Model Development Phase Template

Date	22 June 2024
Team ID	740770
Project Title	Disease Prediction Using Machine Learning
Maximum Marks	4 Marks

Initial Model Training Code, Model Validation and Evaluation Report

The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include classification reports, accuracy, and confusion matrices for multiple models, presented through respective screenshots.

Initial Model Training Code:

Paste the screenshot of the model training code

```
[71] # Split the new data into training, validation, and testing sets
     X1_train, X1_val, y1_train, y1_val = train_test_split(X_new, y, test_size=0.3, random_state=42)
     x1_test=x_test.drop(to_drop,axis=1)
```

```
[89] from sklearn.svm import SVC
     svm1=SVC(C=1)
     svm1.fit(X1_train,y1_train)
     y_pred_svc = svm1.predict(X1_val)
     y_pred = svm1.predict(X1_val)
     yt_pred = svm1.predict(X1_train)
     y_pred1 = svm1.predict(x1_test)
     print('the Training Accuracy of the algorithm is',accuracy_score(y1_train,yt_pred))
     print('the Validation Accuracy of the algorithm is',accuracy_score(y1_val,y_pred))
     print('the Testing Accuracy of the algorithm is',accuracy_score(y_test,y_pred1))
```

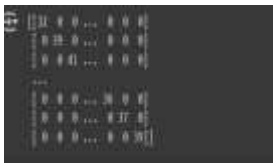
```
[73] # Train a Random Forest Classifier and calculate accuracy
     rfc = RandomForestClassifier(random_state=42)
     rfc.fit(X1_train, y1_train)
     y_pred_rfc = rfc.predict(X1_val)
```

```
[74] y_pred = rfc.predict(X1_val)
     yt_pred = rfc.predict(X1_train)
     y_pred1 = rfc.predict(x1_test)
```

```
[75] knn=KNeighborsClassifier()
     knn.fit(X1_train, y1_train)
     y_pred_knn = knn.predict(X1_val)
```

```
     y_pred = rfc.predict(X1_val)
     yt_pred = rfc.predict(X1_train)
     y_pred1 = rfc.predict(x1_test)
```

Model Validation and Evaluation Report:

Model	Classification Report	Accuracy	
SVM	the Training Accuracy of the algorithm is 0.9930313588850174 the Validation Accuracy of the algorithm is 0.9959349593495935 the Testing Accuracy of the algorithm is 1.0	100%	**Confusion Matrix** 

Random Forest	the Training Accuracy of the algorithm is 0.9930313588850174 the Validation Accuracy of the algorithm is 0.9959340959405935 the Testing Accuracy of the algorithm is 1.0	100%	[[32 0 0 ... 0 0 0] [0 39 0 ... 0 0 0] [0 0 41 ... 0 0 0] ... [0 0 0 ... 36 0 0] [0 0 0 ... 0 37 0] [0 0 0 ... 0 0 39]]

knn	the Training Accuracy of the algorithm is 0.9930313588850174 the Validation Accuracy of the algorithm is 0.9959340959405935 the Testing Accuracy of the algorithm is 1.0	100%	[[32 0 0 ... 0 0 0] [0 39 0 ... 0 0 0] [0 0 41 ... 0 0 0] ... [0 0 0 ... 36 0 0] [0 0 0 ... 0 37 0] [0 0 0 ... 0 0 39]]