Our first attempt at improvement was to try non-normalized data on the neural network. We reasoned that since our results were not desirable with normalized data, we should try non-normalized data to see if its results would be any better. We found that the neural net's predictions were worse with non-normalized data. The predictions for revenue had a correlation coefficient of 0.1889, and root relative squared error of 146.0285%. Predictions for gross profit had a correlation coefficient of –0.0608 and a root relative squared error of 193.357%. Thus, the predictions correlated less with the actual results than our first predictions had.

Next, we tried running the MLP for twice as many epochs, 1000 rather than 500. We reasoned that this might give the model more time to learn patterns in the data that were useful. After testing, we found that running for 1000 epochs brought no improvement. The predictions for revenue had a correlation coefficient of 0.1889, and root relative squared error of 146.0285%. Predictions for gross profit had a correlation coefficient of –0.0608 and a root relative squared error of 193.357%. The results were exactly the same as the results we saw for non-normalized data for 500 epochs.

Hidden layers are a powerful tool for learning complex problems. Since WEKA has a built-in option to select the number of hidden layers, we ran our tests with more than one. With 2 hidden layers predicting revenue, we obtained a much better correlation coefficient of 0.8463, with 67.169% root relative squared error. We found that using 4 hidden layers to predict revenue did not perform as well, with a correlation coefficient of 0.7038, and root relative squared error of 102.4885%. These results were obviously much better than our initial ones. We decided next to try trimming down the features we used, thereby making better use of our small pool of data.

In an attempt to shrink our number of features, we decided to try using 30 different financial ratios as attributes, rather than financial statement values. Our reasoning was that each ratio often encapsulates the same information as two or more statement values. We hoped that using ratios would give us more information with less attributes. Since the ratio values are proportionally much smaller than revenue and gross profit, we decided to use percent growth of profit and revenue as the targets, rather than the unaltered values. This would put the targets into a closer range with the instance values.

Using ratios only did not perform well. The predictions for revenue had a correlation coefficient of -0.0009, and root relative squared error of 154.0149%. Predictions for gross profit had a correlation coefficient of 0.1205 and a root relative squared error of 109.5771%. With those inputs, predicting revenue was essentially as bad as random guessing, and while predicting gross profit fared better, it was still much worse than predictions with hidden layers.

We tried using our own form of ratios next. We took just over 30 values from financial statements, and divided them all by the total assets of the corresponding company. The intuition was that this would put the values somewhere in between zero and one, since total assets is quite often the largest amount in a company’s financial statements. We went with percent growth of revenue and gross profit for the targets once again, so that their range was closer to that of the inputs. The predictions for revenue had a correlation coefficient of -0.0075, and root relative squared error of 302.339%. Predictions for gross profit had a correlation coefficient of 0.0032 and a root relative squared error of 100.8712%. These results were worse than our original results.

The decision was made that it would be wise for us to use some kind of feature selection algorithm. If the algorithm could detect the most important features, then perhaps we could use just a few and get better results. We used WEKA’s best-first attribute selection to algorithmically select only the most important features. For revenue, the algorithm selected **revenue, SGA expense, accounts receivable, and operating cash flow** as the most important features (all from the year before the revenue we were trying to predict). We didn’t give it a limit on the number of features to select, but it only chose four. When the MLP was run with just those four features, the results were the best we had seen. The predictions for revenue had a correlation coefficient of 0.994, and root relative squared error of 12.1536%. Thus, the model’s predictions were 99.4% correlated with the actual target.

We decided to try wrapper-style attribute selection, to see if we could get even better results. We used WEKA’s wrapper algorithm. It goes through and tries each attribute individually, picks the best, and then does the same to pick a second attribute, and so on. This algorithm took much more time to run than WEKA’s best-first attribute selection. It took over an hour. When the algorithm completed, it had selected **revenue, inventory, stockholder’s equity, days sales outstanding, free cash flow, working capital, and enterprise value at end of period** in order to predict next year’s revenue. When those features were used to predict, we got even better results. The predictions for revenue had a correlation coefficient of 0.9965, and root relative squared error of 12.3258%. These attributes correlated slightly better with the targets than the previous four had.

We wondered if the results would improve by using both the attributes obtained by using a wrapper and the attributes obtained by the best-first algorithm. However, in testing those 10 attributes, the correlation coefficient was 0.9915, with 16.755% root relative squared error. Since this was worse than either set of features also, we didn’t pursue the thought of combining the two sets any further.

For gross profit, we went straight to trying the wrapper attribute selection that had worked well for revenue. The wrapper algorithm chose the following features (from the previous year) to predict gross profit: **gross profit, SGA expense, stockholder’s equity, investing cash flow, free cash flow, and enterprise value at end of period.** It was very successful, obtaining 0.9965 as a correlation coefficient and a root relative squared error of 12.3258%. We made some tweaks to our predicting of revenue, and were able to obtain a final correlation of 0.9976 with a root relative squared error of 7.7937%. We were thus convinced that both revenue and gross profit could generally be reasonably predicted with only a few items from the financial statements.