**Cleaning and Transforming the Dataset**

* [The dataset](https://www.kaggle.com/datasets/shelvigarg/wine-quality-dataset?resource=download) has 6497 samples and 12 features.
* I had to do minimal cleaning to the dataset. There were some samples with blank values, which were dropped prior to the training of the model.
* In terms of the classifying feature, the dataset is imbalanced. There are approximately 4100 good wines and 2400 bad wines. Therefore, we will use multiple metrics to evaluate the model.
* The dataset is not set up for binary classification. Each wine has been rated on a scale of 3 to 9. We will classify ratings/qualities >= 6 as “good” wine (else, wine is “bad”).
* The only non-numerical feature is “type”, which is either “red” or “white”. We will drop this feature after transforming it to 2 new features - “isWhite” and “isRed” that hold values of either 0 or 1.

**Optimizing the Model**

At first, I tried x, y, and z and was not been able to achieve over 80% accuracy

I tried different **optimizers** in SGD and Adam. The only glaring difference between the two is that Adam seems to be slightly faster. Both optimizers result in nearly identical metrics.

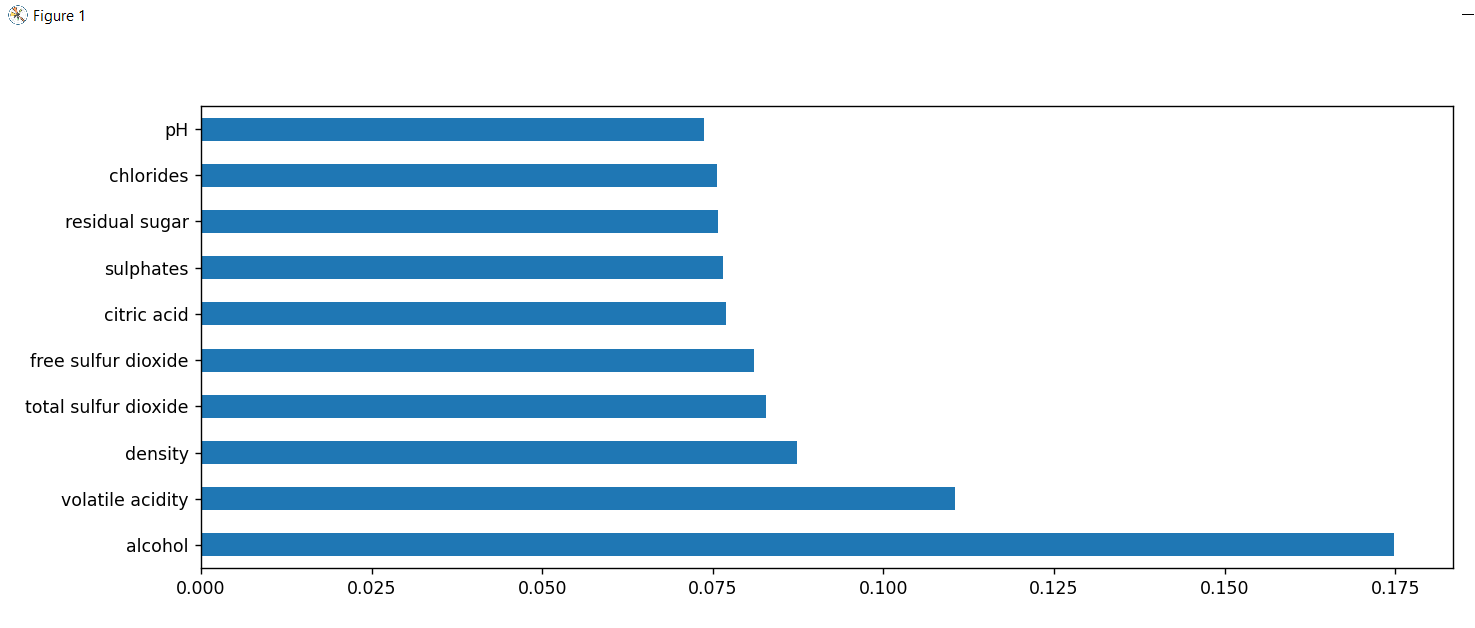
Doubling the number of **epochs** does not seem to have an impact on results.

Then, I saw a reasonably large jump in metrics (4% boost to accuracy) when the **number of layers** increased (from 4 to 7). The first layer was given 50 **nodes** even though there are 13 features in the dataset.

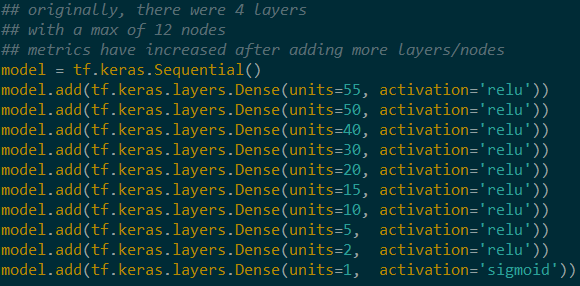
**Feature Ranking**

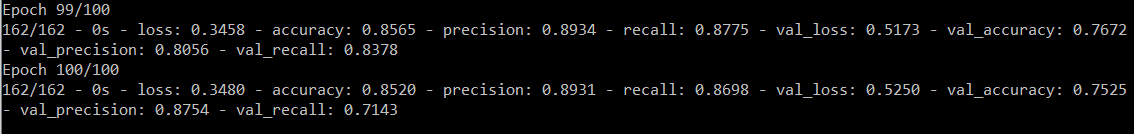
**Top 10 most important features**

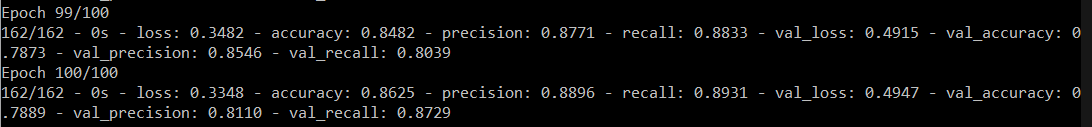
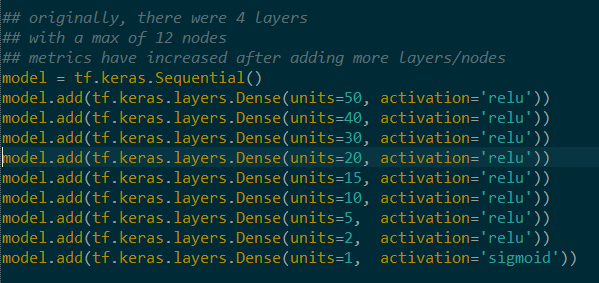
**Visualized with matplotlib**

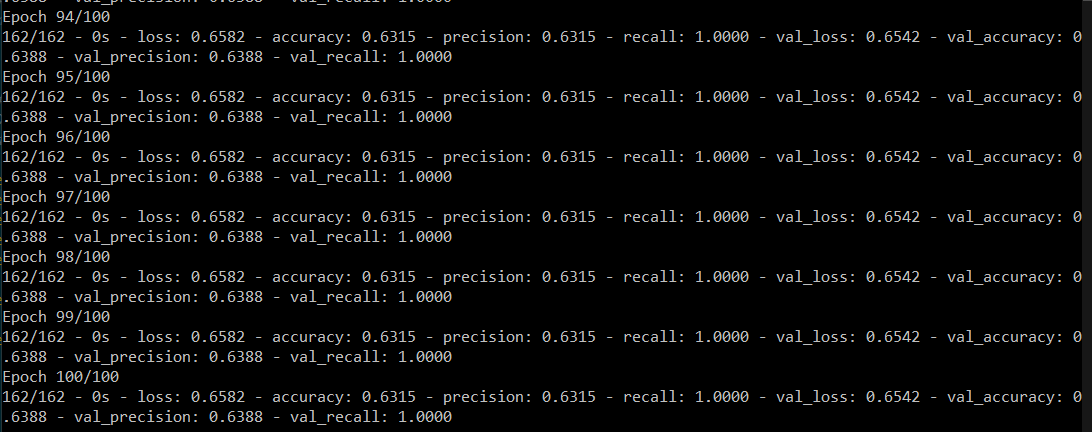
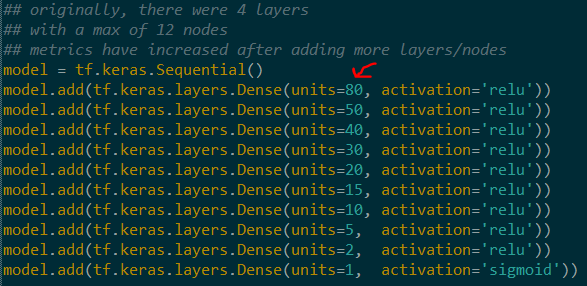
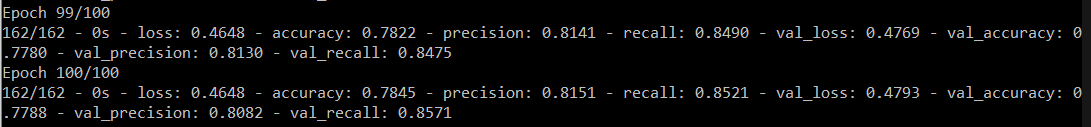
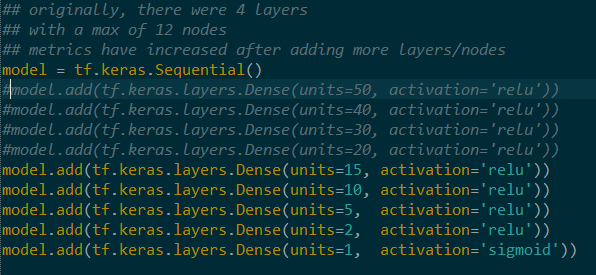


# Tests

****

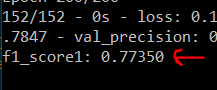
****

****

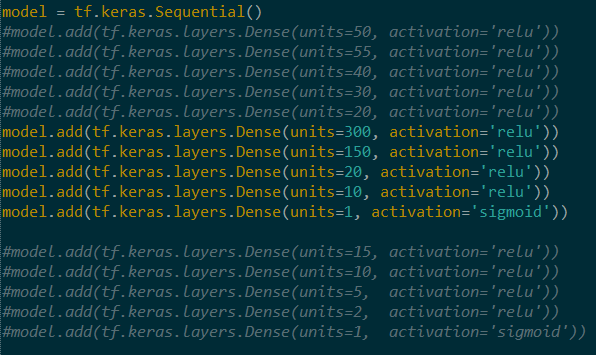
****

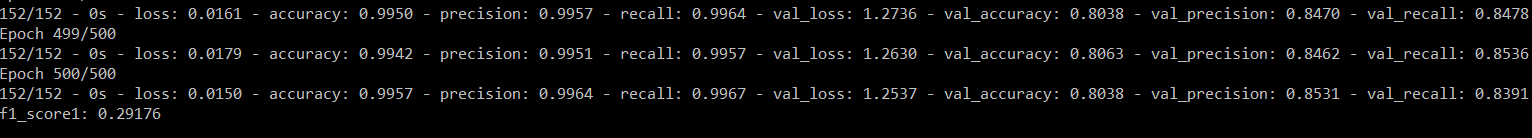
Adding a layer of 80 neurons had a significantly negative impact on metrics

**F1 Score**



Below, I have completely changed the architecture.





Decent training metrics are through the roof, but at the cost of the validation metrics. This was just to experiment. The architecture has been reverted to the original configuration.