Example 1 – Wine quality prediction:

Problem statement:

The following problem used a decision tree algorithm to predict the quality of wine from 3 to 8 using the following attributes, fixed-acidity, volatile-acidity, citric-acid, residual-sugar, chlorides, free-sulfur-dioxide, total-sulfur-dioxide, density, pH, sulphates, alcohol, and quality. An example of the data set, visualised using python, is as follows:

	fixed-acidity	volatile-acidity	citric-acid	residual-sugar	chlorides	free-sulfur-dioxide	total-sulfur-dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

The full data set is made up of 1600 rows (source: https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009).

Description of AI technique:

The artificial intelligence (AI) technique used was a decision tree. A decision tree is used to classify examples such as types of flowers, car brands or in this case the quality of wine. A decision tree starts off with a root node and tests the most important attribute first, this is to immediately shorten the length of the decision tree and prevent overfitting. If the most important attribute is tested first, some attributes won't need to be necessarily tested for the algorithm to decide on the correct classification. As the decision tree continues each new input goes into a new node until it reaches a leaf, which returns the classification of that leaf. Once the leaf is reached the decision tree has its output.

If a decision tree is not shallow and in fact has a lot of nodes leading to different leaves, the decision tree may need to be pruned. Pruning means to remove any unnecessary attributes from the tree and can be done using a significance test of all the attributes in the data set. A significance test involves looking at the decision tree nodes that only has leaf nodes as descendants. If the test appears to be irrelevant, then the input node is replaced by a leaf node, meaning it goes straight to classification and provides an output.

How the AI technique is used to solve the problem:

The data set is fed into the decision tree and broken up into training and testing data sets. The training data generates the decision tree, and the testing data is then used to test its accuracy. At the end, the decision tree is tested with two examples not seen in the training or testing data. As stated before, the attributes of the data set include fixed-acidity, volatile-acidity, citric-acid, residual-sugar, chlorides, free-sulfur-dioxide, total-sulfur-dioxide, density, pH, sulphates, alcohol, and quality. The attributes excluding 'quality' are used to create the nodes of the decision tree and the 'quality' attribute is set as the target value. This process of setting the target and decision tree attributes will be displayed in the demo video to help visualize what is going on in this step. Once the attributes and target are set and the decision tree is trained and tested as discussed above the decision tree will hypothetically be able to predict the quality of the wine from only being told the values of the attributes in the data set such as fixed-acidity, volatile-acidity, etc.

Solution Results:

The results of the decision tree were somewhat successful in predicting the quality of wine as it provided a solution that is significantly better than purely guessing, however, the accuracy of the decision tree was only 50-57 percent accurate. However, it did manage to provide a prediction when new data that the decision tree hadn't seen before was inputted. The most probable cause for the drop in accuracy would be the decision tree overfitting the data as typically, the decision tree would start with the most significant attributes. However, using the programming skills developed from lab 2 of introduction to artificial intelligence no significance testing or pruning was explicitly coded in the Jupyter notebook. This would have resulted in a less accurate decision tree; however, it still provided a prediction that is far superior to just guessing and by using a decision tree the accuracy of prediction was significantly better than using a model like linear regression which only provided an accuracy of 36%. Overall, the decision tree with an accuracy of 50-57 percent was able to make predictions of the quality of wine from the given data attributes.

References:

Learning, U. M. (2017, November 27). Red Wine Quality. Retrieved from

https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009