**Project Development Phase - Sprint Delivery Plan**

**Sprint 3 – Model Building**

| | |
|---|---|
| Date | 18 November 2022 |
| Team ID | PNT2022TMID52974 |
| Project Name | Statistical Machine Learning Approaches to Liver Disease Prediction |

## Machine learning Algorithms for Model Building:

Train and Test the model using classification algorithm:

- Import the classification algorithm:

```
In [17]: from sklearn.svm import SVC
         from sklearn.ensemble import RandomForestClassifier
         from sklearn.neighbors import KNeighborsClassifier
         from sklearn.preprocessing import LabelEncoder
         le = LabelEncoder()
         le.fit(xtrain['Gender'].astype(str))
         xtrain['Gender'] = le.transform(xtrain['Gender'].astype(str))
         xtest['Gender'] = le.transform(xtest['Gender'].astype(str))
```

- Initialize the model

```
In [18]: svm=SVC()
         RFmodel=RandomForestClassifier()
         KNNmodel=KNeighborsClassifier()
```

- Training model with our data
  - SVC Model
  - Random Forest Model
  - K-Nearest Neighbors

```
In [19]: from sklearn.svm import SVC
         svm=SVC(gamma='auto')
         svm.fit(xtrain,ytrain)

Out[19]: SVC(gamma='auto')

In [20]: from sklearn.ensemble import RandomForestClassifier
         RFmodel=RandomForestClassifier()
         RFmodel.fit(xtrain,ytrain)

Out[20]: RandomForestClassifier()

In [21]: from sklearn.neighbors import KNeighborsClassifier
         KNN=KNeighborsClassifier()
         KNN.fit(xtrain,ytrain)

Out[21]: KNeighborsClassifier()
```

- Evaluation Metrics:

In [22]:
```
import warnings
warnings.filterwarnings("ignore")
#accuracies
from sklearn.metrics import accuracy_score
SVMpred=svm.predict(xtest)
SVMaccuracy=accuracy_score(SVMpred,ytest)
print('SVM accuracy:',SVMaccuracy*100)

RFpred=RFmodel.predict(xtest)
RFaccuracy=accuracy_score(ytest,RFpred)
print('RFC accuracy:',RFaccuracy*100)

KNNpred=KNN.predict(xtest)
KNNaccuracy=accuracy_score(KNNpred,ytest)
print('KNN accuracy:',KNNaccuracy*100)
```

```
SVM accuracy: 73.50427350427351
RFC accuracy: 75.21367521367522
KNN accuracy: 75.21367521367522
```

In [23]:
```
#confusion matrix
from sklearn.metrics import confusion_matrix
cm1=confusion_matrix(ytest,SVMpred)
cm2=confusion_matrix(ytest,RFpred)
cm3=confusion_matrix(ytest,KNNpred)
print('SVM Confusion matrix',cm1)
print('RFC Confusion matrix',cm2)
print('KNN Confusion matrix',cm3)
```

```
SVM Confusion matrix [[86  0]
 [31  0]]
RFC Confusion matrix [[73 13]
 [16 15]]
KNN Confusion matrix [[69 17]
 [12 19]]
```

- Save the model:

In [24]:
```
import pickle
pickle.dump(svm,open('liver_analysis.pkl','wb'))
```