PIMA DIABETES PREDICTION USING ARTIFICIAL INTELLIGENCE

Gloria Hawkins, Media Design School 2023
Contact: gloria.hawkinsroberts@mds.ac.nz

Image retrieved from https://www.cdc.gov/diabetes/library/spotlights/diabetes-facts-stats.html

CONTENTS

-TOPIC

-BACKGROUND

-VISUALISATIONS

-MODEL SELECTION

-MODEL PREPARATION

-SOFTWARE IMPLEMENTATION

-MODEL REVIEW

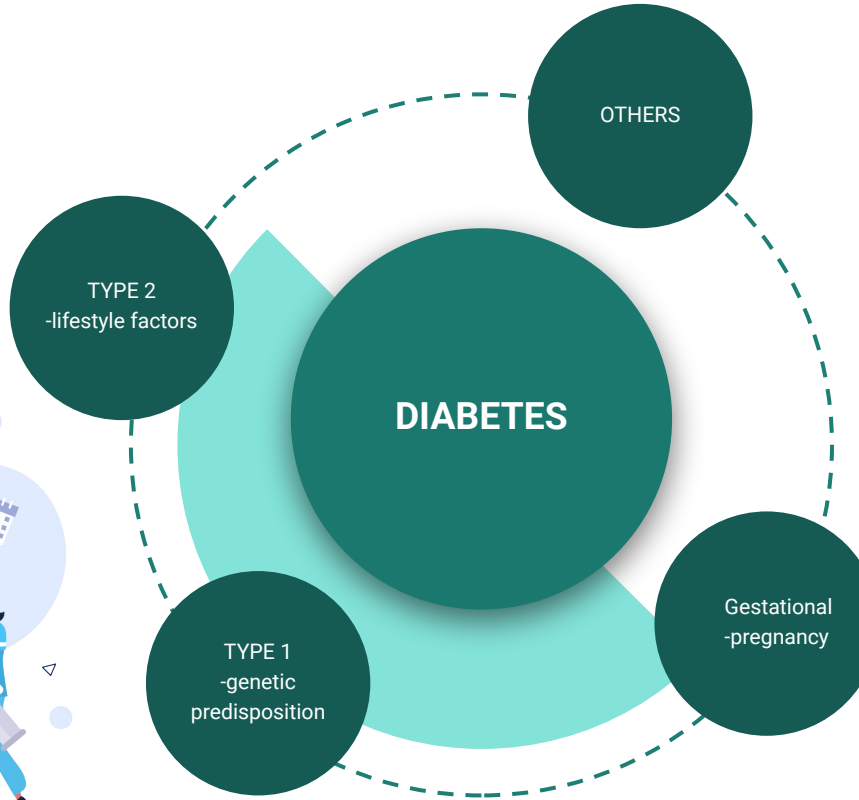
-QUESTiONS

PROJECT TOPIC:

-DIABETES

-TYPES

-SYMPTOMS



Images retrieved from:
https://www.phistar.in/why-clinical-trials-are-necessary-possible-advantages-and-disadvantages-of-clinical-trial/

DATA BACKGROUND:

-SOURCE

-PIMA INDIANS

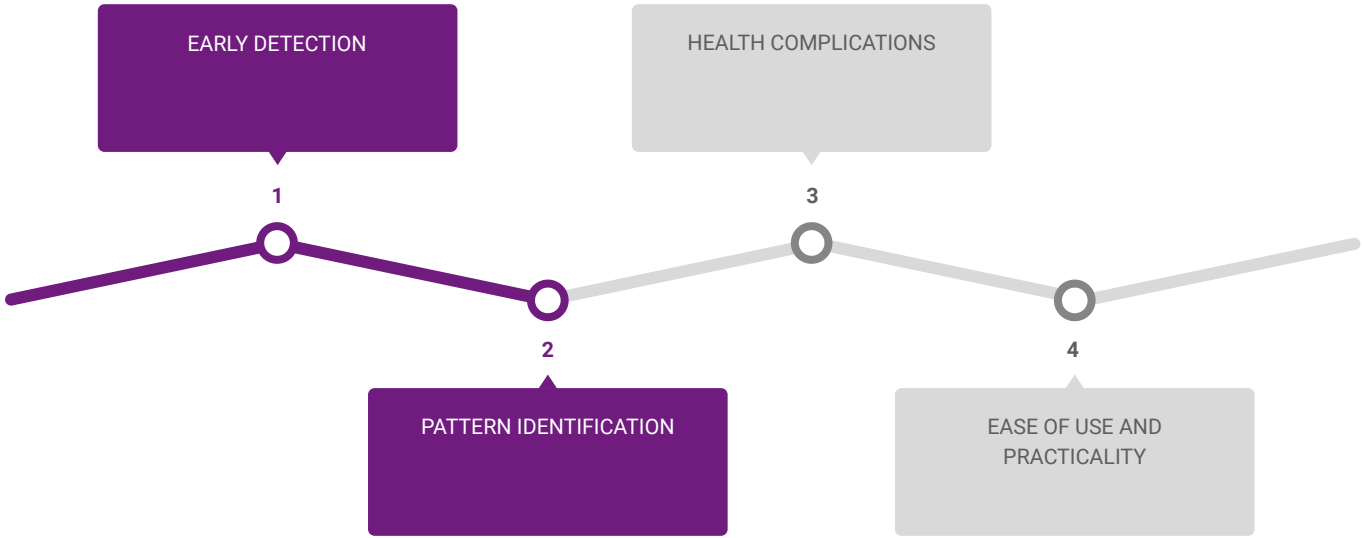
-9 VARIABLES

-INDEPENDENT AND DEPENDENT VARIABLES

NIH
NIDDK

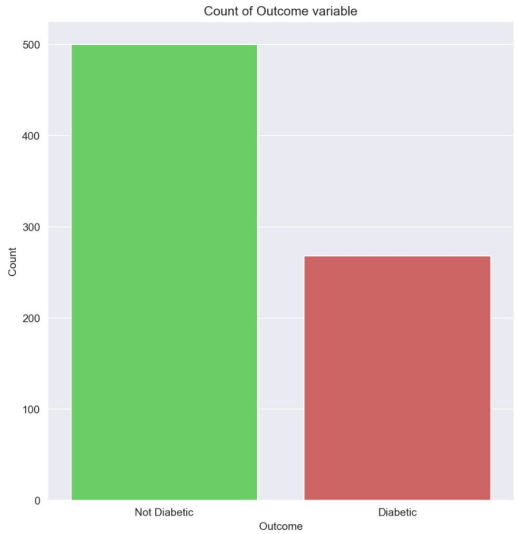
Images retrieved from: https://accessgenealogy.com/arizona/pima-tribe.html,, https://www.mapsofworld.com/answers/united-states/many-states-mexico-border/, https://www.nih.gov/.

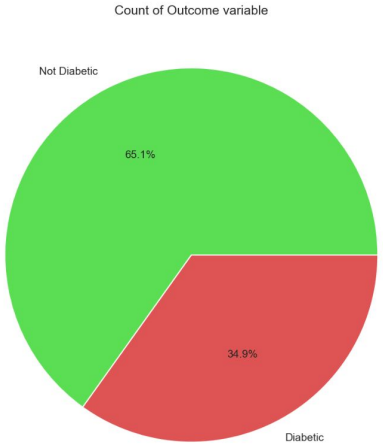
WHY CHOOSE AI?



VISUALISATIONS

PLOTTING COUNTS



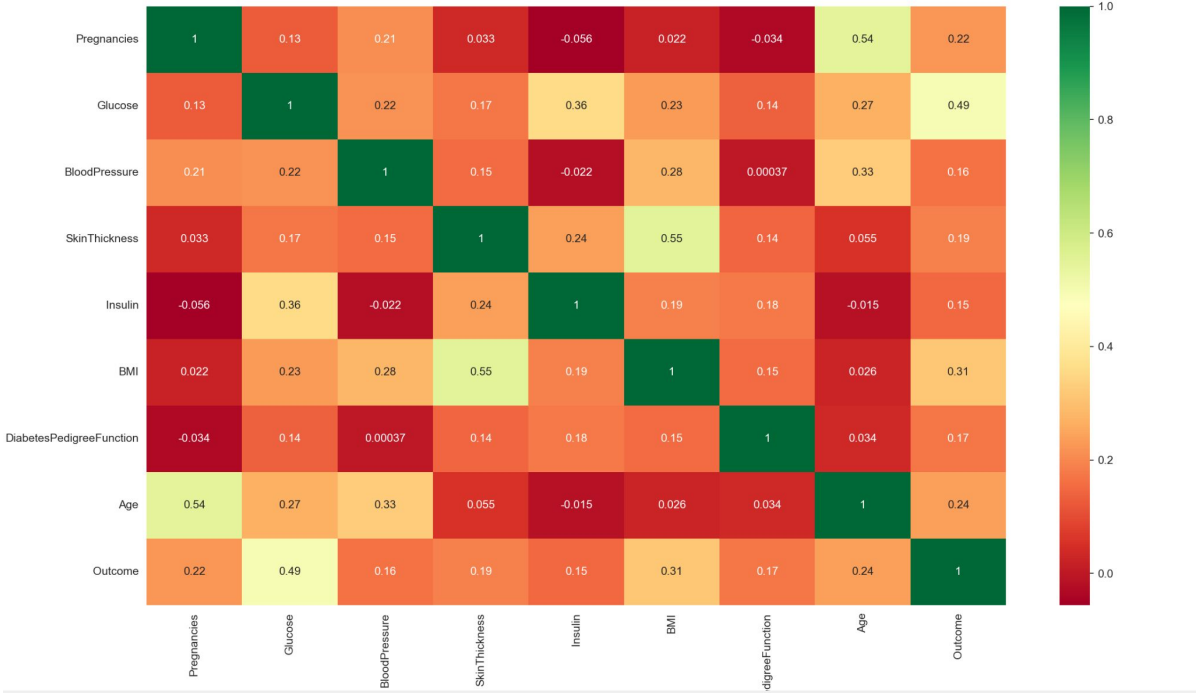


VISUALISATIONS

HEATMAP

-CORRELATION

-GRID MATRIX

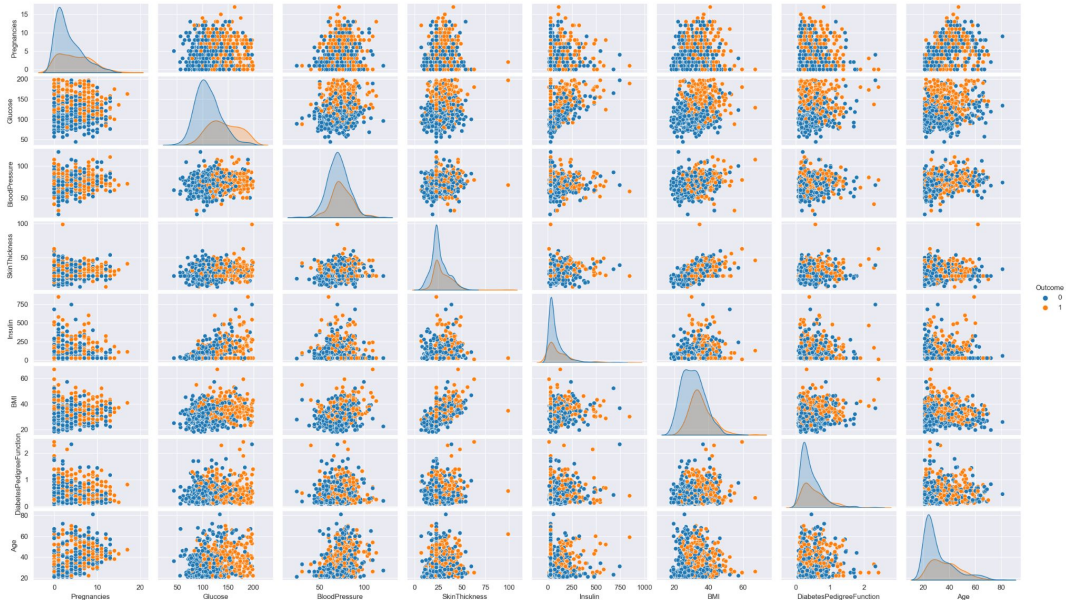


VISUALISATIONS

ALL FEATURES PAIR PLOT

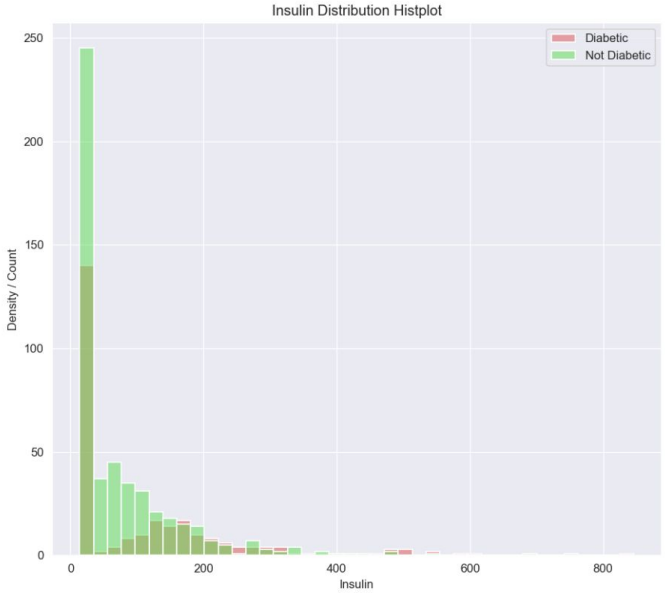
-GLUCOSE
-INSULIN

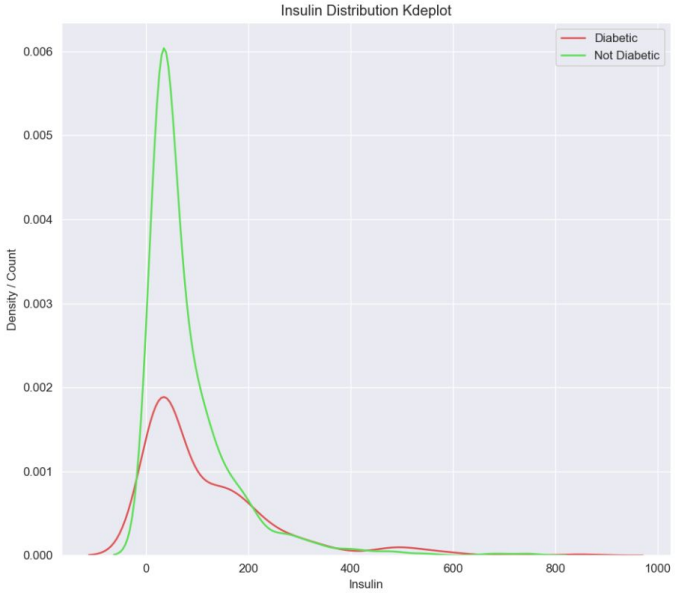
-LOW CORRELATION



VISUALISATIONS

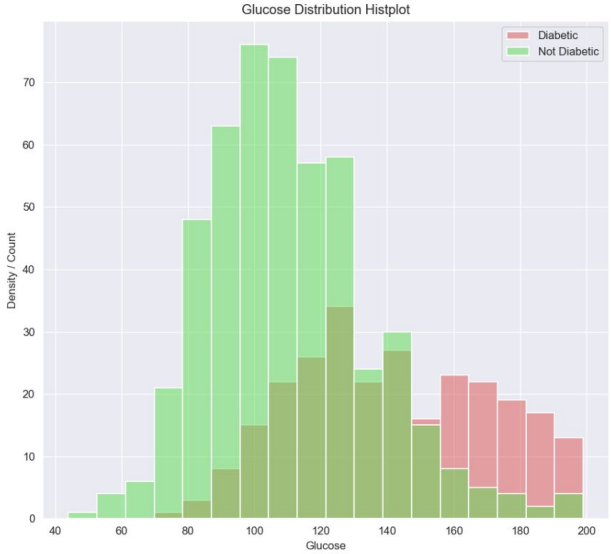
VARIABLE: INSULIN

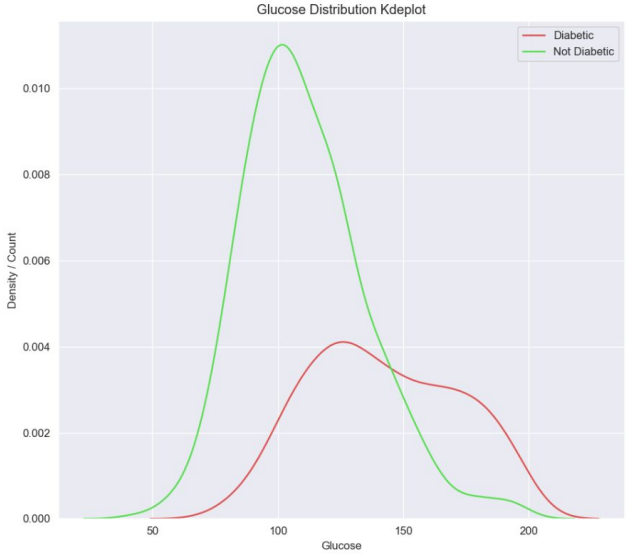




VISUALISATIONS

VARIABLE GLUCOSE



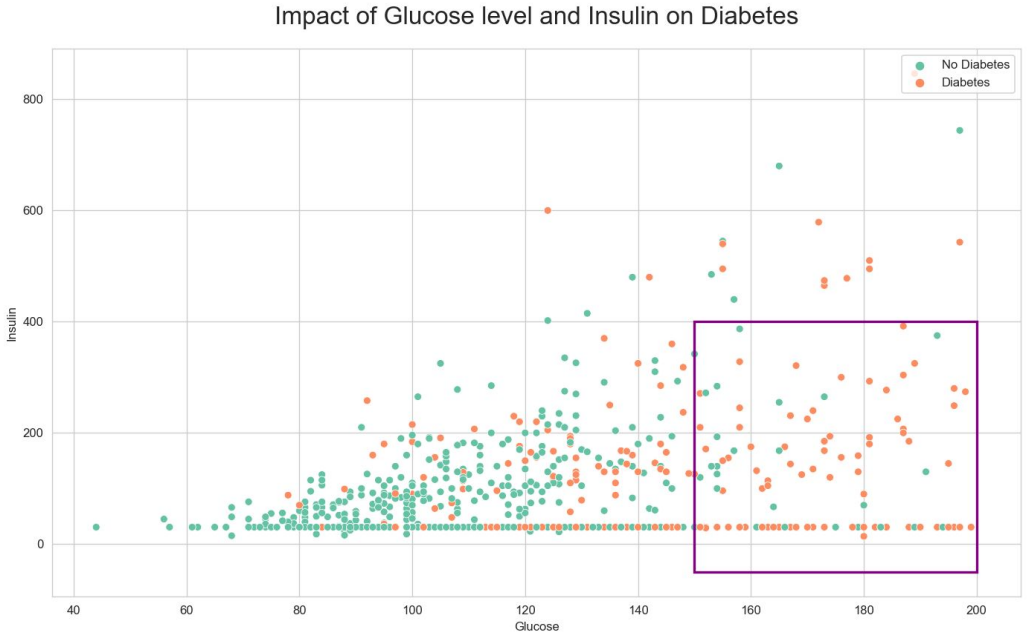


VISUALISATIONS

GLUCOSE AND INSULIN

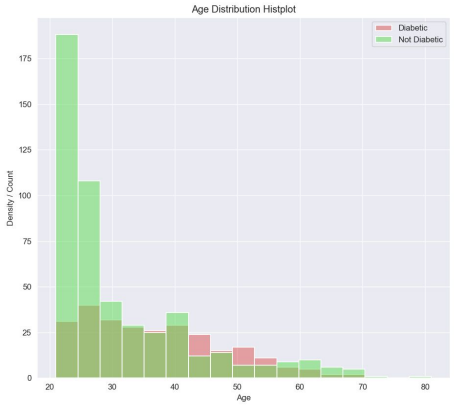
-HIDDEN PATTERN

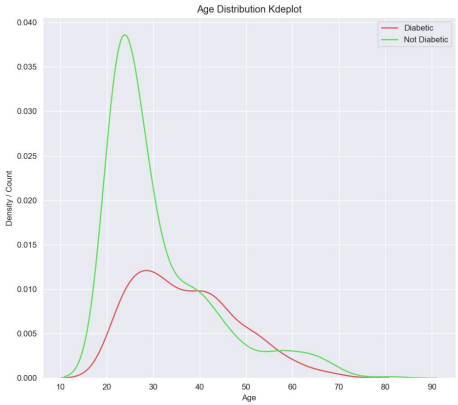
-CORRELATION

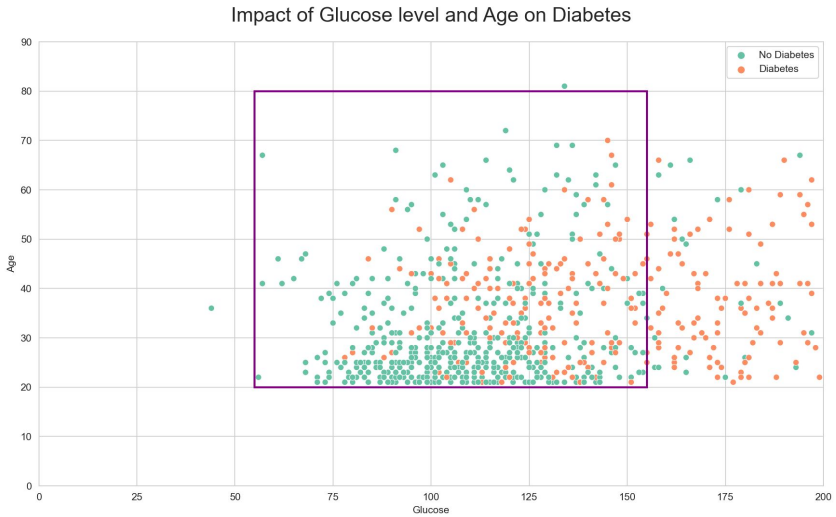


VISUALISATIONS

VARIABLE AGE







MODELLING

MODEL CHOICE:

-RANDOM FOREST

-ABOUT THE MODEL

-TESTING DIFFERENT MODELS

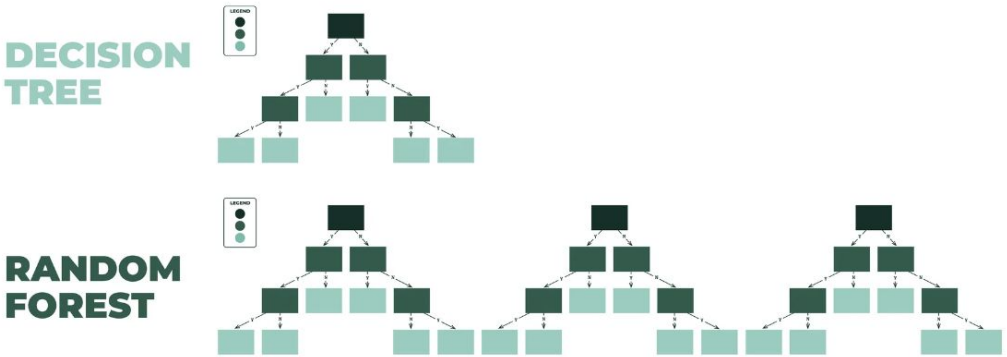
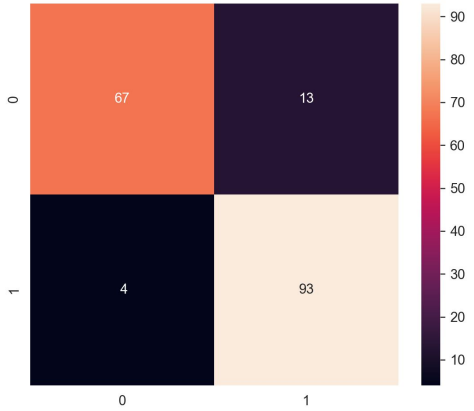


Image retrieved from: https://godatadrive.com/blog/random-forests

VISUALISATIONS

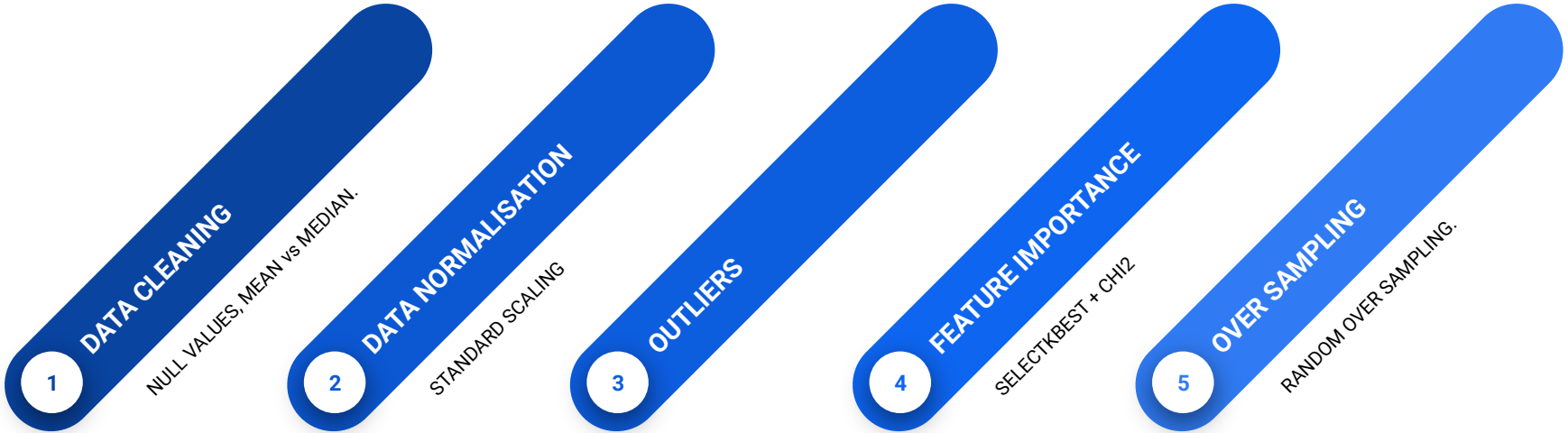
MODEL CHOICE:

```
Train Accuracy of Logistic Regression 77.68729641693811
Accuracy (test) score of Logistic Regression 79.22077922077922
Accuracy (test) score of Logistic Regression 79.22077922077922
Train Accuracy of KNN 79.96742671009773
Accuracy (test) score of KNN 70.77922077922078
Accuracy (test) score of KNN 70.77922077922078
Train Accuracy of naive-bayes 76.0586319218241
Accuracy (test) score of naive-bayes 75.32467532467533
Accuracy (test) score of naive-bayes 75.32467532467533
Train Accuracy of Random Forest 99.8371335504886
Accuracy (test) score of Random Forest 81.81818181818183
Accuracy (test) score of Random Forest 81.81818181818183
```



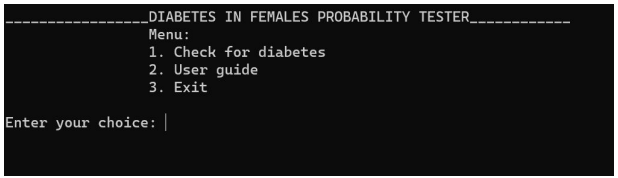

```
0.903954802259887
              precision    recall  f1-score   support

           0       0.94      0.84      0.89        80
           1       0.88      0.96      0.92        97

    accuracy                           0.90       177
   macro avg       0.91      0.90      0.90       177
weighted avg       0.91      0.90      0.90       177
```

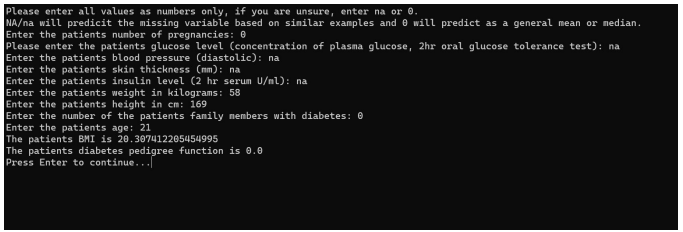
MODEL PREPARATION

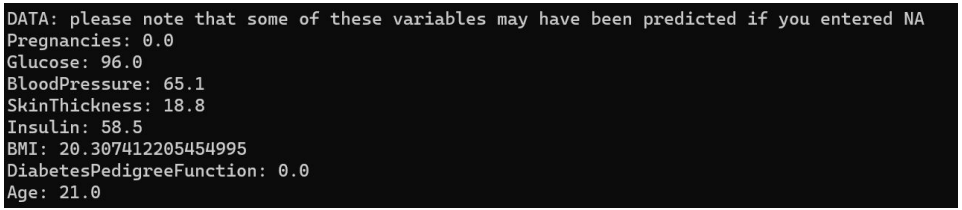


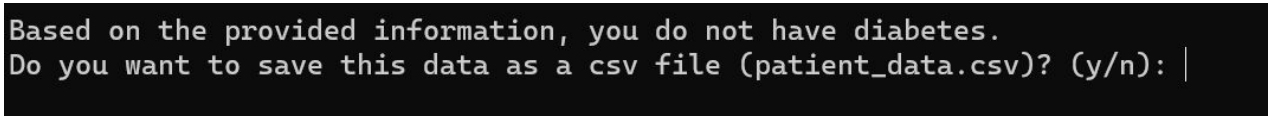
Variable :	No. of zero values
Glucose	5
Blood Pressure	35
Skin Thickness	227
Insulin	384
BMI	11

SOFTWARE IMPLEMENTATION AND USAGE









MODEL CONCLUSION

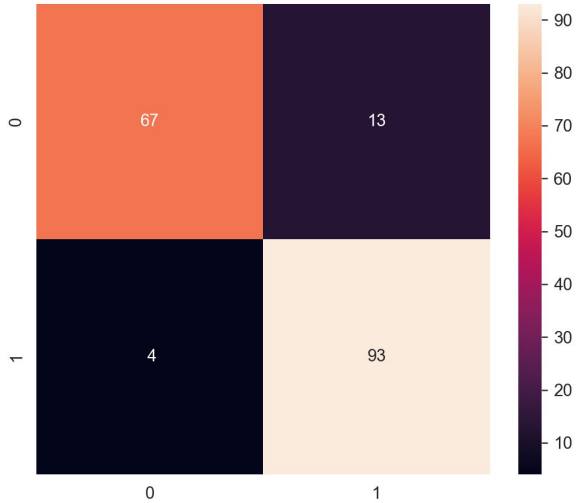
-TRUE VS FALSE POSITIVES AND NEGATIVES

-PRECISION

-RECALL

-F1

-SUPPORT



0.903954802259887

	precision	recall	f1-score	support
0	0.94	0.84	0.89	80
1	0.88	0.96	0.92	97
accuracy			0.90	177
macro avg	0.91	0.90	0.90	177
weighted avg	0.91	0.90	0.90	177

QUESTIONS!

Images retrieved from:
https://www.phistar.in/why-clinical-trials-are-necessary-possible-advantages-and-disadvantages-of-clinical-trial/