National University
of computer and emerging sciences

Assignment # 02

Classification

DS5004 Machine Learning For Data Science

Muhammad Usman

21P-8743

Submitted to:

Dr. M.Taimoor Khan

Department of Computer Science MS(DS)

FAST-NUCES Peshawar

Classification:

I have implemented the following 5 classification algorithms on

Skin Segmentation Data Set.

1) Logistic Regression
2) Decision Tree
3) K Nearest Neighbor
4) Naive Bayes
5) Logistic Regression from Scratch Without Sklearn

Evaluations:

To Classify the data set I have used Stratified K Fold Cross Validation.

Stratified k fold cross validation is an extension of regular k fold cross validation but specifically for classification problems where rather than the splits being completely random, the ratio between the target classes is the same in each fold as it is in the full dataset.

For Evaluations purpose Calculated Precision,Recall and F1 Score.

Logistic Regression Evaluation Score on Stratified 5 fold Cross Validation.

	Precision	Recall	F1 Score
F1	0.792909	0.815670	0.804129
F2	0.791106	0.820193	0.805387
F3	0.795030	0.824027	0.809269
F4	0.799148	0.830105	0.814333
F5	0.787606	0.827254	0.806944

Note: Each row represents precision recall and f1 score for each fold (F1-F5).

Decision Tree Evaluation Score on Stratified 5 fold Cross Validation.

	Precision	Recall	F1 Score
F1	0.998132	0.998132	0.998132
F2	0.998132	0.998329	0.998231
F3	0.998134	0.998919	0.998526
F4	0.997349	0.998624	0.997986
F5	0.997252	0.999115	0.998183

K Nearest Neighbor Evaluation Score on Stratified 5 fold Cross Validation.

	Precision	Recall	F1 Score
F1	0.998135	0.999705	0.998919
F2	0.997939	0.999803	0.998871
F3	0.997842	1.000000	0.998920
F4	0.997352	0.999803	0.998576
F5	0.997744	0.999902	0.998821

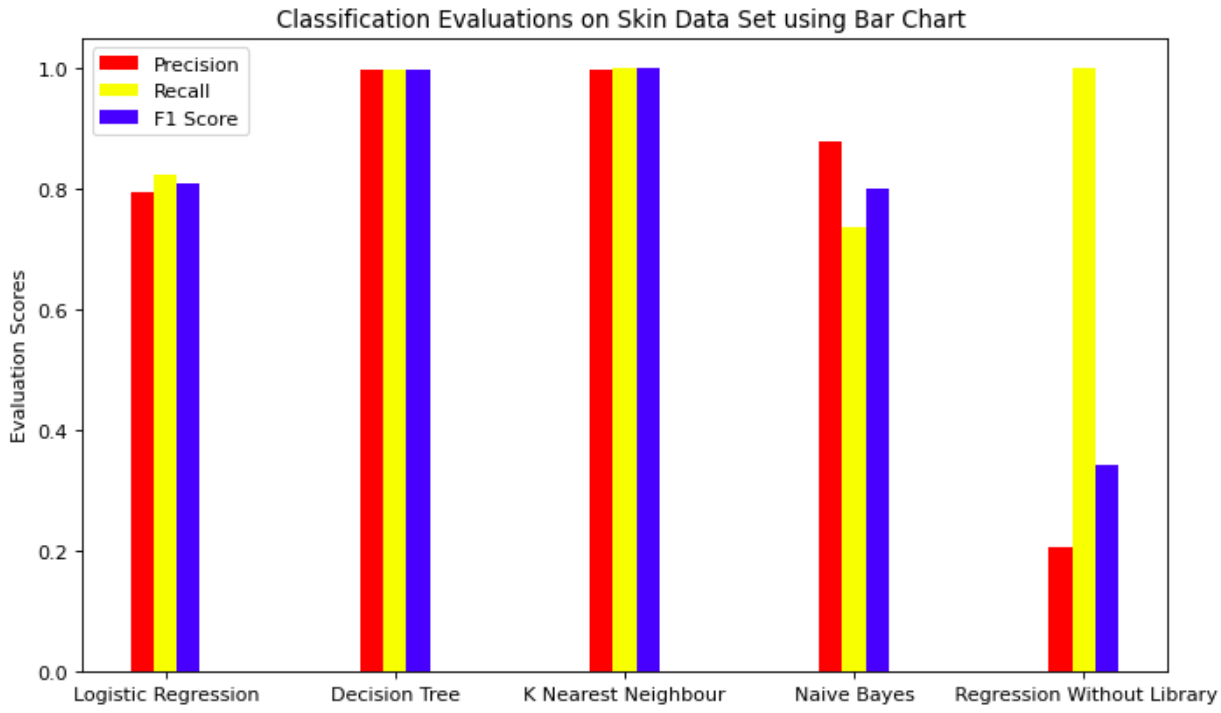
Naive Bayes Evaluation Score on Stratified 5 fold Cross Validation.

	Precision	Recall	F1 Score
F1	0.876707	0.725619	0.794040
F2	0.876141	0.735745	0.799829
F3	0.882471	0.735942	0.802573
F4	0.881810	0.741618	0.805661
F5	0.872533	0.738964	0.800213

Logistic Regression Algorithm (without Library functions sklearn) Evaluation Score on Stratified 5 fold Cross Validation.

	Precision	Recall	F1 Score
F1	0.207541	1.0	0.343742
F2	0.207545	1.0	0.343747
F3	0.207545	1.0	0.343747
F4	0.207525	1.0	0.343719
F5	0.207525	1.0	0.343719

It is evident from the above tables that Decision Tree and K Nearest Neighbor have the highest evaluation score as compared to other classifiers. To conclude the above tabular evaluations graphically calculated the mean in each algorithm for the 5 Folds to get a single numerical number for precision, recall and f1 score for each classifier.



Highest Evaluation Score Values (Mean):

Classifier	Precision	Recall	F1 Score
Decision Tree	0.9977998415476769	0.9986236283654888	0.9982114469164621
K Nearest Neighbor	0.9978023478160484	0.9998426996666094	0.9988214592455253