**For my machine learning I used the health data that was download from Kaggel. The data was cleaned by a team member.**

Table

Description automatically generated

**I then imported and explored the data. From the data I found out that data is not balanced.**

Table

Description automatically generated

**Graphical user interface, text, application, chat or text message

Description automatically generated**

Table

Description automatically generated

Table

Description automatically generated

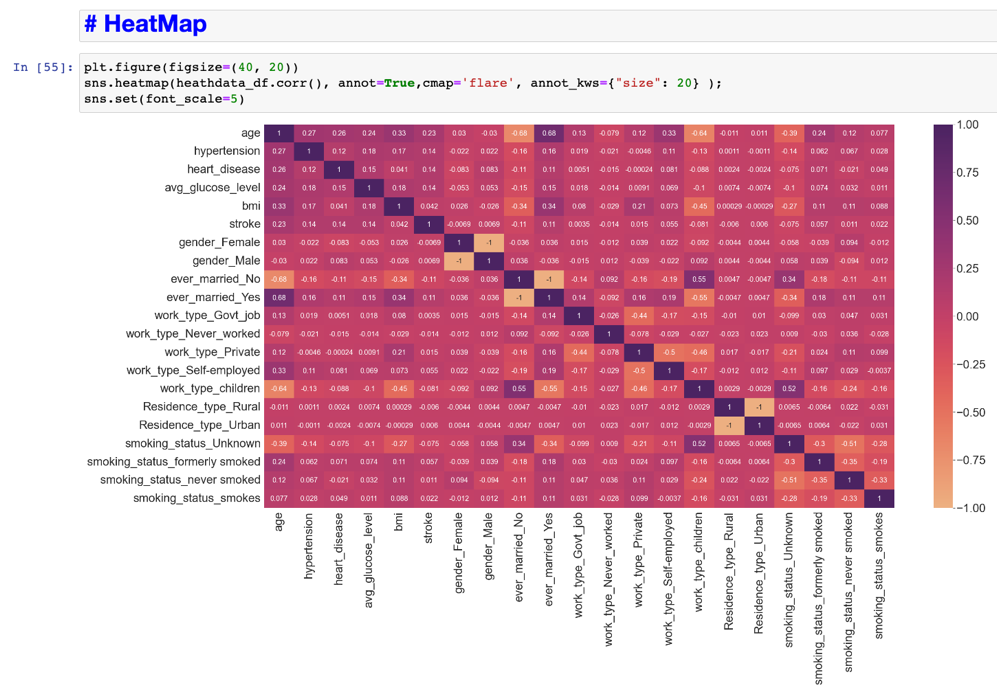
Table

Description automatically generated

**Next I converted the categorical data to numeric using pd.get.dummies.**

Table

Description automatically generated

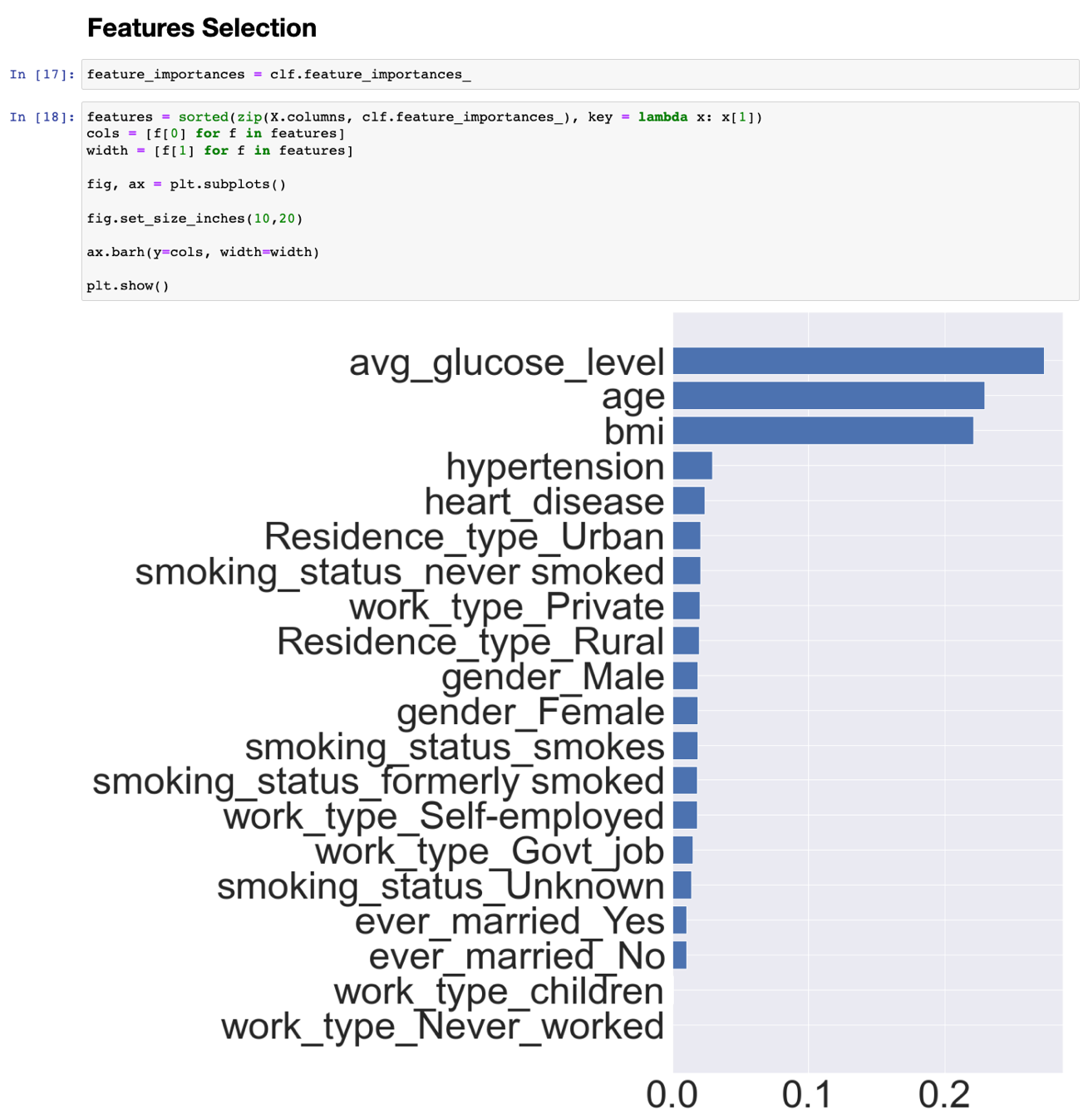


Graphical user interface

Description automatically generated

**In this model random forest classifier was used, and forest classifier with features selection, from this the results showed that the data was over fitting.**





**Then gridsearchCV was used but still gave an unbalanced confusion matrix.**

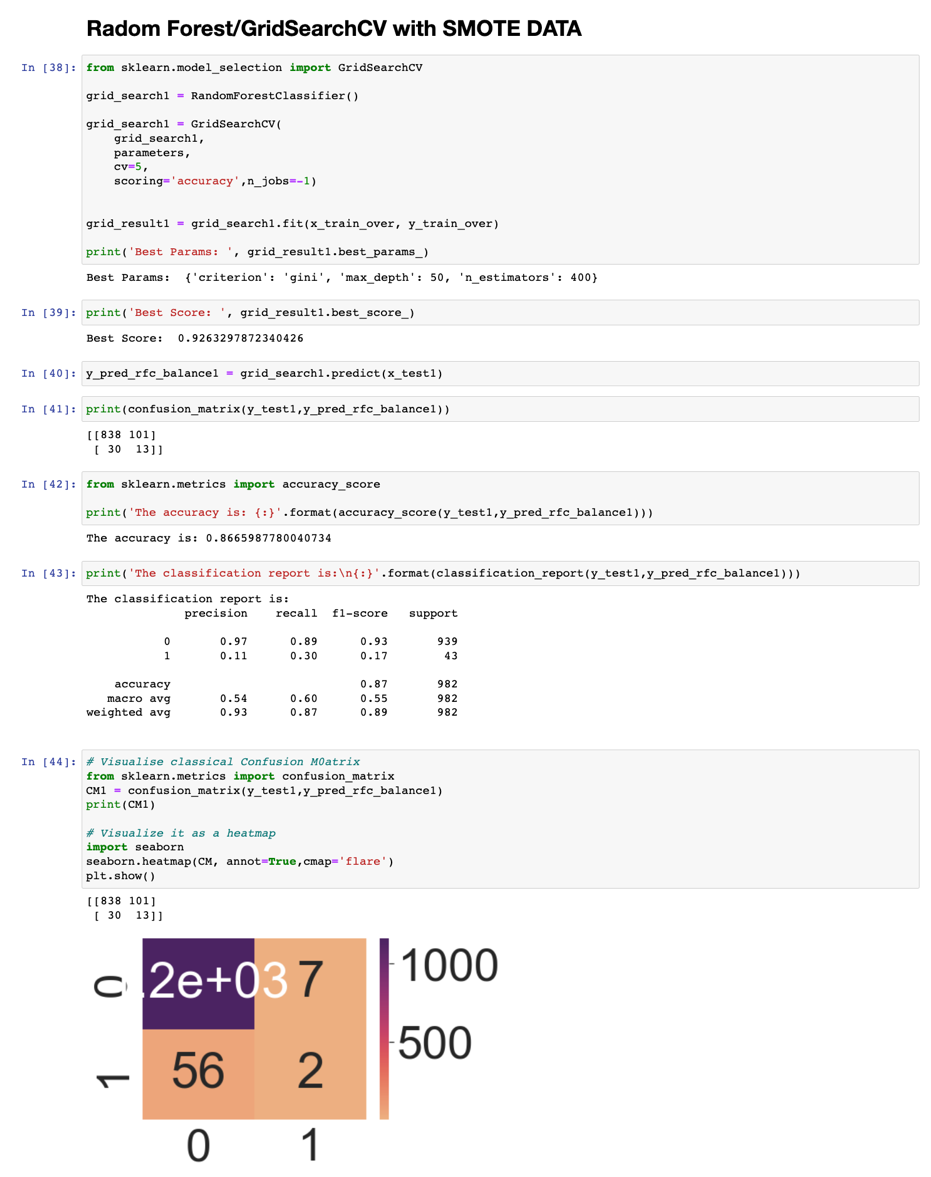
Graphical user interface, application

Description automatically generated

The data was then over sampled using the SMOTE however it produces an unbalanced confusion matrix.

Text

Description automatically generated



The data was undersampling using RandomUnderSampler which produce a better/ balanced confusion matrix.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated