Report

Assignment 3

DS-312-A-Application of Data Science

Group Members:

Muhammad Nadeem

Bilal Butt

Mehar Hamid Ishfaq

Problem Statement:

In this Assignment we generate some dummy text related to Domain and Implement Different Algorithm's and Find the Useful Information from it.

What Can we do.

In this Assignment first we install Library that are called Faker. This Library use for generate some dummy text. Attributes that we used are insulin, glucose, age, family_history, physical_activity and diet.

In which we assign labels as shown in figure below.

```
# Assign labels based on the high sugar threshold
labels = ['High' if g >= high_sugar_threshold else 'Low' for g in glucose]
```

And finally data will be shown as.

	Name	Email	Insulin_Level	Fasting_Glucose	Age	Family_History	Physical_Activity	Diet	Label
0	Faith Mccullough	catherine18@example.org	35.149173	237.214865	71	Yes	1.958651	Medium	High
1	Faith Mccullough	catherine18@example.org	31.720625	135.028612	65	No	1.234302	High	Low
2	Faith Mccullough	catherine18@example.org	35.247386	215.555406	88	Yes	2.349152	Medium	High
3	Faith Mccullough	catherine18@example.org	15.369787	219.451127	62	Yes	3.347901	High	High
4	Faith Mccullough	catherine18@example.org	12.533210	199.556730	67	Yes	1.130143	Medium	High

In which some columns are categorical features so we need to Encode it.

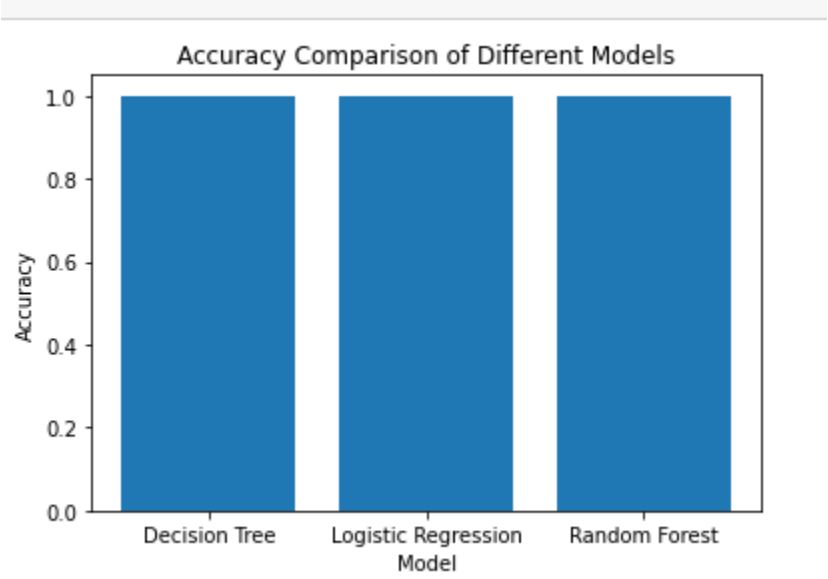
We encode these columns from LabelEncoder library using this Built in Library ' from sklearn.preprocessing import LabelEncoder ' Finally we have data like Below

```
     Name  Email  Insulin_Level  Fasting_Glucose  Age  Family_History
0       0      0      35.149173       237.214865   71               1
1       0      0      31.720625       135.028612   65               0
2       0      0      35.247386       215.555406   88               1
3       0      0      15.369787       219.451127   62               1
4       0      0      12.533210       199.556730   67               1
..    ...    ...            ...              ...  ...             ...
995     0      0      42.623465       195.316449   56               0
996     0      0      33.588734       133.882890   55               1
997     0      0       5.023597       228.187132   92               0
998     0      0      30.674663        78.264077   94               0
999     0      0      37.077999        80.109185   46               1

     Physical_Activity  Diet  Label
0             1.958651     2      0
1             1.234302     0      1
2             2.349152     2      0
3             3.347901     0      0
4             1.130143     2      0
..                 ...   ...    ...
995           2.742678     1      0
996           1.924178     1      1
997           3.345028     1      0
998           4.981729     1      1
999           1.159088     0      1

[1000 rows x 9 columns]
```

The We Implement Three models Decision Tree, Logistic Regression, Random Forest. Accuracies of these three models is

1.0 . I think it not overfit because accuracy 1 is on test data not on Training data.

Finally we Compare these three models that are:



I can Resubmit This assignment For Submit this Report. Code of This Assignment is Submit before Due Date. Thank You for Late Submit Permission.

The End