

Project Development Phase

SPRINT 2

Date	03.11.2022
Team ID	PNT2022TMID17773
Project Name	Project - Statistical Machine Learning Approaches To Liver Disease Prediction

Executable Program

Model Building:

```
y = df.Dataset
x = df.drop('Dataset', axis=1)
x_train, x_test, y_train, y_test = train_test_split(x, y,
                                                    test_size=0.2,
                                                    random_state=42,
                                                    )
```

```
!pip install imblearn
from imblearn.over_sampling import SMOTE
smote=SMOTE()
x_train,y_train=smote.fit_resample(x_train,y_train)
```

K-NEIGHBORS:

```
m1=KNeighborsClassifier(n_neighbors=21)
m1.fit(x_train,y_train)
ans=m1.predict(x_test)
print(accuracy_score(y_test,ans)*float(100))
69.02654867256636
a1=confusion_matrix(y_test,ans)
a1
array([[24,  8],
       [27, 54]], dtype=int64)
sns.heatmap(a1,annot=True)
```

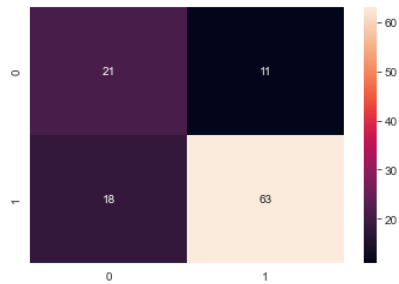


RANDOM FOREST:

```
m3=RandomForestClassifier(n_estimators=20)
m3.fit(x_train,y_train)
res1=m3.predict(x_test)
print(accuracy_score(y_test,res1)*float(100))
```

```
0.7433628318584071
```

```
a3=(confusion_matrix(y_test,res1))  
sns.heatmap(a3,annot=True)
```



SVM:

```
m4=SVC(C=1, gamma=1)  
m4.fit(x_train, y_train)  
res2=m4.predict(x_test)  
print(accuracy_score(y_test,res2)*float(100))  
71.68141592920354  
print(confusion_matrix(y_test,res2))  
[[ 0 32]  
 [ 0 81]]
```

LOGISTIC REGRESSION:

```
m5=LogisticRegression()  
m5.fit(x_train,y_train)  
res4=m5.predict(x_test)  
a5=accuracy_score(y_test,res4)  
a5  
0.6460176991150443  
print(confusion_matrix(y_test,res4))  
[[26  6]  
 [34 47]]
```

DECISION TREE:

```
m6= DecisionTreeClassifier()  
m6.fit(x_train, y_train)  
res5=m6.predict(x_test)  
a6=accuracy_score(y_test,res5)*float(100)  
a6  
71.68141592920354  
  
print(confusion_matrix(y_test,res5))  
[[18 14]  
 [18 63]]
```

SAVING MODEL

```
import joblib  
joblib.dump(m3,'disease.pkl')  
['disease.pkl']
```