"IT'S WINE-O-CLOCK!"

-Middle-aged woman who's youngest child just left for college

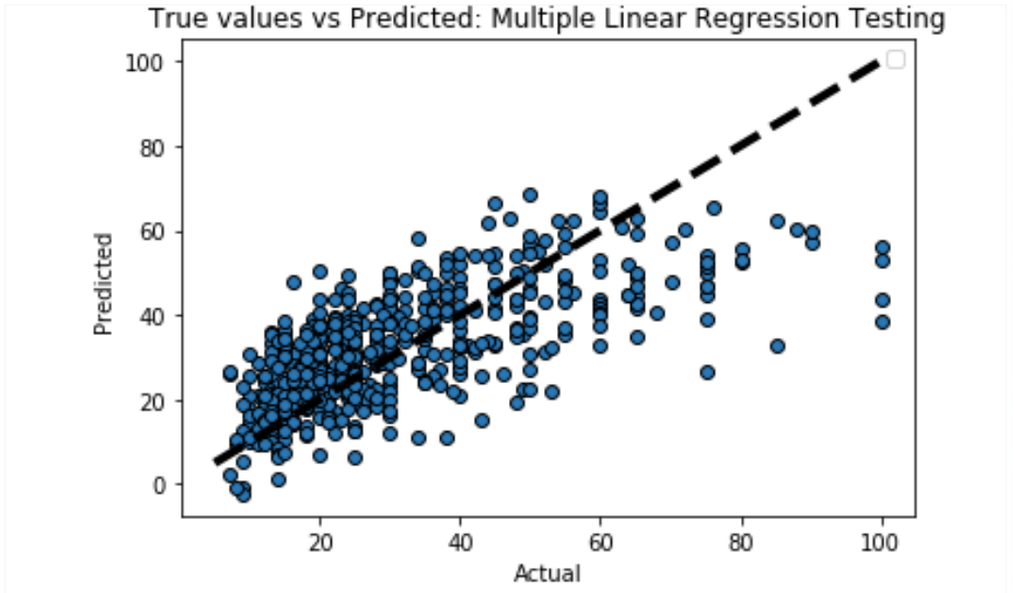
A DEEP AND MACHINE LEARNING APPROACH TO UNDERSTANDING HOW MUCH OUR FAVORITE DRINK IS WORTH

The Dataset

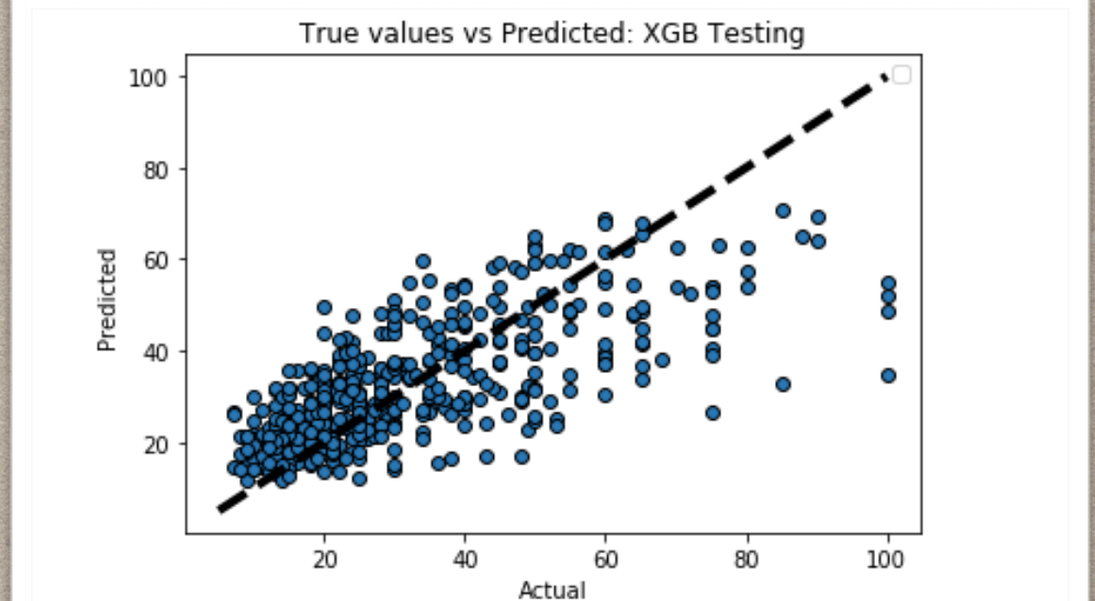
- *Variety: The grape of the wine. This dataset includes observations from 60 different grapes*
- *Country: the country the wine is from. This dataset includes observations from 40 different countries*
- *Province: the province or the state the wine is from. This dataset includes observations from 50 provinces*
- *Price: The price of the wine*
- *Points: Rating assigned to the wine by Wine Enthusiast Magazine*
- *Score: A number generated by running a sentiment analysis on descriptions of the wines given by sommeliers*
- *In total this dataset included over 98,000 observations*

The metric used to measure the accuracy of the models was the mean-squared-error, the square root of which gives us an average of the error

Traditional Machine Learning Methods



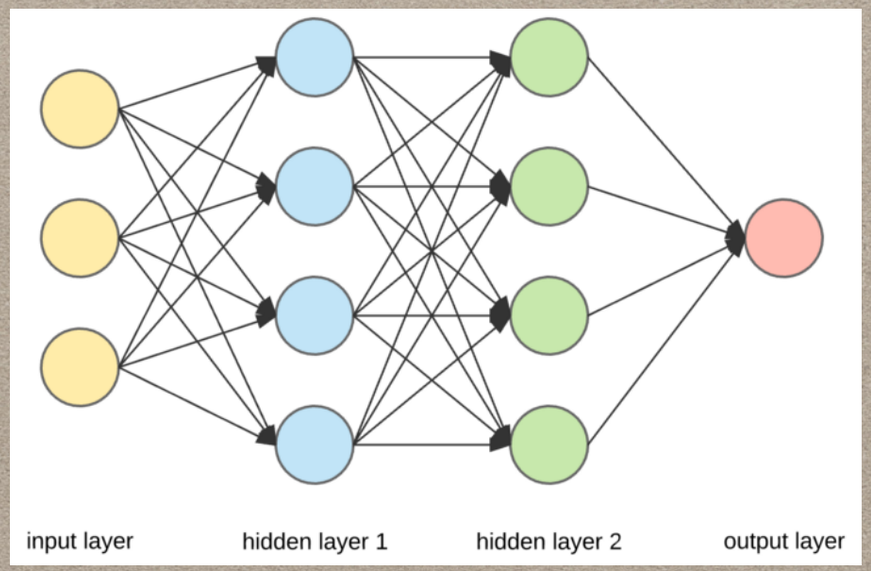
On average this model was able to predict the price of wine within $14 of the actual price



On average this model was able to predict the price of wine within $13 from the actual price

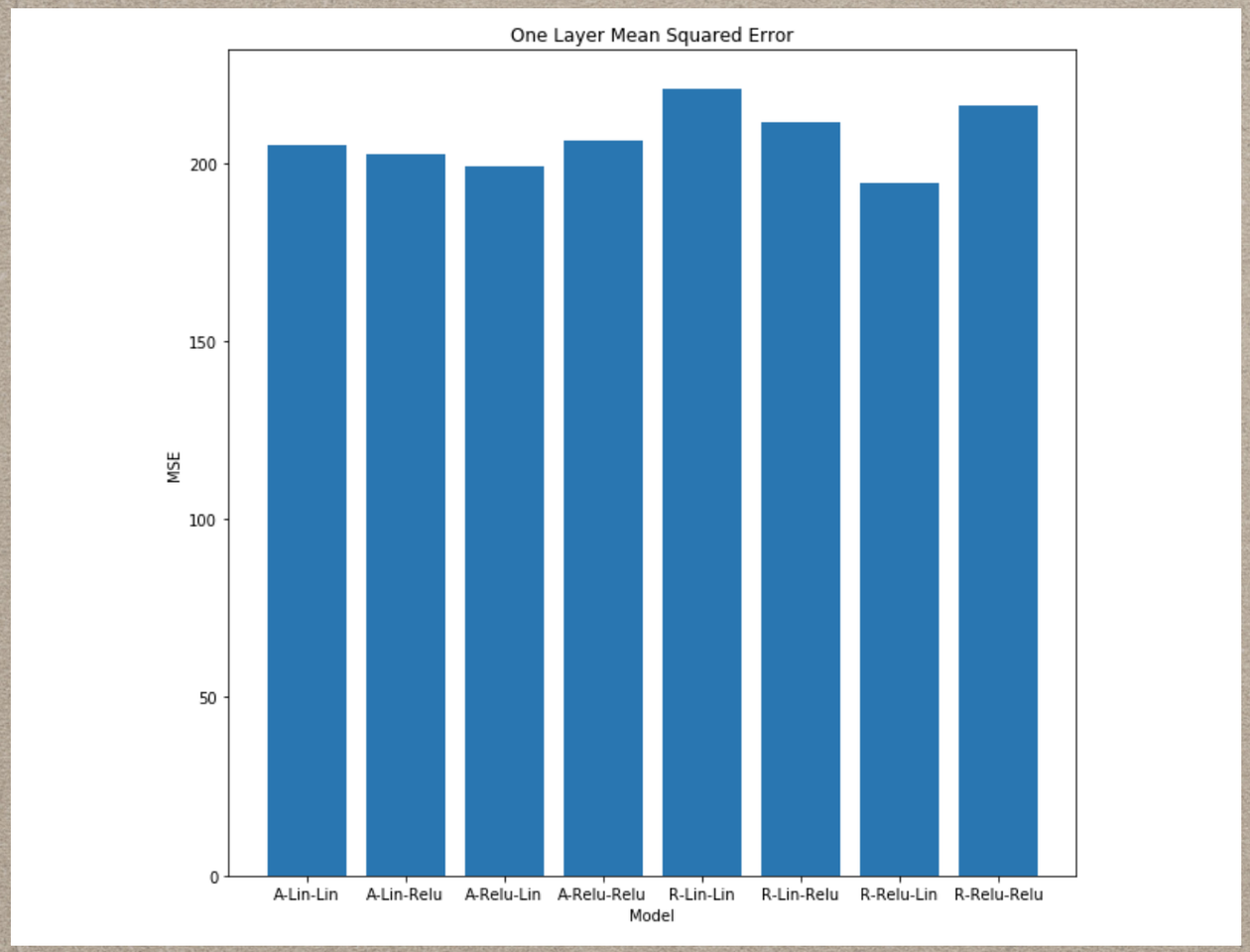
WHAT IS DEEP LEARNING AND WHAT IS A NEURAL NETWORK?

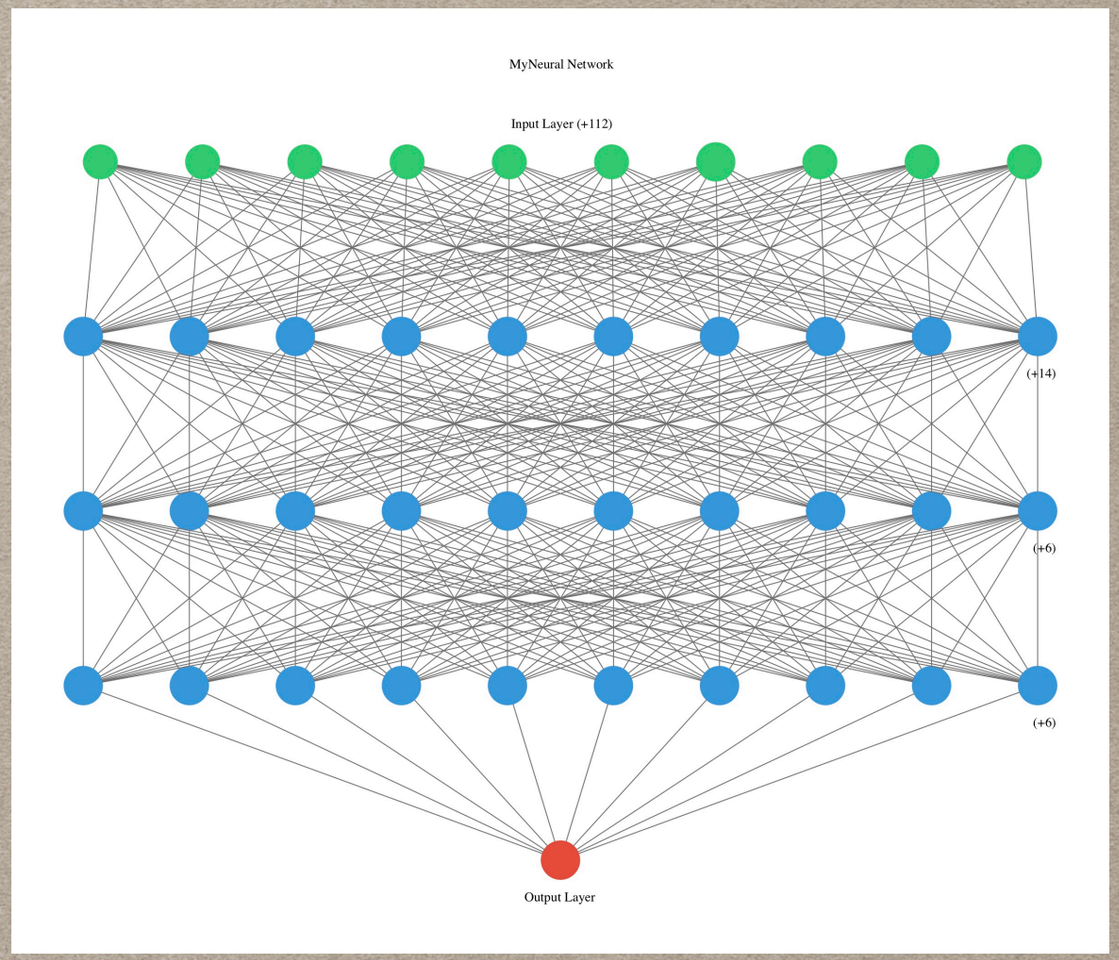
"Neural networks are a set of algorithms, modeled after the human brain. They are sensors: a form of machine perception. Deep learning is a name for a certain type of stacked neural network composed of several node layers. Each layer's output is simultaneously the subsequent layer's input, starting from an initial input layer."



Source: https://d4datascience.wordpress.com/2017/09/03/intro-to-deep-learning/
Picture source: https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6

In total, 10 neural networks were built. First, 8 two layer neural networks were built and the error margins were observed

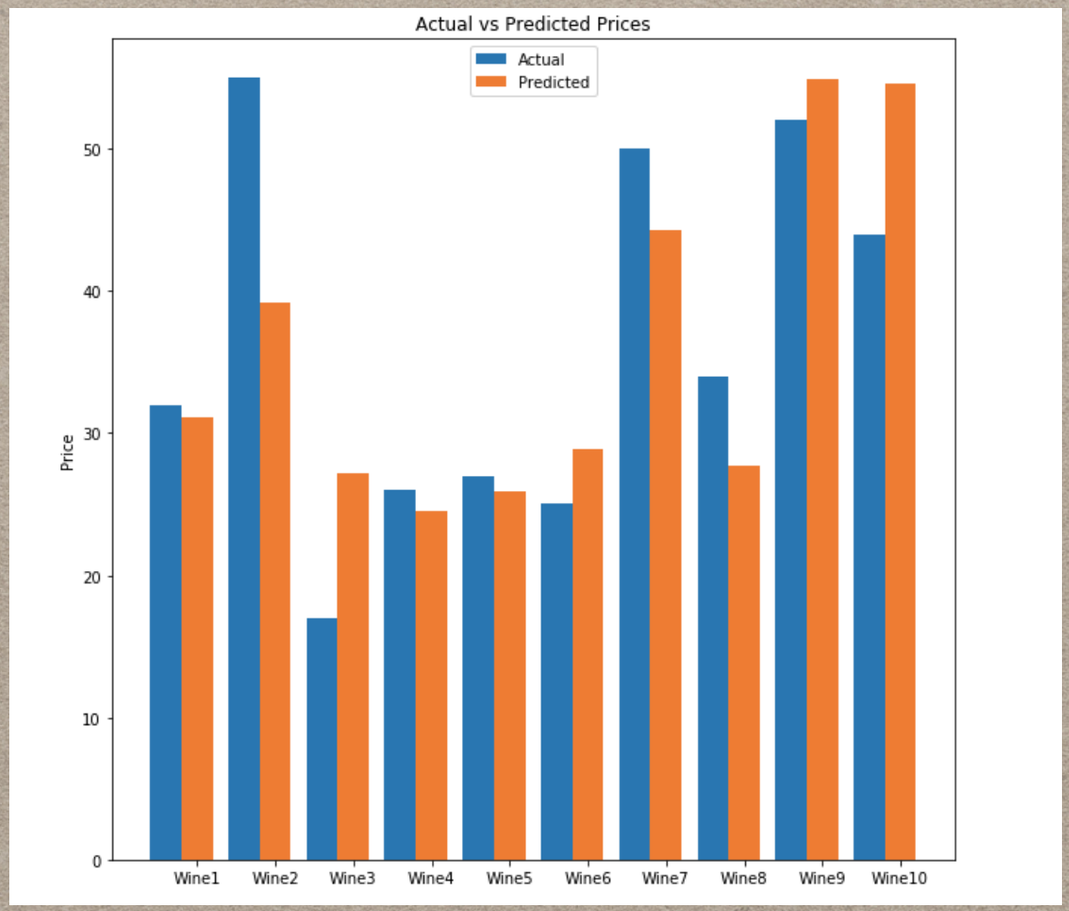




Next, I chose the best performing double layer network and added a third layer to it to see if a lower error margin could be achieved

Model_Name	Mean_Squared_Error
Multiple Linear Regression	202.637958
XGB_MSE	193.262657
A-Lin-Lin	204.855454
A-Lin-Relu	202.214236
A-Relu-Lin	199.109278
A-Relu-Relu	206.274836
R-Lin-Lin	220.658750
R-Lin-Relu	211.212185
R-Relu-Lin	194.199329
R-Relu-Relu	215.810161
R-Relu-Relu-Lin	194.642690
R-Relu-Linear-Linear	209.333455

After building a total of 12 models I found the model that yielded the lowest error margin was the XGBRegressor model. On average the error in predicting the price of wine is $13



	Actual	Predicted
115294	32	31.155443
115295	55	39.198372
115299	17	27.154383
115300	26	24.545677
115302	27	25.852600
115303	25	28.884472
115304	50	44.260384
115305	34	27.754457
115306	52	54.852261
115307	44	54.582882

CONCLUSION

- After building a total of 12 models using both machine learning and deep learning methods, I found that the lowest instance of error was achieved using classic machine learning methods.
- The best performing model was able to predict the price of a wine based on the province, country, grape, sentiment analysis, and points of the wine.
- The average of error in prediction for our best model was $13