machine learning

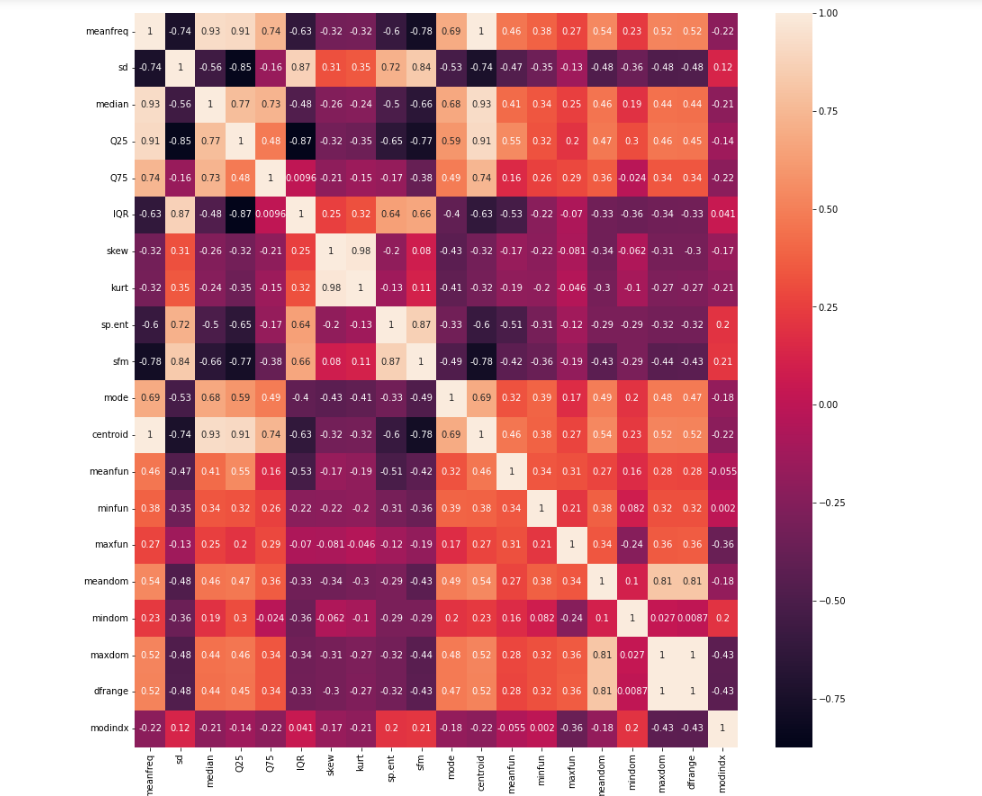
Name: Kareem Abdelhady

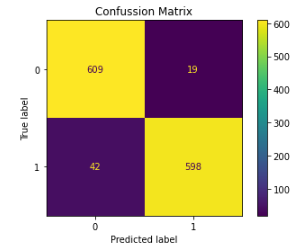
Email: [Kareem197155@bue.edu.eg](mailto:Kareem197155@bue.edu.eg)

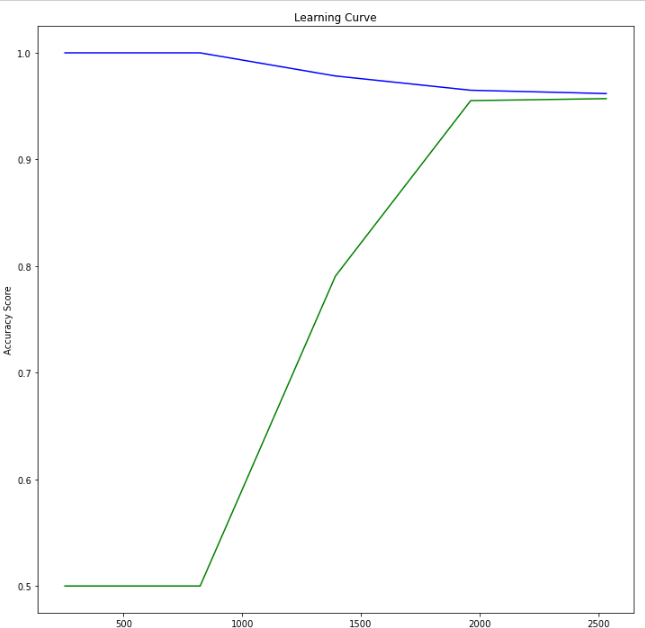
Phone numbers: +201158806629/+201030791740

# Classification:

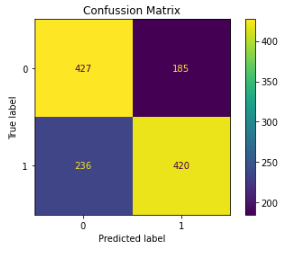
1. Data set: Gender Recognition by Voice: <https://www.kaggle.com/primaryobjects/voicegender>
2. Description: this data contains values for the voice through it we can predict if the person is Male or female. Contains 20 feature (columns) and only one Target which name is (Label). It also has 3166 row or data records. Features name are not clear so after reading description of the data I found that:
3. meanfreq is mean frequency (in kHz)
4. sd is standard deviation of frequency
5. median is median frequency (in kHz)
6. Q25 is first quantile (in kHz)
7. Q75 is third quantile (in kHz)
8. IQR is interquantile range (in kHz)
9. skew is skewness (see note in specprop description)
10. kurt is kurtosis (see note in specprop description)
11. sp.ent is spectral entropy
12. sfm is spectral flatness
13. mode is mode frequency
14. centroid is frequency centroid (see specprop)
15. peakf is peak frequency (frequency with highest energy)
16. meanfun is average of fundamental frequency measured across acoustic signal
17. minfun is minimum fundamental frequency measured across acoustic signal
18. maxfun is maximum fundamental frequency measured across acoustic signal
19. meandom is average of dominant frequency measured across acoustic signal
20. mindom is minimum of dominant frequency measured across acoustic signal
21. maxdom is maximum of dominant frequency measured across acoustic signal
22. dfrange is range of dominant frequency measured across acoustic signal
23. modindx is modulation index. Calculated as the accumulated absolute difference between adjacent measurements of fundamental frequencies divided by the frequency range
24. label is male or female
25. Data Preprocessing:
    1. Check if there are any null’s in the data to drop it but there are no nulls
    2. Check if there are any duplicates in the data, found two and has been dropped
    3. Remove the outliers
26. Visualizations: using heat map to identify the values in every class according to correlation.

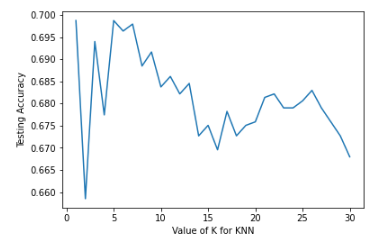


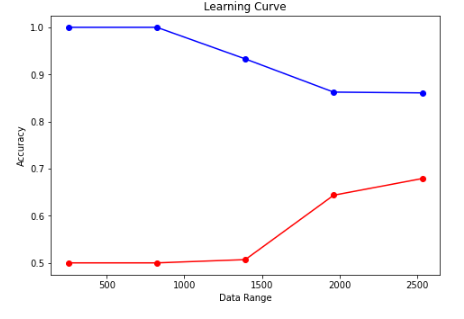
1. Confusion Matrix (decision tree classifier): This matrix represents the performance of the decision tree classifier, each row in a confusion matrix represents an actual class, while each column represents a predicted class.
2. Learning curve according to the decision Tree: it represent the gap between learning and testing curve, it represent the accuracy.

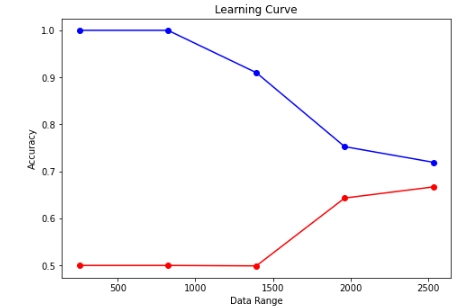


1. Confusion Matrix (KNN classifier):

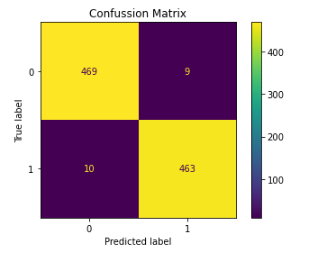


1. KKN plot: to identify the testing accuracy with the value of K for Knn.
2. Learning curve according to the KNN:

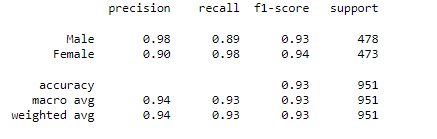


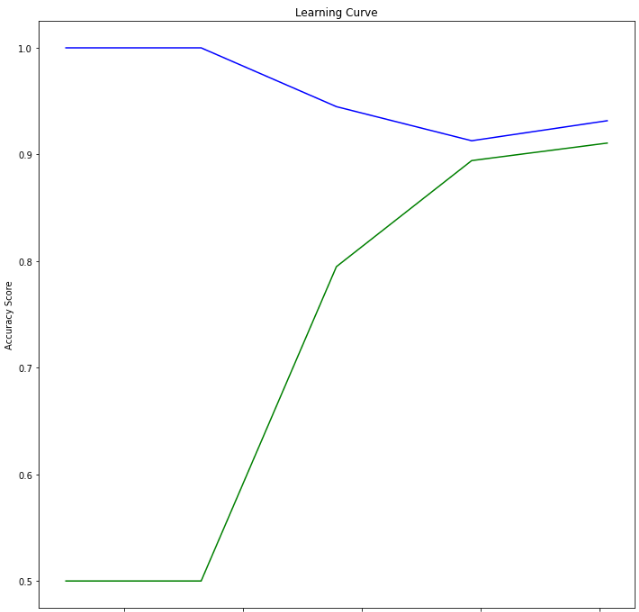


1. Confusion Matrix (Random forest classifier):



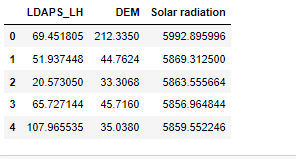
1. Classification report (Random forest classifier):



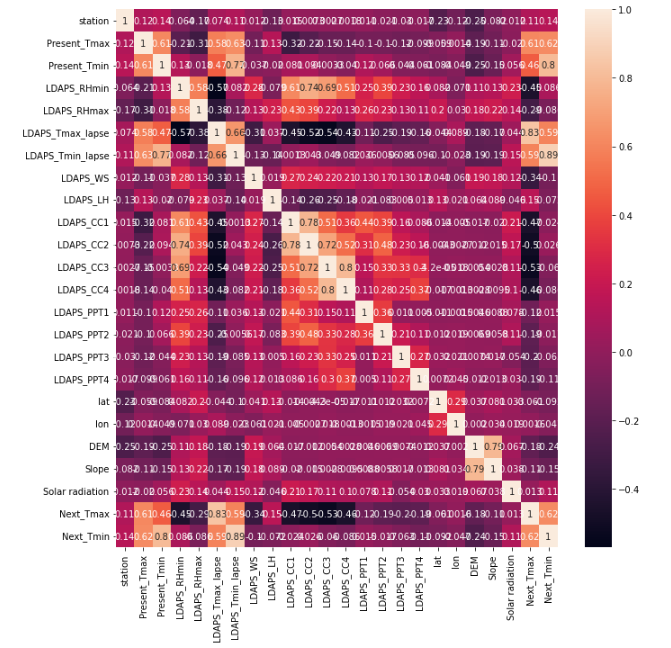
1. Learning curve according to the Random forest:

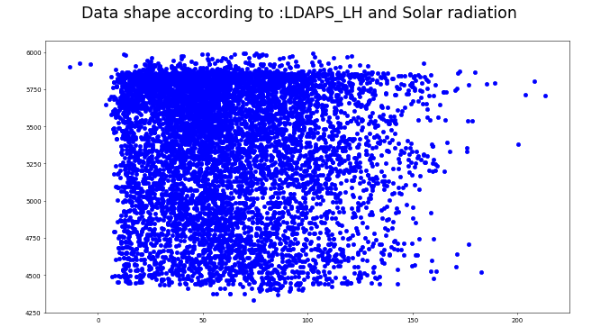
# Regression data set:

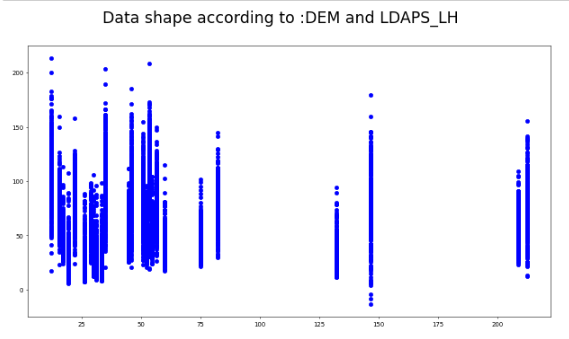
1. Data set: Bias correction of numerical prediction model temperature forecast Data Set: [https://archive.ics.uci.edu/ml/datasets/Bias+correction+of+numerical+prediction+model+temperature+forecast#](https://archive.ics.uci.edu/ml/datasets/Bias+correction+of+numerical+prediction+model+temperature+forecast)
2. Data description: This data is for the purpose of bias correction of next-day maximum and minimum air temperatures; it contains 7752 rows and 25 columns (features).
3. Data Preprocessing:
   1. Check if there any null values in the data and drop it
   2. Check if there any duplicate values in the data and remove it
   3. Divide the data into two classes to be a binary classification
4. Visualizations:
5. Find which feature in the data set effect the result’s using the feature selection

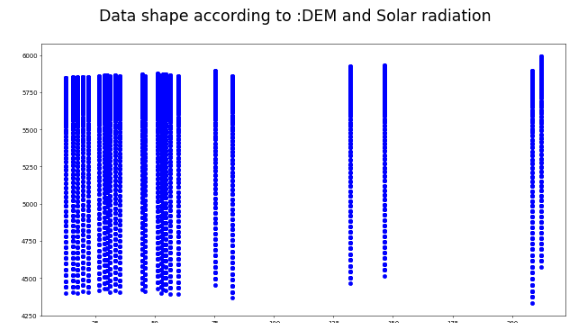


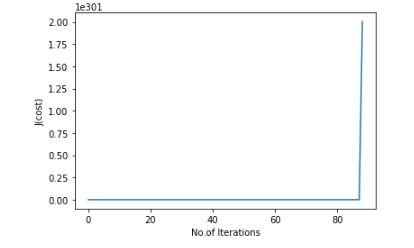
1. Using the heat map to classify every value in classes according to correlation:

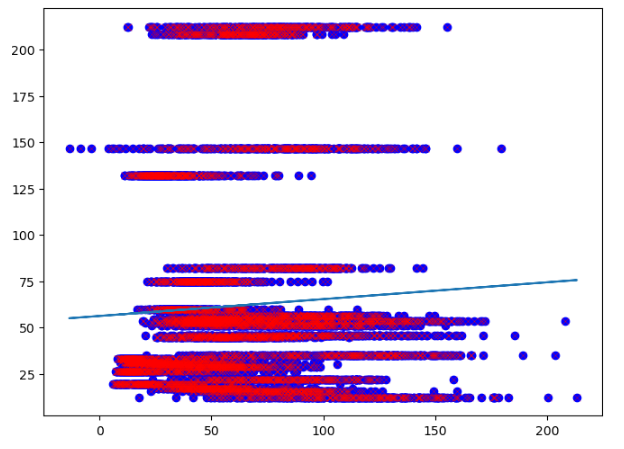


1. Data shape according to :LDAPS\_LH and Solar radiation
2. Data shape according to: DEM and LDAPS\_LH:



1. Data shape according to: DEM and Solar radiation:
2. Gradient Descent:



1. LinearRegression: