234604
SVC	0.7690615835777126
DecisionTree	0.8687683284457478

5.) Simple SMOTE technique

Model	Accuracy
LogisticRegression	0.7983870967741935
KNN	0.8797653958944281
SVC	0.782991202346041
DecisionTree	0.8717008797653959

Conclusion/Learning:

From the given practical we got an idea about how to deal with the class imbalance problem with various techniques performed above. Also how accuracy changes with change in preprocessing is been depicted here. Exploring inbuilt library for performing the task was great exposure.