In [18]:
```
from pyspark.sql.types import BooleanType
from pyspark.ml.feature import StringIndexer, VectorAssembler
from pyspark.ml.classification import LinearSVC
from pyspark.sql.session import SparkSession
from pyspark.sql.functions import expr
from pyspark.ml.evaluation import BinaryClassificationEvaluator
from helpers.helper_functions import translate_to_file_string
from pyspark.sql import DataFrameReader
from pyspark.sql import SparkSession
from pyspark.ml.feature import IndexToString, Normalizer, StringInd
from pyspark.ml.evaluation import BinaryClassificationEvaluator
from pyspark.ml.classification import DecisionTreeClassifier
from pyspark.ml.classification import RandomForestClassifier
from pyspark.ml.evaluation import MulticlassClassificationEvaluator
from pyspark.ml.classification import MultilayerPerceptronClassific
from pyspark.ml.evaluation import RegressionEvaluator
from pyspark.ml.tuning import CrossValidator, ParamGridBuilder
from pyspark.ml import Pipeline
from helpers.helper_functions import translate_to_file_string
from pyspark.ml.classification import MultilayerPerceptronClassific
from sklearn.metrics import roc_curve, auc
import seaborn as sns
import pandas as pd
import os
import warnings
import matplotlib.pyplot as plt
warnings.filterwarnings('ignore')
```

In [19]:
```
inputFile = translate_to_file_string("../data/heart_val.csv")
```

In [20]:
```
spark = (SparkSession
       .builder
       .appName("HeartDiseaseAnalNeuralNw")
       .getOrCreate())
```

In [21]:
```
# load data file.
# create a DataFrame using an ifered Schema
df = spark.read.option("header", "true") \
       .option("inferSchema", "true") \
       .option("delimiter", ";") \
       .csv(inputFile)
```

In [22]:
```
#transform labels
labelIndexer = StringIndexer().setInputCol("target").setOutputCol("
sexIndexer = StringIndexer().setInputCol("sex").setOutputCol("sex_n
```

In [23]:
```
#feature columns for evaluation
featureCols = df.columns.copy()
featureCols.remove("target")
featureCols.remove("sex")
featureCols = featureCols + ["sex_num"]
```

In [24]:
```
#create vector assembler of feature columns
assembler =  VectorAssembler(outputCol="features", inputCols=featur
```

In [25]:
```
#Build feauture Indexer
featureIndexer = VectorIndexer(inputCol="features",outputCol="index
```

In [26]:
```
#Create scaler for scaled output
scaler = StandardScaler(inputCol="features", outputCol="scaledFeatu
                        withStd=True, withMean=False)
```

In [27]:
```
#Convert Indexed labels back to original labels
predConverter = IndexToString(inputCol="prediction",outputCol="pred
```

In [28]:
```
#create Classifier
nn = MultilayerPerceptronClassifier(seed=1234, featuresCol="scaledF

# build network parameters grid


paramGrid =  ParamGridBuilder().addGrid(nn.layers, [[ len(featureCo
                .addGrid(nn.blockSize,  [128 ]) \
                .addGrid(nn.maxIter,[ 100, 1000, 5000 ] )\
                .addGrid(nn.stepSize, [0.003, 0.03, 0.3 ])\
                .addGrid(nn.tol, [ 0.05, 0.1, 0.2 ]) \
                .build()
```

In [29]:
```
#split data for etsting

splits = df.randomSplit([0.7, 0.3 ], 5433)
train = splits[0]
test = splits[1]
```

In [30]:
```
#Pipelining of all steps
pipeline = Pipeline(stages= [labelIndexer,sexIndexer,  assembler, f
```

In [32]:
```
#build evaluator
evaluator =  BinaryClassificationEvaluator(labelCol="label",rawPred
```

In [33]:
```
#Cross validator
cv = CrossValidator(estimator=pipeline, evaluator=evaluator,estimato
```

In [34]:
```
#train model
nwModel = cv.fit(train)
```

In [35]:
```
#Find out the best model
bestModel = nwModel.bestModel.stages[5]
print("Layers: " , bestModel.layers)
print(bestModel.explainParams())
```

```
Layers:  MultilayerPerceptronClassifier_b844c5104df1__layers
blockSize: block size for stacking input data in matrices. Data is
stacked within partitions. If block size is more than remaining da
ta in a partition then it is adjusted to the size of this data. (d
efault: 128, current: 128)
featuresCol: features column name. (default: features, current: sc
aledFeatures)
initialWeights: The initial weights of the model. (undefined)
labelCol: label column name. (default: label)
layers: Sizes of layers from input layer to output layer E.g., Arr
ay(780, 100, 10) means 780 inputs, one hidden layer with 100 neuro
ns and output layer of 10 neurons. (current: [13, 10, 5, 2])
maxIter: max number of iterations (>= 0). (default: 100, current:
100)
predictionCol: prediction column name. (default: prediction)
probabilityCol: Column name for predicted class conditional probab
ilities. Note: Not all models output well-calibrated probability e
stimates! These probabilities should be treated as confidences, no
t precise probabilities. (default: probability)
rawPredictionCol: raw prediction (a.k.a. confidence) column name.
(default: rawPrediction)
seed: random seed. (default: -3151049751922601847, current: 1234)
solver: The solver algorithm for optimization. Supported options:
l-bfgs, gd. (default: l-bfgs)
stepSize: Step size to be used for each iteration of optimization
(>= 0). (default: 0.03, current: 0.003)
thresholds: Thresholds in multi-class classification to adjust the
probability of predicting each class. Array must have length equal
to the number of classes, with values > 0, excepting that at most
one value may be 0. The class with largest value p/t is predicted,
where p is the original probability of that class and t is the cla
ss's threshold. (undefined)
tol: the convergence tolerance for iterative algorithms (>= 0). (d
efault: 1e-06, current: 0.05)
+---+---+---+--------+----+---+-------+-------+-----+-------+-----
+---+----+------+-----+--------+--------------------+-------------
------+--------------------+--------------------+-----------------
---+----------+--------------+
|age|sex| cp|trestbps|chol|fbs|restecg|thalach|exang|oldpeak|slope
| ca|thal|target|label|sex_num|            features|     indexedFe
atures|      scaledFeatures|       rawPrediction|         probabil
ity|prediction|predictedLabel|
+---+---+---+--------+----+---+-------+-------+-----+-------+-----
+---+----+------+-----+--------+--------------------+-------------
```

```
------+--------------------+--------------------+-----------------
---+----------+--------------+
| 35|  f|  0|     138| 183|  0|      1|    182|    0|    1.4|    2
|  0|   2|     y|  0.0|    1.0|[35.0,0.0,138.0,1...|[35.0,0.0,138.
0,1...|[3.90265392446968...|[0.75600387078402...|[0.56268825971894
...|       0.0|             y|
| 35|  m|  0|     120| 198|  0|      1|    130|    1|    1.6|    1
|  0|   3|     n|  1.0|    0.0|[35.0,0.0,120.0,1...|[35.0,0.0,120.
0,1...|[3.90265392446968...|[0.80397153514321...|[0.56223118838589
...|       0.0|             y|
| 35|  m|  1|     122| 192|  0|      1|    174|    0|    0.0|    2
|  0|   2|     y|  0.0|    0.0|[35.0,1.0,122.0,1...|[35.0,1.0,122.
0,1...|[3.90265392446968...|[0.77422147463516...|[0.56210469355355
...|       0.0|             y|
| 40|  m|  0|     152| 223|  0|      1|    181|    0|    0.0|    2
|  0|   3|     n|  1.0|    0.0|(13,[0,2,3,5,6,9,...|(13,[0,2,3,5,6
,9,...|(13,[0,2,3,5,6,9,...|[0.76989833218852...|[0.56344962222457
...|       0.0|             y|
| 41|  f|  2|     112| 268|  0|      0|    172|    1|    0.0|    2
|  0|   2|     y|  0.0|    1.0|[41.0,2.0,112.0,2...|[41.0,2.0,112.
0,2...|[4.57168031152162...|[0.83698058803820...|[0.56443458710346
...|       0.0|             y|
| 42|  m|  0|     136| 315|  0|      1|    125|    1|    1.8|    1
|  0|   1|     n|  1.0|    0.0|[42.0,0.0,136.0,3...|[42.0,0.0,136.
0,3...|[4.68318470936361...|[0.84205018979597...|[0.56503943757065
...|       0.0|             y|
| 44|  m|  0|     110| 197|  0|      0|    177|    0|    0.0|    2
|  1|   2|     n|  1.0|    0.0|(13,[0,2,3,6,9,10...|(13,[0,2,3,6,9
,10...|(13,[0,2,3,6,9,10...|[0.80015998461802...|[0.56499295072504
...|       0.0|             y|
| 44|  m|  0|     112| 290|  0|      0|    153|    0|    0.0|    2
|  1|   2|     n|  1.0|    0.0|(13,[0,2,3,6,9,10...|(13,[0,2,3,6,9
,10...|(13,[0,2,3,6,9,10...|[0.81689851934968...|[0.56619207987703
...|       0.0|             y|
| 44|  m|  1|     130| 219|  0|      0|    188|    0|    0.0|    2
|  0|   2|     y|  0.0|    0.0|(13,[0,1,2,3,6,9,...|(13,[0,1,2,3,6
,9,...|(13,[0,1,2,3,6,9,...|[0.80024768624755...|[0.56575376098334
...|       0.0|             y|
| 44|  m|  2|     140| 235|  0|      0|    180|    0|    0.0|    2
|  0|   2|     y|  0.0|    0.0|(13,[0,1,2,3,6,9,...|(13,[0,1,2,3,6
,9,...|(13,[0,1,2,3,6,9,...|[0.79893682105947...|[0.56589432516363
...|       0.0|             y|
| 45|  m|  0|     142| 309|  0|      0|    147|    1|    0.0|    1
|  3|   3|     n|  1.0|    0.0|[45.0,0.0,142.0,3...|[45.0,0.0,142.
0,3...|[5.01769790288959...|[0.81345684063538...|[0.56450352631033
...|       0.0|             y|
| 45|  m|  1|     128| 308|  0|      0|    170|    0|    0.0|    2
|  0|   2|     y|  0.0|    0.0|(13,[0,1,2,3,6,9,...|(13,[0,1,2,3,6
,9,...|(13,[0,1,2,3,6,9,...|[0.81499673822548...|[0.56651269547490
...|       0.0|             y|
| 46|  f|  2|     142| 177|  0|      0|    160|    1|    1.4|    0
|  0|   2|     y|  0.0|    1.0|[46.0,2.0,142.0,1...|[46.0,2.0,142.
0,1...|[5.12920230073158...|[0.73399915218625...|[0.55942737551260
```

```
...|       0.0|             y|
| 46|  m|  0|     120| 249|  0|      0|    144|    0|    0.8|    2
|  0|   3|     n|  1.0|    0.0|(13,[0,2,3,6,8,9,...|(13,[0,2,3,6,8
,9,...|(13,[0,2,3,6,8,9,...|[0.81677825679023...|[0.56632178808091
...|       0.0|             y|
| 46|  m|  1|     101| 197|  1|      1|    156|    0|    0.0|    2
|  0|   3|     y|  0.0|    0.0|[46.0,1.0,101.0,1...|[46.0,1.0,101.
0,1...|[5.12920230073158...|[0.73580666727620...|[0.56117191633421
...|       0.0|             y|
| 46|  m|  2|     150| 231|  0|      1|    147|    0|    3.6|    1
|  0|   2|     n|  1.0|    0.0|[46.0,2.0,150.0,2...|[46.0,2.0,150.
0,2...|[5.12920230073158...|[0.79225281477384...|[0.56592138619937
...|       0.0|             y|
| 47|  m|  2|     130| 253|  0|      1|    179|    0|    0.0|    2
|  0|   2|     y|  0.0|    0.0|[47.0,2.0,130.0,2...|[47.0,2.0,130.
0,2...|[5.24070669857357...|[0.79279580045762...|[0.56421452354560
...|       0.0|             y|
| 48|  f|  2|     130| 275|  0|      1|    139|    0|    0.2|    2
|  0|   2|     y|  0.0|    1.0|[48.0,2.0,130.0,2...|[48.0,2.0,130.
0,2...|[5.35221109641556...|[0.78935461046518...|[0.56323778540989
...|       0.0|             y|
| 49|  f|  1|     134| 271|  0|      1|    162|    0|    0.0|    1
|  0|   2|     y|  0.0|    1.0|[49.0,1.0,134.0,2...|[49.0,1.0,134.
0,2...|[5.46371549425755...|[0.73064126606378...|[0.56003367652533
...|       0.0|             y|
| 49|  m|  2|     120| 188|  0|      1|    139|    0|    2.0|    1
|  3|   3|     n|  1.0|    0.0|[49.0,2.0,120.0,1...|[49.0,2.0,120.
0,1...|[5.46371549425755...|[0.77175989062351...|[0.56062631843779
...|       0.0|             y|
+---+---+---+--------+----+---+-------+-------+-----+-------+-----
+---+----+------+-----+-------+--------------------+--------------
------+--------------------+--------------------+-----------------
---+----------+--------------+
only showing top 20 rows

Test Error =  0.3524970963995354
```

In [36]:

```
#test model
predictions = nwModel.transform(test)
predictions.show()
```

```
+---+---+---+--------+----+---+-------+-------+-----+-------+-----
+---+----+------+-----+-------+--------------------+--------------
------+--------------------+--------------------+-----------------
---+----------+--------------+
|age|sex| cp|trestbps|chol|fbs|restecg|thalach|exang|oldpeak|slope
| ca|thal|target|label|sex_num|            features|     indexedFe
atures|      scaledFeatures|       rawPrediction|            probabil
ity|prediction|predictedLabel|
+---+---+---+--------+----+---+-------+-------+-----+-------+-----
+---+----+------+-----+-------+--------------------+--------------
------+--------------------+--------------------+-----------------
---+----------+--------------+
| 35|  f|  0|     138| 183|  0|      1|    182|    0|    1.4|    2
|  0|   2|     y|  0.0|    1.0|[35.0,0.0,138.0,1...|[35.0,0.0,138.
0,1...|[3.90265392446968...|[0.75600387078402...|[0.56268825971894
...|       0.0|              y|

| 35|  m|  0|     120| 198|  0|       1|    130|    1|   1.6|    1
|  0|   3|     n|  1.0|    0.0|[35.0,0.0,120.0,1...|[35.0,0.0,120.
0,1...|[3.90265392446968...|[0.80397153514321...|[0.56223118838589
...|       0.0|              y|
| 35|  m|  1|     122| 192|  0|       1|    174|    0|   0.0|    2
|  0|   2|     y|  0.0|    0.0|[35.0,1.0,122.0,1...|[35.0,1.0,122.
0,1...|[3.90265392446968...|[0.77422147463516...|[0.56210469355355
...|       0.0|              y|
| 40|  m|  0|     152| 223|  0|       1|    181|    0|   0.0|    2
|  0|   3|     n|  1.0|    0.0|(13,[0,2,3,5,6,9,...|(13,[0,2,3,5,6
,9,...|(13,[0,2,3,5,6,9,...|[0.76989833218852...|[0.56344962222457
...|       0.0|              y|
| 41|  f|  2|     112| 268|  0|       0|    172|    1|   0.0|    2
|  0|   2|     y|  0.0|    1.0|[41.0,2.0,112.0,2...|[41.0,2.0,112.
0,2...|[4.57168031152162...|[0.83698058803820...|[0.56443458710346
...|       0.0|              y|
| 42|  m|  0|     136| 315|  0|       1|    125|    1|   1.8|    1
|  0|   1|     n|  1.0|    0.0|[42.0,0.0,136.0,3...|[42.0,0.0,136.
0,3...|[4.68318470936361...|[0.84205018979597...|[0.56503943757065
...|       0.0|              y|
| 44|  m|  0|     110| 197|  0|       0|    177|    0|   0.0|    2
|  1|   2|     n|  1.0|    0.0|(13,[0,2,3,6,9,10...|(13,[0,2,3,6,9
,10...|(13,[0,2,3,6,9,10...|[0.80015998461802...|[0.56499295072504
...|       0.0|              y|
| 44|  m|  0|     112| 290|  0|       0|    153|    0|   0.0|    2
|  1|   2|     n|  1.0|    0.0|(13,[0,2,3,6,9,10...|(13,[0,2,3,6,9
,10...|(13,[0,2,3,6,9,10...|[0.81689851934968...|[0.56619207987703
...|       0.0|              y|
| 44|  m|  1|     130| 219|  0|       0|    188|    0|   0.0|    2
|  0|   2|     y|  0.0|    0.0|(13,[0,1,2,3,6,9,...|(13,[0,1,2,3,6
,9,...|(13,[0,1,2,3,6,9,...|[0.80024768624755...|[0.56575376098334
...|       0.0|              y|
| 44|  m|  2|     140| 235|  0|       0|    180|    0|   0.0|    2
|  0|   2|     y|  0.0|    0.0|(13,[0,1,2,3,6,9,...|(13,[0,1,2,3,6
,9,...|(13,[0,1,2,3,6,9,...|[0.79893682105947...|[0.56589432516363
...|       0.0|              y|
| 45|  m|  0|     142| 309|  0|       0|    147|    1|   0.0|    1
|  3|   3|     n|  1.0|    0.0|[45.0,0.0,142.0,3...|[45.0,0.0,142.
0,3...|[5.01769790288959...|[0.81345684063538...|[0.56450352631033
...|       0.0|              y|
| 45|  m|  1|     128| 308|  0|       0|    170|    0|   0.0|    2
|  0|   2|     y|  0.0|    0.0|(13,[0,1,2,3,6,9,...|(13,[0,1,2,3,6
,9,...|(13,[0,1,2,3,6,9,...|[0.81499673822548...|[0.56651269547490
...|       0.0|              y|
| 46|  f|  2|     142| 177|  0|       0|    160|    1|   1.4|    0
|  0|   2|     y|  0.0|    1.0|[46.0,2.0,142.0,1...|[46.0,2.0,142.
0,1...|[5.12920230073158...|[0.73399915218625...|[0.55942737551260
...|       0.0|              y|
| 46|  m|  0|     120| 249|  0|       0|    144|    0|   0.8|    2
|  0|   3|     n|  1.0|    0.0|(13,[0,2,3,6,8,9,...|(13,[0,2,3,6,8
,9,...|(13,[0,2,3,6,8,9,...|[0.81677825679023...|[0.56632178808091
...|       0.0|               y|
| 46|  m|  1|     101| 197|  1|      1|    156|    0|    0.0|    2

|  0|   3|     y|  0.0|    0.0|[46.0,1.0,101.0,1...|[46.0,1.0,101.
0,1...|[5.12920230073158...|[0.73580666727620...|[0.56117191633421
...|       0.0|               y|
| 46|  m|  2|     150| 231|  0|      1|    147|    0|    3.6|    1
|  0|   2|     n|  1.0|    0.0|[46.0,2.0,150.0,2...|[46.0,2.0,150.
0,2...|[5.12920230073158...|[0.79225281477384...|[0.56592138619937
...|       0.0|               y|
| 47|  m|  2|     130| 253|  0|      1|    179|    0|    0.0|    2
|  0|   2|     y|  0.0|    0.0|[47.0,2.0,130.0,2...|[47.0,2.0,130.
0,2...|[5.24070669857357...|[0.79279580045762...|[0.56421452354560
...|       0.0|               y|
| 48|  f|  2|     130| 275|  0|      1|    139|    0|    0.2|    2
|  0|   2|     y|  0.0|    1.0|[48.0,2.0,130.0,2...|[48.0,2.0,130.
0,2...|[5.35221109641556...|[0.78935461046518...|[0.56323778540989
...|       0.0|               y|
| 49|  f|  1|     134| 271|  0|      1|    162|    0|    0.0|    1
|  0|   2|     y|  0.0|    1.0|[49.0,1.0,134.0,2...|[49.0,1.0,134.
0,2...|[5.46371549425755...|[0.73064126606378...|[0.56003367652533
...|       0.0|               y|
| 49|  m|  2|     120| 188|  0|      1|    139|    0|    2.0|    1
|  3|   3|     n|  1.0|    0.0|[49.0,2.0,120.0,1...|[49.0,2.0,120.
0,1...|[5.46371549425755...|[0.77175989062351...|[0.56062631843779
...|       0.0|               y|
+---+---+---+--------+----+---+-------+-------+-----+-------+-----
+---+----+------+-----+-------+--------------------+-------------
------+--------------------+--------------------+-----------------
---+----------+--------------+
only showing top 20 rows
```

In [37]:

```
accuracy = evaluator.evaluate(predictions)
print("Test Error = " ,(1.0 - accuracy))
```

```
Test Error =  0.3524970963995354
```

In [38]:

```
spark.stop()
```