HEALTHCARE ANALYTICS

Pima Indian Diabetes

PRESENTED BY :

ABHISHEK KANGALE
PRAVEEN KUMAR
VIVEK JOHN
MANO ARAVIND M

AKASHNIL ROY
AMEYA KANAWADE
MOUMITA GHOSH

Hypelane Clothing Co. 2020

Agenda

- Diabetes - A global health problem
  - Types of diabetes
- Missing values & their significance
- EDA Observations
  - Glucose vs Age
  - Pregnancy vs Age
  - Glucose vs Blood pressure
- Hyperparameter tuned DL model
- Classfication report
- Conclusion

Diabetes - a global health problem

DEFINITION

Diabetes is a disease that occurs when your blood glucose, also called blood sugar, is too high

AMONG TOP 10 CAUSE OF DEATHS

2016 - 1.6 million dead
2017 - 4 milion dead (estimated)

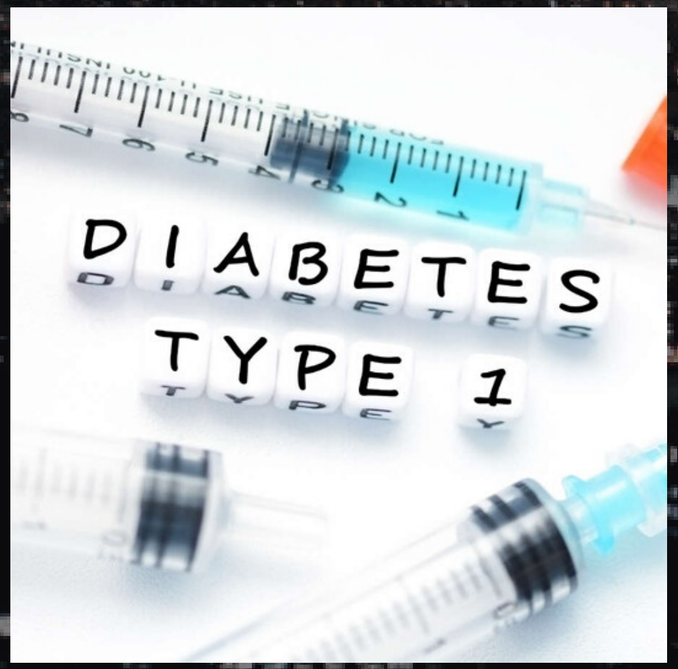
COMMON CONSEQUENCES OF DIABETES

Over time, it can damage the heart, blood vessels, eyes, kidneys and nerves

PREVENTION

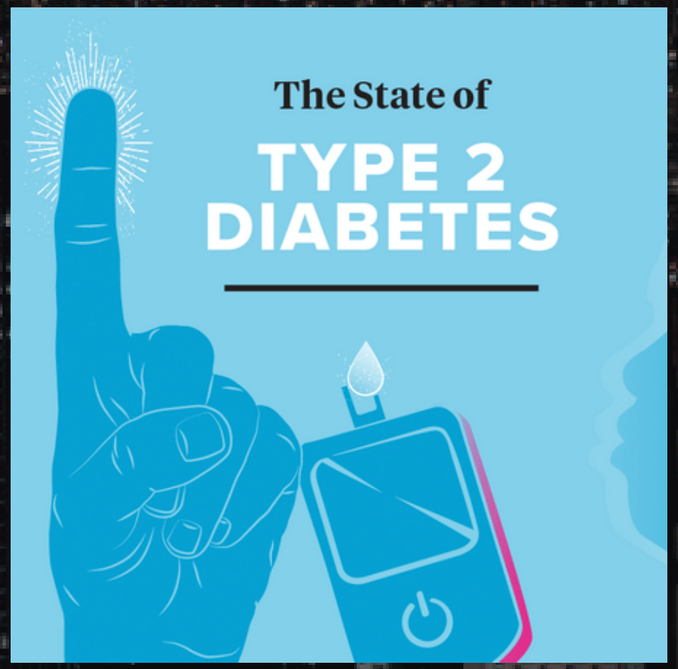
- achieve and maintain healthy body weight
- be physically active – at least 30 minutes of regular exercise daily
- eat a healthy diet, avoiding sugar and saturated fats intake
- avoid tobacco use – smoking increases the risk of diabetes and cardiovascular diseases

TYPES OF DIABETES



TYPE 1

characterized by deficient insulin production and requires daily administration of insulin



TYPE 2

formerly called non-insulin-dependent, results from the body's ineffective use of insulin

GESTATIONAL

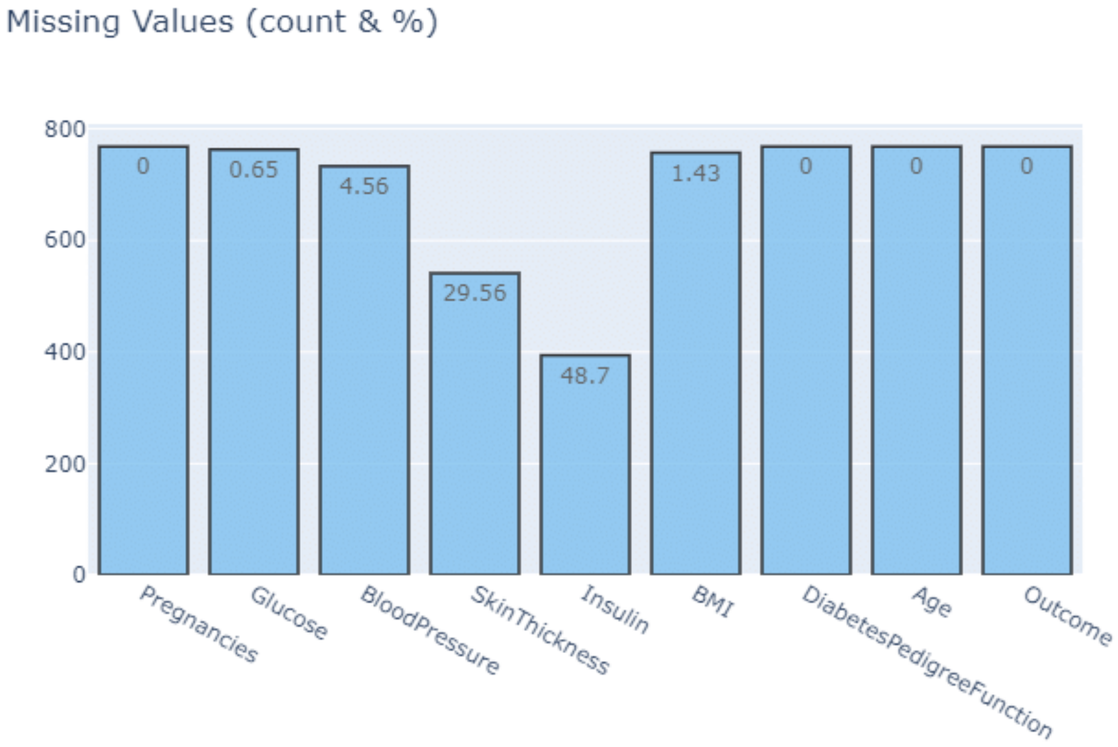
characterised by blood glucose values above normal but below those diagnostic of diabetes, occurring during pregnancy

MISSING VALUES & THEIR SIGNIFICANCE

We saw that some features contain 0 as a value, but it doesn't make sense as neither Insulin nor Skin Thickness should have 0 as a value and hence this indicates a missing value

Medain impute values :

Variable	Diabetic	Healthy
Insulin	169.5	102.5
SkinThickness	32	27
BloodPressure	74.5	70
BMI	34	30
Glucose	140	107



Missing values :

Variable	Percent	Count
Insulin	48.7%	374
SkinThickness	29.56%	227
BloodPressure	4.56%	35
BMI	1.43%	11
Glucose	0.65%	5

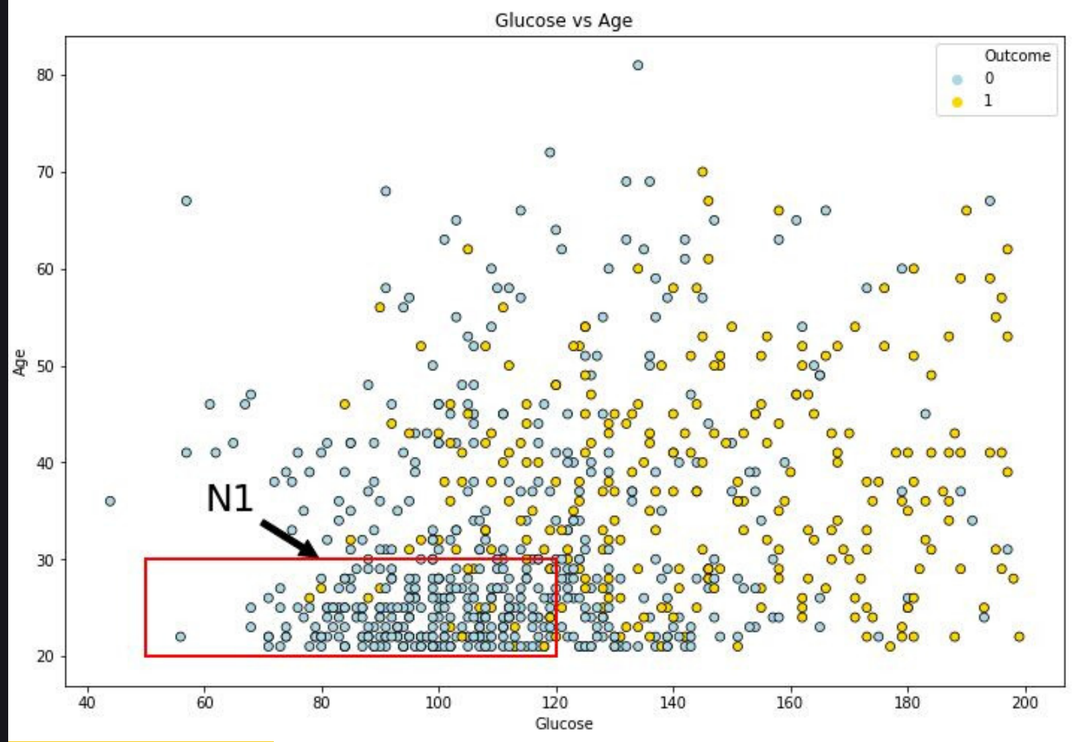
EDA Observations :

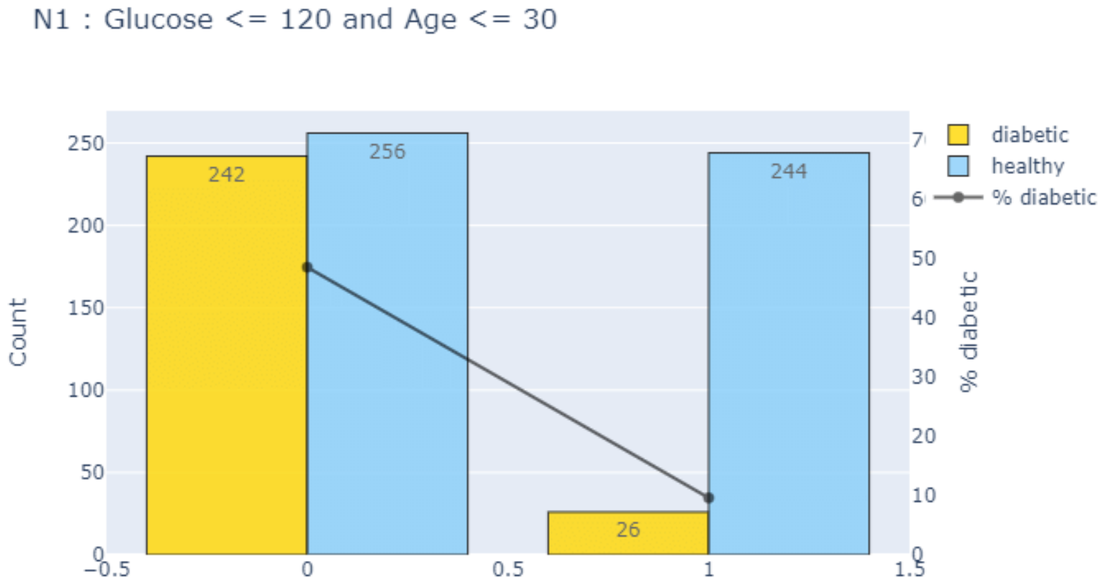
Glucose vs Age

The output of a scatter plot shows us a certain region where individuals are much healthier, which is Glucose level <= 120 & Age <= 30

Percentage of diabetic patients in a population where :

- Glucose level <= 120 & Age <= 30 - 9.62 %
- Rest - 48.59 %





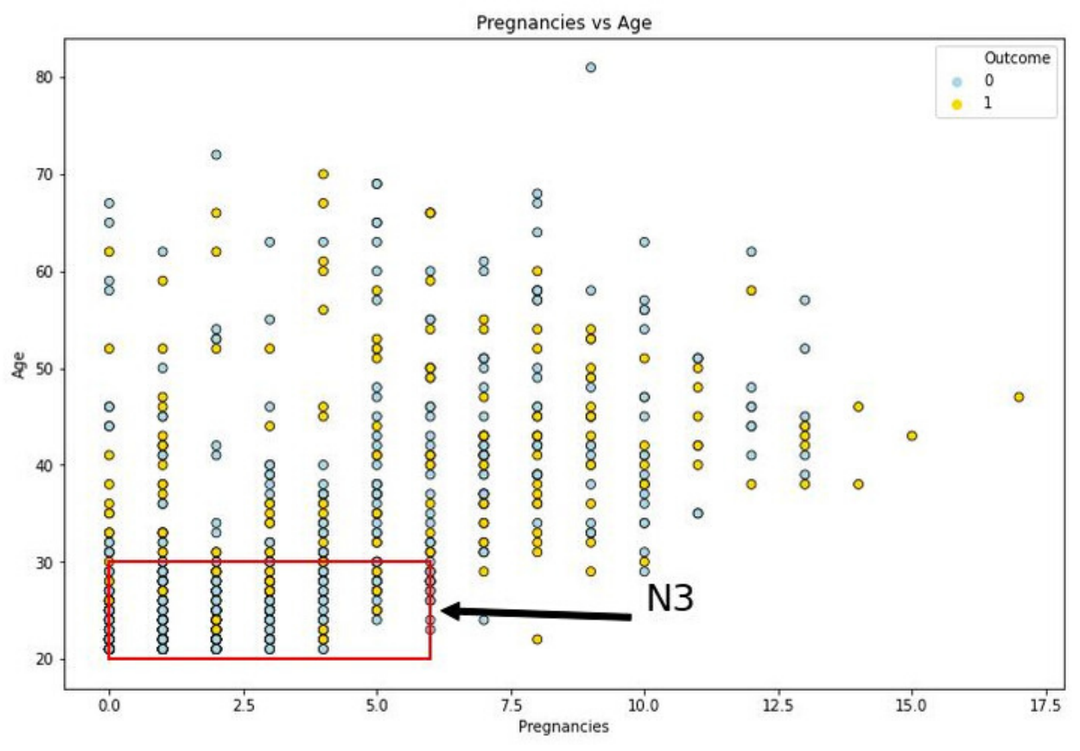
EDA Observations :

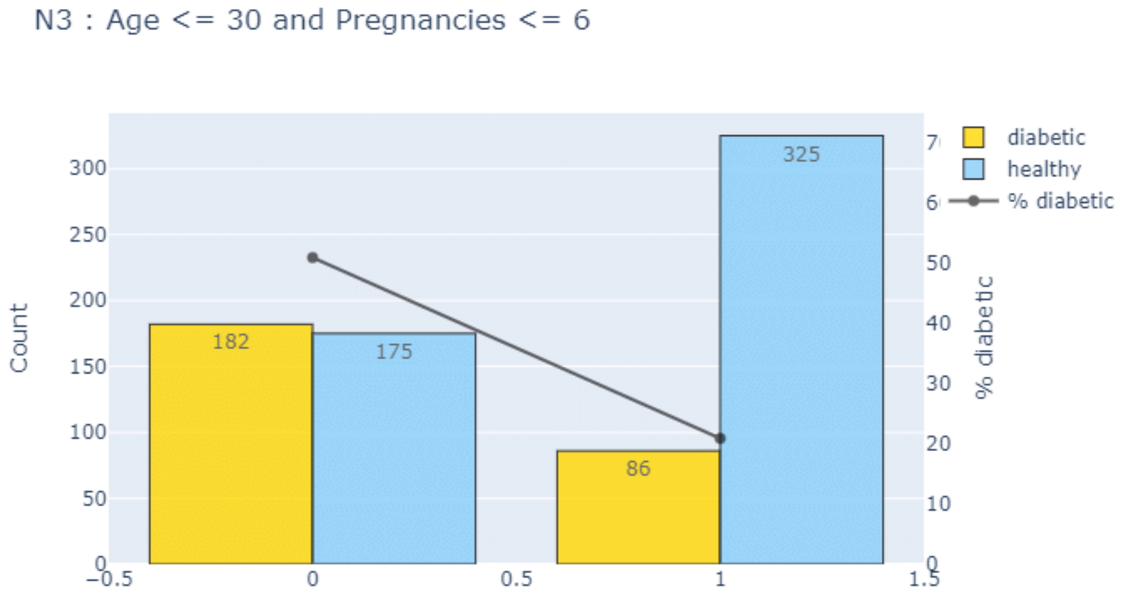
Pregnancy vs Age

The output of a scatter plot shows us a certain region where individuals are much healthier, which is Pregnancy <= 6 & Age <= 30

Percentage of diabetic patients in a population where :

- Pregnancy <= 6 & Age <= 30 - 21 %
- Rest - 51 %





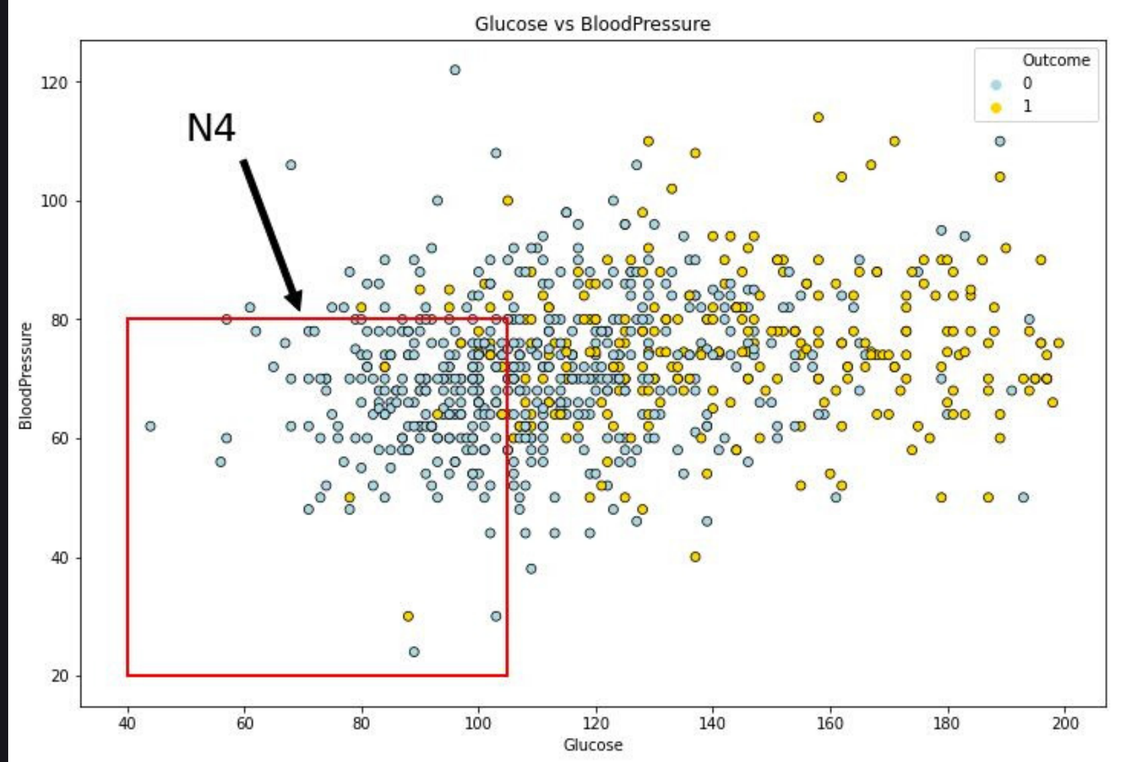
EDA Observations :

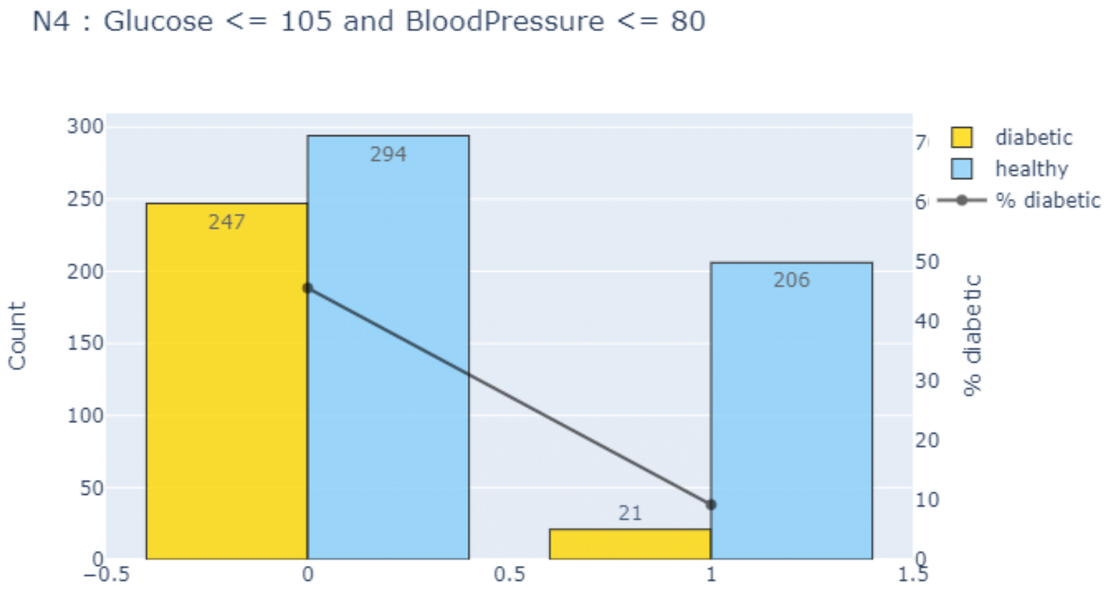
Glucose vs Blood pressure

The output of a scatter plot shows us a certain region where individuals are much healthier, which is Glucose level <= 105 & Blood pressure <= 80

Percentage of diabetic patients in a population where :

- Glucose level <= 105
  & Blood pressure <= 80 - 9.25 %
- Rest - 45.65 %





Hyper parameter tuned DL model

CLASS IMBALANCE

The data is imbalanced. The number of non-diabetic is 268 the number of diabetic patients is 500.
This has been resolved through upsampling.

SCALING

The dataset is standardized using Standard Scaler to prevent any biases creeping in due to the skewness in the values of different features

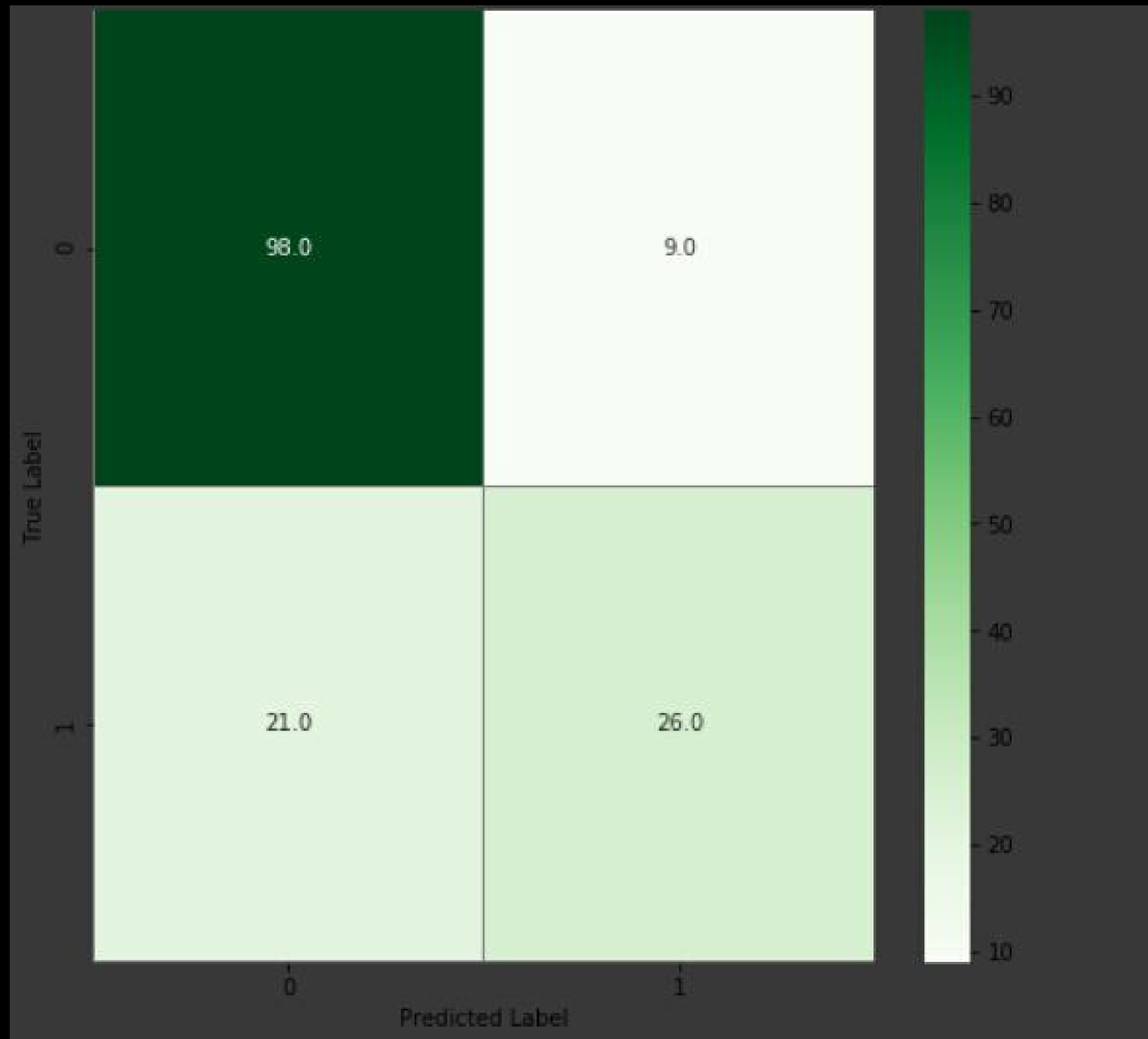
GRID SEARCH

All the best values for hyperparameters - batch size, no. of epochs, learn rate, dropout rate have been identified with the help of Grid search

Classification report

	precision	recall	f1-score	support
0	0.82	0.92	0.87	107
1	0.74	0.55	0.63	47
accuracy			0.81	154
macro avg	0.78	0.73	0.75	154
weighted avg	0.80	0.81	0.80	154

Confusion matrix



Thank you