**REPORT ON DIABETES CLASSIFICATION**

Prepared by:

Mats’eliso Thamae

Tumelo Moleleki

Puseletso Ntikoane

**Introduction**

This report seeks to evaluate the performance of our model on classifying whether a person has diabetes or not. If the person has diabetes, the model will indicate that as ‘1’ or ‘0’ if the person is negative.

**List of libraries used**

In this project, a couple of libraries were used including:

Tensorflow: Model configuration

Pandas : Data manipulation

Matplotlib: Data visualization

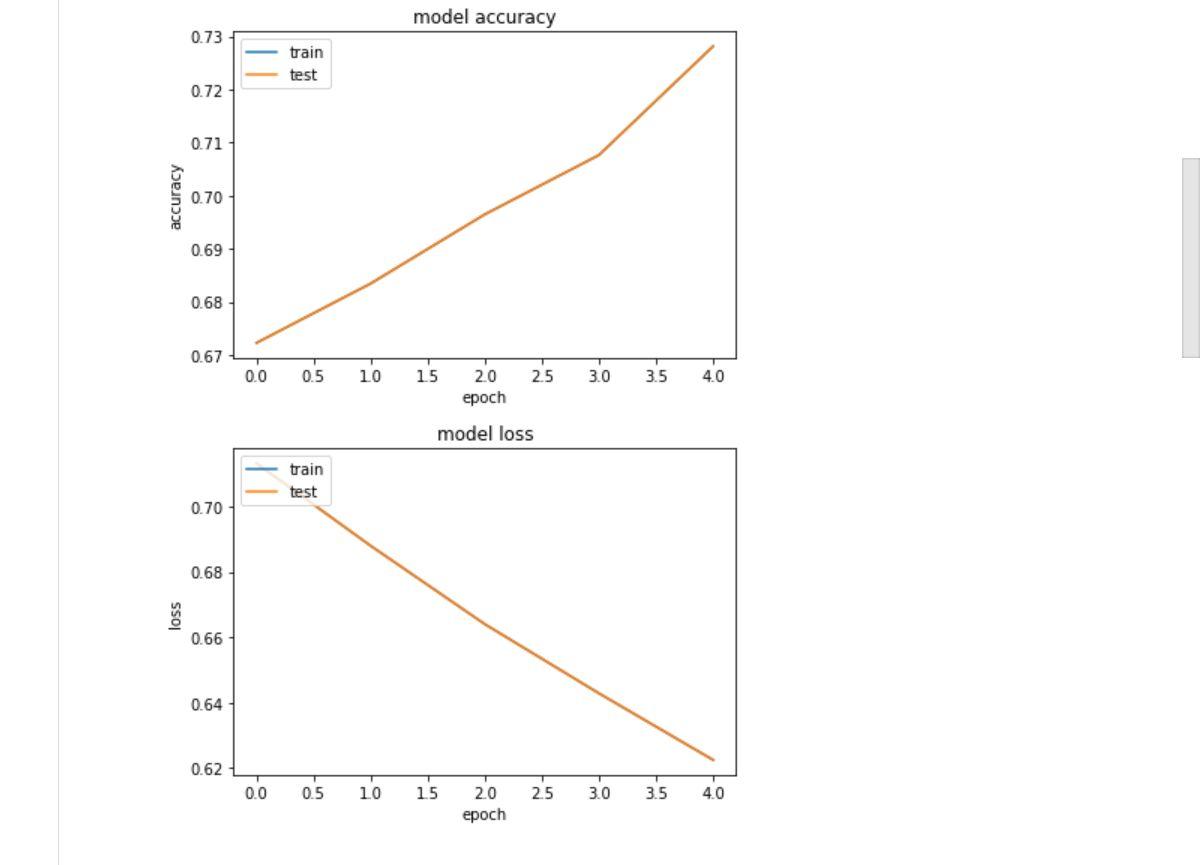
Seaborn: To visualize the data for better cleaning before it could be fed into the model

Ipywidgets : To add interactivity to the code and visualization for the user

**Performance**

80% accuracy is not bad for the model that iterates over the data only 10 times(This is also called number of epochs in machine learning lingo).However with over 2000 epochs the efficiency of the model starts hitting the plateau which is bad for fitting because the model has reached its peak performance and it is no longer necessary to continue training it. And with a little over 10000 epochs, the performance of the model seems to be steadily declining, the performance gap between training and testing is huge. There seems to be an issue of over-fitting.

**Model training after 5 epochs**



**How the model could be improved**

To discourage an issue of overfitting, more data could be used so that there is a relatively huge amount of data used for training the model. This means that deep neural networks are more efficient on large amounts of data. Moreover, overfitting occurs when the model goes over the same data too many times until it memorizes the patterns and this is not desirable because the model will perform badly when tested.

**Keys takeaways**

* A huge percentage of accuracy on the training dataset is not always the best indication of the performance of the model as a whole.
* Data cleaning is the stage where the most time should be spent because it affects the performance of the model by an order of a magnitude.