Kyle Montera

MISY 408

12/13/19

Final Project

**Predicting Poisonous Mushrooms using Neural Networks**

**Executive Analysis**

Using a neural network, the probability of a mushroom being poisonous could be accurately predicted. Using the most effective model, the accuracy of predicting the training set of data was 98.95%, and the testing data accuracy was 99.26%.

* Reducing the complexity of the model made the model more effective
* A random mushroom has a strong chance of being poisonous, and should be avoided

**Preliminary Model Analysis**

The preliminary model consists of a neural network with varying number of input layers, neurons per layer, and number of epochs run. The data provided consisted of categorical variables describing attributes of different mushrooms. In order to select appropriate features to use, the ‘crosstab’ function was used to examine mushroom attributes, and the counts of which were either poisonous or not. (Figures 1-4) The first variable “class” was encoded using “get\_dummies(drop\_first = True)” in order to create a binary variable. This variable will signify whether a mushroom is poisonous or not. (Figure 5) The features selected avoided having a high correlation to the poisonous class. This was done in order to continue feeding the network new data, and prevent the model from memorizing data. The features selected include a ‘Silky’ color above ring, ‘Narrow’ gills, a ‘Wooded’ habitat, an ‘Evanescent’ ring, ‘Solitary’ populations, a ‘Bulbous’ stalk root, and an ‘Enlarging’ stalk shape. These features were then encoded into binary variables using get\_dummies(drop\_first = False). (Figures 6-7) The data was split into a train set containing 80% of the data, and a test set containing 20% of the data (Figure 8)

**Model Architectures**

The architecture for the neural networks tested can be found in the table provided. (Figure 9) Relu was used as the activation function. Relu gives an output if the prediction is positive, and 0 otherwise. Sigmoid was used as the output function as it is effective when a network outputs a probability. Adam was the used optimizer. Different numbers of epochs were tested for each model in order to minimize overfit and balance the train/test model accuracy. The accuracy and MSE for Model 3 over the course of these epochs can be found in (Figures 10-11)

**Results**

The results can be found in Figure 9. Model 3 proved to be the most effective. (Figures 12-13) The model was highly accurate in predicting whether a mushroom is poisonous or not, and finding the probability of being so. Additionally, when a mushroom is edible, the model does not predict the mushroom being poisonous. Additionally, the model was the most parsimonious of the three. While the models output similar accuracy scores, Model 3 contained the least complexity. The third model contained the smallest number of hidden layers, the smallest number of neurons, and the fewest epochs needed to accurately make predictions. The confusion matrix of the testing set for the third model presents the predictions made by the model from the testing set. (Figure 14) The top row indicates classifications as non-poisonous, the left most value being the number of correct classifications. The bottom row indicates classifications as poisonous, the rightmost value being the number of correct classifications. Overall, the model predicted the testing set with an accuracy of 99.26%. Three predictions of hypothetical mushrooms were made to test the model. The first test predicted a mushroom that was in a wooded environment, and had a bulbous stalk root shape. (Figure 15) The model predicted that there would be a 99% chance this mushroom was poisonous. The second test predicted a mushroom that has a silky stalk surface above its ring. The model predicted the mushroom not to be poisonous, with a 12.67% chance of being so. (Figure 16) For the third prediction, a mushroom that had broad gills, a solitary population, and a bulbous stalk root was tested. The model predicted that based on the learning it had done, this mushroom was poisonous, and had a 98.6% chance of being so. (Figure 17)

**Appendix**

Figure 1

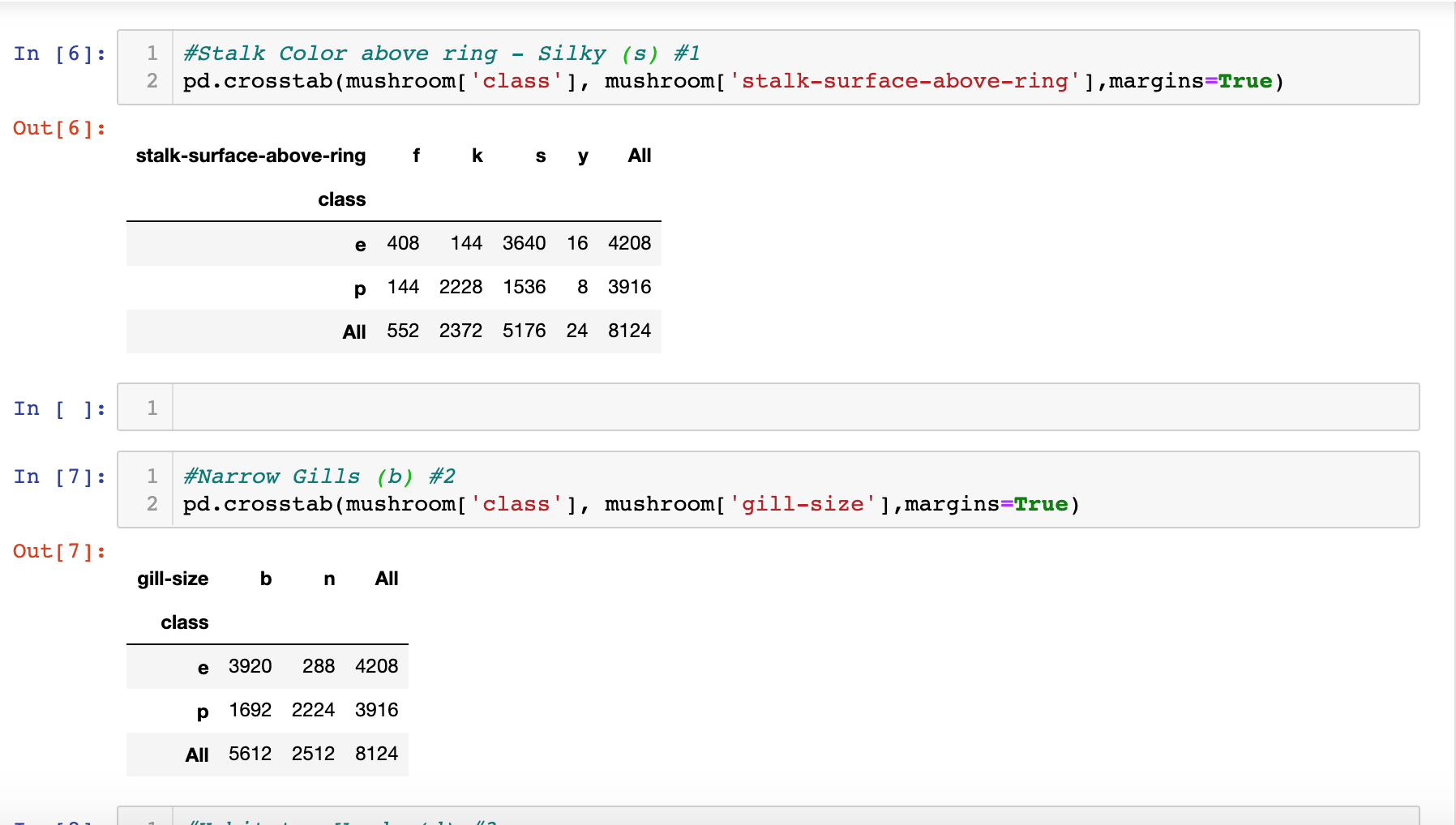
****

Figure 2

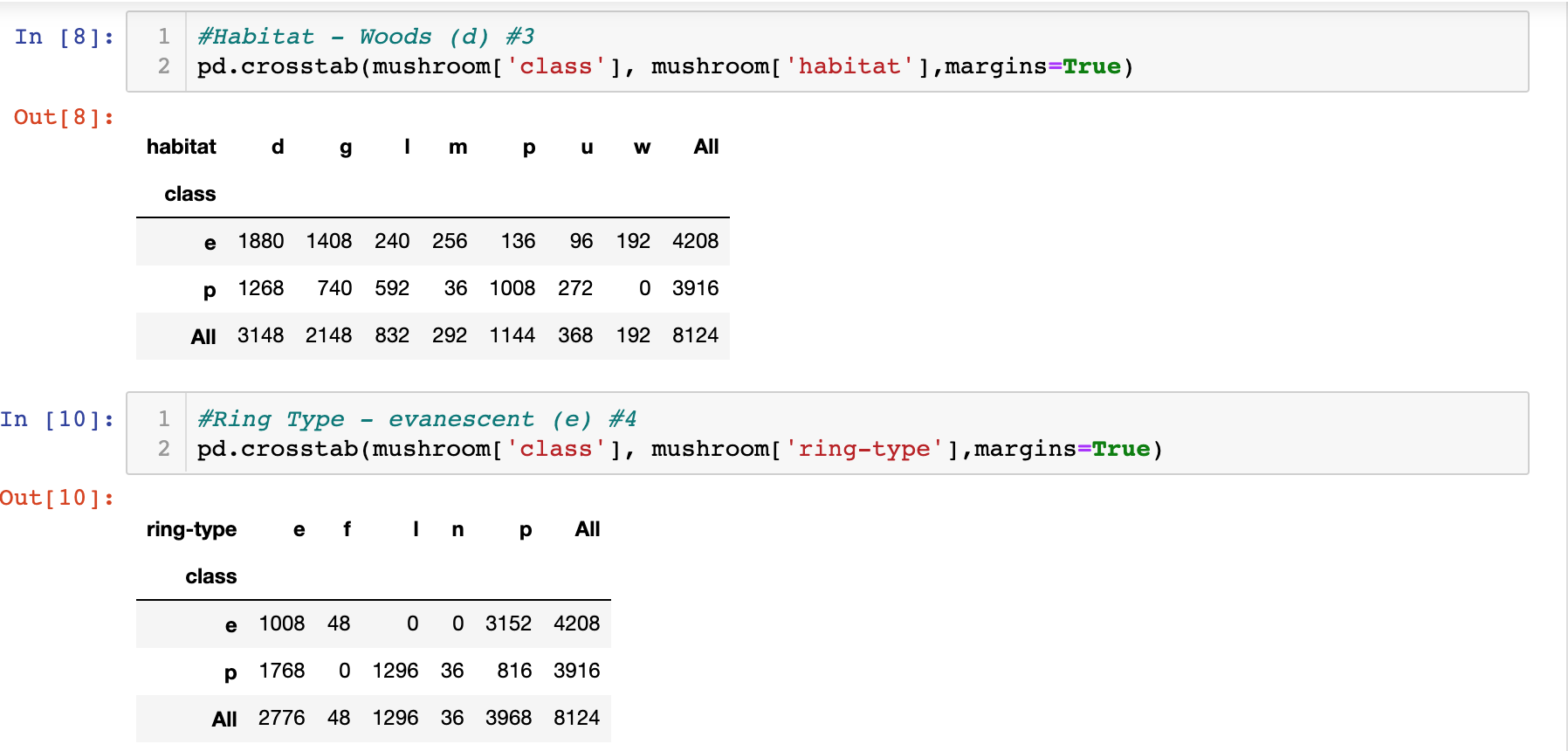


Figure 3

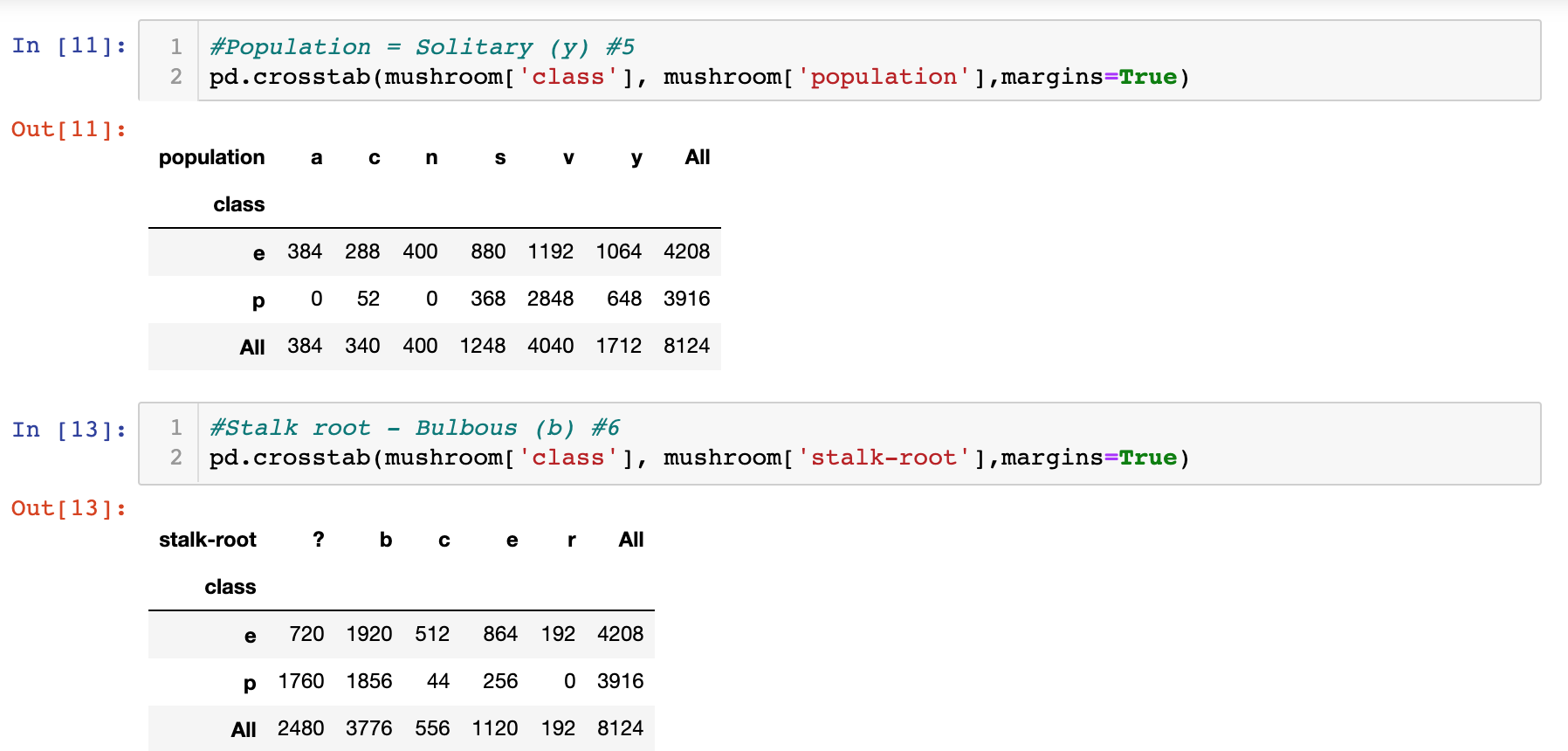


Figure 4



Figure 5

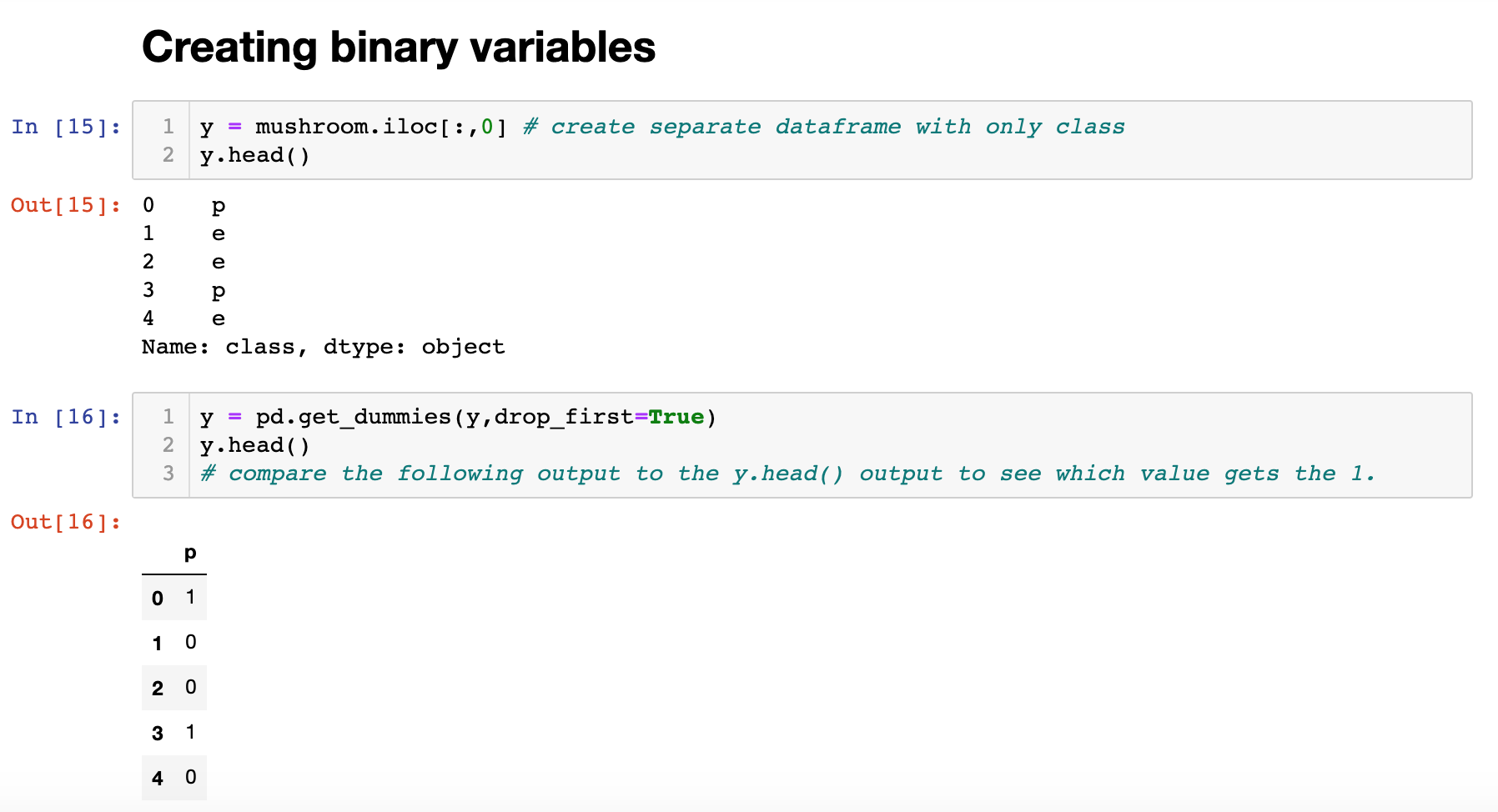


Figure 6



Figure 7

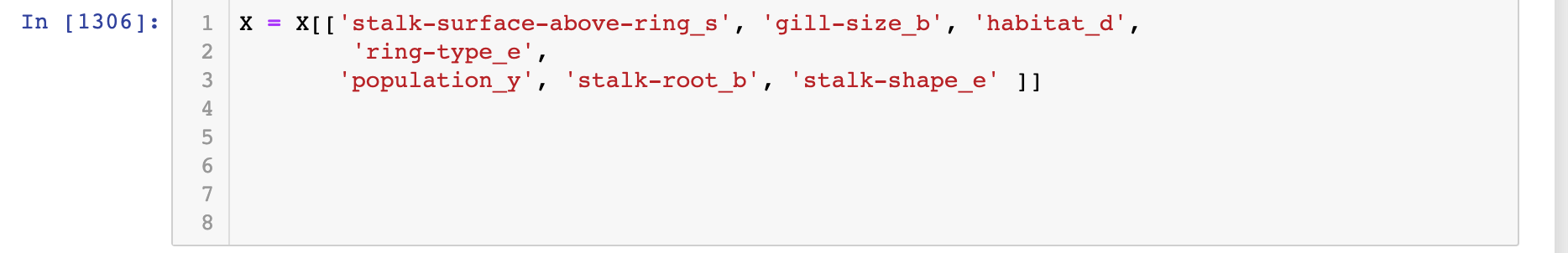


Figure 8

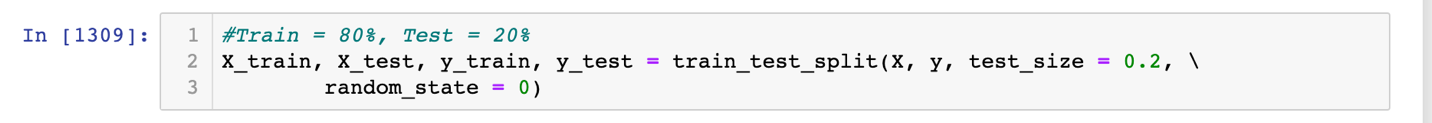


Figure 9

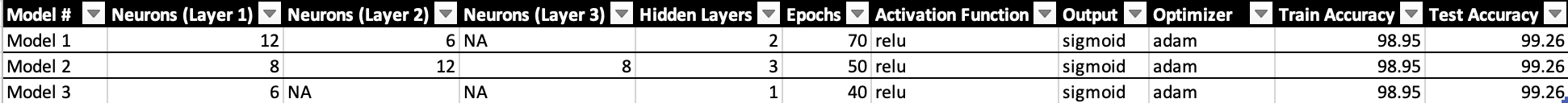


Figure 10

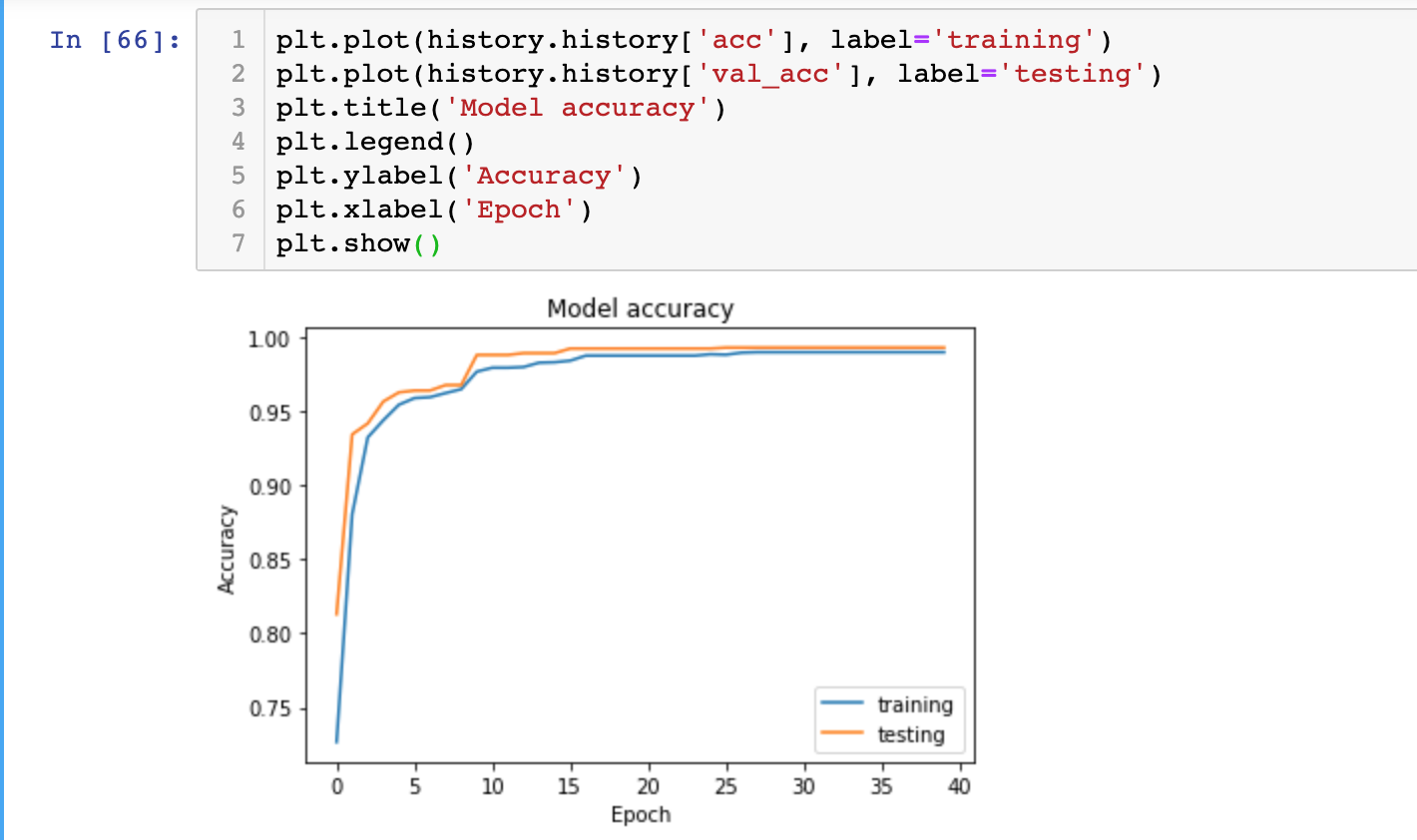


Figure 11

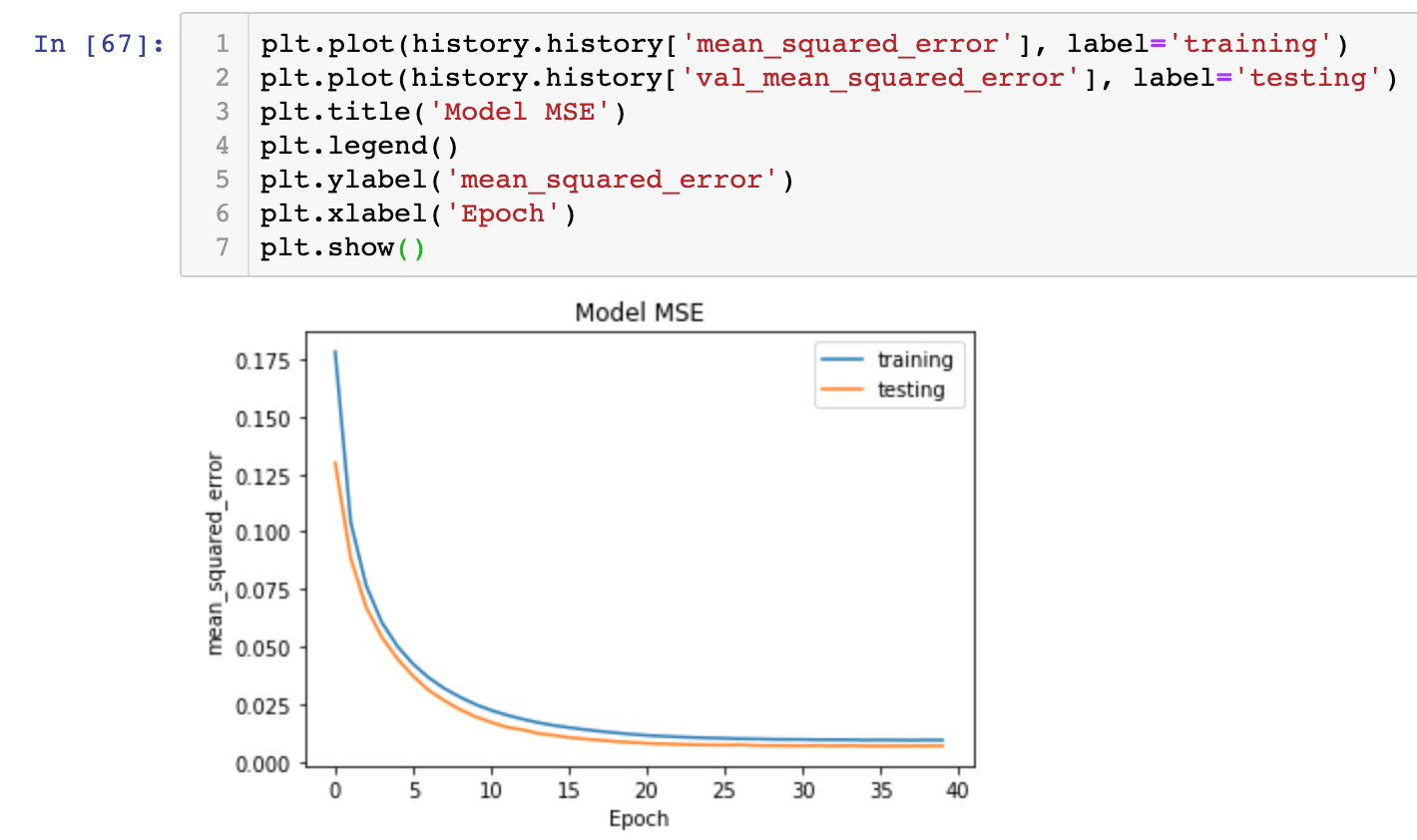


Figure 12 (Train data)

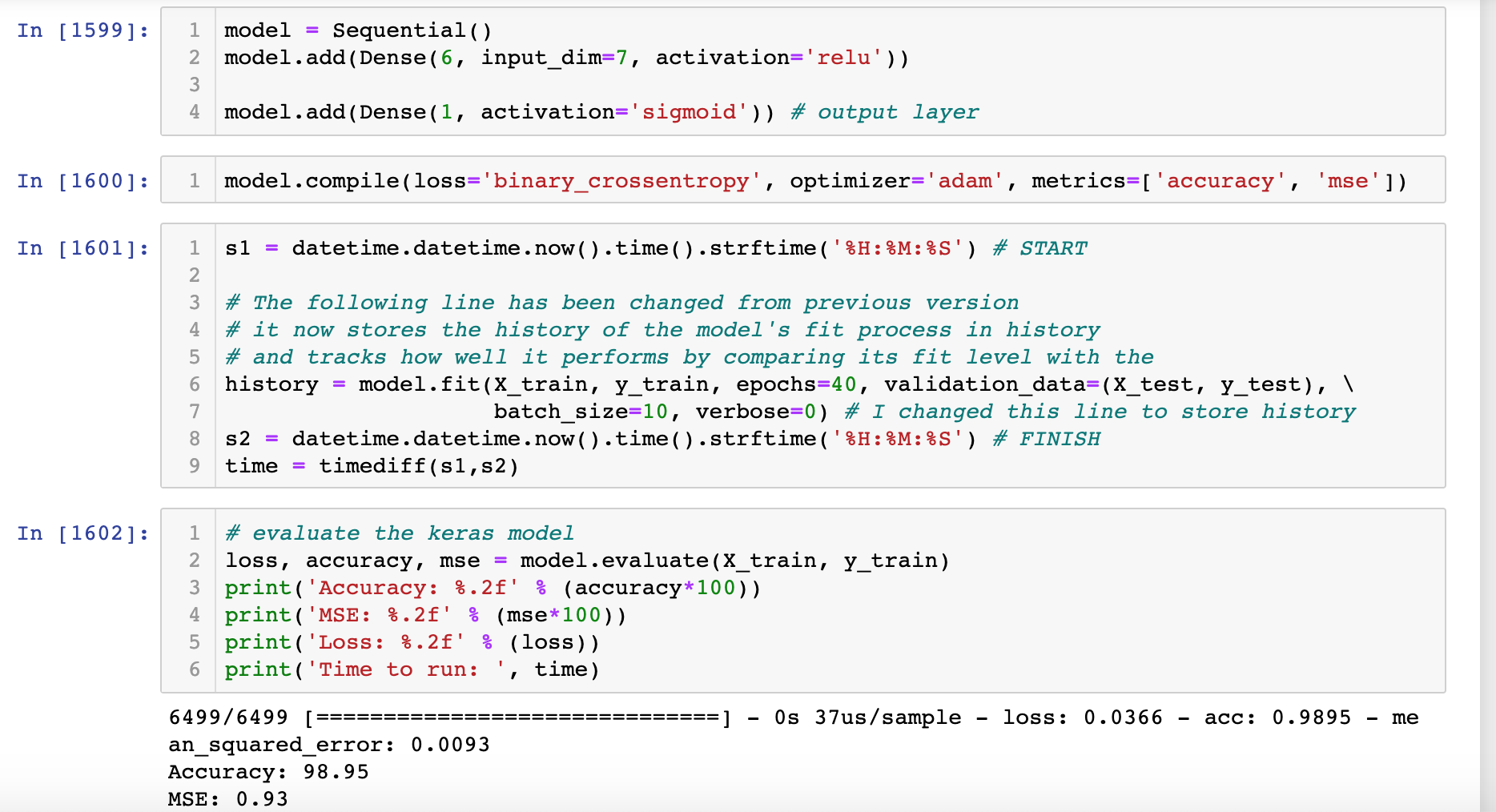


Figure 13

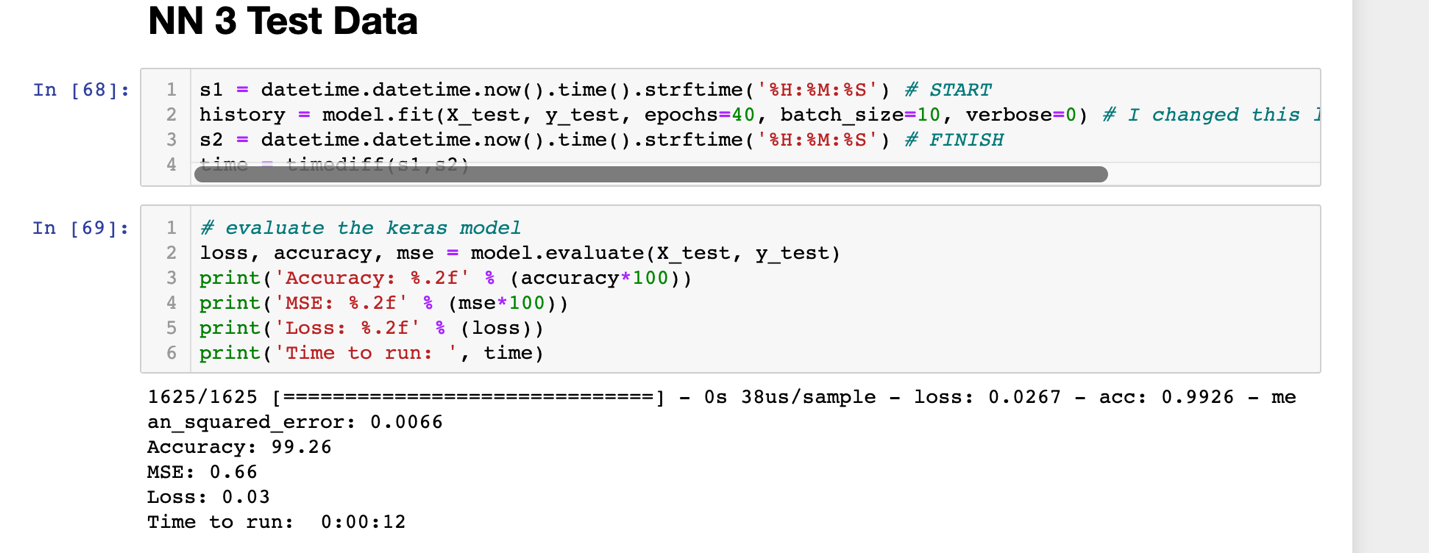


Figure 14



Figure 15

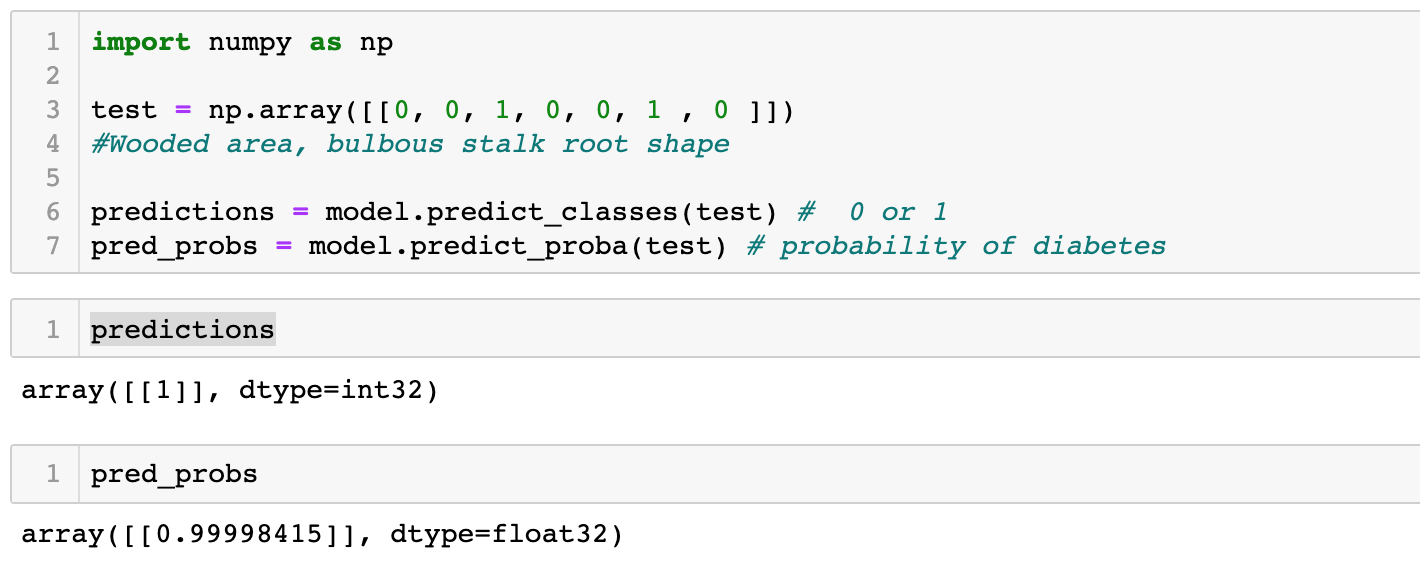


Figure 16

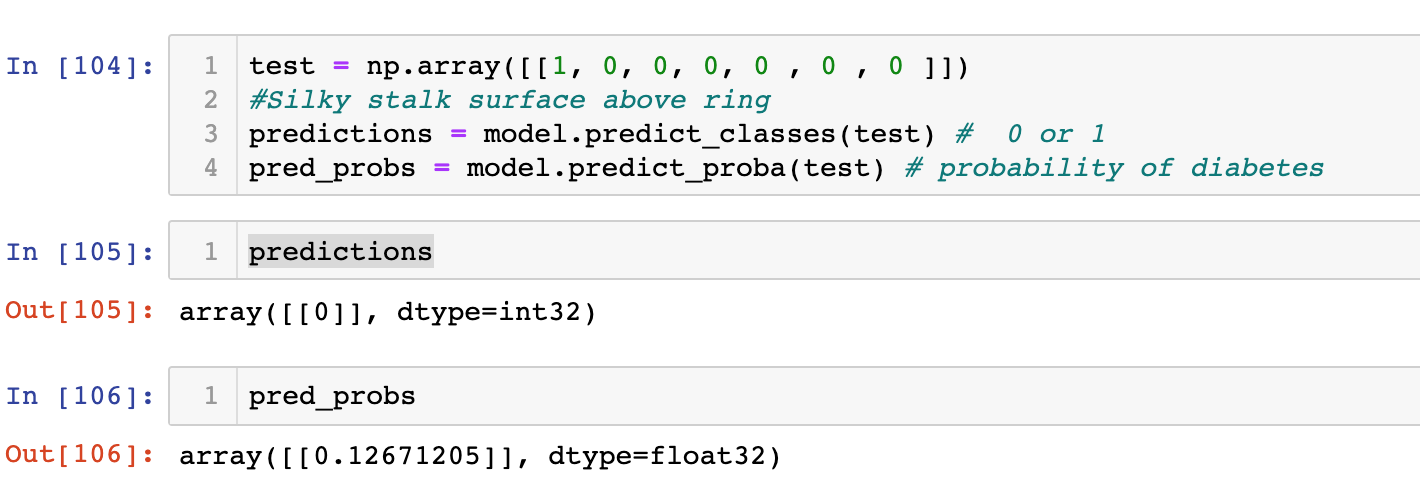


Figure 17

