Classification of patients on the Indian liver patient dataset

Submitted in partial fulfillment of the requirements for the degree of

Bachelor of Technology

in

Information Technology(IT)

by

SAURAV AGRAWAL

16BIT0176

Under the guidance of

Prof. USHA DEVI G.

SITE,

VIT, Vellore.

May, 2020

DECLARATION

I hereby declare that the thesis entitled “Classification of Patients on the Indian Patients Liver Dataset" submitted by me, for the award of the degree of *Bachelor of Technology in Programme* to VIT is a record of bonafide work carried out by me under the supervision of Prof. Usha Devi G..

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place : Vellore

Date :

Signature of the Candidate

CERTIFICATE

This is to certify that the thesis entitled "Classification of Patients on the Indian Liver Patient Dataset" submitted by **Saurav Agrawal &16BIT0176**, **SITE**, VIT, for the award of the degree of *Bachelor of Technology in Information Technology*, is a record of bonafide work carried out by him under my supervision during the period, 01. 12. 2018 to 30.04.2019, as per the VIT code of academic and research ethics.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The thesis fulfills the requirements and regulations of the University and in my opinion meets the necessary standards for submission.

Place : Vellore

Date : **Signature of the Guide**

Internal Examiner **External Examiner**

Head of the Department

Programme

ACKNOWLEDGEMENTS

It is my pleasure to express with deep sense of gratitude to Prof Usha Devi G., for her constant guidance, continual encouragement, understanding; more than all, she taught me patience in my endeavor. She provided me with all the necessary elements regarding the project which helped me in ways I cannot express and helped me in completion of this project

I would like to express my gratitude to Dr. G. Viswanathan, Mr. Sankar Viswanathan, Dr. Sekar Viswanathan, Mr. G V Selvam, Dr. Anand A. Smauel, Dr. S. Narayanan, and Dr. Balakrushna Tripathy, School of Information Technology and Engineering, for providing with an environment to work in and for his inspiration during the tenure of the course.

In jubilant mood I express ingeniously my whole-hearted thanks to Dr. Jasmine Norman, Head of the Department and Senior Associate Professor (SITE) , all teaching staff and members working as limbs of our university for their not-self-centered enthusiasm coupled with timely encouragements showered on me with zeal, which prompted the acquirement of the requisite knowledge to finalize my course study successfully. I would like to thank my parents for their support.

It is indeed a pleasure to thank my friends who persuaded and encouraged me to take up and complete this task. At last but not least, I express my gratitude and appreciation to all those who have helped me directly or indirectly toward the successful completion of this project.

Place: Vellore
Date:

Name of the student
SAURAV AGRAWAL

Executive Summary

This project is an attempt to better classification models to predict liver patients on the Indian Liver Patient Dataset. This project uses various classifications algorithms ranging from individual classifiers to groups of classifiers to even neural networks. Various resampling techniques are also used in this project. K-Fold cross validation is also used to cross validate the testing data on the entire dataset. Principal Component Analysis is also used to reduce dimensionality. Insights from research papers are used along with new developments to enhance the classification.

TABLE OF CONTENTS

List of Figures

List of Table

List of Abbreviations

ILPD	Indian Liver Patient Dataset
SMOTE	Synthetic Minority Over-sampling Technique
ENN	Edited Nearest Neighbour
PCA	Principal Component Analysis
SVM	Support Vector Machine
LR	Logistic Regression
RF	Random Forest
KNN	K Nearest Neighbour
RUS	Random Under Sampler
ROS	Random Over Sampler
CC	Cluster Centroids

1. INTRODUCTION

1.1. OBJECTIVE

As the largest internal and glandular organ of the human body, the liver plays a vital role in detoxification of toxins, metabolism such as protein, carbohydrates and fats, synthesis of bile, and storage of vitamins and glycogens. The liver weighs about 1.36 kg. It is vital for our survival. Smoking, Alcohol consumption, obesity and diabetes results in liver damage and diseases.

Common liver disorders are

- Fatty liver is a revocable condition where large vacuoles of triglyceride fat acquire in liver cells via the process of limit. It can occur in people with a high level of alcohol consumption as well as in people who never had alcohol.
- Hepatitis (usually caused by a virus spread by excess contamination or direct contact with infected body fluids).
- Cirrhosis of the liver is one of the most serious liver diseases. It is an action used to indicate all forms of diseases of the liver characterized by the significant loss of cells. The liver gradually contracts in size and becomes leathery and hard. The regenerative action continues under liver cirrhosis but the progressive loss of liver cells exceeds cell replacement.
- Liver cancer. The risk of liver cancer is higher in those who have cirrhosis or who had valid types of viral hepatitis; but more often, the liver is the site of secondary (metastatic) cancers spread from other organs.

Classification algorithm is one of the greatest significant and applicable data mining techniques used to apply in disease prediction. Classification algorithm is the most common in several automatic medical health diagnoses.This project uses and compares individual classifiers and groups of classifiers (ensemble learning) for the classification of Indian liver patient dataset.

The main objective of this project is to use different individual classifiers and groups of classifiers (ensemble learning) for the correct classification of patients on the Indian liver patient dataset. The classification methodology in this project draws inspiration from previous works and methodology done by various people and tries to obtain a better methodology.

1.2. MOTIVATION

Liver disease either shows vague symptoms like weakness or no symptoms at the initial stage. Liver disease is tricky to diagnose given the subtlety of symptoms while in early stage. Even when the liver is partially damaged it continues to work and when the problems of liver diseases are discovered it is too late for prevention. An early diagnosis of liver problems will increase a patient's survival rate. At the early stage, liver disease is diagnosed by the liver functional test.

As there is increase in liver patients in India, it is estimated that India may become the world capital of liver disease by 2025.[10]

1.3. BACKGROUND

In recent years, there were a number of existing works that tried to classify patients on the ILPD. In a study conducted by Gulia et. al. uses different data mining techniques such as Decision Tree, Preceptron, Random Forest and Support Vector Machine with WEKA tool on the ILPD. It compares the performance of different algorithms before and after feature selection yielding the highest accuracy of 71.8696 % with SVM after feature selection.[1]

In another paper authored by Aiswarya et. al. which presents the difference between Random Forest and Naive Bayes algorithm on the ILPD using the WEKA tool shows that Random Forest works better than Naive Bayes algorithm in the prediction of liver diseases. [2]

In another research paper authored by Kefelegn and Kamat different data mining techniques such as Naive Bayes, C4.5, and Support Vector Machines are used with feature selection and k-fold cross validation on the ILPD. This paper also proposes methodology with different algorithms.[3]

In another research paper authored by Nagaraj and Sridhar which have combined the Support Vector Machine and Artificial Neural Network on the ILPD leading to the creation of a graphical user interface called NeuroSVM. This model produces accuracy of 98.83%.[4]

In another research paper authored by Swapna and Babu which uses Analysis of Variance. This paper focuses on gender and age differences of the dataset thereby concluding which features are more important in a particular gender and age group on the ILPD. This paper divides the dataset into four populations (0-20, 21-40, 41-60 and 61-80) based on the age. [5]

In another research paper authored by Pahareeya et. al. uses different data mining algorithms such as C4.5, Multi Layer Perceptron, SVM, Multiple Linear Regression, Random Forest, Genetic Programming along with K-Fold Cross Validation. This paper focuses on the unbalanced nature of the ILPD and uses oversampling and undersampling, This paper

achieves the highest accuracy of 89.10% with Random Forest when the dataset is oversampled at 200%. [6]

In another research paper authored by Kumar and Thakur uses algorithms such as Support Vector Machine and K Nearest Neighbor along with K Fold Cross validation. The ILPD is made to be somewhat balanced by the usage of Synthetic Minority Over-sampling Technique. This technique achieves highest accuracy of 74.67% with K Nearest Neighbor. [7]

In another research paper authored by Pathan et. al. uses different algorithms such as Naive Bayes, C4.5, Bagging, AdaBoost, Random Forest along with Feature Selection on the ILPD with the help of WEKA tool. This paper achieves 100% accuracy with Random Forest. However the feature selected is only age. Therefore this paper does age based classification. [8]

In another research paper authored by Banu describes different machine learning approaches such as Supervised, Unsupervised and Reinforcement Learning in order to solve liver diseases or disorders. [9]

In another research paper authored by Idris and Bhoite uses algorithms such as SVM, Logistic Regression, Random Forest , AdaBoost and Bagging on the ILPD. The usage of Bagging and Boosting ensemble learning is noteworthy. The highest accuracy of 74.36% is achieved by using Logistic Regression first followed by AdaBoost Classifier. [10]

In another research paper authored by Priya et. al. uses algorithms such as C4.5, Multi Layer Perceptron, Support Vector Machine, Random Forest, Naive Bayes along with Particle Swarm Optimization Feature Selection on the ILPD, This paper achieves highest accuracy of 95.04% with C4.5.[11]

In another research paper authored by Wadhwa and Juneja uses algorithms such as Support Vector Machine, Artificial Neural Network along with Hyperparameter Tuning and Cross Validated on the ILPD. The highest accuracy of 70.94% is achieved by Artificial Neural Network. The usage of hyperparameter tuning is noteworthy.[12]

2. PROJECT DESCRIPTION AND GOALS

a. Dataset Description -

This project uses Indian Liver Patient Dataset(ILPD). The Indian Liver Patient Dataset comprises 10 different attributes of 583 patients. The patients were described as either ‘1’ or ‘2’ on the basis of liver disease. Out of 583 patients, there are 416 liver patient records and 167 non liver patient records. Liver patient records are attributed as 1 and non liver patient records are attributed as 2.

Table 1.1. Indian Liver Patient Dataset Metadata

Sl. No.	Attribute Name	Attribute Type	Attribute Description
1.	Age	Numeric	Age of the patient
2.	Sex	Nominal	Gender of the patient
3.	Total Bilirubin	Numeric	Quantity of total bilirubin in patient
4.	Direct Bilirubin	Numeric	Quantity of direct bilirubin in patient
5.	Alkphos Alkaline Phosphotase	Numeric	Amount of ALP enzyme in patient
6.	Sgpt Alamine Aminotransferase	Numeric	Amount of SGPT in patient
7.	Sgot Aspartate Aminotransferase	Numeric	Amount of SGOT in patient
8.	Total Proteins	Numeric	Protein content in patient
9.	Albumin	Numeric	Amount of albumin in patient
10.	Albumin and Globulin Ratio	Numeric	Fraction of albumin and globulin in patient
11.	Class	Numeric {1, 2}	Status of liver disease in patient

b. System Modules -

I. Data Preprocessing

As the dataset contains missing values, it is necessary to replace them. The dataset contains 4 missing values which is replaced by the mean of the dataset. As most of the algorithms deal with numbers it is important to encode the categorical values. As the dataset features scale differently, it is necessary to normalize them. In the project, all the dataset values are scaled between [0,1] using min-max normalization. Dataset is then segregated into features and label values. The patient column is the label. Training and testing data is also spilt in this step.

II. Resampling the training dataset -

The Metric Trap: Simple metric scores like accuracy_score can be misleading for unbalanced datasets. For example - If email receives 1% spam of the total emails and the classifier produces 99% accuracy on the dataset of spam classification, then the classifier is predicting the most common class and therefore falling it's functionality as spam classifier.

As the count of liver patients and non liver patients is in the ratio of 3.1:1 the algorithms might become biased towards the majority label. So, in order to avoid this different resampling strategies are used.

- Undersampling and Oversampling -

Undersampling causes deletion of majority class and Oversampling causes duplication of minority class. However undersampling causes loss of information and oversampling causes overfitting. In this project, undersampling and oversampling is implemented through DataFrame.sample method.

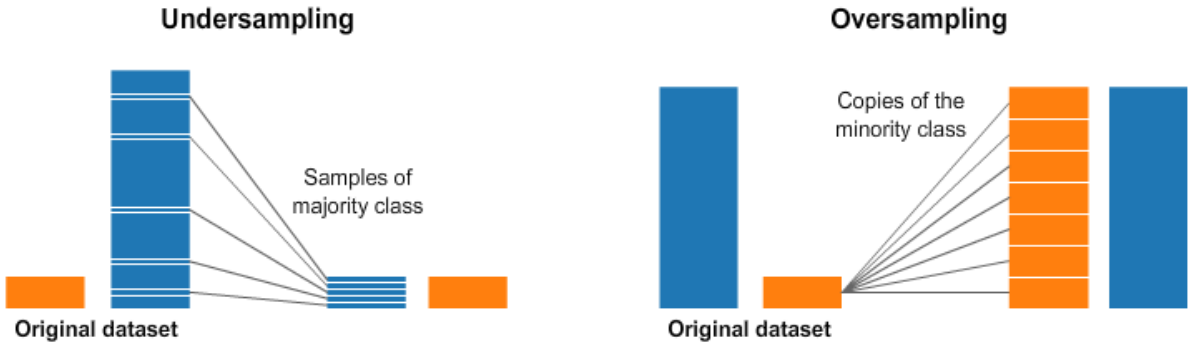


Figure 1.1. Visualization of undersampling and oversampling

- Random Undersampling and Random Oversampling -

Random Undersampling and Random Oversampling is implemented by the

RandomunderSampler class and RandomOverSampler class from the imblearn library. The concept generally is the same with Undersampling and Oversampling but with added tuning to the mode.

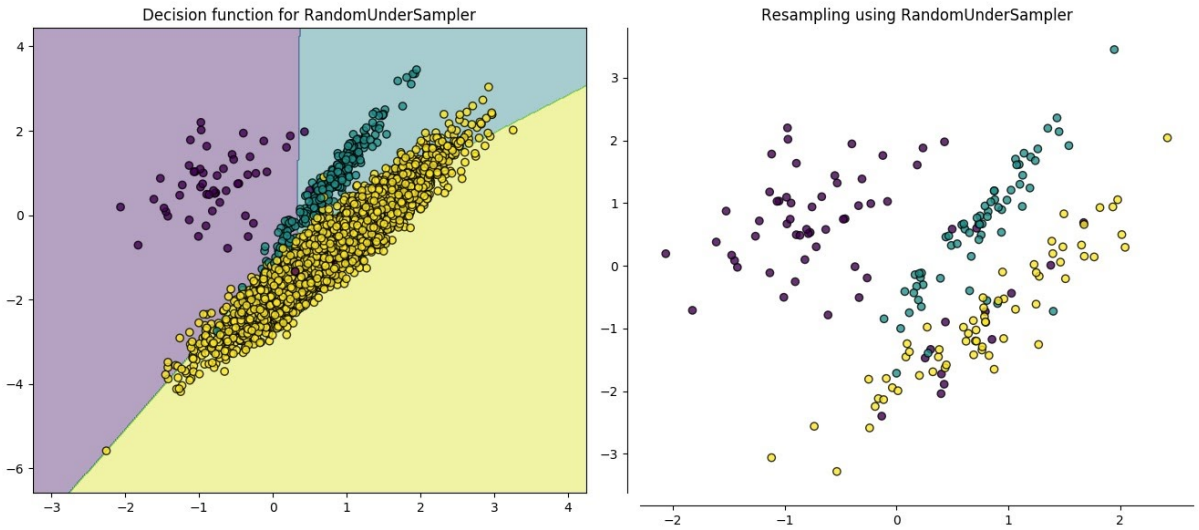


Figure 1.2. Visualization of a decision function of RandomUnderSampler on a dataset and dataset after resampling.

- TomekLink Undersampling -

TomekLink undersampling is implemented through the TomekLinks class from imblearn library. TomekLinks can also allow resample the majority class or resample all the classes but the minority class or resample all classes but the majority class or resample all classes.

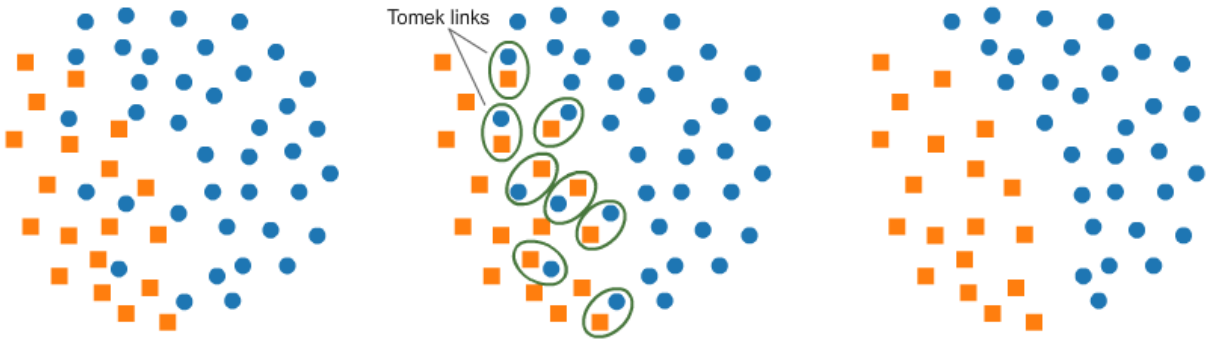


Figure 1.3. Visualization of TomekLinks on a dataset and dataset after resampling

- Cluster Centroid Undersampling -

Cluster Centroids makes use of K-Means to reduce the number of samples, It generates centroid based on the clustering methods. It undersamples the majority class by replacing a cluster of majority samples by the cluster centroid of a KMeans algorithm. This algorithm keeps N majority samples by fitting the KMeans algorithm with N cluster to the majority class and using the coordinates of the N cluster centroids as the new majority samples.
This algorithm is implemented using the ClusterCentroids class from imblearn library.

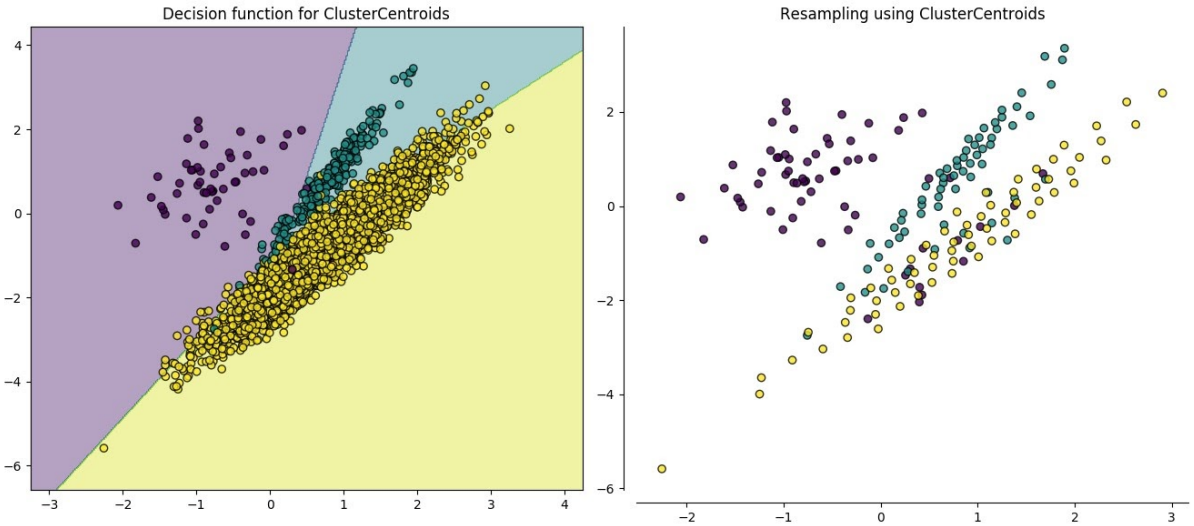


Figure 1.4. Visualization of a decision function of Cluster Centroid on a dataset and dataset after resampling.

- SMOTE Oversampling (Synthetic Minority Oversampling Technique) -

SMOTE (Synthetic Minority Oversampling Technique) consists of synthesizing elements for the minority class, based on those that already exist. It works randomly picking a point from the minority class and computing the k-nearest neighbors for this point. The synthetic points are added between the chosen point and its neighbors.

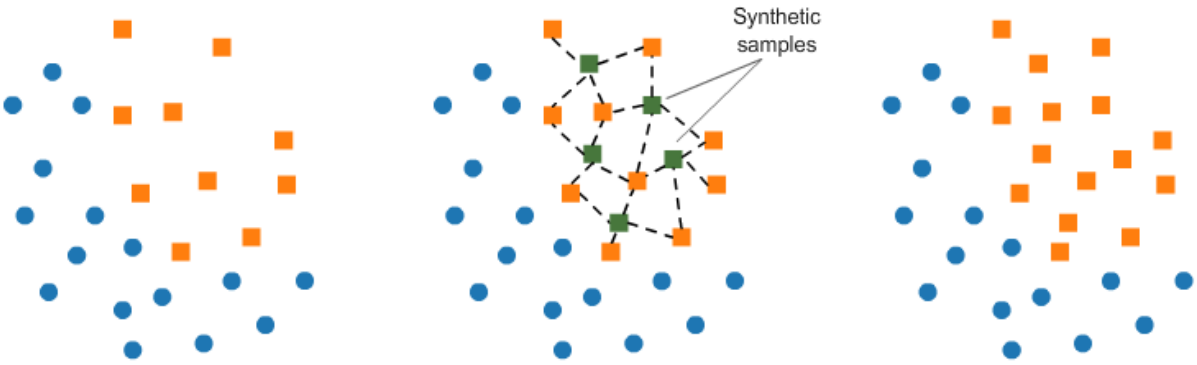


Figure 1.5. Visualization of SMOTE on a dataset and dataset after resampling

- ENN Undersampling (Edited Nearest Neighbours) -

EditedNearestNeighbours applies a nearest-neighbors algorithm and “edit” the dataset by removing samples which do not agree “enough” with their neighborhood. For each sample in the class to be under-sampled, the nearest-neighbours are computed and if the selection criterion is not fulfilled, the sample is removed.

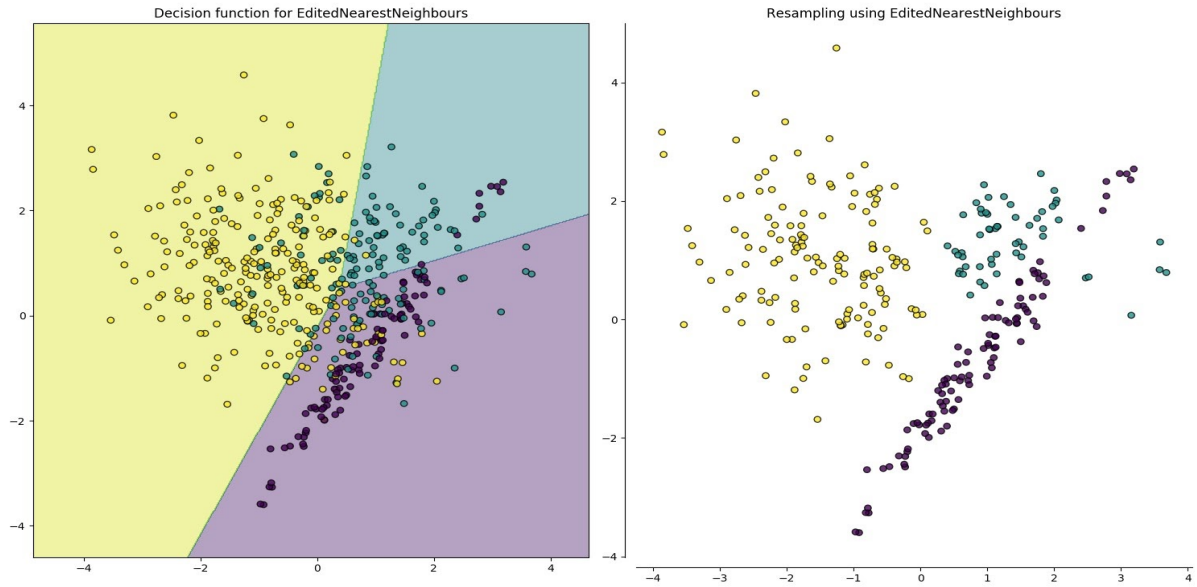


Figure 1.6. Visualization of a decision function of ENN on a dataset and dataset after resampling.

- SMOTEENN combined -

Combination of over- sampling and under-sampling using SMOTE and Edited Nearest Neighbours.

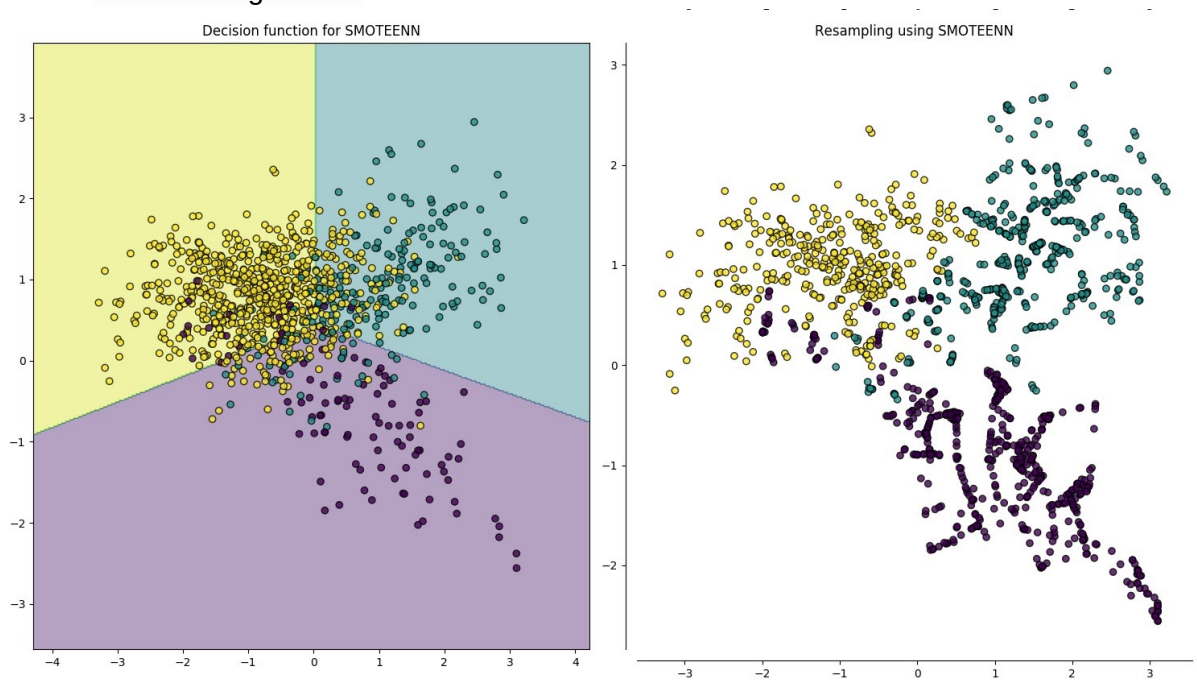


Figure 1.7. Visualization of a decision function of SMOTEENNon a dataset and dataset after resampling.

- SMOTETomek combined -

Combination of over-sampling and under-sampling using SMOTE and TomekLinks.

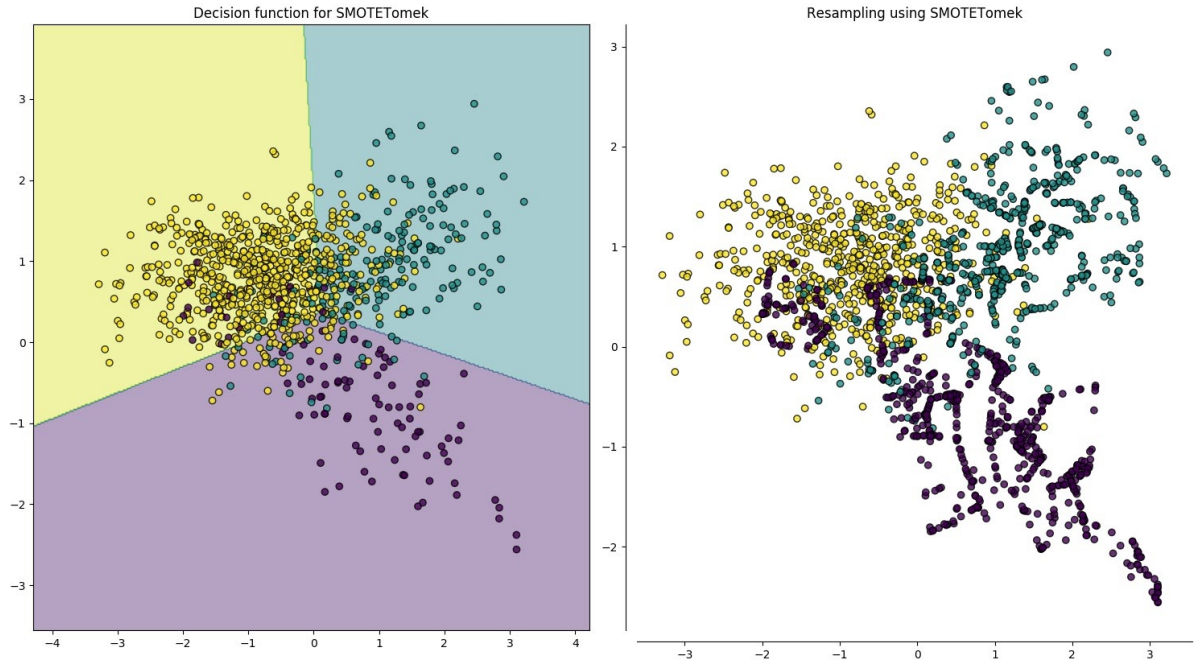


Figure 1.8. Visualization of a decision function of SMOTETomek on a dataset and dataset after resampling.

III. Classification -

Different algorithms are used in this project for classification of patients on the Indian Liver Patient Dataset. They are as follows -

- **Naive Bayes -**

Naive Bayes is an algorithm which is based on the Bayes' Theorem. Bayes' theorem is the probability of an event happening given that another event has already happened. The equation is as follows -

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Figure 1.9. Bayes Theorem

A, B - Events
P(A|B) - Probability of event A occurring given that event B has already occurred. It is also called Posterior Probability.
P(B|A) - Probability of event B occurring given that event A has already occurred.
P(A) - Probability of event A. It is also called Prior Probability.
P(B) - Probability of event B. It is also called Evidence.

In this project, Naive Bayes' is implemented by sklearn.naive_bayes.GaussianNB.

- **Support Vector Machine -**

Support vector machine is a robust algorithm which can efficiently perform linear as well as non linear classification by mapping input values into high dimensional feature spaces. In this project, support vector machine is implemented by sklearn.svm.SVC.

- **Logistic Regression -**

Logistic regression is just like linear regression with sigmoid function as activation function. It is also a supervised classification algorithm. In this project, logistic regression is implemented by sklearn.linear_model.LogisticRegression.

- **K-Nearest Neighbors -**

K-Nearest neighbors is a widely used and popular supervised learning algorithm. It finds euclidean distance from the given k points and assigns them to cluster them together with the nearest point among the given k points. It then calculates the centroid of that cluster and the process is repeated till we find repeated clusters in the following iterations. In this project, k-nearest neighbors is implemented by sklearn.neighbors.KNeighborsClassifier.

- **Random Forest -**

Random forest is an ensemble learning algorithm. It creates a forest of decision trees by randomly selecting subsets of training data and majority voting to get the output. In this project, it is implemented by sklearn.ensemble.RandomForestClassifier.

- **Voting Classifier -**

Voting classifier is also an ensemble learning algorithm. It uses many different classification algorithms as base classifiers and then applies majority voting to decide the output. However, the effectiveness of voting classifier lies mostly between worst performing classification algorithm and best performing classification algorithm. In this project, it is implemented by sklearn.ensemble.VotingClassifier.

- **Bagging Classifier-**

A Bagging classifier is an ensemble meta-estimator that fits base classifiers each on random subsets of the original dataset and then aggregate their individual predictions (either by voting or by averaging) to form a final prediction.
In this project, I have used different algorithms as base estimators -

➢ Decision Tree

- ➢ Perceptron
- ➢ KNN
- ➢ SVC
- ➢ Random Forest
- ➢ Logistic Regression

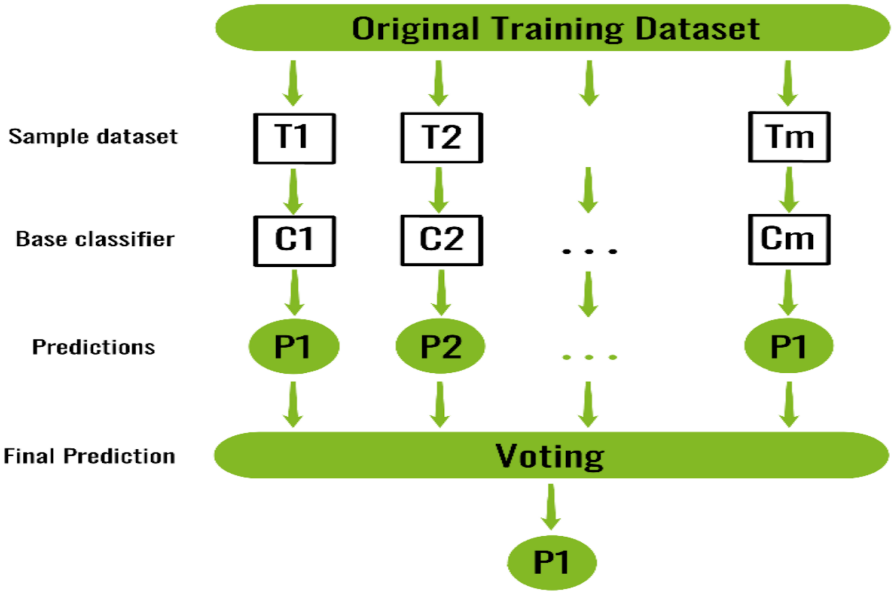


Figure 1.10. Bagging Classifier

- **Boosting -**

Boosting is an ensemble modeling technique which attempts to build a strong classifier from the number of weak classifiers. It is done building a model by using weak models in series.

In this project, I have used following boosting algorithms -

- ➢ Gradient Boosting
- ➢ AdaBoost with base estimator
  - ○ Decision Tree
  - ○ Random Forest
  - ○ Logistic Regression
  - ○ SVC

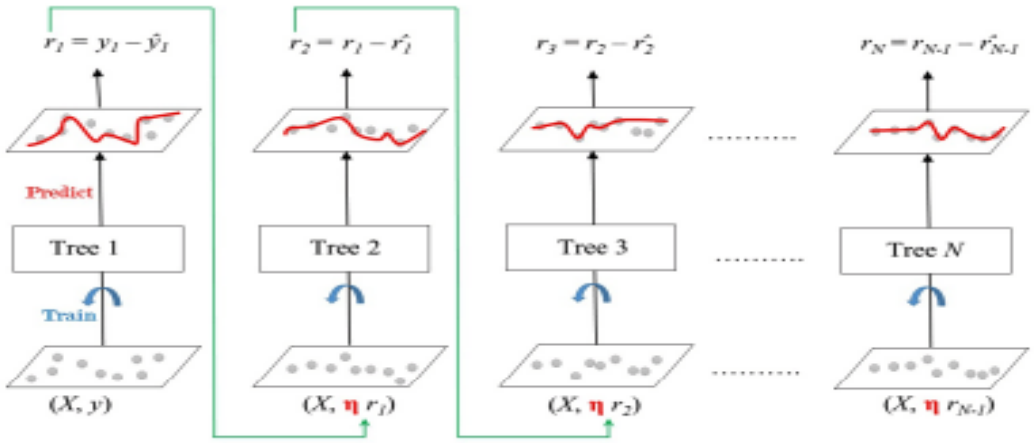


Figure 1.11. Boosting Classifier

- **Extreme Gradient Boosting (XGBoost) -**

XGBoost is one of the implementations of Gradient Boosting algorithm. It follows gradient boosting but also generalizes them by optimizing with a loss function which allows it to control over fitting. It is one of the most robust algorithms when it comes to noisy and real time data. In this project, XGBoost is implemented by xgboost.XGBClassifier.

- **Perceptron -**

Perceptron is a single layered neural network. It is a linear classifier and also a supervised learning algorithm. Perceptron is also sometimes called a linear binary classifier as it is mostly used for binary classification.

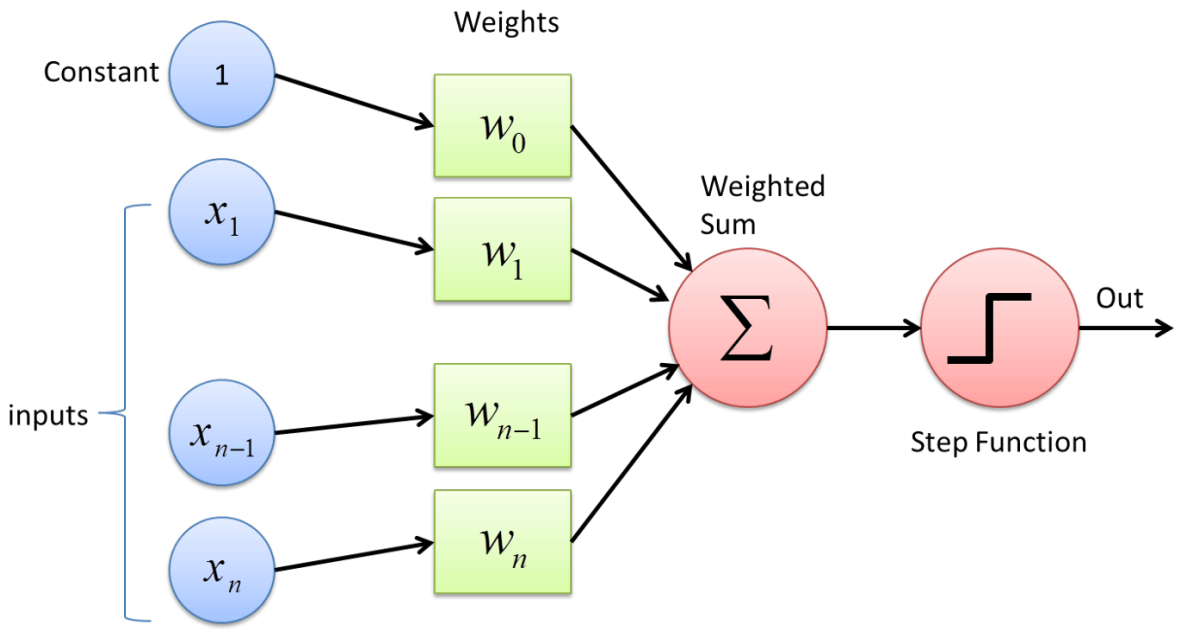


Figure 1.12. Perceptron

In this project, perceptron is implemented sklearn.linear_model.LogisticRegression.

- **Simple Neural Network -**

Simple neural network or artificial neural network is just like a multi layer perceptron. In this project, there are 3 hidden layers with 16, 8 and 4 nodes along with input layer and output layer. This neural network is implemented with tensorflow and keras. In order to avoid bias towards certain features Dropout is implemented on every hidden layer which drops a certain percentage of nodes randomly at the layer. Therefore, the information of that nodes is not passed to further layers.

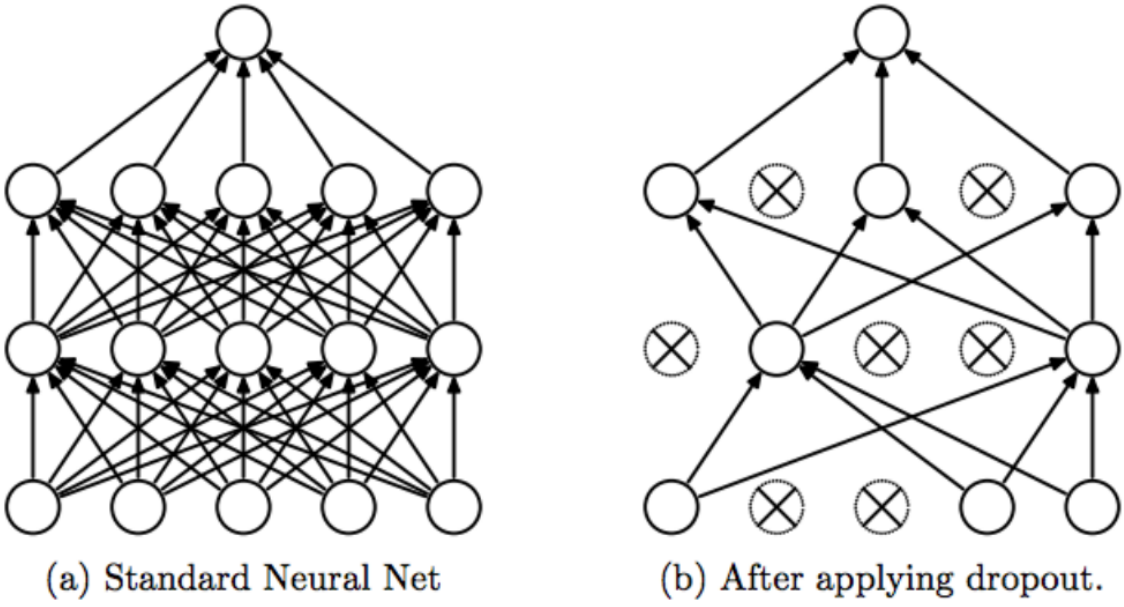


Figure 1.13. Neural Network with and without Dropout.

IV. Evaluation -

In order to find how well the classifier is performing it is necessary to evaluate it on the testing data. The classifiers make a prediction on the features of the testing data generating predicting labels then it is evaluated with the labels of the testing data. In order to perform an efficient evaluation confusion matrix is used.

- **Confusion Matrix -**

It is a table that is used to describe the performance of supervised classification models.

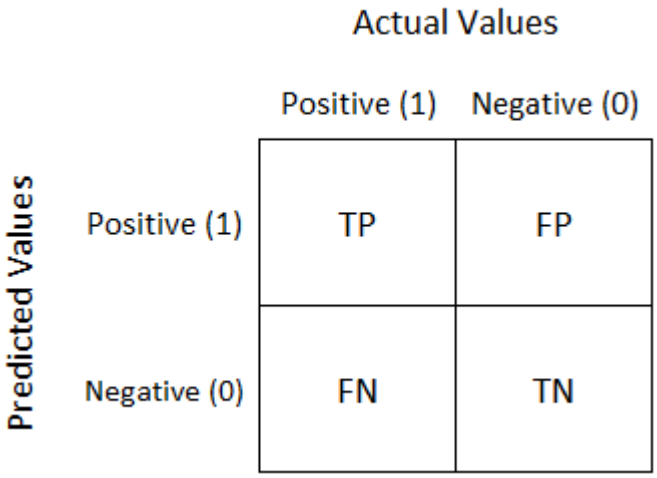


Figure 1.14. Confusion Matrix

The following things can be inferred from the confusion matrix -

- ➢ Accuracy = (TP + TN) / (TP + TN + FP + FN)
- ➢ Recall = TP / (TP +FN)
- ➢ Precision = TP / (TP + FP)
- ➢ Specificity = TN / (TN + FP)
- ➢ F1 Score = 2*Recall*Precision / (Recall + Precision)

So in the case where algorithms predict all the test dataset to a single label like 0 or 1 then it might be possible for Recall or Precision to be NaN(Not a Number) as the denominator becomes zero.

V. PCA (Principal Component Analysis) -

PCA is one of the methods to do feature extraction. Feature Extraction is a process of taking the already existing feature and generating a new feature which is a combination of each of the individual existing features. It combines the input features in a specific way such that it creates Principal Components (PC) and the user can take the most important Principal Components leaving/dropping the least important Principal Components. In the project out of 10 Principal Components which are produced from the dataset, usually the first 5th or 6th gives 95% variance of the original dataset.

VI. K-Fold Cross Validation -

Cross validation (CV) is one of the techniques used to test the effectiveness of machine learning models by taking the entire dataset as test dataset in chunks. K-Fold cross validation is very useful as it ensures that every instance of the dataset will appear as the test dataset at one point of time.

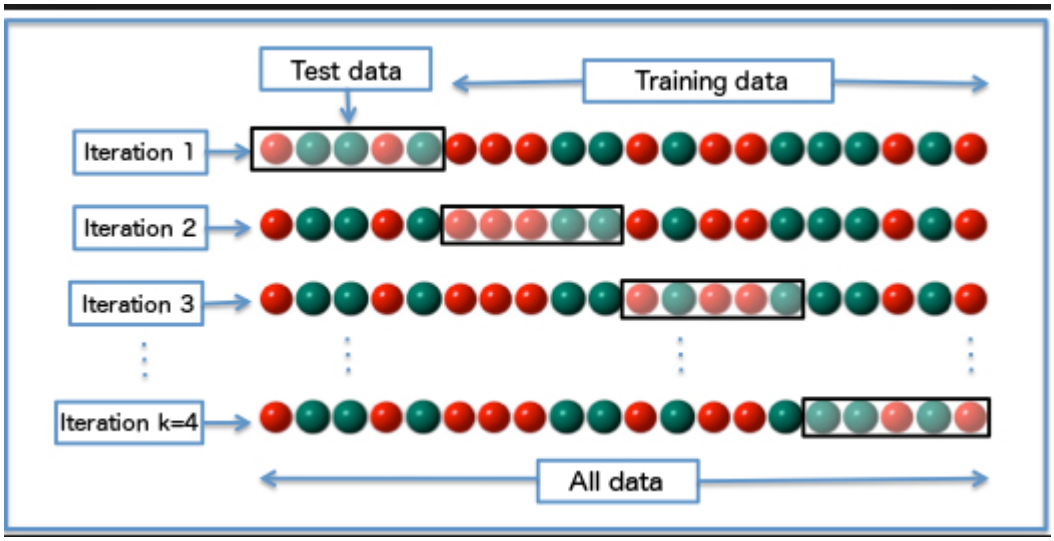


Figure 1.15. Test and Train data at different iterations according to K-Fold Cross Validation.

In the project, I am using K=4 for cross validation.

VII. Hyperparameter Tuning -

Hyperparameters are parameters that are fixed before the training process begins. These parameters express important properties of the model such as its complexity or how fast it should learn. Hyperparameter Tuning is the process of choosing a set of optimal hyperparameters for the learning algorithm.
Types of hyperparameter tuning -

- GridSearchCV :

In this approach the model searches for the best set of hyperparameters from a

grid of hyperparameter values. This approach goes through all the combinations of hyperparameter values making it computationally expensive.

In the following example -

grid={‘C’:[0.1, 0.2, 0.3, 0.4, 0.5],’Alpha’:[0.1, 0.2, 0.3, 0.4]}

GridSearchCV searches for 20models(5X4) and if Cross Validation=10 then it takes 200 fits which drastically increases the computation time to find the model with best hyperparameters within the grid values.

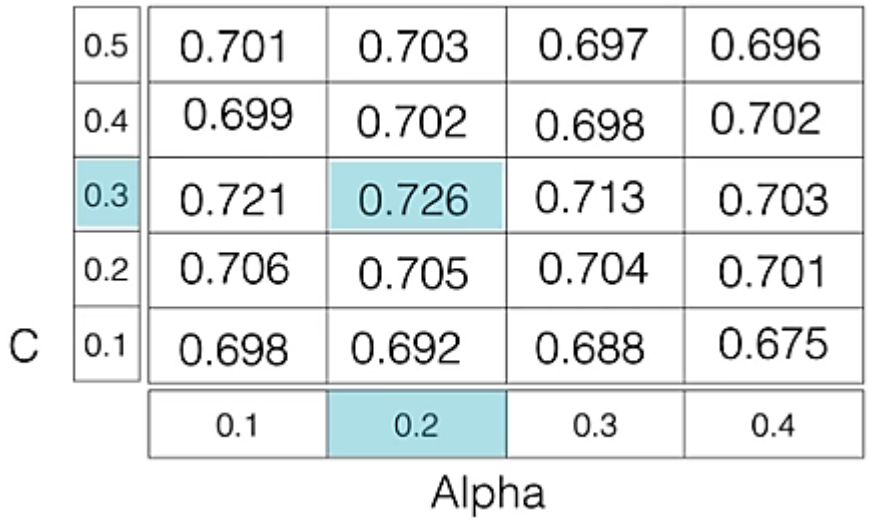


Figure 1.16. Grid of hyperparameters C and Alpha

- RandomSearchCV :

RandomizedSearchCV solves the drawbacks of GridSearchCV, as it goes through only a fixed number of hyperparameter settings. It moves within the grid in random fashion to find the best set hyperparameters. This approach reduces unnecessary computation. In the project I have used RandomizedSearchCV and saved the model.

3. TECHNICAL SPECIFICATION-

Hardware - 4 GB RAM minimum, 20GB Hard Disk Space
Software - Anaconda4.7.12 , Python3.7.5
Packages used –

- Numpy

NumPy stands for Numerical Python. It is a python library consisting of multidimensional array objects and a collection of procedures for processing those arrays. Using NumPy, logical operations and mathematical operations on arrays can be performed.

- Matplotlib

Matplotlib is a plotting library for Python. I used this library to print plots like scatterplot, barplot, line plot, pie chart etc.

- Sklearn

Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machines, random forests, and k-neighbours, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

- Pandas

Pandas library is used for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

- Imblearn

Imblearn library is used for imbalanced dataset in machine learning. Apart from that it also provides batch generators, pipelines, metrics, utilities.

- Tensorflow

TensorFlow is an open source Deep Learning library. It provides excellent functionalities and services when compared to other popular deep learning frameworks. These high-level operations are important for carrying out complex parallel computations and for creating advanced neural network models.

- Keras

Keras is a powerful and simple free open source Python library for evaluating and developing deep learning models and neural networks.

4. DESIGN APPROACH AND DETAILS

4.1. Design approach

- **Label Encoding -**

Label Encoding is a technique to convert categorical values to a number. For example - colors_of_fruits contain 4 different values, label encoding will encode it like this -

- ➢ Orange -> 1
- ➢ Red -> 2
- ➢ Green -> 3
- ➢ Yellow -> 4

In the project, “Gender” is label encoded by using .map({“Male”:0, “Female”:1}). This method is used in research papers [10] and [12].

- **One Hot Encoding -**

Label Encoding has a disadvantage that the numeric values can be misinterpreted by the algorithms. For example - the value of 1 is less than the value of 4 but in real life “Orange” color of fruit is less than “Yellow” color of fruit. One Hot Encoding is a technique to convert each categorical value into a new column of 0 or 1(False/True).

Table 1.2. One Hot Encoding Description

colors_of_fruits	is_orange	is_red	is_green	is_yellow
Orange	1	0	0	0
Red	0	1	0	0
Green	0	0	1	0
Yellow	0	0	0	1

One Hot Encoding is implemented by .get_dummies() function in the project.

- **System Architecture -**

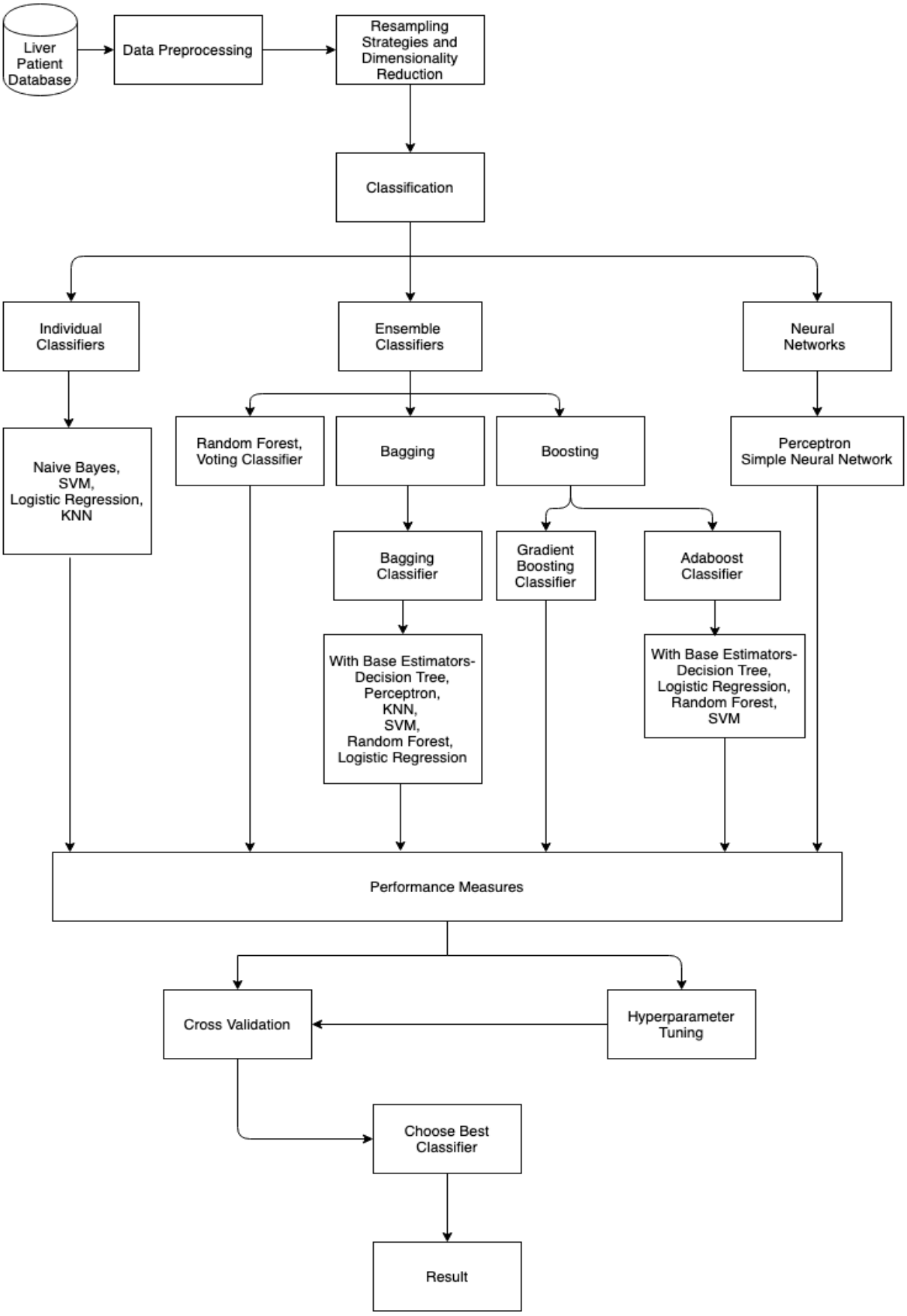


Figure 1.17. Architecture of the model

4.2. Codes and Standards-

Base functions -

```
def clfFitPredict(clf, X_train, X_test, y_train, y_test):
    clf.fit(X_train, y_train.values.ravel())
    y_pred = clf.predict(X_test)
    confusionMatrix(y_test, y_pred)

def confusionMatrix(y_test, y_pred):
    co = confusion_matrix(y_test, y_pred, labels=[0, 1])
    print('\nConfusion Matrix:')
    index = [0, 1]
    columns = [0, 1]
    co_df = pd.DataFrame(co,columns,index)
    plt.figure(figsize=(6,6))
    sns.heatmap(co_df, annot=True, fmt='g')
    plt.ylabel("Actual")
    plt.xlabel("Predicted")
    plt.show()

    total=co[0,0]+co[1,1]+co[0,1]+co[1,0]
    accuracy=(co[0,0]+co[1,1])/total
    print('\nAccuracy : '+ str(round(accuracy, 3)))

    sensitivity = co[0,0]/(co[0,0]+co[0,1])
    print('Sensitivity : '+ str(round(sensitivity, 3)))

    precision = co[0,0]/(co[0,0]+co[1,0])
    print('Precision: ' + str(round(precision, 3)))

    specificity = co[1,1]/(co[1,0]+co[1,1])
    print('Specificity : ' + str(round(specificity, 3)))

    fscore = 2*precision*sensitivity/(precision+sensitivity)
    print('F-Score : ' + str(round(fscore, 3)))

    print("\n",classification_report(y_test, y_pred), "\n")


def params(confusion_matrix):
    co = confusion_matrix
    print('\nConfusion Matrix By taking mean of all individual confusion matrix folds:')
```

```
    index = [0, 1]
    columns = [0, 1]
    co_df = pd.DataFrame(co,columns,index)
    plt.figure(figsize=(6,6))
    sns.heatmap(co_df, annot=True, fmt='g')
    plt.ylabel("Actual")
    plt.xlabel("Predicted")
    plt.show()

    total=co[0,0]+co[1,1]+co[0,1]+co[1,0]
    accuracy=(co[0,0]+co[1,1])/total
    print('\nAccuracy : '+ str(round(accuracy, 3)))

    sensitivity = co[0,0]/(co[0,0]+co[0,1])
    print('Sensitivity : '+ str(round(sensitivity, 3)))

    precision = co[0,0]/(co[0,0]+co[1,0])
    print('Precision: ' + str(round(precision, 3)))

    specificity = co[1,1]/(co[1,0]+co[1,1])
    print('Specificity : ' + str(round(specificity, 3)))

    fscore = 2*precision*sensitivity/(precision+sensitivity)
    print('F-Score : ' + str(round(fscore, 3)))

def crossValidation(clf, X, y, folds):
    #https://stackoverflow.com/questions/41458834/how-is-scikit-learn-cross-val-predict-
accurac#y-score-calculated
    scoreclf = cross_val_score(clf, X, y.values.ravel(), cv=folds)
    print(scoreclf,"\n")
    print(np.mean(scoreclf))
    y_pred = cross_val_predict(clf, X, y.values.ravel(), cv=folds)
    conf_mat = confusion_matrix(y.values.ravel(), y_pred)
    params(conf_mat)
```

Algorithms applied on Original Dataset -

```
print("\n----------------------------------------------------------------------\n")

#1 Naive Bayes On Original dataset
print("Naive Bayes on Original dataset:")
clfFitPredict(GaussianNB(), X_train, X_test, y_train, y_test)

#Cross Validation on Naive Bayes on Original datset
print("\nCross Validation of Naive Bayes on Original dataset:")
crossValidation(GaussianNB(), X, y, 4)

print("\n----------------------------------------------------------------------\n")
print("\n----------------------------------------------------------------------\n")

#2 SVM Classifier On the Original Dataset
print("SVM Classifier on Original dataset:")
clfFitPredict(LinearSVC(), X_train, X_test, y_train, y_test)

#Cross Validation on SVM Classifier on Orginial datset
print("\nCross Validation of SVM Classifier on Original dataset:")
crossValidation(LinearSVC(), X, y, 4)

print("\n----------------------------------------------------------------------\n")
print("\n----------------------------------------------------------------------\n")

#3 Logistic Regressor Classifier On the Original Dataset
print("Logistic Regressor Classifier on Original dataset:")
clfFitPredict(LogisticRegression(), X_train, X_test, y_train, y_test)

#Cross Validation on Logistic Regressor Classifier on Orginial datset
print("\nCross Validation of Logistic Regressor Classifier on Original dataset:")
crossValidation(LogisticRegression(), X, y, 4)

print("\n----------------------------------------------------------------------\n")
print("\n----------------------------------------------------------------------\n")

#4 KNN Classifier On the Original Dataset
print(" KNN Classifier on Original dataset:")
clfFitPredict(KNeighborsClassifier(n_neighbors = 2), X_train, X_test, y_train, y_test)

#Cross Validation on KNN Classifier on Orginial datset
print("\nCross Validation of  KNN Classifier  on Original dataset:")
```

```
crossValidation(KNeighborsClassifier(n_neighbors = 2), X, y, 4)

print("\n----------------------------------------------------------------------\n")
print("\n----------------------------------------------------------------------\n")

#5 RandomForest Classifier On the Original Dataset
print("RandomForest Classifier on Original dataset:")
clfFitPredict(RandomForestClassifier(max_depth=4, random_state=0), X_train, X_test,
y_train, y_test)

#Cross Validation on RandomForest Classifier on Orginial datset
print("\nCross Validation of  RandomForest Classifier on Original dataset:")
crossValidation(RandomForestClassifier(max_depth=4, random_state=0), X, y, 4)

print("\n----------------------------------------------------------------------\n")
print("\n----------------------------------------------------------------------\n")

#6 Voting Classifier for Original Dataset
#https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html
print("Voting Classifier on Original dataset:")
clfs = [('rf', RandomForestClassifier(max_depth=4,random_state=1)),
('lr',LogisticRegression()), ('svm',LinearSVC())]
vclf = VotingClassifier(estimators=clfs, voting='hard')
clfFitPredict(vclf, X_train, X_test, y_train, y_test)

#Cross Validation on Voting Classifier on Orginial datset
print("\nCross Validation of Voting Classifier on Original dataset:")
crossValidation(vclf, X, y, 4)

print("\n----------------------------------------------------------------------\n")
print("\n----------------------------------------------------------------------\n")

#AdaBoostClassifier as base estimator
adb_clf = AdaBoostClassifier(n_estimators=100, random_state=0)
print("\nAdaBoost Classifier on Original dataset:")
clfFitPredict(adb_clf, X_train, X_test, y_train, y_test)

#Cross Validation on AdaBoost Classifier on Orginial datset
print("\nCross Validation of AdaBoost Classifier on Original dataset:")
crossValidation(adb_clf, X, y, 4)

print("\n----------------------------------------------------------------------\n")
```

```
base_estimators = [SVC(),
                   RandomForestClassifier(),
                   LogisticRegression()
                  ]

base_estimator_names = ['SVC', 'RF', 'LR']
index = 0
for base_estimator in base_estimators:
   print("AdaBoost Classifier with {} as base estimator On Original
Dataset".format(base_estimator_names[index]))
   clf_adaboost_temp = AdaBoostClassifier(base_estimator=base_estimator,
n_estimators=100, algorithm='SAMME', random_state=0)
   clfFitPredict(clf_adaboost_temp, X_train, X_test, y_train, y_test)
   print("Cross Validation of AdaBoost Classifier with {} as base estimator On Original
Dataset".format(base_estimator_names[index]))
   crossValidation(clf_adaboost_temp, X, y, 4)
   print("\n-------------------------------------------------------------------------\n")
   index+=1
print("\n-------------------------------------------------------------------------\n")

#GradientBoostingClassifier On Original Dataset

gbc = GradientBoostingClassifier(max_depth=1, subsample=0.8, max_features=0.2,
n_estimators=300, random_state=0)
print("\nGradientBoostingClassifier on Original dataset :")
clfFitPredict(gbc, X_train, X_test, y_train, y_test)

print("\n-------------------------------------------------------------------------\n")

#Cross Validation on GradientBoostingClassifier on Orginial datset
print("\nCross Validation of GradientBoostingClassifier on Original dataset:")
crossValidation(gbc, X, y, 4)

print("\n-------------------------------------------------------------------------\n")
print("\n-------------------------------------------------------------------------\n")

#XGBClassifier on the Original Dataset

xgb_clf = XGBClassifier(objective='binary:logistic', booster='gblinear', n_estimators=10,
seed=1)
print("XGBClassifier on Original dataset :")
clfFitPredict(xgb_clf, X_train, X_test, y_train, y_test)
```

```
#Cross Validation on XGBClassifier on Orginial datset
print("\nCross Validation of XGBClassifier on Original dataset:")
crossValidation(xgb_clf, X, y, 4)

print("\n----------------------------------------------------------------------\n")
#Bagging Classifier On the Original Dataset

clf_bagging = BaggingClassifier(random_state=0)
print("Bagging Classifier on Original dataset :")
clfFitPredict(clf_bagging, X_train, X_test, y_train, y_test)

#Cross Validation on Bagging Classifier on Orginial datset
print("\nCross Validation of Bagging Classifier on Original dataset :")
crossValidation(clf_bagging, X, y, 4)

print("\n----------------------------------------------------------------------\n")

base_estimators = [Perceptron(),
                   KNeighborsClassifier(),
                   SVC(),
                   RandomForestClassifier(),
                   LogisticRegression()
                  ]

base_estimator_names = ['Perceptron', 'KNN', 'SVC', 'RF', 'LR']
index = 0
for base_estimator in base_estimators:
   print("Bagging Classifier with {} as base estimator On Original
Dataset".format(base_estimator_names[index]))
   clf_bagging_temp = BaggingClassifier(base_estimator=base_estimator, random_state=0)
   clfFitPredict(clf_bagging_temp, X_train, X_test, y_train, y_test)
   print("Cross Validation of Bagging Classifier with {} as base estimator On Original
Dataset".format(base_estimator_names[index]))
   crossValidation(clf_bagging_temp, X, y, 4)
   print("\n----------------------------------------------------------------------\n")
   index+=1
#Perceptron On The Original Dataset

clf_percept = Perceptron(tol=0.001, random_state=0)
print("Perceptron on Original dataset :")
clfFitPredict(clf_percept, X_train, X_test, y_train, y_test)

#Cross Validation on Perceptron on Orginial datset
```

```
print("\nCross Validation of Perceptron on Original dataset :")
crossValidation(clf_percept, X, y, 4)

print("\n----------------------------------------------------------------------\n")
#Neural Networks
model=Sequential()
model.add(Dense(16,input_shape=(10,)))
model.add(Dense(8,activation='relu'))
model.add(Dense(4,activation='relu'))
model.add(Dense(1, activation='sigmoid'))
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
history = model.fit(X_train, y_train, batch_size=10, epochs=10)
#print(history.history['loss'])
preds = model.predict(X_test)
#print(preds)
print("\nModel on Training Data -", model.evaluate(X_train, y_train))
print("\nModel on Testing Data -", model.evaluate(X_test, y_test))

pred_classes = model.predict_classes(X_test)
confusionMatrix(y_test, pred_classes)

#Neural Network with Dropout
model=Sequential()
model.add(Dense(16,input_shape=(10,)))
model.add(Dropout(0.2))
model.add(Dense(8,activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(4,activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(1, activation='sigmoid'))
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
history = model.fit(X_train, y_train, batch_size=10, epochs=10)
#print(history.history['loss'])
preds = model.predict(X_test)
#print(preds)
print("\nModel on Training Data -", model.evaluate(X_train, y_train))
print("\nModel on Testing Data -", model.evaluate(X_test, y_test))

pred_classes = model.predict_classes(X_test)
confusionMatrix(y_test, pred_classes)
print("\n----------------------------------------------------------------------\n")
```

The original dataset will go through different resampling techniques and same algorithms as above producing different results.

4.3. Constraints, Alternatives and Tradeoffs -

- **Random_state constraint -**

  The choice of random_state will impact the prediction result. The prediction result might improve or degrade according to the choice of random_state. In order to have reproducible results throughout the code, random_state has been stated wherever necessary. For example - in splitting testing and training dataset, train_test_split(X, y) is used. If random_state is defined as None as follows train_test_split(X, y, random_state=None) then every time the dataset is split the training and testing dataset will be different. So for every value of random_state there will be a difference in training and testing dataset.

- **Hyperparameter Tuning constraint -**

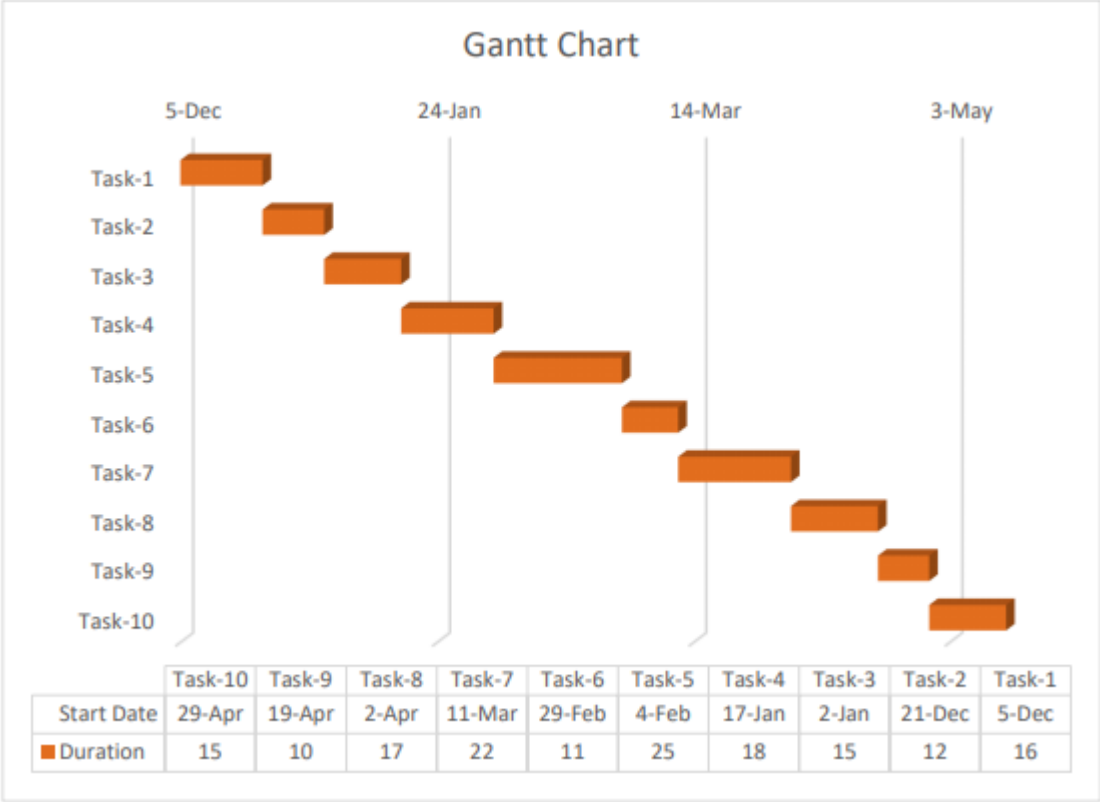
  Hyperparameter tuning is the process of choosing optimal hyperparameter for the algorithms. It is based on the grid of values. The larger the grid of values to choose hyperparameters from, the higher the complexity of hyperparameter tuning and higher training accuracy can get. However, higher training accuracy does not mean higher testing accuracy. Many times it leads to overfitting the training dataset thereby leading to lower testing accuracy than training accuracy.

- **GridSearchCV and RandomizedSearchCV tradeoff -**

  GridSearchCV performs exhaustive search of all the grid values of hyperparameters leading to higher training time of the model. For more values on the grid the training time of the model rises extensively. So, in order to avoid this higher training time we use RandomizedSearchCV in which not all the hyperparameters are tried out. Instead, it tries random combinations of hyperparameters to train the model. This allows us to set the number of attempts that we want the RandomizedSearchCV to search.

5. SCHEDULE, TASKS AND MILESTONES

I. Gantt chart –



	Task-10	Task-9	Task-8	Task-7	Task-6	Task-5	Task-4	Task-3	Task-2	Task-1
Start Date	29-Apr	19-Apr	2-Apr	11-Mar	29-Feb	4-Feb	17-Jan	2-Jan	21-Dec	5-Dec
Duration	15	10	17	22	11	25	18	15	12	16

Figure 1.18. Gantt chart

The above tasks in the schedule are -

- Task-1:

  Proof of concept, Literature survey and Research.

- Task-2:

  Technology choosing and project development plan

- Task-3:

  Data Preprocessing, Correlation, splitting of training and testing dataset. Initial use of classifier algorithms.

- Task-4:

  Normalization, Ensemble classifier algorithms.

- Task-5:

  Resampling again the dataset creation of new training and testing datasets.

- Task-6:

  Usage of Principal Component Analysis with resampled technique to create further new training and testing datasets.

- Task-7:

  Using all the datasets on the algorithms.

- Task-8:

  Hyperparameter Tuning.

- Task-9:

  Evaluation and comparative analysis.

- Task-10:

  Report.

II. Milestones

I followed strict discipline and standards to meet all deadlines. I wanted to complete before the deadline and planned everything accordingly. I faced a few hiccups during the development stage which were solved during the march month.

Following are the milestones respective to task -

- Task-1

I worked on the proof of concept for the project during this period. I went through a few journals which did prior research close to required fields. Due to these journals I realised that the target set was actually achievable.

- Task-3

After using all sorts of techniques and still not achieving considerate results, my attention was drawn to data preprocessing where encoding of categorical values is used. After watching many tutorials and changing label encoding to one hot encoding

- Task-9

In comparative analysis of classifiers, I finally got the desired results. I found that the milestone of Task-3 was not covered by many research papers. Therefore, using proper standards is essential for any project development.

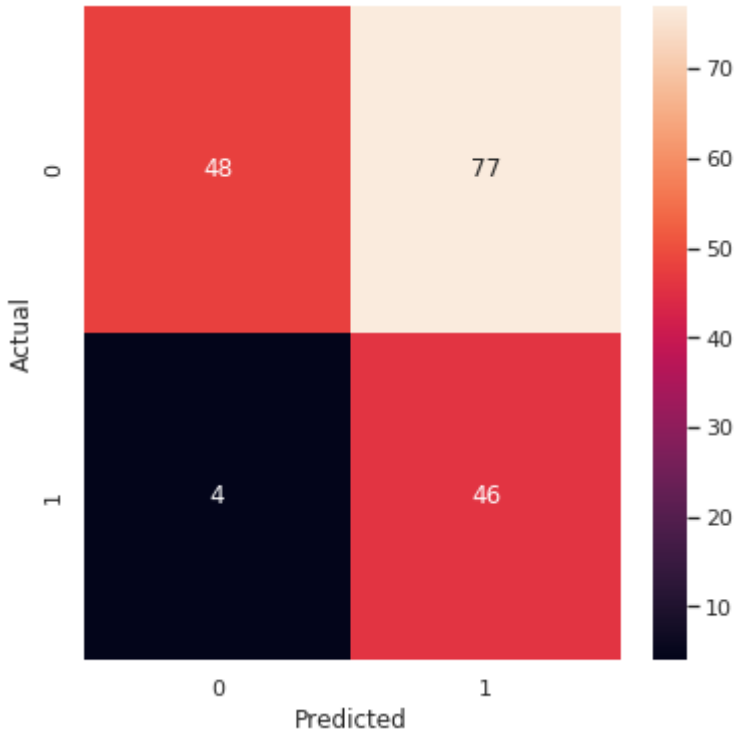
6. PROJECT DEMONSTRATION -

Code and output of naive bayes on the original dataset with Label Encoding-

```
print("\n----------------------------------------------------------------------\n")
#1 Naive Bayes On Original dataset
print("Naive Bayes on Original dataset:")
clfFitPredict(GaussianNB(), X_train, X_test, y_train, y_test)
#Cross Validation on Naive Bayes on Original dataset
print("\nCross Validation of Naive Bayes on Original dataset:")
crossValidation(GaussianNB(), X, y, 4)
print("\n----------------------------------------------------------------------\n")
```

OUTPUT

```
----------------------------------------------------------------------
Naive Bayes on Original dataset:
Confusion Matrix:
```



Accuracy : 0.537

Sensitivity : 0.384

Precision: 0.923

Specificity : 0.92

F-Score : 0.542

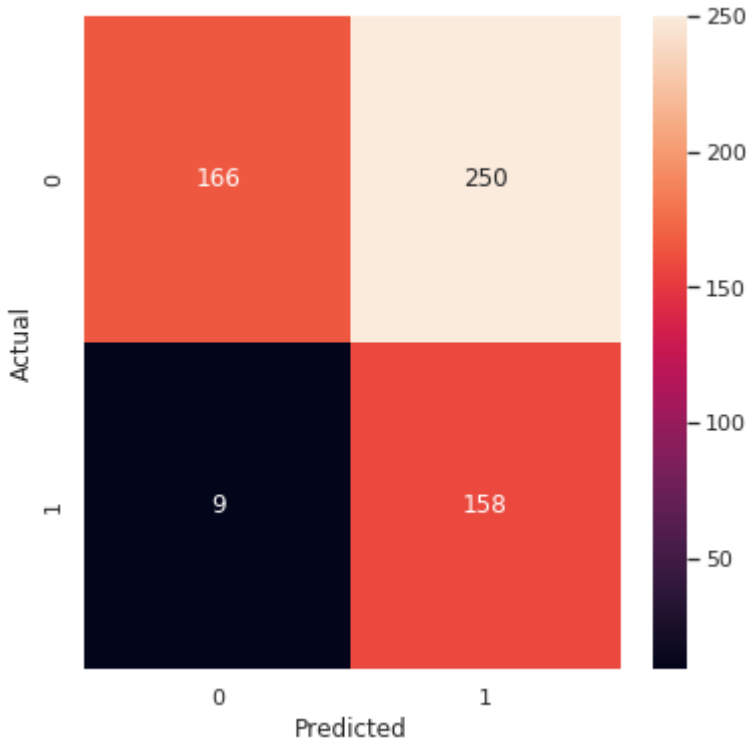
	precision	recall	f1-score	support
0.0	0.92	0.38	0.54	125
1.0	0.37	0.92	0.53	50
accuracy			0.54	175
macro avg	0.65	0.65	0.54	175
weighted avg	0.77	0.54	0.54	175

Cross Validation of Naive Bayes on Original dataset:

[0.60273973 0.54109589 0.52739726 0.55172414]

0.5557392536608408

Confusion Matrix By taking mean of all individual confusion matrix folds:



Accuracy : 0.556

Sensitivity : 0.399

Precision: 0.949

Specificity : 0.946

F-Score : 0.562

---

7. RESULT & DISCUSSION -

- **PCA -**

As the dataset has 10 features and 1 label, we cannot visualize it in a normal method of plotting. So PCA was used to extract the features into 2 main components namely PC1 and PC2. The visualization(scatterplot) is as follows -

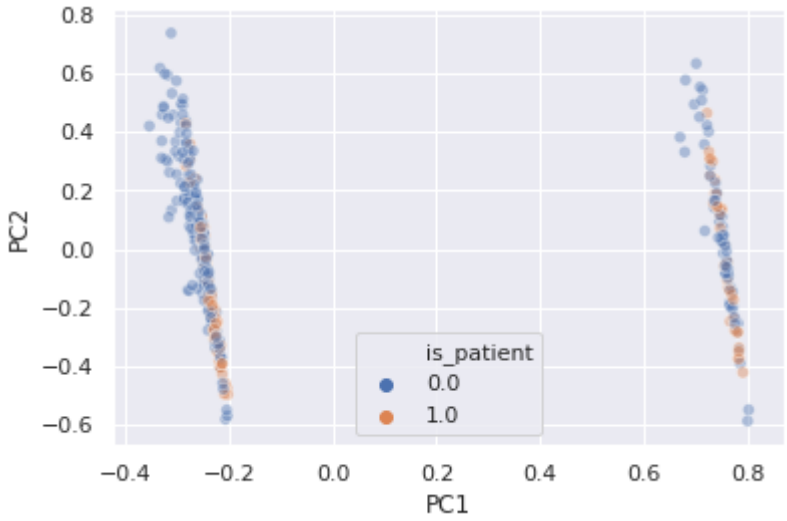


Figure 1.19. Visualization of PC1 and PC2 on the Original Dataset after Label Encoding

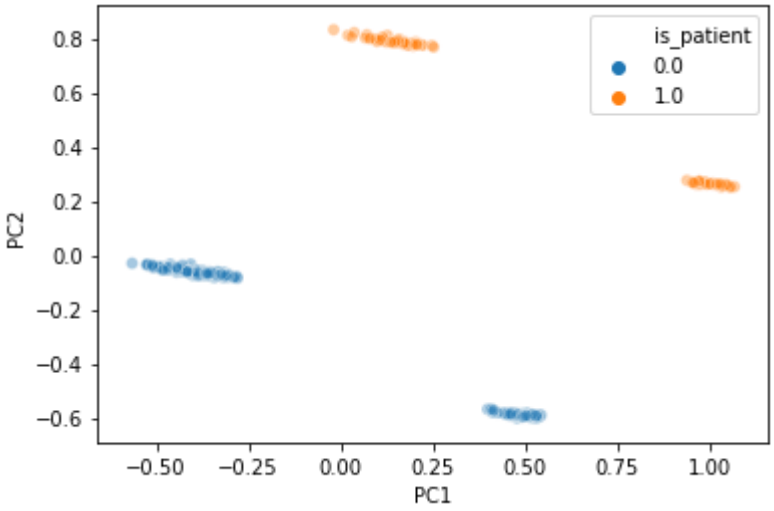


Figure 1.20. Visualization of PC1 and PC2 on the Original Dataset after One Hot Encoding

As it can be seen that plot with label encoding will require complex non-linear classification function to classify whereas with One Hot Encoding it will require simpler linear classification function.

- **Evaluation of Algorithms -**

Naive Bayes with Label Encoding -

Table 1.3. Performance of Naive Bayes with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.537	0.384	0.923	0.92
Original Dataset + K-Fold	0.556	0.399	0.949	0.946
PCA of Original Dataset	0.64	0.624	0.83	0.68
PCA of Original Dataset + K-Fold	0.605	0.555	0.837	0.731
UnderSampled Dataset	0.543	0.384	0.941	0.94
UnderSampled Dataset + K-Fold	0.623	0.397	0.923	0.952
PCA of UnderSampled Dataset	0.514	0.44	0.786	0.7
PCA of UnderSampled Dataset + K-Fold	0.614	0.463	0.8	0.832
OverSampled Dataset	0.543	0.384	0.941	0.94
OverSampled Dataset + K Fold	0.651	0.394	0.932	0.965
PCA of OverSampled Dataset	0.554	0.472	0.831	0.76
PCA of OverSampled Dataset + K-Fold	0.605	0.459	0.721	0.783
RandomUnderSampled Dataset	0.543	0.384	0.941	0.94
RandomUnderSampled Dataset + K-Fold	0.628	0.405	0.925	0.952
PCA of RandomUnderSampled Dataset	0.526	0.432	0.818	0.76

PCA of RandomUnderSampled Dataset + K-Fold	0.609	0.517	0.744	0.743
RandomOverSampled Dataset	0.537	0.376	0.94	0.94
RandomOverSampled Dataset + K-Fold	0.653	0.397	0.932	0.965
PCA of RandomOverSampled Dataset	0.549	0.48	0.811	0.72
PCA of RandomOverSampled Dataset + K-Fold	0.621	0.474	0.743	0.801
TomekLinked Dataset	0.543	0.384	0.941	0.94
TomekLinked Dataset + K-Fold	0.586	0.423	0.958	0.958
PCA of TomekLinked Dataset	0.617	0.592	0.822	0.68
PCA of TomekLInked Dataset + K-Fold	0.613	0.554	0.834	0.749
Cluster Centroid Dataset	0.52	0.352	0.936	0.94
Cluster Centroid Dataset + K-Fold	0.648	0.438	0.93	0.952
PCA of Cluster Centroid Dataset	0.491	0.36	0.833	0.82
PCA of Cluster Centroid Dataset + K-Fold	0.648	0.483	0.86	0.886
SMOTE OverSampled Dataset	0.571	0.448	0.903	0.88
SMOTE OverSampled Dataset + K-Fold	0.671	0.435	0.928	0.959
PCA of SMOTE OverSampled Dataset	0.52	0.472	0.766	0.64
PCA of SMOTE OverSampled Dataset + K-Fold	0.625	0.495	0.736	0.783
ENN UnderSampled Dataset	0.526	0.36	0.938	0.94

ENN UnderSampled Dataset + K-Fold	0.702	0.527	0.94	0.952
PCA of ENN UnderSampled Dataset	0.503	0.352	0.88	0.88
PCA of ENN UnderSampled Dataset + K-Fold	0.663	0.49	0.886	0.91
SMOTEENN Combined Sampled Dataset	0.531	0.368	0.939	0.94
SMOTEENN CombinedSampled Dataset + K-Fold	0.756	0.545	0.948	0.97
PCA of SMOTEENN CombinedSampled Dataset	0.514	0.392	0.845	0.82
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.698	0.498	0.83	0.901
SMOTETomek Combined Sampled Dataset	0.554	0.416	0.912	0.9
SMOTETomekCombinedSampled Dataset + K-Fold	0.669	0.432	0.93	0.96
PCA of SMOTETomek CombinedSampled Dataset	0.531	0.488	0.772	0.64
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.652	0.465	0.83	0.882

Naive Bayes with One Hot Encoding -

Table 1.4. Performance of Naive Bayes with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	0.998	1.0	0.998	0.993
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	0.994	1.0	0.993	0.974
PCA of RandomUnderSampled Dataset + K-Fold	0.992	1.0	0.998	0.979
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	0.996	1.0	0.993	0.991
TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	0.998	1.0	0.998	0.993
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	0.983	1.0	0.978	0.923
PCA of Cluster Centroid Dataset + K-Fold	0.99	1.0	0.984	0.992
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	0.983	1.0	0.978	0.923
PCA of SMOTE OverSampled Dataset + K-Fold	0.994	1.0	0.989	0.985
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	0.998	1.0	0.998	0.993
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN	1.0	1.0	1.0	1.0

CombinedSampled Dataset + K-Fold				
PCA of SMOTEENN CombinedSampled Dataset	0.983	1.0	0.978	0.923
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.992	1.0	0.987	0.983
SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	0.983	1.0	0.978	0.923
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.994	1.0	0.989	0.985

SVM with Label Encoding -

Table 1.5. Performance of SVM with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.726	1.0	0.723	0.04
Original Dataset + K-Fold	0.715	0.969	0.725	0.084
PCA of Original Dataset	0.714	1.0	0.714	0.0
PCA of Original Dataset + K-Fold	0.708	0.978	0.717	0.036
UnderSampled Dataset	0.634	0.544	0.907	0.86
UnderSampled Dataset + K-Fold	0.665	0.756	0.701	0.533
PCA of UnderSampled Dataset	0.623	0.544	0.883	0.82
PCA of UnderSampled Dataset + K-Fold	0.655	0.752	0.692	0.515
OverSampled Dataset	0.623	0.544	0.883	0.82
OverSampled Dataset + K Fold	0.688	0.651	0.749	0.733
PCA of OverSampled Dataset	0.646	0.576	0.889	0.82
PCA of OverSampled Dataset + K-Fold	0.659	0.637	0.712	0.686
RandomUnderSampled Dataset	0.611	0.528	0.88	0.82
RandomUnderSampled Dataset + K-Fold	0.672	0.752	0.711	0.557
PCA of RandomUnderSampled Dataset	0.6	0.496	0.899	0.86
PCA of RandomUnderSampled Dataset + K-Fold	0.663	0.744	0.703	0.545

RandomOverSampled Dataset	0.606	0.528	0.868	0.8
RandomOverSampled Dataset + K-Fold	0.696	0.618	0.784	0.792
PCA of RandomOverSampled Dataset	0.623	0.544	0.883	0.82
PCA of RandomOverSampled Dataset + K-Fold	0.663	0.603	0.736	0.736
TomekLinked Dataset	0.72	0.952	0.735	0.14
TomekLinked Dataset + K-Fold	0.704	0.934	0.722	0.8
PCA of TomekLinked Dataset	0.709	0.96	0.723	0.08
PCA of TomekLInked Dataset + K-Fold	0.69	0.94	0.709	0.12
Cluster Centroid Dataset	0.577	0.432	0.947	0.94
Cluster Centroid Dataset + K-Fold	0.689	0.686	0.765	0.695
PCA of Cluster Centroid Dataset	0.571	0.424	0.946	0.94
PCA of Cluster Centroid Dataset + K-Fold	0.685	0.678	0.763	0.695
SMOTE OverSampled Dataset	0.611	0.528	0.88	0.82
SMOTE OverSampled Dataset + K-Fold	0.695	0.627	0.774	0.777
PCA of SMOTE OverSampled Dataset	0.64	0.56	0.897	0.84
PCA of SMOTE OverSampled Dataset + K-Fold	0.666	0.606	0.739	0.739
ENN UnderSampled Dataset	0.577	0.44	0.932	0.92

ENN UnderSampled Dataset + K-Fold	0.739	0.741	0.801	0.737
PCA of ENN UnderSampled Dataset	0.571	0.424	0.946	0.94
PCA of ENN UnderSampled Dataset + K-Fold	0.734	0.736	0.796	0.731
SMOTEENN Combined Sampled Dataset	0.566	0.408	0.962	0.96
SMOTEENN CombinedSampled Dataset + K-Fold	0.79	0.694	0.862	0.888
PCA of SMOTEENN CombinedSampled Dataset	0.566	0.4	0.98	0.98
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.794	0.689	0.876	0.901
SMOTETomek Combined Sampled Dataset	0.623	0.56	0.864	0.78
SMOTETomekCombinedSampled Dataset + K-Fold	0.689	0.638	0.76	0.752
PCA of SMOTETomek CombinedSampled Dataset	0.623	0.544	0.883	0.82
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.674	0.618	0.748	0.74

SVM with One Hot Encoding -

Table 1.6. Performance of SVM with One Hot Encoding on different Evaluation Metrics

	Accuracy	**Sensitivity**	**Precision**	**Specificity**
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled	1.0	1.0	1.0	1.0

Dataset + K-Fold				
TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

LR with Label Encoding -

Table 1.7. Performance of LR with Label Encoding on different Evaluation Metrics

	Accuracy	**Sensitivity**	**Precision**	**Specificity**
Original Dataset	0.714	1.0	0.714	0.0
Original Dataset + K-Fold	0.717	0.988	0.72	0.042
PCA of Original Dataset	0.714	1.0	0.714	0.0
PCA of Original Dataset + K-Fold	0.719	0.995	0.719	0.03
UnderSampled Dataset	0.674	0.64	0.87	0.76
UnderSampled Dataset + K-Fold	0.636	0.806	0.657	0.389
PCA of UnderSampled Dataset	0.674	0.632	0.878	0.78
PCA of UnderSampled Dataset + K-Fold	0.638	0.822	0.655	0.371
OverSampled Dataset	0.651	0.592	0.881	0.8
OverSampled Dataset + K Fold	0.664	0.697	0.694	0.625
PCA of OverSampled Dataset	0.657	0.608	0.874	0.78
PCA of OverSampled Dataset + K-Fold	0.642	0.678	0.673	0.598
RandomUnderSampled Dataset	0.674	0.6	0.915	0.86
RandomUnderSampled Dataset + K-Fold	0.655	0.806	0.675	0.437
PCA of RandomUnderSampled Dataset	0.68	0.616	0.906	0.84
PCA of RandomUnderSampled Dataset + K-Fold	0.658	0.802	0.678	0.449
RandomOverSampled Dataset	0.651	0.6	0.872	0.78

RandomOverSampled Dataset + K-Fold	0.672	0.675	0.713	0.669
PCA of RandomOverSampled Dataset	0.657	0.6	0.882	0.8
PCA of RandomOverSampled Dataset + K-Fold	0.633	0.668	0.665	0.589
TomekLinked Dataset	0.726	1.0	0.723	0.04
TomekLinked Dataset + K-Fold	0.69	0.966	0.701	0.06
PCA of TomekLinked Dataset	0.72	1.0	0.718	0.02
PCA of TomekLInked Dataset + K-Fold	0.693	0.974	0.701	0.054
Cluster Centroid Dataset	0.554	0.416	0.912	0.9
Cluster Centroid Dataset + K-Fold	0.682	0.769	0.715	0.557
PCA of Cluster Centroid Dataset	0.554	0.416	0.912	0.9
PCA of Cluster Centroid Dataset + K-Fold	0.682	0.769	0.715	0.557
SMOTE OverSampled Dataset	0.651	0.6	0.872	0.78
SMOTE OverSampled Dataset + K-Fold	0.667	0.688	0.701	0.642
PCA of SMOTE OverSampled Dataset	0.64	0.592	0.86	0.76
PCA of SMOTE OverSampled Dataset + K-Fold	0.641	0.673	0.673	0.601
ENN UnderSampled Dataset	0.589	0.456	0.934	0.92
ENN UnderSampled Dataset + K-Fold	0.702	0.774	0.734	0.599
PCA of ENN UnderSampled Dataset	0.589	0.456	0.934	0.92

PCA of ENN UnderSampled Dataset + K-Fold	0.712	0.778	0.744	0.617
SMOTEENN Combined Sampled Dataset	0.56	0.408	0.944	0.94
SMOTEENN CombinedSampled Dataset + K-Fold	0.786	0.706	0.843	0.866
PCA of SMOTEENN CombinedSampled Dataset	0.554	0.4	0.943	0.94
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.784	0.711	0.835	0.858
SMOTETomek Combined Sampled Dataset	0.68	0.64	0.879	0.78
SMOTETomekCombinedSampled Dataset + K-Fold	0.671	0.693	0.706	0.644
PCA of SMOTETomek CombinedSampled Dataset	0.68	0.632	0.888	0.8
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.67	0.691	0.705	0.644

LR with One Hot Encoding -

Table 1.8. Performance of LR with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled	1.0	1.0	1.0	1.0

Dataset + K-Fold				
TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN	1.0	1.0	1.0	1.0

CombinedSampled Dataset				
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

KNN with Label Encoding -

Table 1.9. Performance of KNN with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.697	0.928	0.725	0.12
Original Dataset + K-Fold	0.695	0.916	0.727	0.144
PCA of Original Dataset	0.697	0.936	0.722	0.1
PCA of Original Dataset + K-Fold	0.691	0.911	0.726	0.144
UnderSampled Dataset	0.651	0.784	0.742	0.32
UnderSampled Dataset + K-Fold	0.611	0.818	0.633	0.311
PCA of UnderSampled Dataset	0.634	0.784	0.726	0.26
PCA of UnderSampled Dataset + K-Fold	0.604	0.818	0.627	0.293
OverSampled Dataset	0.686	0.824	0.757	0.34
OverSampled Dataset + K Fold	0.765	0.796	0.781	0.727
PCA of OverSampled Dataset	0.72	0.864	0.771	0.36
PCA of OverSampled Dataset + K-Fold	0.761	0.82	0.763	0.689
RandomUnderSampled Dataset	0.646	0.728	0.765	0.44
RandomUnderSampled Dataset + K-Fold	0.606	0.798	0.633	0.329
PCA of RandomUnderSampled Dataset	0.674	0.792	0.762	0.38
PCA of RandomUnderSampled Dataset + K-Fold	0.616	0.814	0.638	0.329
RandomOverSampled Dataset	0.651	0.776	0.746	0.34

RandomOverSampled Dataset + K-Fold	0.729	0.779	0.741	0.669
PCA of RandomOverSampled Dataset	0.68	0.824	0.752	0.32
PCA of RandomOverSampled Dataset + K-Fold	0.741	0.81	0.742	0.657
TomekLinked Dataset	0.686	0.872	0.736	0.22
TomekLinked Dataset + K-Fold	0.699	0.927	0.72	0.18
PCA of TomekLinked Dataset	0.68	0.872	0.732	0.2
PCA of TomekLInked Dataset + K-Fold	0.708	0.934	0.725	0.192
Cluster Centroid Dataset	0.611	0.712	0.736	0.36
Cluster Centroid Dataset + K-Fold	0.609	0.793	0.636	0.341
PCA of Cluster Centroid Dataset	0.589	0.664	0.735	0.4
PCA of Cluster Centroid Dataset + K-Fold	0.592	0.785	0.623	0.311
SMOTE OverSampled Dataset	0.686	0.816	0.761	0.36
SMOTE OverSampled Dataset + K-Fold	0.738	0.805	0.741	0.657
PCA of SMOTE OverSampled Dataset	0.686	0.824	0.757	0.34
PCA of SMOTE OverSampled Dataset + K-Fold	0.732	0.825	0.725	0.619
ENN UnderSampled Dataset	0.594	0.568	0.807	0.66
ENN UnderSampled Dataset + K-Fold	0.682	0.816	0.696	0.491

PCA of ENN UnderSampled Dataset	0.617	0.568	0.845	0.74
PCA of ENN UnderSampled Dataset + K-Fold	0.667	0.82	0.681	0.449
SMOTEENN Combined Sampled Dataset	0.583	0.56	0.795	0.64
SMOTEENN CombinedSampled Dataset + K-Fold	0.797	0.826	0.782	0.767
PCA of SMOTEENN CombinedSampled Dataset	0.577	0.552	0.793	0.64
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.816	0.847	0.799	0.784
SMOTETomek Combined Sampled Dataset	0.669	0.776	0.764	0.4
SMOTETomekCombinedSampled Dataset + K-Fold	0.759	0.837	0.753	0.663
PCA of SMOTETomek CombinedSampled Dataset	0.686	0.808	0.765	0.38
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.759	0.847	0.749	0.65

KNN with One Hot Encoding -

Table 1.10. Performance of KNN with One Hot Encoding on different Evaluation Metrics

	Accuracy	**Sensitivity**	**Precision**	**Specificity**
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled	1.0	1.0	1.0	1.0

Dataset + K-Fold				
TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0

PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

Random Forest with Label Encoding -

Table 1.11. Performance of Random Forest with Label Encoding on different Evaluation Metrics

	Accuracy	**Sensitivity**	**Precision**	**Specificity**
Original Dataset	0.731	0.968	0.738	0.14
Original Dataset + K-Fold	0.705	0.938	0.728	0.126
PCA of Original Dataset	0.709	0.968	0.72	0.06
PCA of Original Dataset + K-Fold	0.703	0.974	0.714	0.03
UnderSampled Dataset	0.623	0.544	0.883	0.82
UnderSampled Dataset + K-Fold	0.687	0.698	0.754	0.671
PCA of UnderSampled Dataset	0.577	0.568	0.78	0.6
PCA of UnderSampled Dataset + K-Fold	0.638	0.715	0.687	0.527
OverSampled Dataset	0.68	0.64	0.879	0.78
OverSampled Dataset + K Fold	0.742	0.69	0.813	0.806
PCA of OverSampled Dataset	0.583	0.544	0.81	0.68
PCA of OverSampled Dataset + K-Fold	0.707	0.635	0.79	0.795
RandomUnderSampled Dataset	0.611	0.528	0.88	0.82
RandomUnderSampled Dataset + K-Fold	0.677	0.69	0.746	0.659
PCA of RandomUnderSampled Dataset	0.589	0.496	0.873	0.82
PCA of RandomUnderSampled Dataset + K-Fold	0.636	0.603	0.734	0.683
RandomOverSampled Dataset	0.646	0.592	0.871	0.78

RandomOverSampled Dataset + K-Fold	0.734	0.668	0.815	0.815
PCA of RandomOverSampled Dataset	0.646	0.664	0.806	0.6
PCA of RandomOverSampled Dataset + K-Fold	0.705	0.644	0.781	0.78
TomekLinked Dataset	0.686	0.776	0.782	0.46
TomekLinked Dataset + K-Fold	0.706	0.861	0.752	0.353
PCA of TomekLinked Dataset	0.686	0.896	0.727	0.16
PCA of TomekLInked Dataset + K-Fold	0.699	0.948	0.713	0.132
Cluster Centroid Dataset	0.629	0.56	0.875	0.8
Cluster Centroid Dataset + K-Fold	0.702	0.694	0.778	0.713
PCA of Cluster Centroid Dataset	0.56	0.48	0.833	0.76
PCA of Cluster Centroid Dataset + K-Fold	0.699	0.698	0.772	0.701
SMOTE OverSampled Dataset	0.629	0.6	0.833	0.7
SMOTE OverSampled Dataset + K-Fold	0.704	0.642	0.781	0.78
PCA of SMOTE OverSampled Dataset	0.571	0.528	0.805	0.68
PCA of SMOTE OverSampled Dataset + K-Fold	0.676	0.601	0.76	0.768
ENN UnderSampled Dataset	0.606	0.488	0.924	0.9
ENN UnderSampled Dataset + K-Fold	0.734	0.711	0.813	0.766

PCA of ENN UnderSampled Dataset	0.617	0.52	0.903	0.86
PCA of ENN UnderSampled Dataset + K-Fold	0.749	0.762	0.802	0.731
SMOTEENN Combined Sampled Dataset	0.606	0.488	0.924	0.9
SMOTEENN CombinedSampled Dataset + K-Fold	0.771	0.689	0.827	0.853
PCA of SMOTEENN CombinedSampled Dataset	0.589	0.528	0.835	0.74
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.803	0.728	0.859	0.879
SMOTETomek Combined Sampled Dataset	0.657	0.616	0.865	0.76
SMOTETomekCombinedSampled Dataset + K-Fold	0.712	0.671	0.776	0.762
PCA of SMOTETomek CombinedSampled Dataset	0.611	0.64	0.777	0.54
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.681	0.691	0.72	0.669

Random Forest with One Hot Encoding -

Table 1.12. Performance of Random Forest with One Hot Encoding on different Evaluation Metrics

	Accuracy	**Sensitivity**	**Precision**	**Specificity**
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

Voting Classifier as Label Encoding -

Table 1.13. Performance of Voting Classifier with Label Encoding on different Evaluation Metrics

	Accuracy	**Sensitivity**	**Precision**	**Specificity**
Original Dataset	0.726	1.0	0.723	0.04
Original Dataset + K-Fold	0.719	0.976	0.725	0.078
PCA of Original Dataset	0.714	1.0	0.714	0.0
PCA of Original Dataset + K-Fold	0.715	0.99	0.718	0.03
UnderSampled Dataset	0.646	0.568	0.899	0.84
UnderSampled Dataset + K-Fold	0.648	0.76	0.681	0.485
PCA of UnderSampled Dataset	0.646	0.592	0.871	0.78
PCA of UnderSampled Dataset + K-Fold	0.648	0.773	0.678	0.467
OverSampled Dataset	0.651	0.592	0.881	0.8
OverSampled Dataset + K Fold	0.692	0.673	0.743	0.716
PCA of OverSampled Dataset	0.651	0.584	0.89	0.82
PCA of OverSampled Dataset + K-Fold	0.661	0.644	0.711	0.68
RandomUnderSampled Dataset	0.651	0.6	0.872	0.78
RandomUnderSampled Dataset + K-Fold	0.645	0.798	0.668	0.425
PCA of RandomUnderSampled Dataset	0.674	0.608	0.905	0.84
PCA of RandomUnderSampled Dataset + K-Fold	0.648	0.802	0.669	0.425
RandomOverSampled Dataset	0.623	0.56	0.864	0.78

RandomOverSampled Dataset + K-Fold	0.703	0.639	0.78	0.78
PCA of RandomOverSampled Dataset	0.64	0.584	0.869	0.78
PCA of RandomOverSampled Dataset + K-Fold	0.671	0.632	0.733	0.718
TomekLinked Dataset	0.731	0.968	0.738	0.14
TomekLinked Dataset + K-Fold	0.704	0.95	0.717	0.144
PCA of TomekLinked Dataset	0.714	0.984	0.719	0.04
PCA of TomekLInked Dataset + K-Fold	0.695	0.966	0.705	0.078
Cluster Centroid Dataset	0.583	0.456	0.919	0.9
Cluster Centroid Dataset + K-Fold	0.694	0.719	0.753	0.659
PCA of Cluster Centroid Dataset	0.56	0.448	0.875	0.84
PCA of Cluster Centroid Dataset + K-Fold	0.702	0.748	0.748	0.635
SMOTE OverSampled Dataset	0.64	0.576	0.878	0.8
SMOTE OverSampled Dataset + K-Fold	0.707	0.656	0.776	0.768
PCA of SMOTE OverSampled Dataset	0.629	0.584	0.849	0.74
PCA of SMOTE OverSampled Dataset + K-Fold	0.654	0.642	0.703	0.669
ENN UnderSampled Dataset	0.6	0.48	0.923	0.9
ENN UnderSampled Dataset + K-Fold	0.741	0.762	0.791	0.713

PCA of ENN UnderSampled Dataset	0.6	0.48	0.923	0.9
PCA of ENN UnderSampled Dataset + K-Fold	0.732	0.774	0.771	0.671
SMOTEENN Combined Sampled Dataset	0.583	0.448	0.933	0.92
SMOTEENN CombinedSampled Dataset + K-Fold	0.805	0.723	0.867	0.888
PCA of SMOTEENN CombinedSampled Dataset	0.571	0.456	0.891	0.86
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.79	0.736	0.828	0.845
SMOTETomek Combined Sampled Dataset	0.651	0.6	0.872	0.78
SMOTETomekCombinedSampled Dataset + K-Fold	0.702	0.676	0.758	0.734
PCA of SMOTETomek CombinedSampled Dataset	0.623	0.648	0.786	0.56
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.677	0.696	0.712	0.653

Voting Classifier with One Hot Encoding -

Table 1.14. Performance of Voting Classifier with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	0.997	1.0	0.996	0.993
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

AdaBoost Classifier with Label Encoding -

Table 1.15. Performance of Adaboost Classifier with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.726	0.816	0.803	0.5
Original Dataset + K-Fold	0.672	0.784	0.763	0.395
PCA of Original Dataset	0.686	0.808	0.765	0.38
PCA of Original Dataset + K-Fold	0.688	0.82	0.761	0.359
UnderSampled Dataset	0.657	0.608	0.874	0.78
UnderSampled Dataset + K-Fold	0.636	0.653	0.709	0.611
PCA of UnderSampled Dataset	0.594	0.6	0.781	0.58
PCA of UnderSampled Dataset + K-Fold	0.616	0.682	0.673	0.521
OverSampled Dataset	0.686	0.696	0.837	0.66
OverSampled Dataset + K Fold	0.748	0.714	0.805	0.789
PCA of OverSampled Dataset	0.657	0.696	0.798	0.56
PCA of OverSampled Dataset + K-Fold	0.742	0.724	0.79	0.765
RandomUnderSampled Dataset	0.64	0.624	0.83	0.68
RandomUnderSampled Dataset + K-Fold	0.663	0.711	0.717	0.593
PCA of RandomUnderSampled Dataset	0.577	0.544	0.8	0.66
PCA of RandomUnderSampled Dataset + K-Fold	0.633	0.694	0.689	0.545
RandomOverSampled Dataset	0.691	0.712	0.832	0.64

RandomOverSampled Dataset + K-Fold	0.753	0.724	0.807	0.789
PCA of RandomOverSampled Dataset	0.634	0.72	0.756	0.42
PCA of RandomOverSampled Dataset + K-Fold	0.737	0.714	0.788	0.765
TomekLinked Dataset	0.669	0.72	0.796	0.54
TomekLinked Dataset + K-Fold	0.681	0.78	0.765	0.455
PCA of TomekLinked Dataset	0.663	0.752	0.77	0.44
PCA of TomekLInked Dataset + K-Fold	0.672	0.811	0.741	0.353
Cluster Centroid Dataset	0.583	0.472	0.894	0.86
Cluster Centroid Dataset + K-Fold	0.697	0.727	0.752	0.653
PCA of Cluster Centroid Dataset	0.531	0.48	0.779	0.66
PCA of Cluster Centroid Dataset + K-Fold	0.677	0.71	0.735	0.629
SMOTE OverSampled Dataset	0.697	0.736	0.821	0.6
SMOTE OverSampled Dataset + K-Fold	0.741	0.736	0.781	0.748
PCA of SMOTE OverSampled Dataset	0.514	0.552	0.704	0.42
PCA of SMOTE OverSampled Dataset + K-Fold	0.678	0.663	0.726	0.695
ENN UnderSampled Dataset	0.606	0.504	0.9	0.86
ENN UnderSampled Dataset + K-Fold	0.751	0.766	0.803	0.731

PCA of ENN UnderSampled Dataset	0.6	0.504	0.887	0.84
PCA of ENN UnderSampled Dataset + K-Fold	0.714	0.757	0.757	0.653
SMOTEENN Combined Sampled Dataset	0.549	0.504	0.788	0.66
SMOTEENN CombinedSampled Dataset + K-Fold	0.777	0.745	0.799	0.81
PCA of SMOTEENN CombinedSampled Dataset	0.611	0.6	0.806	0.64
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.773	0.74	0.795	0.806
SMOTETomek Combined Sampled Dataset	0.663	0.712	0.795	0.54
SMOTETomekCombinedSampled Dataset + K-Fold	0.738	0.746	0.771	0.728
PCA of SMOTETomek CombinedSampled Dataset	0.64	0.656	0.804	0.6
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.688	0.69	0.729	0.684

AdaBoost Classifier with One Hot Encoding -

Table 1.16. Performance of Adaboost Classifier with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	0.989	0.993	0.993	0.974
PCA of Cluster Centroid Dataset + K-Fold	0.995	0.996	0.996	0.993
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

AdaBoost Classifier with SVC as Base Estimator with Label Encoding -

Table 1.17. Performance of Adaboost Classifier with SVC as Base Estimator with Label Encoding on different Evaluation Metrics

	Accuracy	**Sensitivity**	**Precision**	**Specificity**
Original Dataset	0.714	1.0	0.714	0.0
Original Dataset + K-Fold	0.714	1.0	0.714	0.0
PCA of Original Dataset	0.714	1.0	0.714	0.0
PCA of Original Dataset + K-Fold	0.714	1.0	0.714	0.0
UnderSampled Dataset	0.617	0.624	0.796	0.6
UnderSampled Dataset + K-Fold	0.592	1.0	0.592	0.0
PCA of UnderSampled Dataset	0.617	0.616	0.802	0.62
PCA of UnderSampled Dataset + K-Fold	0.592	1.0	0.592	0.0
OverSampled Dataset	0.6	0.6	0.789	0.6
OverSampled Dataset + K Fold	0.55	1.0	0.55	0.0
PCA of OverSampled Dataset	0.6	0.56	0.824	0.7
PCA of OverSampled Dataset + K-Fold	0.55	1.0	0.55	0.0
RandomUnderSampled Dataset	0.571	0.456	0.891	0.86
RandomUnderSampled Dataset + K-Fold	0.592	1.0	0.592	0.0
PCA of RandomUnderSampled Dataset	0.543	0.408	0.895	0.88
PCA of RandomUnderSampled Dataset + K-Fold	0.592	1.0	0.592	0.0
RandomOverSampled Dataset	0.623	0.584	0.839	0.72

RandomOverSampled Dataset + K-Fold	0.55	1.0	0.55	0.0
PCA of RandomOverSampled Dataset	0.571	0.488	0.847	0.78
PCA of RandomOverSampled Dataset + K-Fold	0.55	1.0	0.55	0.0
TomekLinked Dataset	0.714	1.0	0.714	0.0
TomekLinked Dataset + K-Fold	0.695	1.0	0.695	0.0
PCA of TomekLinked Dataset	0.714	1.0	0.714	0.0
PCA of TomekLInked Dataset + K-Fold	0.695	1.0	0.695	0.0
Cluster Centroid Dataset	0.457	0.264	0.917	0.94
Cluster Centroid Dataset + K-Fold	0.592	1.0	0.592	0.0
PCA of Cluster Centroid Dataset	0.497	0.328	0.911	0.92
PCA of Cluster Centroid Dataset + K-Fold	0.592	1.0	0.592	0.0
SMOTE OverSampled Dataset	0.634	0.632	0.84	0.64
SMOTE OverSampled Dataset + K-Fold	0.55	1.0	0.55	0.0
PCA of SMOTE OverSampled Dataset	0.623	0.576	0.847	0.74
PCA of SMOTE OverSampled Dataset + K-Fold	0.55	1.0	0.55	0.0
ENN UnderSampled Dataset	0.286	0.0	nan	1.0
ENN UnderSampled Dataset + K-Fold	0.589	1.0	0.589	0.0
PCA of ENN UnderSampled	0.286	0.0	nan	1.0

Dataset				
PCA of ENN UnderSampled Dataset + K-Fold	0.589	1.0	0.589	0.0
SMOTEENN Combined Sampled Dataset	0.286	0.0	nan	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	0.503	1.0	0.503	0.0
PCA of SMOTEENN CombinedSampled Dataset	0.286	0.0	nan	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.503	1.0	0.503	0.0
SMOTETomek Combined Sampled Dataset	0.629	0.576	0.857	0.76
SMOTETomekCombinedSampled Dataset + K-Fold	0.552	1.0	0.552	0.0
PCA of SMOTETomek CombinedSampled Dataset	0.606	0.544	0.85	0.76
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.552	1.0	0.552	0.0

AdaBoost Classifier with SVC as Base Estimator with One Hot Encoding -

Table 1.18. Performance of Adaboost Classifier with SVC as Base Estimator with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.77	1.0	0.77	0.0
Original Dataset + K-Fold	0.756	1.0	0.756	1.0
PCA of Original Dataset	0.77	1.0	0.77	0.0
PCA of Original Dataset + K-Fold	0.756	1.0	0.756	1.0
UnderSampled Dataset	0.994	0.993	1.0	1.0
UnderSampled Dataset + K-Fold	0.627	1.0	0.627	0.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	0.627	1.0	0.627	0.0
OverSampled Dataset	0.989	0.985	1.0	1.0
OverSampled Dataset + K Fold	0.562	1.0	0.562	0.0
PCA of OverSampled Dataset	0.985	0.985	1.0	1.0
PCA of OverSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
RandomUnderSampled Dataset	0.994	0.993	1.0	1.0
RandomUnderSampled Dataset + K-Fold	0.627	1.0	0.627	0.0
PCA of RandomUnderSampled Dataset	0.994	0.993	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	0.627	1.0	0.627	0.0
RandomOverSampled Dataset	0.989	0.985	1.0	1.0
RandomOverSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
PCA of RandomOverSampled Dataset	0.989	0.985	1.0	1.0
PCA of RandomOverSampled	0.562	1.0	0.562	0.0

Dataset + K-Fold				
TomekLinked Dataset	0.777	1.0	0.777	0.0
TomekLinked Dataset + K-Fold	0.756	1.0	0.756	0.0
PCA of TomekLinked Dataset	0.777	1.0	0.777	0.0
PCA of TomekLInked Dataset + K-Fold	0.756	1.0	0.756	0.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	0.627	1.0	0.627	0.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	0.627	1.0	0.627	0.0
SMOTE OverSampled Dataset	0.994	0.993	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
PCA of SMOTE OverSampled Dataset	0.994	0.993	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
ENN UnderSampled Dataset	0.777	1.0	0.777	0.0
ENN UnderSampled Dataset + K-Fold	0.756	1.0	0.756	0.0
PCA of ENN UnderSampled Dataset	0.777	1.0	0.777	0.0
PCA of ENN UnderSampled Dataset + K-Fold	0.756	1.0	0.756	0.0
SMOTEENN Combined Sampled Dataset	0.994	0.993	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
PCA of SMOTEENN CombinedSampled Dataset	0.989	0.985	1.0	1.0
PCA of SMOTEENN	0.562	1.0	0.562	0.0

CombinedSampled Dataset + K-Fold				
SMOTETomek Combined Sampled Dataset	0.994	0.993	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
PCA of SMOTETomek CombinedSampled Dataset	0.989	0.985	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.562	1.0	0.562	0.0

AdaBoost Classifier with RF as Base Estimator with Label Encoding -

Table 1.19. Performance of Adaboost Classifier with RF as Base Estimator with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.714	0.888	0.755	0.28
Original Dataset + K-Fold	0.686	0.851	0.745	0.275
PCA of Original Dataset	0.703	0.88	0.748	0.26
PCA of Original Dataset + K-Fold	0.708	0.877	0.754	0.287
UnderSampled Dataset	0.64	0.608	0.844	0.72
UnderSampled Dataset + K-Fold	0.66	0.715	0.712	0.581
PCA of UnderSampled Dataset	0.583	0.6	0.765	0.54
PCA of UnderSampled Dataset + K-Fold	0.641	0.723	0.686	0.521
OverSampled Dataset	0.697	0.816	0.773	0.4
OverSampled Dataset + K Fold	0.795	0.781	0.835	0.812
PCA of OverSampled Dataset	0.72	0.856	0.775	0.38
PCA of OverSampled Dataset + K-Fold	0.816	0.82	0.842	0.812
RandomUnderSampled Dataset	0.657	0.616	0.865	0.76
RandomUnderSampled Dataset + K-Fold	0.675	0.723	0.726	0.605
PCA of RandomUnderSampled Dataset	0.611	0.544	0.861	0.78
PCA of RandomUnderSampled Dataset + K-Fold	0.643	0.682	0.705	0.587

RandomOverSampled Dataset	0.669	0.792	0.756	0.36
RandomOverSampled Dataset + K-Fold	0.802	0.791	0.839	0.815
PCA of RandomOverSampled Dataset	0.691	0.832	0.759	0.34
PCA of RandomOverSampled Dataset + K-Fold	0.803	0.808	0.83	0.798
TomekLinked Dataset	0.686	0.824	0.757	0.34
TomekLinked Dataset + K-Fold	0.712	0.835	0.77	0.431
PCA of TomekLinked Dataset	0.697	0.824	0.769	0.38
PCA of TomekLInked Dataset + K-Fold	0.728	0.882	0.764	0.377
Cluster Centroid Dataset	0.623	0.56	0.864	0.78
Cluster Centroid Dataset + K-Fold	0.692	0.723	0.748	0.647
PCA of Cluster Centroid Dataset	0.589	0.52	0.844	0.76
PCA of Cluster Centroid Dataset + K-Fold	0.689	0.723	0.745	0.641
SMOTE OverSampled Dataset	0.674	0.752	0.783	0.48
SMOTE OverSampled Dataset + K-Fold	0.793	0.781	0.831	0.806
PCA of SMOTE OverSampled Dataset	0.634	0.696	0.77	0.48
PCA of SMOTE OverSampled Dataset + K-Fold	0.761	0.745	0.805	0.78
ENN UnderSampled Dataset	0.606	0.504	0.9	0.86
ENN UnderSampled Dataset + K-Fold	0.729	0.728	0.795	0.731

PCA of ENN UnderSampled Dataset	0.589	0.488	0.884	0.84
PCA of ENN UnderSampled Dataset + K-Fold	0.764	0.782	0.81	0.737
SMOTEENN Combined Sampled Dataset	0.606	0.544	0.85	0.76
SMOTEENN CombinedSampled Dataset + K-Fold	0.788	0.745	0.818	0.832
PCA of SMOTEENN CombinedSampled Dataset	0.611	0.592	0.813	0.66
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.809	0.796	0.82	0.823
SMOTETomek Combined Sampled Dataset	0.64	0.76	0.742	0.34
SMOTETomekCombinedSampled Dataset + K-Fold	0.781	0.789	0.809	0.771
PCA of SMOTETomek CombinedSampled Dataset	0.64	0.76	0.742	0.34
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.771	0.794	0.792	0.743

AdaBoost Classifier with RF as Base Estimator with One Hot Encoding -

Table 1.20. Performance of Adaboost Classifier with RF as Base Estimator with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	0.994	1.0	0.993	0.974
PCA of Cluster Centroid Dataset + K-Fold	0.997	1.0	0.996	0.993
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

AdaBoost Classifier with LR as Base Estimator with Label Encoding -

Table 1.21. Performance of Adaboost Classifier with LR as Base Estimator with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.72	0.904	0.753	0.26
Original Dataset + K-Fold	0.708	0.904	0.743	0.222
PCA of Original Dataset	0.731	0.936	0.75	0.22
PCA of Original Dataset + K-Fold	0.705	0.897	0.743	0.228
UnderSampled Dataset	0.634	0.576	0.867	0.78
UnderSampled Dataset + K-Fold	0.658	0.686	0.722	0.617
PCA of UnderSampled Dataset	0.611	0.544	0.861	0.78
PCA of UnderSampled Dataset + K-Fold	0.675	0.678	0.749	0.671
OverSampled Dataset	0.646	0.608	0.854	0.74
OverSampled Dataset + K Fold	0.69	0.7	0.726	0.677
PCA of OverSampled Dataset	0.657	0.648	0.835	0.68
PCA of OverSampled Dataset + K-Fold	0.712	0.69	0.763	0.739
RandomUnderSampled Dataset	0.629	0.6	0.833	0.7
RandomUnderSampled Dataset + K-Fold	0.655	0.64	0.742	0.677
PCA of RandomUnderSampled Dataset	0.634	0.592	0.851	0.74
PCA of RandomUnderSampled Dataset + K-Fold	0.663	0.624	0.763	0.719

RandomOverSampled Dataset	0.623	0.536	0.893	0.84
RandomOverSampled Dataset + K-Fold	0.662	0.618	0.726	0.716
PCA of RandomOverSampled Dataset	0.617	0.52	0.903	0.86
PCA of RandomOverSampled Dataset + K-Fold	0.663	0.596	0.74	0.745
TomekLinked Dataset	0.691	0.872	0.741	0.24
TomekLinked Dataset + K-Fold	0.692	0.85	0.743	0.329
PCA of TomekLinked Dataset	0.691	0.832	0.759	0.34
PCA of TomekLInked Dataset + K-Fold	0.692	0.853	0.742	0.323
Cluster Centroid Dataset	0.657	0.6	0.882	0.8
Cluster Centroid Dataset + K-Fold	0.697	0.715	0.759	0.671
PCA of Cluster Centroid Dataset	0.669	0.648	0.853	0.72
PCA of Cluster Centroid Dataset + K-Fold	0.707	0.707	0.777	0.707
SMOTE OverSampled Dataset	0.657	0.64	0.842	0.7
SMOTE OverSampled Dataset + K-Fold	0.674	0.651	0.727	0.701
PCA of SMOTE OverSampled Dataset	0.617	0.592	0.822	0.68
PCA of SMOTE OverSampled Dataset + K-Fold	0.679	0.649	0.736	0.716
ENN UnderSampled Dataset	0.6	0.488	0.91	0.88
ENN UnderSampled Dataset + K-Fold	0.722	0.674	0.824	0.79

PCA of ENN UnderSampled Dataset	0.6	0.504	0.887	0.84
PCA of ENN UnderSampled Dataset + K-Fold	0.734	0.736	0.796	0.73
SMOTEENN Combined Sampled Dataset	0.617	0.488	0.953	0.94
SMOTEENN CombinedSampled Dataset + K-Fold	0.786	0.74	0.817	0.832
PCA of SMOTEENN CombinedSampled Dataset	0.606	0.496	0.92	0.88
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.773	0.753	0.787	0.793
SMOTETomek Combined Sampled Dataset	0.68	0.616	0.906	0.84
SMOTETomekCombinedSampled Dataset + K-Fold	0.675	0.623	0.747	0.74
PCA of SMOTETomek CombinedSampled Dataset	0.651	0.6	0.872	0.78
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.671	0.616	0.745	0.74

AdaBoost Classifier with LR as Base Estimator with One Hot Encoding -

Table 1.22. Performance of Adaboost Classifier with LR as Base Estimator with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.77	1.0	0.77	0.0
Original Dataset + K-Fold	0.756	1.0	0.756	0.0
PCA of Original Dataset	0.77	1.0	0.77	0.0
PCA of Original Dataset + K-Fold	0.756	1.0	0.756	0.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	0.777	1.0	0.777	0.0
TomekLinked Dataset + K-Fold	0.756	1.0	0.756	0.0
PCA of TomekLinked Dataset	0.777	1.0	0.777	0.0
PCA of TomekLInked Dataset + K-Fold	0.756	1.0	0.756	0.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	0.777	1.0	0.777	0.0
ENN UnderSampled Dataset + K-Fold	0.756	1.0	0.756	0.0
PCA of ENN UnderSampled Dataset	0.777	1.0	0.777	0.0
PCA of ENN UnderSampled Dataset + K-Fold	0.756	1.0	0.756	0.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

Gradient Boosting with Label Encoding -

Table 1.23. Performance of Gradient Boosting Classifier with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.76	0.896	0.794	0.42
Original Dataset + K-Fold	0.719	0.873	0.766	0.335
PCA of Original Dataset	0.691	0.88	0.738	0.22
PCA of Original Dataset + K-Fold	0.693	0.892	0.735	0.198
UnderSampled Dataset	0.611	0.568	0.835	0.72
UnderSampled Dataset + K-Fold	0.677	0.727	0.727	0.605
PCA of UnderSampled Dataset	0.6	0.584	0.802	0.64
PCA of UnderSampled Dataset + K-Fold	0.636	0.694	0.691	0.551
OverSampled Dataset	0.68	0.672	0.848	0.7
OverSampled Dataset + K Fold	0.713	0.692	0.764	0.739
PCA of OverSampled Dataset	0.589	0.592	0.779	0.58
PCA of OverSampled Dataset + K-Fold	0.712	0.68	0.769	0.751
RandomUnderSampled Dataset	0.611	0.544	0.861	0.78
RandomUnderSampled Dataset + K-Fold	0.68	0.727	0.73	0.61
PCA of RandomUnderSampled Dataset	0.583	0.544	0.81	0.68
PCA of RandomUnderSampled Dataset + K-Fold	0.655	0.69	0.717	0.605
RandomOverSampled Dataset	0.657	0.64	0.842	0.7

RandomOverSampled Dataset + K-Fold	0.716	0.697	0.765	0.739
PCA of RandomOverSampled Dataset	0.64	0.688	0.782	0.52
PCA of RandomOverSampled Dataset + K-Fold	0.707	0.69	0.755	0.727
TomekLinked Dataset	0.703	0.8	0.787	0.46
TomekLinked Dataset + K-Fold	0.708	0.832	0.768	0.425
PCA of TomekLinked Dataset	0.64	0.776	0.735	0.3
PCA of TomekLInked Dataset + K-Fold	0.692	0.864	0.738	0.299
Cluster Centroid Dataset	0.617	0.536	0.882	0.82
Cluster Centroid Dataset + K-Fold	0.682	0.702	0.746	0.653
PCA of Cluster Centroid Dataset	0.526	0.432	0.818	0.76
PCA of Cluster Centroid Dataset + K-Fold	0.682	0.719	0.737	0.629
SMOTE OverSampled Dataset	0.629	0.656	0.788	0.56
SMOTE OverSampled Dataset + K-Fold	0.72	0.719	0.759	0.721
PCA of SMOTE OverSampled Dataset	0.583	0.584	0.777	0.58
PCA of SMOTE OverSampled Dataset + K-Fold	0.659	0.635	0.714	0.689
ENN UnderSampled Dataset	0.583	0.48	0.882	0.84
ENN UnderSampled Dataset + K-Fold	0.717	0.736	0.772	0.689
PCA of ENN UnderSampled	0.594	0.496	0.886	0.84

Dataset				
PCA of ENN UnderSampled Dataset + K-Fold	0.714	0.762	0.755	0.647
SMOTEENN Combined Sampled Dataset	0.577	0.472	0.881	0.84
SMOTEENN CombinedSampled Dataset + K-Fold	0.773	0.715	0.812	0.832
PCA of SMOTEENN CombinedSampled Dataset	0.537	0.48	0.789	0.68
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.784	0.74	0.813	0.828
SMOTETomek Combined Sampled Dataset	0.629	0.64	0.8	0.6
SMOTETomekCombinedSampled Dataset + K-Fold	0.725	0.726	0.765	0.724
PCA of SMOTETomek CombinedSampled Dataset	0.611	0.624	0.788	0.58
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.7	0.686	0.75	0.718

Gradient Boosting with One Hot Encoding -

Table 1.24. Performance of Gradient Boosting Classifier with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	0.989	0.993	0.993	0.974
PCA of Cluster Centroid Dataset + K-Fold	0.995	0.996	0.996	0.993
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

XGB Classifier with Label Encoding -

Table 1.25. Performance of XGB Classifier with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.714	1.0	0.714	0.0
Original Dataset + K-Fold	0.714	1.0	0.714	0.0
PCA of Original Dataset	0.714	1.0	0.714	0.0
PCA of Original Dataset + K-Fold	0.714	1.0	0.714	0.0
UnderSampled Dataset	0.646	0.64	0.825	0.66
UnderSampled Dataset + K-Fold	0.592	1.0	0.592	0.0
PCA of UnderSampled Dataset	0.657	0.792	0.744	0.32
PCA of UnderSampled Dataset + K-Fold	0.592	1.0	0.592	0.0
OverSampled Dataset	0.634	0.632	0.814	0.64
OverSampled Dataset + K Fold	0.55	1.0	0.55	0.0
PCA of OverSampled Dataset	0.663	0.776	0.758	0.38
PCA of OverSampled Dataset + K-Fold	0.55	1.0	0.55	0.0
RandomUnderSampled Dataset	0.617	0.616	0.802	0.62
RandomUnderSampled Dataset + K-Fold	0.592	1.0	0.592	0.0
PCA of RandomUnderSampled Dataset	0.629	0.696	0.763	0.46
PCA of RandomUnderSampled Dataset + K-Fold	0.592	1.0	0.592	0.0
RandomOverSampled Dataset	0.629	0.576	0.857	0.76
RandomOverSampled Dataset + K-Fold	0.55	1.0	0.55	0.0

PCA of RandomOverSampled Dataset	0.663	0.752	0.77	0.44
PCA of RandomOverSampled Dataset + K-Fold	0.55	1.0	0.55	0.0
TomekLinked Dataset	0.714	1.0	0.714	0.0
TomekLinked Dataset + K-Fold	0.695	1.0	0.695	0.0
PCA of TomekLinked Dataset	0.714	1.0	0.714	0.0
PCA of TomekLInked Dataset + K-Fold	0.695	1.0	0.695	0.0
Cluster Centroid Dataset	0.549	0.408	0.911	0.9
Cluster Centroid Dataset + K-Fold	0.592	1.0	0.592	0.0
PCA of Cluster Centroid Dataset	0.623	0.568	0.855	0.76
PCA of Cluster Centroid Dataset + K-Fold	0.592	1.0	0.592	0.0
SMOTE OverSampled Dataset	0.64	0.608	0.844	0.72
SMOTE OverSampled Dataset + K-Fold	0.55	1.0	0.55	0.0
PCA of SMOTE OverSampled Dataset	0.663	0.784	0.754	0.36
PCA of SMOTE OverSampled Dataset + K-Fold	0.55	1.0	0.55	0.0
ENN UnderSampled Dataset	0.377	0.136	0.944	0.98
ENN UnderSampled Dataset + K-Fold	0.589	1.0	0.589	0.0
PCA of ENN UnderSampled Dataset	0.4	0.168	0.955	0.98
PCA of ENN UnderSampled Dataset + K-Fold	0.589	1.0	0.589	0.0
SMOTEENN Combined Sampled Dataset	0.286	0.0	nan	1.0

SMOTEENN CombinedSampled Dataset + K-Fold	0.67	0.834	0.63	0.504
PCA of SMOTEENN CombinedSampled Dataset	0.286	0.0	nan	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.636	0.864	0.595	0.405
SMOTETomek Combined Sampled Dataset	0.623	0.568	0.855	0.76
SMOTETomekCombinedSampled Dataset + K-Fold	0.552	1.0	0.552	0.0
PCA of SMOTETomek CombinedSampled Dataset	0.651	0.72	0.776	0.48
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.552	1.0	0.552	0.0

XGB Classifier with One Hot Encoding -

Table 1.26. Performance of XGB Classifier with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.77	1.0	0.77	0.0
Original Dataset + K-Fold	0.756	1.0	0.756	0.0
PCA of Original Dataset	0.77	1.0	0.77	0.0
PCA of Original Dataset + K-Fold	0.756	1.0	0.756	0.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	0.627	1.0	0.627	0.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	0.627	1.0	0.627	0.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	0.562	1.0	0.562	0.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	0.627	1.0	0.627	0.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	0.627	1.0	0.627	0.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	0.562	1.0	0.562	0.0

TomekLinked Dataset	0.777	1.0	0.777	0.0
TomekLinked Dataset + K-Fold	0.756	1.0	0.756	0.0
PCA of TomekLinked Dataset	0.777	1.0	0.777	0.0
PCA of TomekLInked Dataset + K-Fold	0.756	1.0	0.756	0.0
Cluster Centroid Dataset	0.954	1.0	0.993	0.974
Cluster Centroid Dataset + K-Fold	0.627	1.0	0.627	0.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	0.627	1.0	0.627	0.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
ENN UnderSampled Dataset	0.777	1.0	0.777	0.0
ENN UnderSampled Dataset + K-Fold	0.756	1.0	0.756	0.0
PCA of ENN UnderSampled Dataset	0.777	1.0	0.777	0.0
PCA of ENN UnderSampled Dataset + K-Fold	0.756	1.0	0.756	0.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.562	1.0	0.562	0.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	0.562	1.0	0.562	0.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.562	1.0	0.562	0.0

BaggingClassifier with Label Encoding -

Table 1.27. Performance of Bagging Classifier with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.709	0.888	0.75	0.26
Original Dataset + K-Fold	0.683	0.846	0.744	0.275
PCA of Original Dataset	0.68	0.84	0.745	0.28
PCA of Original Dataset + K-Fold	0.705	0.849	0.764	0.347
UnderSampled Dataset	0.674	0.696	0.821	0.62
UnderSampled Dataset + K-Fold	0.67	0.748	0.71	0.557
PCA of UnderSampled Dataset	0.674	0.696	0.821	0.62
PCA of UnderSampled Dataset + K-Fold	0.67	0.748	0.71	0.557
OverSampled Dataset	0.697	0.824	0.769	0.38
OverSampled Dataset + K Fold	0.801	0.815	0.821	0.783
PCA of OverSampled Dataset	0.68	0.816	0.756	0.34
PCA of OverSampled Dataset + K-Fold	0.811	0.837	0.823	0.78
RandomUnderSampled Dataset	0.629	0.648	0.794	0.58
RandomUnderSampled Dataset + K-Fold	0.692	0.802	0.713	0.533
PCA of RandomUnderSampled Dataset	0.669	0.648	0.853	0.72
PCA of RandomUnderSampled Dataset + K-Fold	0.636	0.736	0.677	0.491

RandomOverSampled Dataset	0.72	0.864	0.771	0.36
RandomOverSampled Dataset + K-Fold	0.797	0.81	0.818	0.78
PCA of RandomOverSampled Dataset	0.686	0.784	0.778	0.44
PCA of RandomOverSampled Dataset + K-Fold	0.802	0.815	0.823	0.786
TomekLinked Dataset	0.703	0.832	0.77	0.38
TomekLinked Dataset + K-Fold	0.695	0.832	0.755	0.383
PCA of TomekLinked Dataset	0.703	0.832	0.77	0.38
PCA of TomekLInked Dataset + K-Fold	0.719	0.866	0.762	0.383
Cluster Centroid Dataset	0.64	0.616	0.837	0.7
Cluster Centroid Dataset + K-Fold	0.692	0.764	0.728	0.587
PCA of Cluster Centroid Dataset	0.611	0.592	0.813	0.66
PCA of Cluster Centroid Dataset + K-Fold	0.689	0.748	0.733	0.605
SMOTE OverSampled Dataset	0.669	0.768	0.768	0.42
SMOTE OverSampled Dataset + K-Fold	0.773	0.803	0.788	0.736
PCA of SMOTE OverSampled Dataset	0.634	0.744	0.744	0.36
PCA of SMOTE OverSampled Dataset + K-Fold	0.727	0.767	0.744	0.677
ENN UnderSampled Dataset	0.617	0.544	0.872	0.8
ENN UnderSampled Dataset + K-Fold	0.717	0.774	0.752	0.635

PCA of ENN UnderSampled Dataset	0.594	0.512	0.865	0.8
PCA of ENN UnderSampled Dataset + K-Fold	0.736	0.778	0.775	0.677
SMOTEENN Combined Sampled Dataset	0.594	0.552	0.821	0.7
SMOTEENN CombinedSampled Dataset + K-Fold	0.764	0.787	0.755	0.741
PCA of SMOTEENN CombinedSampled Dataset	0.623	0.608	0.817	0.66
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.764	0.787	0.755	0.741
SMOTETomek Combined Sampled Dataset	0.64	0.76	0.742	0.34
SMOTETomekCombinedSampled Dataset + K-Fold	0.734	0.794	0.742	0.659
PCA of SMOTETomek CombinedSampled Dataset	0.646	0.736	0.76	0.42
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.735	0.786	0.747	0.672

BaggingClassifier with One Hot Encoding -

Table 1.28. Performance of Bagging Classifier with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled	1.0	1.0	1.0	1.0

Dataset + K-Fold				
TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	0.989	0.993	0.993	0.974
PCA of Cluster Centroid Dataset + K-Fold	0.992	0.996	0.992	0.986
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-	1.0	1.0	1.0	1.0

Fold				
SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

BaggingClassifier with Perceptron as Base Estimator with Label Encoding -

Table 1.29. Performance of Bagging Classifier with Perception as Base Estimator with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.72	0.922	0.721	0.04
Original Dataset + K-Fold	0.693	0.916	0.726	0.138
PCA of Original Dataset	0.68	0.872	0.732	0.2
PCA of Original Dataset + K-Fold	0.707	0.957	0.722	0.084
UnderSampled Dataset	0.709	0.824	0.78	0.42
UnderSampled Dataset + K-Fold	0.628	0.888	0.632	0.251
PCA of UnderSampled Dataset	0.709	0.824	0.78	0.42
PCA of UnderSampled Dataset + K-Fold	0.628	0.888	0.632	0.251
OverSampled Dataset	0.691	0.824	0.763	0.36
OverSampled Dataset + K Fold	0.631	0.656	0.667	0.601
PCA of OverSampled Dataset	0.691	0.8	0.775	0.42
PCA of OverSampled Dataset + K-Fold	0.621	0.805	0.619	0.396
RandomUnderSampled Dataset	0.629	0.536	0.905	0.86
RandomUnderSampled Dataset + K-Fold	0.655	0.839	0.666	0.389
PCA of RandomUnderSampled Dataset	0.611	0.512	0.901	0.86
PCA of RandomUnderSampled Dataset + K-Fold	0.653	0.843	0.662	0.377

RandomOverSampled Dataset	0.709	0.776	0.808	0.54
RandomOverSampled Dataset + K-Fold	0.629	0.733	0.642	0.501
PCA of RandomOverSampled Dataset	0.674	0.64	0.87	0.76
PCA of RandomOverSampled Dataset + K-Fold	0.657	0.524	0.779	0.818
TomekLinked Dataset	0.743	0.992	0.738	0.12
TomekLinked Dataset + K-Fold	0.693	0.882	0.732	0.263
PCA of TomekLinked Dataset	0.737	0.984	0.737	0.12
PCA of TomekLInked Dataset + K-Fold	0.682	0.945	0.702	0.084
Cluster Centroid Dataset	0.554	0.424	0.898	0.88
Cluster Centroid Dataset + K-Fold	0.685	0.653	0.778	0.731
PCA of Cluster Centroid Dataset	0.68	0.712	0.817	0.6
PCA of Cluster Centroid Dataset + K-Fold	0.709	0.74	0.762	0.665
SMOTE OverSampled Dataset	0.623	0.536	0.893	0.84
SMOTE OverSampled Dataset + K-Fold	0.61	0.829	0.606	0.343
PCA of SMOTE OverSampled Dataset	0.663	0.64	0.851	0.72
PCA of SMOTE OverSampled Dataset + K-Fold	0.657	0.661	0.698	0.651
ENN UnderSampled Dataset	0.691	0.632	0.908	0.84
ENN UnderSampled Dataset + K-Fold	0.719	0.812	0.738	0.587

PCA of ENN UnderSampled Dataset	0.606	0.472	0.952	0.94
PCA of ENN UnderSampled Dataset + K-Fold	0.7	0.82	0.713	0.527
SMOTEENN Combined Sampled Dataset	0.583	0.448	0.933	0.92
SMOTEENN CombinedSampled Dataset + K-Fold	0.805	0.711	0.879	0.901
PCA of SMOTEENN CombinedSampled Dataset	0.566	0.4	0.98	0.98
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.805	0.711	0.879	0.901
SMOTETomek Combined Sampled Dataset	0.594	0.464	0.935	0.92
SMOTETomekCombinedSampled Dataset + K-Fold	0.653	0.653	0.699	0.653
PCA of SMOTETomek CombinedSampled Dataset	0.64	0.624	0.83	0.68
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.646	0.616	0.706	0.684

BaggingClassifier with Perceptron as Base Estimator with One Hot Encoding -

Table 1.30. Performance of Bagging Classifier with Perception as Base Estimator with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

Bagging Classifier with KNN as Base Estimator with Label Encoding -

Table 1.31. Performance of Bagging Classifier with KNN as Base Estimator with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.646	0.824	0.72	0.2
Original Dataset + K-Fold	0.666	0.83	0.747	0.323
PCA of Original Dataset	0.64	0.84	0.709	0.14
PCA of Original Dataset + K-Fold	0.657	0.805	0.738	0.287
UnderSampled Dataset	0.537	0.504	0.768	0.62
UnderSampled Dataset + K-Fold	0.606	0.599	0.694	0.617
PCA of UnderSampled Dataset	0.537	0.504	0.768	0.62
PCA of UnderSampled Dataset + K-Fold	0.606	0.599	0.694	0.617
OverSampled Dataset	0.64	0.632	0.823	0.66
OverSampled Dataset + K Fold	0.704	0.603	0.81	0.827
PCA of OverSampled Dataset	0.629	0.632	0.806	0.62
PCA of OverSampled Dataset + K-Fold	0.707	0.632	0.792	0.798
RandomUnderSampled Dataset	0.554	0.488	0.813	0.72
RandomUnderSampled Dataset + K-Fold	0.623	0.616	0.71	0.635
PCA of RandomUnderSampled Dataset	0.577	0.552	0.793	0.64
PCA of RandomUnderSampled	0.641	0.665	0.709	0.605

Dataset + K-Fold				
RandomOverSampled Dataset	0.566	0.576	0.758	0.54
RandomOverSampled Dataset + K-Fold	0.662	0.589	0.742	0.751
PCA of RandomOverSampled Dataset	0.611	0.64	0.777	0.54
PCA of RandomOverSampled Dataset + K-Fold	0.675	0.62	0.746	0.742
TomekLinked Dataset	0.657	0.76	0.76	0.4
TomekLinked Dataset + K-Fold	0.664	0.759	0.759	0.449
PCA of TomekLinked Dataset	0.674	0.768	0.774	0.44
PCA of TomekLInked Dataset + K-Fold	0.679	0.787	0.759	0.431
Cluster Centroid Dataset	0.52	0.408	0.836	0.8
Cluster Centroid Dataset + K-Fold	0.633	0.62	0.721	0.653
PCA of Cluster Centroid Dataset	0.537	0.416	0.867	0.84
PCA of Cluster Centroid Dataset + K-Fold	0.655	0.616	0.756	0.713
SMOTE OverSampled Dataset	0.606	0.568	0.826	0.7
SMOTE OverSampled Dataset + K-Fold	0.664	0.579	0.753	0.768
PCA of SMOTE OverSampled Dataset	0.611	0.6	0.806	0.64
PCA of SMOTE OverSampled Dataset + K-Fold	0.687	0.63	0.759	0.757
ENN UnderSampled Dataset	0.537	0.4	0.893	0.88
ENN UnderSampled Dataset + K-Fold	0.732	0.707	0.812	0.766
PCA of ENN UnderSampled	0.537	0.384	0.923	0.92

Dataset				
PCA of ENN UnderSampled Dataset + K-Fold	0.734	0.707	0.816	0.772
SMOTEENN Combined Sampled Dataset	0.543	0.384	0.941	0.94
SMOTEENN CombinedSampled Dataset + K-Fold	0.803	0.715	0.87	0.892
PCA of SMOTEENN CombinedSampled Dataset	0.537	0.4	0.893	0.88
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.803	0.715	0.87	0.892
SMOTETomek Combined Sampled Dataset	0.56	0.544	0.773	0.6
SMOTETomekCombinedSampled Dataset + K-Fold	0.685	0.583	0.792	0.811
PCA of SMOTETomek CombinedSampled Dataset	0.566	0.576	0.758	0.54
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.698	0.613	0.792	0.802

Bagging Classifier with KNN as Base Estimator with One Hot Encoding -

Table 1.32. Performance of Bagging Classifier with KNN as Base Estimator with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
TomekLinked Dataset	1.0	1.0	1.0	1.0

TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

Bagging Classifier with SVC as Base Estimator with Label Encoding -

Table 1.33. Performance of Bagging Classifier with SVC as Base Estimator with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.714	1.0	0.714	0.0
Original Dataset + K-Fold	0.714	1.0	0.714	0.0
PCA of Original Dataset	0.714	1.0	0.714	0.0
PCA of Original Dataset + K-Fold	0.71	0.995	0.713	0.0
UnderSampled Dataset	0.629	0.592	0.841	0.72
UnderSampled Dataset + K-Fold	0.645	0.773	0.675	0.461
PCA of UnderSampled Dataset	0.629	0.592	0.841	0.72
PCA of UnderSampled Dataset + K-Fold	0.645	0.773	0.675	0.4611
OverSampled Dataset	0.663	0.608	0.884	0.8
OverSampled Dataset + K Fold	0.668	0.62	0.735	0.727
PCA of OverSampled Dataset	0.663	0.608	0.884	0.8
PCA of OverSampled Dataset + K-Fold	0.661	0.659	0.704	0.663
RandomUnderSampled Dataset	0.623	0.544	0.883	0.82
RandomUnderSampled Dataset + K-Fold	0.663	0.707	0.718	0.599
PCA of RandomUnderSampled Dataset	0.617	0.512	0.914	0.88
PCA of RandomUnderSampled Dataset + K-Fold	0.628	0.69	0.684	0.539
RandomOverSampled Dataset	0.606	0.528	0.868	0.8

RandomOverSampled Dataset + K-Fold	0.69	0.627	0.765	0.765
PCA of RandomOverSampled Dataset	0.6	0.536	0.848	0.76
PCA of RandomOverSampled Dataset + K-Fold	0.659	0.63	0.716	0.695
TomekLinked Dataset	0.72	0.976	0.726	0.08
TomekLinked Dataset + K-Fold	0.697	1.0	0.697	0.06
PCA of TomekLinked Dataset	0.709	0.936	0.731	0.14
PCA of TomekLInked Dataset + K-Fold	0.704	0.969	0.711	0.102
Cluster Centroid Dataset	0.526	0.4	0.862	0.84
Cluster Centroid Dataset + K-Fold	0.677	0.665	0.759	0.695
PCA of Cluster Centroid Dataset	0.537	0.424	0.855	0.82
PCA of Cluster Centroid Dataset + K-Fold	0.682	0.69	0.752	0.671
SMOTE OverSampled Dataset	0.594	0.48	0.909	0.88
SMOTE OverSampled Dataset + K-Fold	0.688	0.606	0.778	0.789
PCA of SMOTE OverSampled Dataset	0.589	0.52	0.853	0.78
PCA of SMOTE OverSampled Dataset + K-Fold	0.675	0.618	0.747	0.745
ENN UnderSampled Dataset	0.554	0.416	0.912	0.9
ENN UnderSampled Dataset + K-Fold	0.739	0.728	0.809	0.754
PCA of ENN UnderSampled Dataset	0.56	0.416	0.929	0.92

PCA of ENN UnderSampled Dataset + K-Fold	0.736	0.72	0.811	0.76
SMOTEENN Combined Sampled Dataset	0.56	0.424	0.914	0.9
SMOTEENN CombinedSampled Dataset + K-Fold	0.786	0.668	0.877	0.905
PCA of SMOTEENN CombinedSampled Dataset	0.571	0.44	0.917	0.9
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.786	0.668	0.877	0.905
SMOTETomek Combined Sampled Dataset	0.634	0.568	0.877	0.8
SMOTETomekCombinedSampled Dataset + K-Fold	0.669	0.618	0.739	0.731
PCA of SMOTETomek CombinedSampled Dataset	0.589	0.544	0.819	0.7
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.656	0.626	0.716	0.693

BaggingClassifier with SVC as Base Estimator with One Hot Encoding-

Table 1.34. Performance of Bagging Classifier with SVC as Base Estimator with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	0.994	0.993	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

BaggingClassifier with RF as Base Estimator with Label Encoding

Table 1.35. Performance of Bagging Classifier with RF as Base Estimator with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.714	0.912	0.745	0.22
Original Dataset + K-Fold	0.702	0.892	0.742	0.228
PCA of Original Dataset	0.726	0.936	0.745	0.2
PCA of Original Dataset + K-Fold	0.707	0.913	0.738	0.192
UnderSampled Dataset	0.634	0.592	0.851	0.74
UnderSampled Dataset + K-Fold	0.667	0.698	0.728	0.623
PCA of UnderSampled Dataset	0.634	0.592	0.851	0.74
PCA of UnderSampled Dataset + K-Fold	0.667	0.698	0.728	0.623
OverSampled Dataset	0.64	0.712	0.767	0.46
OverSampled Dataset + K Fold	0.777	0.757	0.822	0.801
PCA of OverSampled Dataset	0.663	0.752	0.77	0.44
PCA of OverSampled Dataset + K-Fold	0.794	0.791	0.827	0.798
RandomUnderSampled Dataset	0.646	0.584	0.88	0.8
RandomUnderSampled Dataset + K-Fold	0.667	0.678	0.739	0.653
PCA of RandomUnderSampled Dataset	0.611	0.536	0.87	0.8
PCA of RandomUnderSampled Dataset + K-Fold	0.626	0.632	0.705	0.617

RandomOverSampled Dataset	0.646	0.704	0.779	0.5
RandomOverSampled Dataset + K-Fold	0.786	0.75	0.843	0.83
PCA of RandomOverSampled Dataset	0.669	0.76	0.772	0.44
PCA of RandomOverSampled Dataset + K-Fold	0.777	0.75	0.828	0.809
TomekLinked Dataset	0.663	0.784	0.754	0.36
TomekLinked Dataset + K-Fold	0.704	0.837	0.761	0.401
PCA of TomekLinked Dataset	0.703	0.856	0.759	0.32
PCA of TomekLInked Dataset + K-Fold	0.734	0.906	0.758	0.341
Cluster Centroid Dataset	0.606	0.536	0.859	0.78
Cluster Centroid Dataset + K-Fold	0.707	0.719	0.689	0.744
PCA of Cluster Centroid Dataset	0.571	0.512	0.821	0.72
PCA of Cluster Centroid Dataset + K-Fold	0.699	0.719	0.76	0.671
SMOTE OverSampled Dataset	0.617	0.688	0.754	0.44
SMOTE OverSampled Dataset + K-Fold	0.775	0.75	0.825	0.806
PCA of SMOTE OverSampled Dataset	0.611	0.656	0.766	0.5
PCA of SMOTE OverSampled Dataset + K-Fold	0.731	0.7	0.786	0.768
ENN UnderSampled Dataset	0.611	0.496	0.925	0.9
ENN UnderSampled Dataset + K-	0.727	0.72	0.796	0.737

Fold				
PCA of ENN UnderSampled Dataset	0.606	0.472	0.952	0.94
PCA of ENN UnderSampled Dataset + K-Fold	0.771	0.774	0.826	0.766
SMOTEENN Combined Sampled Dataset	0.617	0.504	0.926	0.9
SMOTEENN CombinedSampled Dataset + K-Fold	0.797	0.74	0.837	0.853
PCA of SMOTEENN CombinedSampled Dataset	0.589	0.536	0.827	0.72
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.797	0.74	0.837	0.853
SMOTETomek Combined Sampled Dataset	0.623	0.704	0.752	0.42
SMOTETomekCombinedSampled Dataset + K-Fold	0.761	0.764	0.796	0.759
PCA of SMOTETomek CombinedSampled Dataset	0.646	0.72	0.769	0.46
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.75	0.761	0.781	0.737

BaggingClassifier with RF as Base Estimator with One Hot Encoding -

Table 1.36. Performance of Bagging Classifier with RF as Base Estimator with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	0.994	1.0	0.993	0.974
PCA of Cluster Centroid Dataset + K-Fold	0.997	1.0	0.996	0.993
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

BaggingClassifier with LR as Base Estimator with Label Encoding -

Table 1.37. Performance of Bagging Classifier with LR as Base Estimator with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.714	1.0	0.714	0.0
Original Dataset + K-Fold	0.717	0.99	0.719	0.036
PCA of Original Dataset	0.714	1.0	0.714	0.0
PCA of Original Dataset + K-Fold	0.72	0.995	0.72	0.036
UnderSampled Dataset	0.686	0.64	0.889	0.8
UnderSampled Dataset + K-Fold	0.633	0.835	0.647	0.341
PCA of UnderSampled Dataset	0.686	0.64	0.889	0.8
PCA of UnderSampled Dataset + K-Fold	0.633	0.835	0.647	0.341
OverSampled Dataset	0.691	0.656	0.882	0.78
OverSampled Dataset + K Fold	0.686	0.7	0.72	0.669
PCA of OverSampled Dataset	0.686	0.648	0.88	0.78
PCA of OverSampled Dataset + K-Fold	0.654	0.68	0.687	0.622
RandomUnderSampled Dataset	0.669	0.616	0.885	0.8
RandomUnderSampled Dataset + K-Fold	0.66	0.789	0.685	0.473
PCA of RandomUnderSampled Dataset	0.68	0.64	0.879	0.78
PCA of RandomUnderSampled Dataset + K-Fold	0.648	0.777	0.676	0.461

RandomOverSampled Dataset	0.651	0.584	0.89	0.82
RandomOverSampled Dataset + K-Fold	0.671	0.683	0.708	0.657
PCA of RandomOverSampled Dataset	0.651	0.592	0.881	0.8
PCA of RandomOverSampled Dataset + K-Fold	0.631	0.666	0.664	0.589
TomekLinked Dataset	0.72	0.968	0.729	0.1
TomekLinked Dataset + K-Fold	0.693	0.961	0.705	0.084
PCA of TomekLinked Dataset	0.726	0.976	0.731	0.1
PCA of TomekLInked Dataset + K-Fold	0.695	0.966	0.705	0.078
Cluster Centroid Dataset	0.589	0.472	0.908	0.88
Cluster Centroid Dataset + K-Fold	0.68	0.748	0.721	0.581
PCA of Cluster Centroid Dataset	0.594	0.472	0.922	0.9
PCA of Cluster Centroid Dataset + K-Fold	0.687	0.752	0.728	0.593
SMOTE OverSampled Dataset	0.651	0.584	0.89	0.82
SMOTE OverSampled Dataset + K-Fold	0.663	0.69	0.695	0.63
PCA of SMOTE OverSampled Dataset	0.64	0.584	0.869	0.78
PCA of SMOTE OverSampled Dataset + K-Fold	0.641	0.685	0.669	0.587
ENN UnderSampled Dataset	0.589	0.456	0.934	0.92
ENN UnderSampled Dataset + K-	0.709	0.766	0.747	0.629

Fold				
PCA of ENN UnderSampled Dataset	0.594	0.464	0.935	0.92
PCA of ENN UnderSampled Dataset + K-Fold	0.722	0.782	0.754	0.635
SMOTEENN Combined Sampled Dataset	0.56	0.408	0.944	0.94
SMOTEENN CombinedSampled Dataset + K-Fold	0.792	0.719	0.845	0.866
PCA of SMOTEENN CombinedSampled Dataset	0.566	0.416	0.945	0.94
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.792	0.719	0.845	0.866
SMOTETomek Combined Sampled Dataset	0.691	0.656	0.882	0.78
SMOTETomekCombinedSampled Dataset + K-Fold	0.664	0.693	0.697	0.628
PCA of SMOTETomek CombinedSampled Dataset	0.674	0.632	0.878	0.78
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.659	0.686	0.693	0.625

BaggingClassifier with LR as Base Estimator with One Hot Encoding -

Table 1.38. Performance of Bagging Classifier with LR as Base Estimator with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	1.0	1.0	1.0	1.0
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-	1.0	1.0	1.0	1.0

Fold				
SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

Perceptron with Label Encoding -

Table 1.39. Performance of Perceptron with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.703	0.856	0.759	0.32
Original Dataset + K-Fold	0.638	0.712	0.765	0.455
PCA of Original Dataset	0.731	0.984	0.732	0.1
PCA of Original Dataset + K-Fold	0.616	0.608	0.806	0.635
UnderSampled Dataset	0.714	0.864	0.766	0.34
UnderSampled Dataset + K-Fold	0.626	0.715	0.673	0.497
PCA of UnderSampled Dataset	0.714	0.864	0.766	0.34
PCA of UnderSampled Dataset + K-Fold	0.626	0.715	0.673	0.497
OverSampled Dataset	0.686	0.736	0.807	0.56
OverSampled Dataset + K Fold	0.606	0.853	0.6	0.305
PCA of OverSampled Dataset	0.714	1.0	0.714	0.0
PCA of OverSampled Dataset + K-Fold	0.637	0.488	0.766	0.818
RandomUnderSampled Dataset	0.503	0.312	0.975	0.98
RandomUnderSampled Dataset + K-Fold	0.631	0.913	0.63	0.222
PCA of RandomUnderSampled Dataset	0.543	0.456	0.826	0.76
PCA of RandomUnderSampled Dataset + K-Fold	0.599	0.798	0.627	0.311
RandomOverSampled Dataset	0.451	0.24	0.968	0.98
RandomOverSampled Dataset + K-	0.634	0.375	0.902	0.95

Fold				
PCA of RandomOverSampled Dataset	0.526	0.456	0.792	0.7
PCA of RandomOverSampled Dataset + K-Fold	0.565	0.666	0.593	0.443
TomekLinked Dataset	0.497	0.304	0.974	0.98
TomekLinked Dataset + K-Fold	0.673	0.808	0.744	0.365
PCA of TomekLinked Dataset	0.566	0.416	0.945	0.94
PCA of TomekLInked Dataset + K-Fold	0.655	0.79	0.734	0.347
Cluster Centroid Dataset	0.474	0.28	0.946	0.96
Cluster Centroid Dataset + K-Fold	0.665	0.715	0.718	0.593
PCA of Cluster Centroid Dataset	0.503	0.408	0.797	0.74
PCA of Cluster Centroid Dataset + K-Fold	0.641	0.748	0.678	0.485
SMOTE OverSampled Dataset	0.669	0.752	0.777	0.46
SMOTE OverSampled Dataset + K-Fold	0.653	0.587	0.728	0.733
PCA of SMOTE OverSampled Dataset	0.68	0.88	0.728	0.18
PCA of SMOTE OverSampled Dataset + K-Fold	0.624	0.416	0.805	0.877
ENN UnderSampled Dataset	0.451	0.24	0.968	0.98
ENN UnderSampled Dataset + K-Fold	0.648	0.669	0.714	0.617
PCA of ENN UnderSampled Dataset	0.497	0.304	0.974	0.98
PCA of ENN UnderSampled Dataset + K-Fold	0.707	0.753	0.75	0.641

SMOTEENN Combined Sampled Dataset	0.611	0.52	0.89	0.84
SMOTEENN CombinedSampled Dataset + K-Fold	0.7	0.672	0.715	0.728
PCA of SMOTEENN CombinedSampled Dataset	0.697	0.872	0.747	0.26
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	0.7	0.672	0.715	0.728
SMOTETomek Combined Sampled Dataset	0.674	0.68	0.833	0.66
SMOTETomekCombinedSampled Dataset + K-Fold	0.631	0.641	0.675	0.619
PCA of SMOTETomek CombinedSampled Dataset	0.731	0.976	0.735	0.12
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.644	0.485	0.788	0.839

Perceptron with One Hot Encoding -

Table 1.40. Performance of Perceptron with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + K-Fold	0.998	0.998	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + K Fold	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + K-Fold	0.997	0.995	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + K-Fold	0.998	0.998	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

TomekLinked Dataset	0.994	0.993	1.0	1.0
TomekLinked Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	0.994	1.0	0.993	0.974
PCA of TomekLInked Dataset + K-Fold	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	1.0	1.0	1.0	1.0
Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset + K-Fold	0.998	0.998	1.0	1.0
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	0.994	1.0	0.993	0.974
SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0

SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + K-Fold	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + K-Fold	0.997	0.995	1.0	1.0

Simple Neural Network With Label Encoding -

Table 1.41. Performance of Artificial Neural Network with Label Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	0.714	1.0	0.714	0.0
Original Dataset + Dropout(0.2)	0.714	1.0	0.714	0.0
PCA of Original Dataset	0.714	1.0	0.714	0.0
PCA of Original Dataset + Dropout(0.2)	0.714	1.0	0.714	0.0
UnderSampled Dataset	0.686	0.824	0.757	0.34
UnderSampled Dataset + Dropout(0.2)	0.651	0.808	0.732	0.26
PCA of UnderSampled Dataset	0.646	0.688	0.789	0.54
PCA of UnderSampled Dataset + Dropout(0.2)	0.646	0.8	0.73	0.26
OverSampled Dataset	0.64	0.576	0.878	0.8
OverSampled Dataset + Dropout(0.2)	0.68	0.624	0.897	0.82
PCA of OverSampled Dataset	0.703	0.68	0.876	0.76
PCA of OverSampled Dataset + Dropout(0.2)	0.651	0.592	0.881	0.8
RandomUnderSampled Dataset	0.491	0.304	0.95	0.96
RandomUnderSampled Dataset + Dropout(0.2)	0.651	0.592	0.881	0.8
PCA of RandomUnderSampled Dataset	0.606	0.56	0.833	0.72
PCA of RandomUnderSampled Dataset + Dropout(0.2)	0.543	0.384	0.941	0.94
RandomOverSampled Dataset	0.68	0.704	0.822	0.62

RandomOverSampled Dataset + Dropout(0.2)	0.646	0.648	0.818	0.64
PCA of RandomOverSampled Dataset	0.617	0.536	0.882	0.82
PCA of RandomOverSampled Dataset + Dropout(0.2)	0.634	0.568	0.877	0.8
TomekLinked Dataset	0.714	1.0	0.714	0.0
TomekLinked Dataset + Dropout(0.2)	0.714	1.0	0.714	0.0
PCA of TomekLinked Dataset	0.714	1.0	0.714	0.0
PCA of TomekLInked Dataset + Dropout(0.2)	0.714	1.0	0.714	0.0
Cluster Centroid Dataset	0.577	0.432	0.947	0.94
Cluster Centroid Dataset + Dropout(0.2)	0.64	0.624	0.83	0.68
PCA of Cluster Centroid Dataset	0.669	0.84	0.734	0.24
PCA of Cluster Centroid Dataset + Dropout(0.2)	0.669	0.672	0.832	0.66
SMOTE OverSampled Dataset	0.663	0.6	0.893	0.82
SMOTE OverSampled Dataset + Dropout(0.2)	0.663	0.624	0.867	0.76
PCA of SMOTE OverSampled Dataset	0.703	0.688	0.869	0.74
PCA of SMOTE OverSampled Dataset + Dropout(0.2)	0.651	0.6	0.872	0.78
ENN UnderSampled Dataset	0.686	0.64	0.889	0.8
ENN UnderSampled Dataset + Dropout(0.2)	0.429	0.216	0.931	0.96
PCA of ENN UnderSampled	0.526	0.344	0.977	0.98

Dataset				
PCA of ENN UnderSampled Dataset + Dropout(0.2)	0.44	0.224	0.966	0.98
SMOTEENN Combined Sampled Dataset	0.566	0.424	0.93	0.92
SMOTEENN CombinedSampled Dataset + Dropout(0.2)	0.451	0.248	0.939	0.96
PCA of SMOTEENN CombinedSampled Dataset	0.6	0.472	0.937	0.92
PCA of SMOTEENN CombinedSampled Dataset + Dropout(0.2)	0.286	0.0	nan	1.0
SMOTETomek Combined Sampled Dataset	0.72	0.736	0.852	0.68
SMOTETomekCombinedSampled Dataset + Dropout(0.2)	0.286	0.0	nan	1.0
PCA of SMOTETomek CombinedSampled Dataset	0.646	0.592	0.871	0.78
PCA of SMOTETomekCombinedSampled Dataset + Dropout(0.2)	0.629	0.56	0.875	0.8

Simple Neural Network With One Hot Encoding -

Table 1.42. Performance of Artificial Neural Network with One Hot Encoding on different Evaluation Metrics

	Accuracy	Sensitivity	Precision	Specificity
Original Dataset	1.0	1.0	1.0	1.0
Original Dataset + Dropout(0.2)	0.994	1.0	0.993	0.974
PCA of Original Dataset	1.0	1.0	1.0	1.0
PCA of Original Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
UnderSampled Dataset	1.0	1.0	1.0	1.0
UnderSampled Dataset + Dropout(0.2)	0.956	0.956	1.0	1.0
PCA of UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of UnderSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
OverSampled Dataset	1.0	1.0	1.0	1.0
OverSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of OverSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
RandomUnderSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of RandomUnderSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
RandomOverSampled Dataset	1.0	1.0	1.0	1.0
RandomOverSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
PCA of RandomOverSampled Dataset	1.0	1.0	1.0	1.0

PCA of RandomOverSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
TomekLinked Dataset	1.0	1.0	1.0	1.0
TomekLinked Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
PCA of TomekLinked Dataset	0.994	1.0	0.993	0.974
PCA of TomekLInked Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
Cluster Centroid Dataset	0.994	1.0	0.993	0.974
Cluster Centroid Dataset + Dropout(0.2)	0.994	1.0	0.993	0.974
PCA of Cluster Centroid Dataset	1.0	1.0	1.0	1.0
PCA of Cluster Centroid Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
SMOTE OverSampled Dataset	0.994	0.993	1.0	1.0
SMOTE OverSampled Dataset + Dropout(0.2)	0.994	1.0	0.993	0.974
PCA of SMOTE OverSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTE OverSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
ENN UnderSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset	1.0	1.0	1.0	1.0
PCA of ENN UnderSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
SMOTEENN Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTEENN CombinedSampled	1.0	1.0	1.0	1.0

Dataset + Dropout(0.2)				
PCA of SMOTEENN CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTEENN CombinedSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
SMOTETomek Combined Sampled Dataset	1.0	1.0	1.0	1.0
SMOTETomekCombinedSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0
PCA of SMOTETomek CombinedSampled Dataset	1.0	1.0	1.0	1.0
PCA of SMOTETomekCombinedSampled Dataset + Dropout(0.2)	1.0	1.0	1.0	1.0

- **Hyperparameter Tuning -**

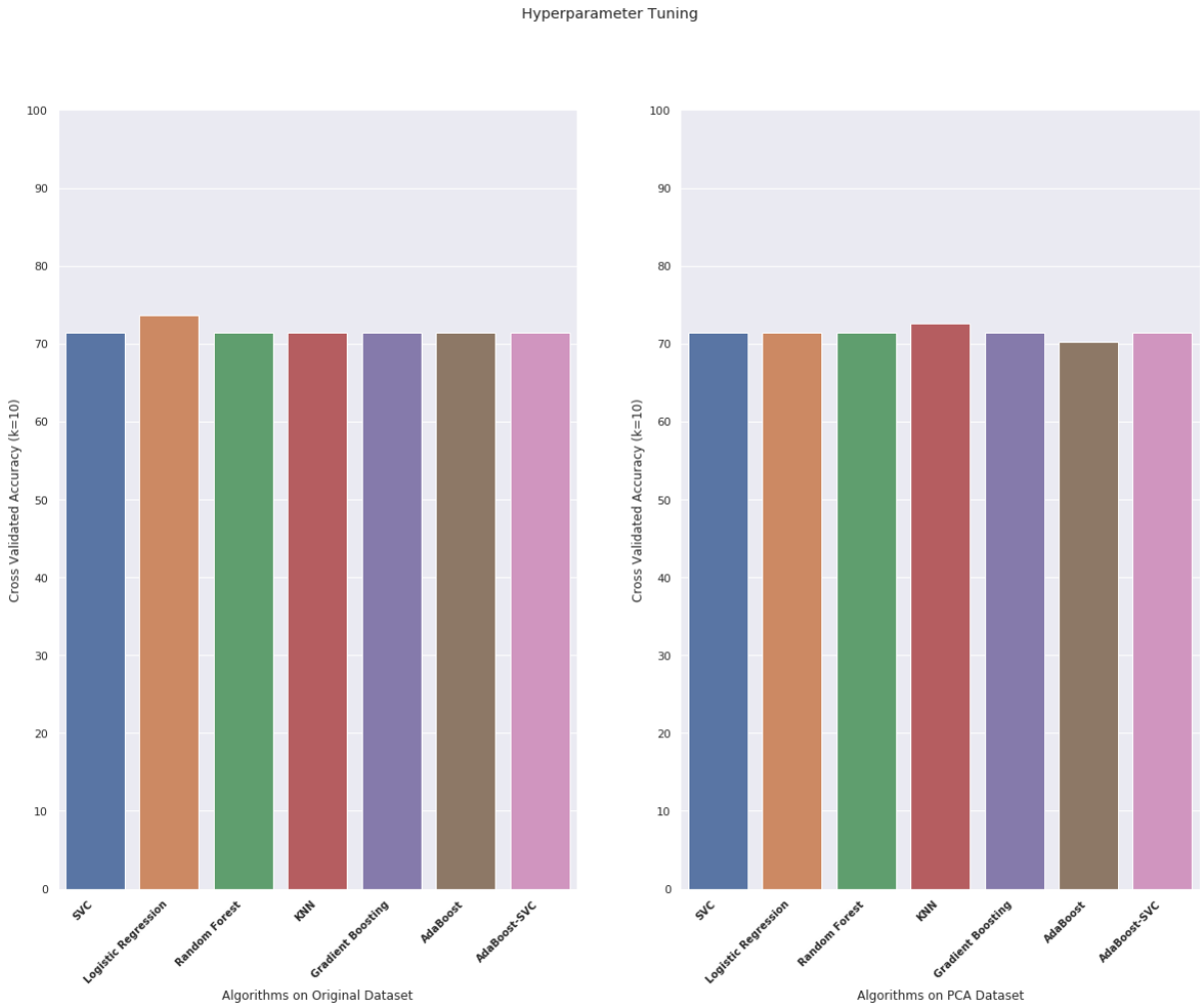


Figure 1.21. Hyperparameter Tuning on the original Dataset with Label Encoding.

Similarly, there are results of other resampling techniques.

8. SUMMARY

In this project, two different methods are used on a set of algorithms along with resampling techniques. It can be seen that there is a wide difference in the evaluation metrics of algorithms with Label Encoding and with One Hot Encoding. So, it can be concluded that usage of One Hot Encoding is better than Label Encoding. With one hot encoding, evaluation metrics performance is high even with simpler individual classifiers. The dimension of this dataset is very small as compared to present day Big Data. Future work can be to get more data on liver functional tests. More datasets with liver functional tests data can further provide enhancement and insights to classify the liver patients.

9. REFERENCES

[1] A. Gulia, P. Rani and Dr. R. Vohra, “Liver Patient Classification Using Intelligent Techniques”, International Journal of Computer Science and Information Technologies, Vol. 5 (4) , (2014), 5110-5115

[2] M. Aiswarya, Swathi Srinivas and A.G. H. Narayanan, “Illustration of Random Forest and Naïve Bayes Algorithms on Indian Liver Patient Data Set”, International Journal of Pure and Applied Mathematics, Vol. 119 No. 10 (2018), 585-595

[3] S. Kefelegn and P. Kamat, “Prediction and Analysis of Liver Disorder Diseases by using Data Mining Technique: Survey”, International Journal of Pure and Applied Mathematics, Vol. 118 No. 9 (2018), 765-770

[4] K. Nagaraj and A. Sridhar, “NeuroSVM: A Graphical User Interface for Identification of Liver Patients”, International Journal of Computer Science and Information Technologies, Vol. 5(6), (2014), 8280-8284

[5] K.Swapna and Prof. M.S. P. Babu, “Critical Analysis of Indian Liver Patients Dataset using ANOVA Method”, International Journal of Engineering & Technology, Vol.17 (3), (2017), 19-33

[6] J. Pahareeya, R. Vohra, J. Makhijani and S. Patsariya“Liver Patient Classification using Intelligence Techniques”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4 (2), (2014), 295-299

[7] P. Kumar and R. S. Thakur, “Early Detection of the Liver Disorder from Imbalance Liver Function Test Datasets”, International Journal of Innovative Technology and Exploring Engineering, Vol. 8 (4), (2019), 179-186

[8] A. Pathan, D. Mhaske, S. Jadhav, R. Bhondave and Dr.K.Rajeswari, “Comparative Study of Different Classification Algorithms on ILPD Dataset to Predict Liver Disorder”, International Journal for Research in Applied Science & Engineering Technology, Vol. 6 (II),

(2018), 388-394

[9] M. Banu H., “Liver Disease Prediction using Machine-Learning Algorithms”, International Journal of Engineering and Advanced Technology, Vol. 8 (6), (2019), 2532-2534

[10] K. Idris and S. Bhoite, “Applications of Machine Learning for Prediction of Liver Disease”, International Journal of Computer Applications Technology and Research, Vol. 8 (9), (2019), 394-396

[11] M. B. Priya, P. L. Juliet and P.R. Tamilselvi, “Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms”, International Research Journal of Engineering and Technology, Vol. 5 (1), (2018), 206-211

[12] M. Wadhwa and S. Juneja, “Comparing Classification Models For Predicting Liver Diseases”, International Journal of Computer Science and Mobile Computing, Vol. 7 (4), (2018), 135 – 140