

Report Of Model And Preprocessing And Postprocessing Operations Performed

Preprocessing operations

- Did feature scaling using standard scalar.
Reason: Since the values might be large, so in that case more computational power and more time will be required by the neural network to train, so it is better to scale the values to lower values.
- Created a custom function which is when passed in random search, it gives random values of :
 1. number of hidden layers in the neural network
 2. number of neurons in each hidden layer
 3. learning rate
- The custom function is passed in random search which creates different models according to data from custom function.
- The random search applies training data to different models and finds the best 10 models.
- Then took parameters of best model and again created neural network and trained it for 100 epochs.
Note: Since the best model from random search had 18 hidden layers so I manually created neural network, so that I could train model for more epochs to get better results. I could have directly took best model from random search but it was trained on less epochs and increasing number of epochs in random search will increase training time of random search very much as it runs on many model combinations(in my case I took 30).

About the model

- The data is trained using Multilayer Perceptron (Artificial Neural network). I used this approach as RNN and LSTMs works well in related data in which the next data is related and dependent on previous data and CNN works well for image data. And ANN seemed suitable for the data.
- Since the data is not very large, the model is trained for 100 epochs.

- Found number of hidden layers, number of neurons in each layer and learning rate using keras random search which will work well on the data.
- Model have 18 hidden layers, with activation function RELU as it works well on hidden layers and kernel initializer 'he uniform'.
- Model have 1 output layer with activation function sigmoid as problem is of classification and kernel initializer 'glorot uniform'.
- The loss function used binary crossentropy as it works well for classification And optimizer is adam.

Postprocessing operations

- After the model is trained, the model is tested on testing data and validation accuracy, loss, f1 score, precision and recall are found.
- Also the values of accuracy and loss during training and validation of model are plotted.