PROJECT DEVELOPMENT PHASE

TEAM MEMBERS

JESWIN W - 917719IT040
JEYA GANESH A V - 917719IT041
RAHUL HARIESH B - 917719IT074
SANJAYKUMAR S - 917719IT082

SPRINT -2

PROJECT ID	PNT2022TMID21545
PROJECT NAME	ANALYTICS FOR HOSPITAL HEALTH CARE DATA

EPIC:

Prediction of LOS

TASKS:

- As a patient, they are able to predict the length of stay in different hospitals.
- As a hospital manager, they are able to predict the length of stay in different hospitals so they plan accordingly.
- As a user, they need an easily understandable UI to get their predictions.

ALGORITHM USED:

Random Forest:

- Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique.
- It can be used for both Classification and Regression problems in ML.
- It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.
- It is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.
- The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Decision Tree Classifier:

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems.
- It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
- The decisions or the test are performed on the basis of features of the given dataset.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

MLP Classifier:

- MLPClassifier stands for Multi-layer Perceptron classifier which in the name itself connects to a Neural Network.
- Unlike other classification algorithms such as Support Vectors or Naive Bayes Classifier, MLPClassifier relies on an underlying Neural Network to perform the task of classification.
- MLP classifier is a very powerful neural network model that enables the learning of non-linear functions for complex data.
- The method uses forward propagation to build the weights and then it computes the loss. Next, back propagation is used to update the weights so that the loss is reduced.

Gaussian NB:

- Naïve Bayes is a probabilistic machine learning algorithm used for many classification functions and is based on the Bayes theorem.
- Gaussian Naïve Bayes is the extension of naïve Bayes.
- While other functions are used to estimate data distribution, Gaussian or normal distribution is the simplest to implement as you will need to calculate the mean and standard deviation for the training data.
- As the name suggests, Gaussian Naïve Bayes classifier assumes that the data from each label is drawn from a simple Gaussian distribution.
- The Scikit-learn provides sklearn.naive_bayes.GaussianNB to implement the Gaussian Naïve Bayes algorithm for classification.

PREDICTION WITH ACCURACY:

RANDOM FOREST:

```
from sklearn. ensemble import RandomForestClassifier
from sklearn .metrics import accuracy_score
rfc = RandomForestClassifier(n_estimators=150)
rfc. fit(x, y)
y_test_preds = rfc. predict(test_data[features])
accuracy = accuracy_score(y_test_preds, test_data["Stay"])
accuracy
```
`[47]`

```
0.8512443727792086
```

The accuracy for prediction of length of stay using the algorithm Random forest is 85.12%

DECISION TREE CLASSIFIER:

```
from sklearn.tree import DecisionTreeClassifier
dtc = DecisionTreeClassifier()
dtc.fit(x, y)
y_pred = dtc.predict(test_data[features])
accuracy=accuracy_score(y_pred,test_data["Stay"])
accuracy
```
`[48]`

```
0.8187615371706662
```

The accuracy for prediction of length of stay using the algorithm Decision tree classifier is 81.87%

MLP CLASSIFIER:

```
from sklearn.neural_network import MLPClassifier
clf = MLPClassifier(random_state=1, max_iter=1000).fit(x, y)
y_pred = clf.predict(test_data[features])
accuracy=accuracy_score(y_pred,test_data["Stay"])
accuracy
```

```
/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:709: U
  warnings.warn("Training interrupted by user.")
```

```
0.5744106466652561
```

The accuracy for prediction of length of stay using the algorithm MLP classifier is 57.44%

GAUSSIAN NB:

```
from sklearn.naive_bayes import GaussianNB
gnb=GaussianNB()
gnbmodel=gnb.fit(x,y)
gnbpred=gnb.predict(test_data[features])
accuracy=accuracy_score(gnbpred,test_data["Stay"])
accuracy
```

```
0.48643265210824693
```

The accuracy for prediction of length of stay using the algorithm Gaussian NB is 48.64%

INFERENCE:

From the above prediction models, Random Forest Algorithm has highest accuracy, so we chose this for our prediction of length of stay.