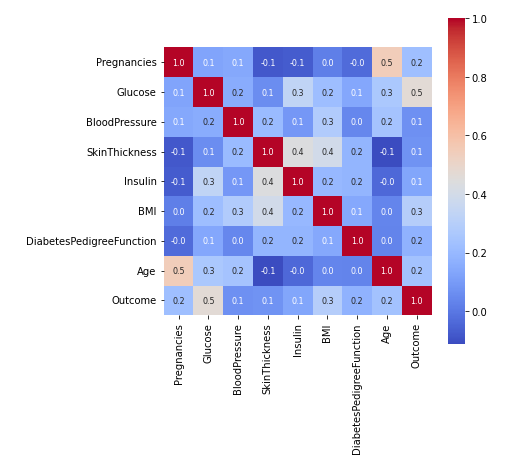
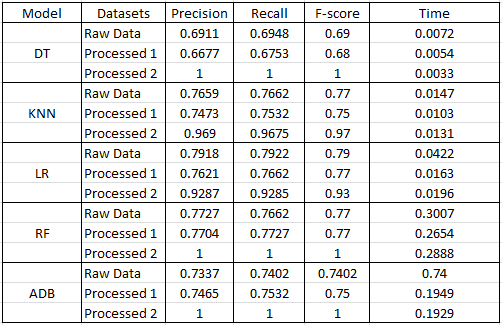
DATASET-1:



* From the heatmap above, we can infer the following:

1. Features (Pregnancies ,Age) and (Glucose,Outcome) have a strong positive correlation with a value of 0.5
2. Features (Insulin,SkinThickness) and (BMI,SkinThickness) also have a string positive correlation with a value of 0.4
3. Another positive correlation is between features (Glucose,Insulin),(BloodPressure,BMI),(BMI,Outcome) and (Age,Glucose). The value being 0.3 for all the paired features.
4. Features (Pregnancies,Insulin),(SkinThickness,Pregnancies) and (SkinThickness,Age) have a string negative correlation with the value of -0.1



* From the above table we can infer that,

1. Decision Tree and AdaBoost Models have the highest precision for processed 2 with the value of 1. Also,KNN Model have the second highest precision for processed 2 with the value of 0.969. Decision Tree Model have the lowest precision value of 0.6677 for processed 1.
2. Regarding Recall values, Decision Tree,Random Forest and AdaBoost Models have the highest recall value of 1 for processed 2. Also, KNN Model have the second highest recall value of 0.9675 for processed 2. Decision Tree Model have lowest recall value of 0.9753 for processed 1.
3. Regarding F-score values, Decision Tree,Random Forest and AdaBoost Models have the highest F-score for processed 2 with the value of 1.Also,KNN Model have the second highest F-score value of 0.97. Decision Tree have the lowest F-score value of 0.68 for processed 1.
4. Regarding Time taken by different models,we can infer that, AdaBoost Model for raw data is taking the longest time of 0.74s when compared to other models.Also, Decision Tree is taking the least time of 0.0033s compared to all other models.

* From the above graph, we can infer that for the raw original data,

1. Linear Regression Model have the highest accuracy with the value of 79.2% compared to other models.
2. Also,both KNN and Random Forest Models have the second highest accuracy with the value of 76.62%.
3. Then,AdaBoost Model have the accuracy value of 74.02% and Decision Tree have the lowest accuracy with the value of 69.48%.

* From the above graph for the processed 1, we can infer that

1. Random Forest Model have the highest accuracy with the value of 77.27% compared to other models.
2. Linear Regression Model have the second highest accuracy with the value of 76.62%.
3. Both KNN and AdaBoost have the moderate accuracy with the value of 75.32%.
4. Decision Tree have the least accuracy value of 67.53%.

* From the above graph for processed 2, we can infer that

1. Decision Tree, Random Forest and AdaBoost Models have the highest accuracy with the value of 100%.
2. KNN Model have the second highest accuracy with the value of 96.75%.
3. Linear Regression Model have the lowest accuracy value of 92.85%.

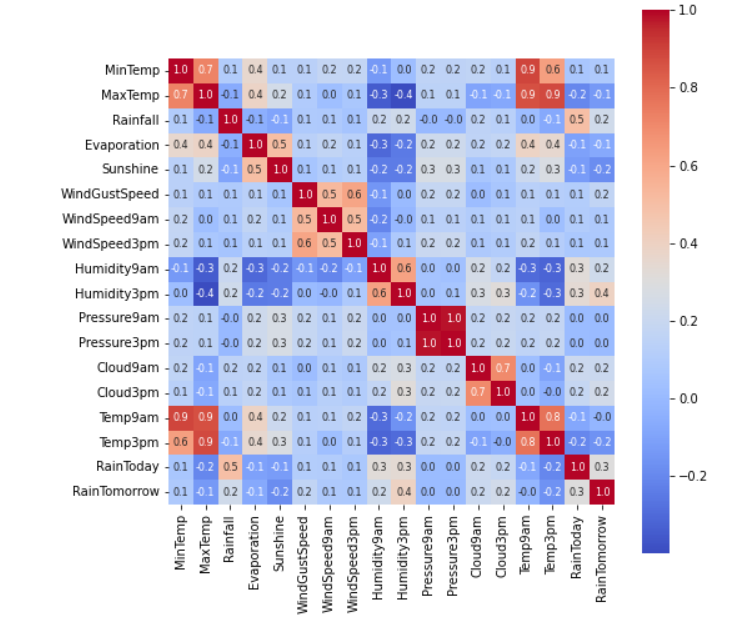
* From the above graph results using Voting classifier for raw data,processed1 and processed2, we can infer that,

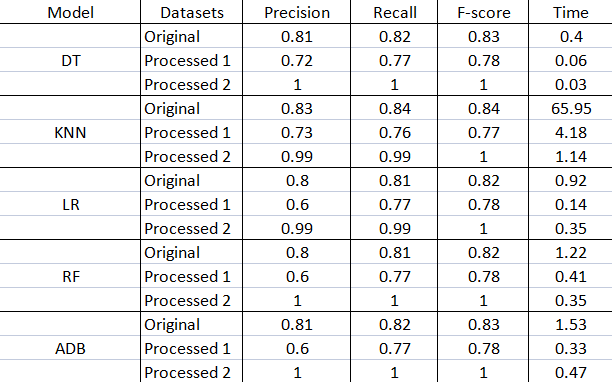
1. Processed2 have the highest accuracy value of 100%.
2. Raw data have the moderate accuracy value of 76.56%.
3. Processed1 have the lowest accuracy value of 74.73%.

**INFERENCE:**

* The accuracies in the processed stage 2 increased from original stage and processed stage 1.
* DT and ADB outperformed the rest with 100 accuracy.
* Using voting ensemble method accuracy of 100 is achieved from DT,KNN,LR,RF,ADB classifiers.

**DATASET-2:**





**INFERENCE:**

* The accuracies in the processed stage 2 increased from original stage and processed stage.
* DT,RF and ADB outperformed the rest with 100 accuracy.
* Using voting ensemble method accuracy of 100 is achieved from DT,KNN,LR,RF,ADB classifiers.