Final Term Project

Data Mining

CS- 634-104

FALL-21

Professor: Yasser Abduallah

Department of Computer Science

NAME: Kavya Basavalingappa Umashankar

UCID – kbu2

GITHUB LINK

Contents **Pg no**

Introduction:

Algorithm Overview:

Computers that learn from data and use algorithms to execute tasks without being explicitly programmed are referred to as machine learning. Deep learning is based on a complicated set of algorithms that are modeled after the human brain. This allows unstructured data, such as documents, photos, and text, to be processed.

K Nearest Neighbours:

The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

The K-NN working can be explained on the basis of the below algorithm:

- **Step-1:** Select the number K of the neighbors
- **Step-2:** Calculate the Euclidean distance of **K number of neighbors**
- **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.
- **Step-4:** Among these k neighbors, count the number of the data points in each category.
- **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.

Random Forest Algorithm:

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set.

Steps involved in Random forests :

- **Step-1:**Randomly pick $\sqrt{total\ number\ of\ attributes}$
- **Step-2:**Calculate the gini impurity
- **Step-3:**Split based on gini impurity
- **Step-4:**Repeat the above process for different attributes

- **Step-5:**Consider the highest gini impurity and build 1 random tree
- **Step-6:**Repeat this process to form N number of trees.

Long Short Term Memory:

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems.This is a behavior required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deep learning.

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can process not only single data points (such as images), but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition, speech recognition and anomaly detection in network traffic or IDSs (intrusion detection systems).

A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs, hidden Markov models and other sequence learning methods in numerous applications.

ASSUMPTIONS:

The implementation has some assumptions to make sure the running of the application is flawless and without exception.

- Dataset sit cleaned.
- Dataset is not imbalance.
- All the attributes have correlation with the output variable

REQUIREMENTS:

- Modern Operating System:
- Windows 7, 8 or 10 / Mac OS X 10.11 or higher, 64-bit / Linux: RHEL 6/7, 64-bit (almost all libraries also work in Ubuntu)
- x86 64-bit CPU (Intel / AMD architecture)
- 4 GB RAM
- 5 GB free disk space
- Python IDE: PyCharm/Spyder with Numpy and Pandas - Reading and usage of data,sklearn, keras and tensorflow- for models, matplotlib and Seaborn -for heatmap installed.

DATASET OVERVIEW:

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. I have selected the redwine variants. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available.This data set was prepossessed and downloaded from the UCI Machine Learning Repository. This data set was simple, cleaned, practice data set for classification modelling.

Attribute Information:

Input variables (based on physicochemical tests):

1 - fixed acidity (float values in range of 4.600 to 15.900)

2 - volatile acidity (float values in range of 0.12 to 1.58)

3 - citric acid (Float values in range of 0 to 1)

4 - residual sugar (float values in range of 0.9 to 15.5)

5 - chlorides (float values in range of 0.012 to 0.6110)

6 - free sulfur dioxide (integer values in range of 1 to 72)

7 - total sulfur dioxide (float values in range of 6 to 289)

8 - density (float values in range of 0.990 to 1.000)

9 - pH (float values in range of 2.74 to 4.01)

10 - sulphates (float values in range of 0.33 to 2.000)

11 - alcohol (float values in range of 8.40 to 14.90)

Output variable (based on sensory data):

12 - quality ('good' or 'bad')

COMPILE AND RUN THE SOURCE CODE:

Python automatically compiles the code before running it. The file is saved as “Evaluation.py” and all required files are present in the directory. In the terminal, the command **python Evaluation.py** in pycharm command prompt and **runfile(“address of the file”, wfile=”address of execution file of sypder”)** in spyder will compile and execute the code. All the assumptions are in place as explained in the “Assumptions” Section.

TESTING IMPLEMENTATION:

1. K Nearest Neighbour:

Training the KNN model from sklearn using 10 fold cross validation and calculating the metrics.

```
Fold =  1
Fold Size =  160
The total positive =  47
The total negative =  113
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[106, 33], [7, 14]]
True Positive rate =  0.2978723404255319
True Negative rate =  0.9380530973451328
False Positive rate =  0.061946902654867256
False negative rate =  0.7021276595744681
Accuracy =  0.75
Precision =  0.6666666666666666
Recall =  0.2978723404255319
F1 Measure =  0.4117647058823529
Error rate =  0.25
Balanced Accuracy =  0.6179627188853323
True Skill statistics =  0.23592543777066466
Heidke Skill Score =  0.28138333707612845
----------------------------------------------------------------------------------------------------
Fold =  2
Fold Size =  160
The total positive =  78
The total negative =  82
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[72, 53], [10, 25]]
True Positive rate =  0.32051282051282054
True Negative rate =  0.8780487804878049
False Positive rate =  0.12195121951219512
False negative rate =  0.6794871794871795
Accuracy =  0.60625
Precision =  0.7142857142857143
Recall =  0.32051282051282054
F1 Measure =  0.4424778761061947
Error rate =  0.39375
Balanced Accuracy =  0.5992808005003127
True Skill statistics =  0.19856160100062542
Heidke Skill Score =  0.20126782884310618
----------------------------------------------------------------------------------------------------
```

```
---------------------------------------------------------------------------------------------------
Fold =  3
Fold Size =  160
The total positive =  105
The total negative =  55
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[40, 59], [15, 46]]
True Positive rate =  0.4380952380952381
True Negative rate =  0.7272727272727273
False Positive rate =  0.2727272727272727
False negative rate =  0.5619047619047619
Accuracy =  0.5375
Precision =  0.7540983606557377
Recall =  0.4380952380952381
F1 Measure =  0.5542168674698795
Error rate =  0.4625
Balanced Accuracy =  0.5826839826839827
True Skill statistics =  0.16536796536796539
Heidke Skill Score =  0.1389090909090909
---------------------------------------------------------------------------------------------------
Fold =  4
Fold Size =  160
The total positive =  91
The total negative =  69
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[43, 50], [26, 41]]
True Positive rate =  0.45054945054945056
True Negative rate =  0.6231884057971014
False Positive rate =  0.37681159420289856
False negative rate =  0.5494505494505495
Accuracy =  0.525
Precision =  0.6119402985074627
Recall =  0.45054945054945056
F1 Measure =  0.5189873417721519
Error rate =  0.475
Balanced Accuracy =  0.536868928173276
True Skill statistics =  0.073737856346552
Heidke Skill Score =  0.07076264710377503
---------------------------------------------------------------------------------------------------
Fold =  5
Fold Size =  160
The total positive =  54
The total negative =  106
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[85, 39], [21, 15]]
True Positive rate =  0.2777777777777778
True Negative rate =  0.8018867924528302
False Positive rate =  0.19811320754716982
False negative rate =  0.7222222222222222
Accuracy =  0.625
Precision =  0.4166666666666667
Recall =  0.2777777777777778
F1 Measure =  0.33333333333333337
Error rate =  0.375
Balanced Accuracy =  0.539832285115304
True Skill statistics =  0.07966457023060797
Heidke Skill Score =  0.0867579908675799
---------------------------------------------------------------------------------------------------
Fold =  6
Fold Size =  160
The total positive =  97
The total negative =  63
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[39, 45], [24, 52]]
True Positive rate =  0.5360824742268041
True Negative rate =  0.6190476190476191
False Positive rate =  0.38095238095238093
False negative rate =  0.4639175257731959
Accuracy =  0.56875
Precision =  0.6842105263157895
Recall =  0.5360824742268041
F1 Measure =  0.6011560693641619
Error rate =  0.43125
Balanced Accuracy =  0.5775650466372115
True Skill statistics =  0.15513009327442318
Heidke Skill Score =  0.14656771799628943
---------------------------------------------------------------------------------------------------
```

```
Fold =  7
Fold Size =  160
The total positive =  117
The total negative =  43
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[34, 56], [9, 61]]
True Positive rate =  0.5213675213675214
True Negative rate =  0.7906976744186046
False Positive rate =  0.20930232558139536
False negative rate =  0.47863247863247865
Accuracy =  0.59375
Precision =  0.8714285714285714
Recall =  0.5213675213675214
F1 Measure =  0.6524064171122995
Error rate =  0.40625
Balanced Accuracy =  0.656032597893063
True Skill statistics =  0.312065195786126
Heidke Skill Score =  0.23190546528803546
--------------------------------------------------------------------------------------------------
Fold =  8
Fold Size =  160
The total positive =  106
The total negative =  54
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[36, 52], [18, 54]]
True Positive rate =  0.5094339622641509
True Negative rate =  0.6666666666666666
False Positive rate =  0.3333333333333333
False negative rate =  0.49056603773584906
Accuracy =  0.5625
Precision =  0.75
Recall =  0.5094339622641509
F1 Measure =  0.6067415730337078
Error rate =  0.4375
Balanced Accuracy =  0.5880503144654088
True Skill statistics =  0.17610062893081763
Heidke Skill Score =  0.15254237288135594
--------------------------------------------------------------------------------------------------
Fold =  9
Fold Size =  160
The total positive =  75
The total negative =  85
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[60, 53], [25, 22]]
True Positive rate =  0.29333333333333333
True Negative rate =  0.7058823529411765
False Positive rate =  0.29411764705882354
False negative rate =  0.7066666666666667
Accuracy =  0.5125
Precision =  0.46808510638297873
Recall =  0.29333333333333333
F1 Measure =  0.36065573770491804
Error rate =  0.4875
Balanced Accuracy =  0.4996078431372549
True Skill statistics =  -0.0007843137254902044
Heidke Skill Score =  -0.0008019246190858059
--------------------------------------------------------------------------------------------------
Fold =  10
Fold Size =  159
The total positive =  85
The total negative =  74
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[48, 44], [26, 41]]
True Positive rate =  0.4823529411764706
True Negative rate =  0.6486486486486487
False Positive rate =  0.35135135135135137
False negative rate =  0.5176470588235295
Accuracy =  0.559748427672956
Precision =  0.6119402985074627
Recall =  0.4823529411764706
F1 Measure =  0.5394736842105263
Error rate =  0.44025157232704404
Balanced Accuracy =  0.5655007949125597
True Skill statistics =  0.13100158982511922
Heidke Skill Score =  0.12897167005791205
--------------------------------------------------------------------------------------------------
**************************************************************************************************
```

```
**********************************************************************************************
Average True Positive rate =  0.4127377859729099
Average True Negative rate =  0.7399392765078312
Average False Positive rate =  0.26006072349216874
Average False negative rate =  0.5872622140270901
Average Accuracy =  0.5840998427672955
Average Precision =  0.6549322209417051
Average Recall =  0.4127377859729099
Average F1 Measure =  0.5021213605989526
Average Error rate =  0.41590015723270446
Average Balanced Accuracy =  0.5763385312403705
Average True Skill statistics =  0.15267706248074112
Average Heidke Skill Score =  0.14382661964041873
**********************************************************************************************
```

Figure 1: 10-fold cross validation for KNN model.

2. Random Forest:

Training the Random forest model from sklearn using 10 fold cross validation and calculating the metrics.

```
Fold =  1
Fold Size =  160
The total positive =  47
The total negative =  113
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[94, 27], [19, 20]]
True Positive rate =  0.425531914893617
True Negative rate =  0.831858407079646
False Positive rate =  0.168141592920354
False negative rate =  0.574468085106383
Accuracy =  0.7125
Precision =  0.5128205128205128
Recall =  0.425531914893617
F1 Measure =  0.4651162790697675
Error rate =  0.2875
Balanced Accuracy =  0.6286951609866315
True Skill statistics =  0.25739032197326306
Heidke Skill Score =  0.270853972657024
----------------------------------------------------------------------------------------------
Fold =  2
Fold Size =  160
The total positive =  78
The total negative =  82
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[68, 34], [14, 44]]
True Positive rate =  0.5641025641025641
True Negative rate =  0.8292682926829268
False Positive rate =  0.17073170731707318
False negative rate =  0.4358974358974359
Accuracy =  0.7
Precision =  0.7586206896551724
Recall =  0.5641025641025641
F1 Measure =  0.6470588235294118
Error rate =  0.3
Balanced Accuracy =  0.6966854283927455
True Skill statistics =  0.3933708567854909
Heidke Skill Score =  0.3958464443045941
----------------------------------------------------------------------------------------------
```

```
Fold =  3
Fold Size =  160
The total positive =  105
The total negative =  55
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[36, 22], [19, 83]]
True Positive rate =  0.7904761904761904
True Negative rate =  0.6545454545454545
False Positive rate =  0.34545454545454546
False negative rate =  0.20952380952380953
Accuracy =  0.74375
Precision =  0.8137254901960784
Recall =  0.7904761904761904
F1 Measure =  0.8019323671497584
Error rate =  0.25625
Balanced Accuracy =  0.7225108225108225
True Skill statistics =  0.445021645021645
Heidke Skill Score =  0.4393162393162393
---------------------------------------------------------------------------------------------------
Fold =  4
Fold Size =  160
The total positive =  91
The total negative =  69
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[49, 21], [20, 70]]
True Positive rate =  0.7692307692307693
True Negative rate =  0.7101449275362319
False Positive rate =  0.2898550724637681
False negative rate =  0.23076923076923078
Accuracy =  0.74375
Precision =  0.7777777777777778
Recall =  0.7692307692307693
F1 Measure =  0.7734806629834255
Error rate =  0.25625
Balanced Accuracy =  0.7396878483835005
True Skill statistics =  0.47937569676700115
Heidke Skill Score =  0.4785373608903021
---------------------------------------------------------------------------------------------------
```

```
Fold =  5
Fold Size =  160
The total positive =  54
The total negative =  106
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[89, 32], [17, 22]]
True Positive rate =  0.4074074074074074
True Negative rate =  0.839622641509434
False Positive rate =  0.16037735849056603
False negative rate =  0.5925925925925926
Accuracy =  0.69375
Precision =  0.5641025641025641
Recall =  0.4074074074074074
F1 Measure =  0.47311827956989244
Error rate =  0.30625
Balanced Accuracy =  0.6235150244584207
True Skill statistics =  0.24703004891684135
Heidke Skill Score =  0.2650918635170604
-----------------------------------------------------------------------------------------------------
Fold =  6
Fold Size =  160
The total positive =  97
The total negative =  63
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[36, 11], [27, 86]]
True Positive rate =  0.8865979381443299
True Negative rate =  0.5714285714285714
False Positive rate =  0.42857142857142855
False negative rate =  0.1134020618556701
Accuracy =  0.7625
Precision =  0.7610619469026548
Recall =  0.8865979381443299
F1 Measure =  0.8190476190476191
Error rate =  0.2375
Balanced Accuracy =  0.7290132547864506
True Skill statistics =  0.4580265095729013
Heidke Skill Score =  0.47936290460695324
-----------------------------------------------------------------------------------------------------

Fold =  7
Fold Size =  160
The total positive =  117
The total negative =  43
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[25, 18], [18, 99]]
True Positive rate =  0.8461538461538461
True Negative rate =  0.5813953488372093
False Positive rate =  0.4186046511627907
False negative rate =  0.15384615384615385
Accuracy =  0.775
Precision =  0.8461538461538461
Recall =  0.8461538461538461
F1 Measure =  0.8461538461538461
Error rate =  0.225
Balanced Accuracy =  0.7137745974955277
True Skill statistics =  0.4275491949910554
Heidke Skill Score =  0.4275491949910555
-----------------------------------------------------------------------------------------------------
Fold =  8
Fold Size =  160
The total positive =  106
The total negative =  54
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[31, 18], [23, 88]]
True Positive rate =  0.8301886792452831
True Negative rate =  0.5740740740740741
False Positive rate =  0.42592592592592593
False negative rate =  0.16981132075471697
Accuracy =  0.74375
Precision =  0.7927927927927928
Recall =  0.8301886792452831
F1 Measure =  0.8110599078341014
Error rate =  0.25625
Balanced Accuracy =  0.7021313766596786
True Skill statistics =  0.4042627533193571
Heidke Skill Score =  0.41365749016803716
-----------------------------------------------------------------------------------------------------
```

```
Fold =  9
Fold Size =  160
The total positive =  75
The total negative =  85
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[66, 18], [19, 57]]
True Positive rate =  0.76
True Negative rate =  0.7764705882352941
False Positive rate =  0.2235294117647059
False negative rate =  0.24
Accuracy =  0.76875
Precision =  0.75
Recall =  0.76
F1 Measure =  0.7549668874172186
Error rate =  0.23125
Balanced Accuracy =  0.7682352941176471
True Skill statistics =  0.5364705882352941
Heidke Skill Score =  0.5360501567398119
-----------------------------------------------------------------------------------------------------
Fold =  10
Fold Size =  159
The total positive =  85
The total negative =  74
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[42, 15], [32, 70]]
True Positive rate =  0.8235294117647058
True Negative rate =  0.5675675675675675
False Positive rate =  0.43243243243243246
False negative rate =  0.17647058823529413
Accuracy =  0.7044025157232704
Precision =  0.6862745098039216
Recall =  0.8235294117647058
F1 Measure =  0.748663101604278
Error rate =  0.29559748427672955
Balanced Accuracy =  0.6955484896661367
True Skill statistics =  0.3910969793322734
Heidke Skill Score =  0.3969983054950375
-----------------------------------------------------------------------------------------------------
*****************************************************************************************************

Average True Positive rate =  0.7103218721418713
Average True Negative rate =  0.693637587349641
Average False Positive rate =  0.3063624126503591
Average False negative rate =  0.28967812785812874
Average Accuracy =  0.7348152515723271
Average Precision =  0.726333013020532
Average Recall =  0.7103218721418713
Average F1 Measure =  0.714059777435932
Average Error rate =  0.265184748427673
Average Balanced Accuracy =  0.7019797297457562
Average True Skill statistics =  0.40395945949151224
Average Heidke Skill Score =  0.41032639326861153
*****************************************************************************************************
```

Figure 2: 10-fold cross validation for Random Forest model.

3. Long Short Term Memory:

Training the LSTM model from Keras using 10 fold cross validation and calculating the metrics.

```
Epoch 1/5
9/9 [==============================] - 1s 11ms/step - loss: 0.7422 - accuracy: 0.4385
Epoch 2/5
9/9 [==============================] - 0s 6ms/step - loss: 0.7001 - accuracy: 0.4538
Epoch 3/5
9/9 [==============================] - 0s 8ms/step - loss: 0.6899 - accuracy: 0.5615
Epoch 4/5
9/9 [==============================] - 0s 6ms/step - loss: 0.6871 - accuracy: 0.5615
Epoch 5/5
9/9 [==============================] - 0s 5ms/step - loss: 0.6858 - accuracy: 0.5615
Fold =  1
Fold Size =  160
The total positive =  47
The total negative =  113
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[0, 0], [113, 47]]
True Positive rate =  1.0
True Negative rate =  0.0
False Positive rate =  1.0
False negative rate =  0.0
Accuracy =  0.29375
Precision =  0.29375
Recall =  1.0
F1 Measure =  0.45410628019323673
Error rate =  0.70625
Balanced Accuracy =  0.5
True Skill statistics =  0.0
Heidke Skill Score =  0.0
--------------------------------------------------------------------------------------------------------
Epoch 1/5
9/9 [==============================] - 1s 11ms/step - loss: 0.6891 - accuracy: 0.5400
Epoch 2/5
9/9 [==============================] - 0s 8ms/step - loss: 0.6871 - accuracy: 0.5400
Epoch 3/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6849 - accuracy: 0.5400
Epoch 4/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6840 - accuracy: 0.5400
Epoch 5/5
9/9 [==============================] - 0s 6ms/step - loss: 0.6817 - accuracy: 0.5400
Fold =  2
Fold Size =  160
The total positive =  78
The total negative =  82
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[74, 59], [8, 19]]
True Positive rate =  0.24358974358974358
True Negative rate =  0.9024390243902439
False Positive rate =  0.0975609756097561
False negative rate =  0.7564102564102564
Accuracy =  0.58125
Precision =  0.7037037037037037
Recall =  0.24358974358974358
F1 Measure =  0.3619047619047619
Error rate =  0.41875
Balanced Accuracy =  0.5730143839899937
True Skill statistics =  0.14602876797998748
Heidke Skill Score =  0.14839529710835717
--------------------------------------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------------------------------
Epoch 1/5
9/9 [==============================] - 1s 10ms/step - loss: 0.7257 - accuracy: 0.4788
Epoch 2/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6994 - accuracy: 0.4788
Epoch 3/5
9/9 [==============================] - 0s 8ms/step - loss: 0.6939 - accuracy: 0.5003
Epoch 4/5
9/9 [==============================] - 0s 8ms/step - loss: 0.6923 - accuracy: 0.5212
Epoch 5/5
9/9 [==============================] - 0s 8ms/step - loss: 0.6913 - accuracy: 0.5212
Fold =  3
Fold Size =  160
The total positive =  105
The total negative =  55
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[36, 35], [19, 70]]
True Positive rate =  0.6666666666666666
True Negative rate =  0.6545454545454545
False Positive rate =  0.34545454545454546
False negative rate =  0.3333333333333333
Accuracy =  0.6625
Precision =  0.7865168539325843
Recall =  0.6666666666666666
F1 Measure =  0.7216494845360824
Error rate =  0.3375
Balanced Accuracy =  0.6606060606060606
True Skill statistics =  0.32121212121212117
Heidke Skill Score =  0.3004048582995951
----------------------------------------------------------------------------------------------------

 ----------------------------------------------------------------------------------------------------
 Epoch 1/5
 9/9 [==============================] - 1s 11ms/step - loss: 0.6867 - accuracy: 0.5316
 Epoch 2/5
 9/9 [==============================] - 0s 8ms/step - loss: 0.6853 - accuracy: 0.5330
 Epoch 3/5
 9/9 [==============================] - 0s 9ms/step - loss: 0.6838 - accuracy: 0.5302
 Epoch 4/5
 9/9 [==============================] - 0s 8ms/step - loss: 0.6823 - accuracy: 0.5351
 Epoch 5/5
 9/9 [==============================] - 0s 8ms/step - loss: 0.6805 - accuracy: 0.5337
 Fold =  4
 Fold Size =  160
 The total positive =  91
 The total negative =  69
 Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[44, 24], [25, 67]]
 True Positive rate =  0.7362637362637363
 True Negative rate =  0.6376811594202898
 False Positive rate =  0.36231884057971014
 False negative rate =  0.26373626373626374
 Accuracy =  0.69375
 Precision =  0.7282608695652174
 Recall =  0.7362637362637363
 F1 Measure =  0.7322404371584701
 Error rate =  0.30625
 Balanced Accuracy =  0.6869724478420131
 True Skill statistics =  0.37394489568402617
 Heidke Skill Score =  0.3746011486917677
```

```
---------------------------------------------------------------------------------------------------
Epoch 1/5
9/9 [==============================] - 1s 12ms/step - loss: 0.6866 - accuracy: 0.5566
Epoch 2/5
9/9 [==============================] - 0s 10ms/step - loss: 0.6844 - accuracy: 0.5566
Epoch 3/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6836 - accuracy: 0.5566
Epoch 4/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6823 - accuracy: 0.5566
Epoch 5/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6810 - accuracy: 0.5566
Fold =  5
Fold Size =  160
The total positive =  54
The total negative =  106
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[0, 0], [106, 54]]
True Positive rate =  1.0
True Negative rate =  0.0
False Positive rate =  1.0
False negative rate =  0.0
Accuracy =  0.3375
Precision =  0.3375
Recall =  1.0
F1 Measure =  0.5046728971962617
Error rate =  0.6625
Balanced Accuracy =  0.5
True Skill statistics =  0.0
Heidke Skill Score =  0.0

---------------------------------------------------------------------------------------------------
Epoch 1/5
9/9 [==============================] - 1s 12ms/step - loss: 0.6871 - accuracy: 0.5372
Epoch 2/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6846 - accuracy: 0.5420
Epoch 3/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6834 - accuracy: 0.5268
Epoch 4/5
9/9 [==============================] - 0s 6ms/step - loss: 0.6812 - accuracy: 0.5288
Epoch 5/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6792 - accuracy: 0.5566
Fold =  6
Fold Size =  160
The total positive =  97
The total negative =  63
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[55, 36], [8, 61]]
True Positive rate =  0.6288659793814433
True Negative rate =  0.873015873015873
False Positive rate =  0.12698412698412698
False negative rate =  0.3711340206185567
Accuracy =  0.725
Precision =  0.8840579710144928
Recall =  0.6288659793814433
F1 Measure =  0.7349397590361446
Error rate =  0.275
Balanced Accuracy =  0.7509409261986582
True Skill statistics =  0.5018818523973163
Heidke Skill Score =  0.46561408835585244
```

```
----------------------------------------------------------------------------------------------------
Epoch 1/5
9/9 [==============================] - 1s 10ms/step - loss: 0.6924 - accuracy: 0.5288
Epoch 2/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6906 - accuracy: 0.5129
Epoch 3/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6896 - accuracy: 0.5455
Epoch 4/5
9/9 [==============================] - 0s 8ms/step - loss: 0.6887 - accuracy: 0.5142
Epoch 5/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6881 - accuracy: 0.5247
Fold =  7
Fold Size =  160
The total positive =  117
The total negative =  43
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[43, 111], [0, 6]]
True Positive rate =  0.05128205128205128
True Negative rate =  1.0
False Positive rate =  0.0
False negative rate =  0.9487179487179487
Accuracy =  0.30625
Precision =  1.0
Recall =  0.05128205128205128
F1 Measure =  0.09756097560975609
Error rate =  0.69375
Balanced Accuracy =  0.5256410256410257
True Skill statistics =  0.05128205128205128
Heidke Skill Score =  0.028233749179251477

----------------------------------------------------------------------------------------------------
Epoch 1/5
9/9 [==============================] - 1s 11ms/step - loss: 0.7162 - accuracy: 0.5205
Epoch 2/5
9/9 [==============================] - 0s 8ms/step - loss: 0.6932 - accuracy: 0.5205
Epoch 3/5
9/9 [==============================] - 0s 8ms/step - loss: 0.6893 - accuracy: 0.5205
Epoch 4/5
9/9 [==============================] - 0s 6ms/step - loss: 0.6880 - accuracy: 0.5462
Epoch 5/5
9/9 [==============================] - 0s 5ms/step - loss: 0.6862 - accuracy: 0.5274
Fold =  8
Fold Size =  160
The total positive =  106
The total negative =  54
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[39, 46], [15, 60]]
True Positive rate =  0.5660377358490566
True Negative rate =  0.7222222222222222
False Positive rate =  0.2777777777777778
False negative rate =  0.4339622641509434
Accuracy =  0.61875
Precision =  0.8
Recall =  0.5660377358490566
F1 Measure =  0.6629834254143646
Error rate =  0.38125
Balanced Accuracy =  0.6441299790356394
True Skill statistics =  0.2882599580712788
Heidke Skill Score =  0.25267993874425726
----------------------------------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------------------------------
Epoch 1/5
9/9 [==============================] - 1s 9ms/step - loss: 0.6874 - accuracy: 0.5420
Epoch 2/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6860 - accuracy: 0.5420
Epoch 3/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6852 - accuracy: 0.5420
Epoch 4/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6838 - accuracy: 0.5420
Epoch 5/5
9/9 [==============================] - 0s 7ms/step - loss: 0.6829 - accuracy: 0.5420
Fold =  9
Fold Size =  160
The total positive =  75
The total negative =  85
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[0, 0], [85, 75]]
True Positive rate =  1.0
True Negative rate =  0.0
False Positive rate =  1.0
False negative rate =  0.0
Accuracy =  0.46875
Precision =  0.46875
Recall =  1.0
F1 Measure =  0.6382978723404256
Error rate =  0.53125
Balanced Accuracy =  0.5
True Skill statistics =  0.0
Heidke Skill Score =  0.0

----------------------------------------------------------------------------------------------------
Epoch 1/5
9/9 [==============================] - 1s 8ms/step - loss: 0.7088 - accuracy: 0.4653
Epoch 2/5
9/9 [==============================] - 0s 5ms/step - loss: 0.6979 - accuracy: 0.4201
Epoch 3/5
9/9 [==============================] - 0s 4ms/step - loss: 0.6950 - accuracy: 0.5264
Epoch 4/5
9/9 [==============================] - 0s 4ms/step - loss: 0.6935 - accuracy: 0.5347
Epoch 5/5
9/9 [==============================] - 0s 5ms/step - loss: 0.6918 - accuracy: 0.5347
Fold =  10
Fold Size =  159
The total positive =  85
The total negative =  74
Confusion Matrix : [[true_negative, false_negative],[false_positive,true_positive]] =  [[1, 3], [73, 82]]
True Positive rate =  0.9647058823529412
True Negative rate =  0.013513513513513514
False Positive rate =  0.9864864864864865
False negative rate =  0.03529411764705882
Accuracy =  0.5220125786163522
Precision =  0.5290322580645161
Recall =  0.9647058823529412
F1 Measure =  0.6833333333333333
Error rate =  0.4779874213836478
Balanced Accuracy =  0.48910969793322734
True Skill statistics =  -0.021780604133545323
Heidke Skill Score =  -0.023200677392040642
----------------------------------------------------------------------------------------------------
****************************************************************************************************
Average True Positive rate =  0.685741179538564
Average True Negative rate =  0.48034172471075964
Average False Positive rate =  0.5196582752892402
Average False negative rate =  0.3142588204614361
Average Accuracy =  0.5209512578616352
Average Precision =  0.6531571656280514
Average Recall =  0.685741179538564
Average F1 Measure =  0.5591689226722838
Average Error rate =  0.4790487421383648
Average Balanced Accuracy =  0.5830414521246619
Average True Skill statistics =  0.1660829042493236
Average Heidke Skill Score =  0.15467284029870404
****************************************************************************************************
```

Figure 3: 10-fold cross validation for LSTM model.

REFERENCES

1. https://www.kaggle.com/nareshbhat/wine-quality-binary-classification -Dataset
2. https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html -KNN classifier
3. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html - Random forest Classifier
4. https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/ -LSTM
5. https://www.tensorflow.org/api_docs/python/tf/keras/layers/LSTM -LSTM
6. https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/ - LSTM