We have built a model which is generic and can create a decision tree for any categorical or numeric data. We import the data and encode the categorical data to numeric form. Then the data is split into training set and test set. The training set is used to create a tree and the testing set is used to test the decision tree. We use a recursive function to create a decision tree and save it in a nested dictionary. Then another recursive function is used for the testing of the decision tree.

We have used numpy package for the calculations, pandas package for loading and slicing the data and sklearn package to encode the categorical to numeric form. The real reason behind changing the categorical data to numeric for is to make our model generic.

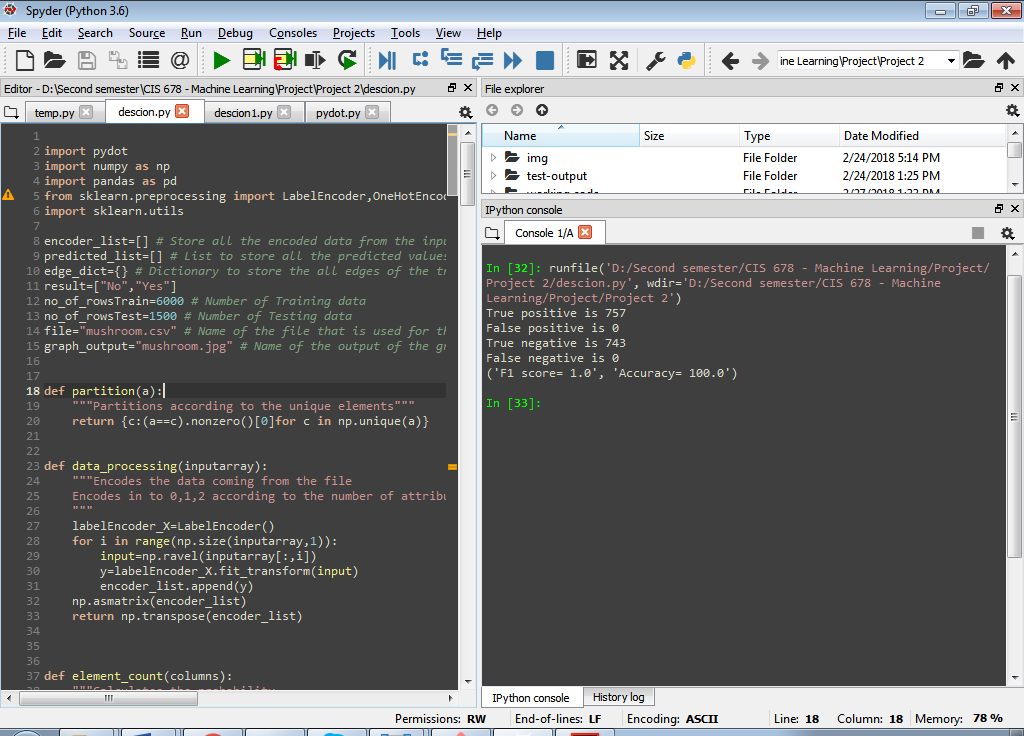
We encountered a couple of problems while creating our model. Our first difficult challenge was to store the tree in a testable format. We tried saving our tree as a Class with its node and its children, but it was really difficult to keep track of the parent node because of the recursive function. So, we moved to saving the tree in a dictionary which was more convenient. Other minor bugs kept delaying our project. Working with numpy array calculations in a recursive function was pretty challenging. Creating a graph of the tree was also very challenging. First we tried graphviz, then we moved to Json parsing and rendering to a web browser. Both did not work for us. Finally we were able to plot the graph with pyplot.

Below is a are the screenshots that shows the true positive, false positive, true negative, false negative, Accuracy and F1 score for each dataset we have experimented with.

Mushroom Data:

True Positive: 757  
False Positive: 0  
True Negative: 743  
False Negative: 0

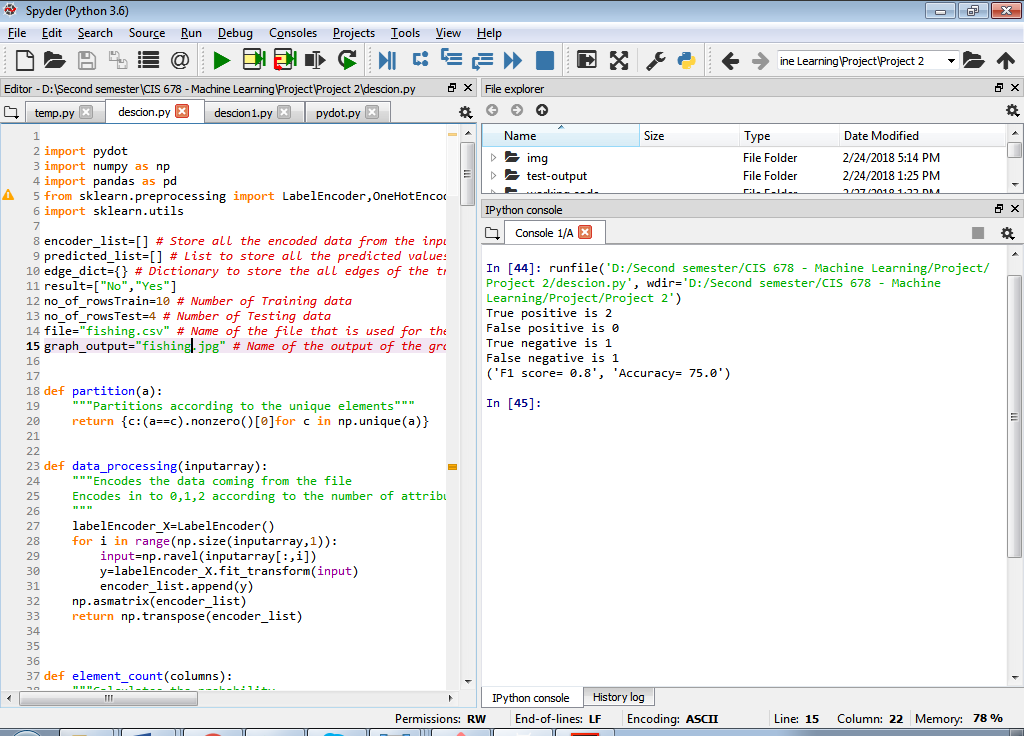
F1 Score: 1  
Accuracy: 100 %



Fishing Data:

True Positive: 2  
False Positive: 0  
True Negative: 1  
False Negative: 1

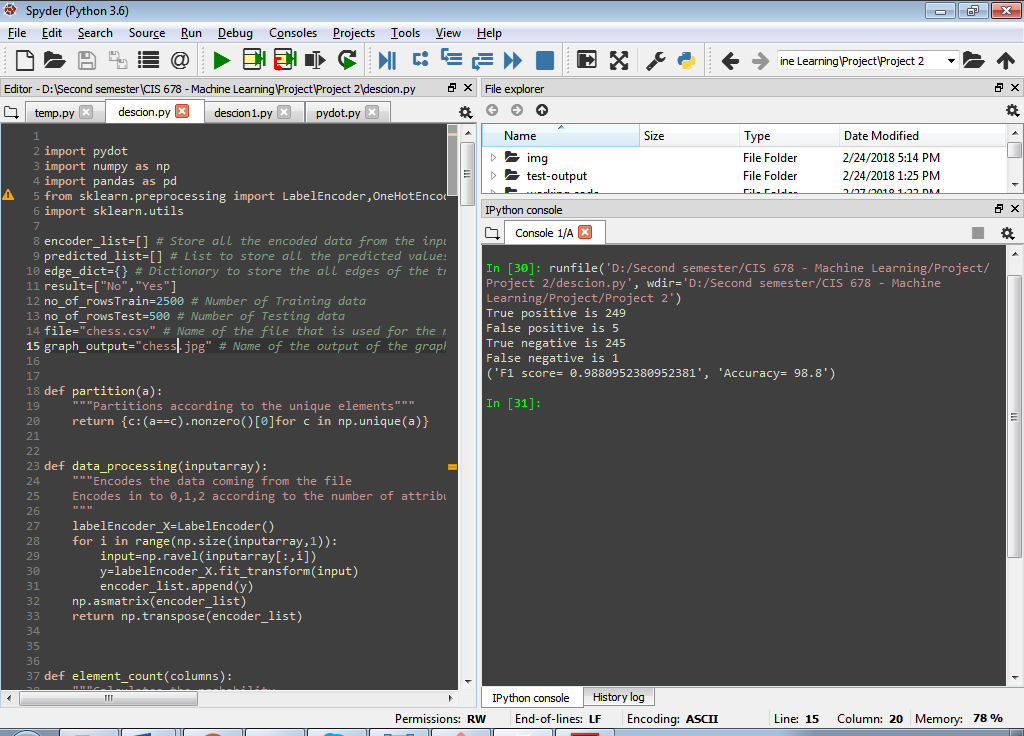
F1 Score: 0.8  
Accuracy: 75 %



**Chess (King-Rook vs. King-Pawn)** Data:

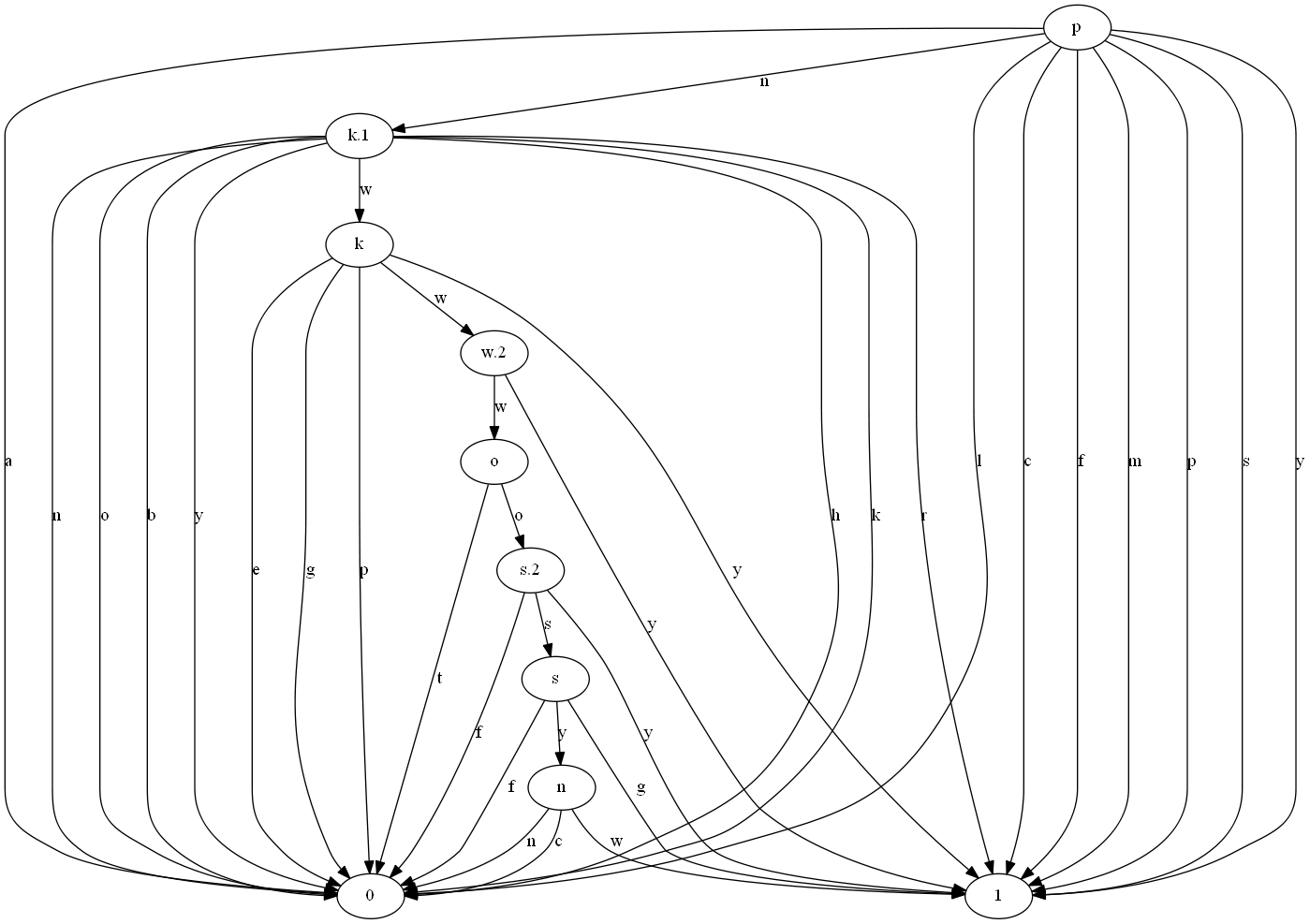
True Positive: 249  
False Positive: 5  
True Negative: 245  
False Negative: 1

F1 Score: 0.988  
Accuracy: 98.8 %

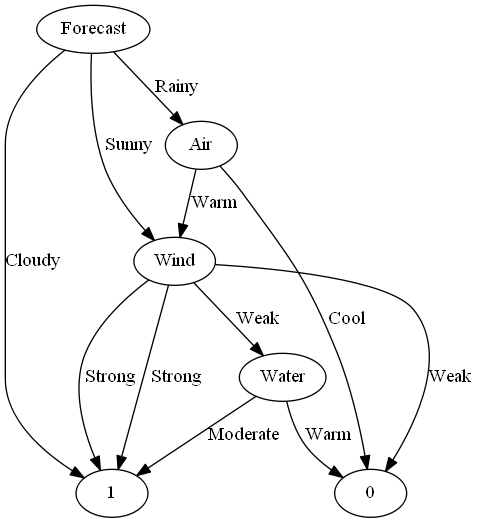


Below are the sereen shots of the tree graph generated for different datasets by our model

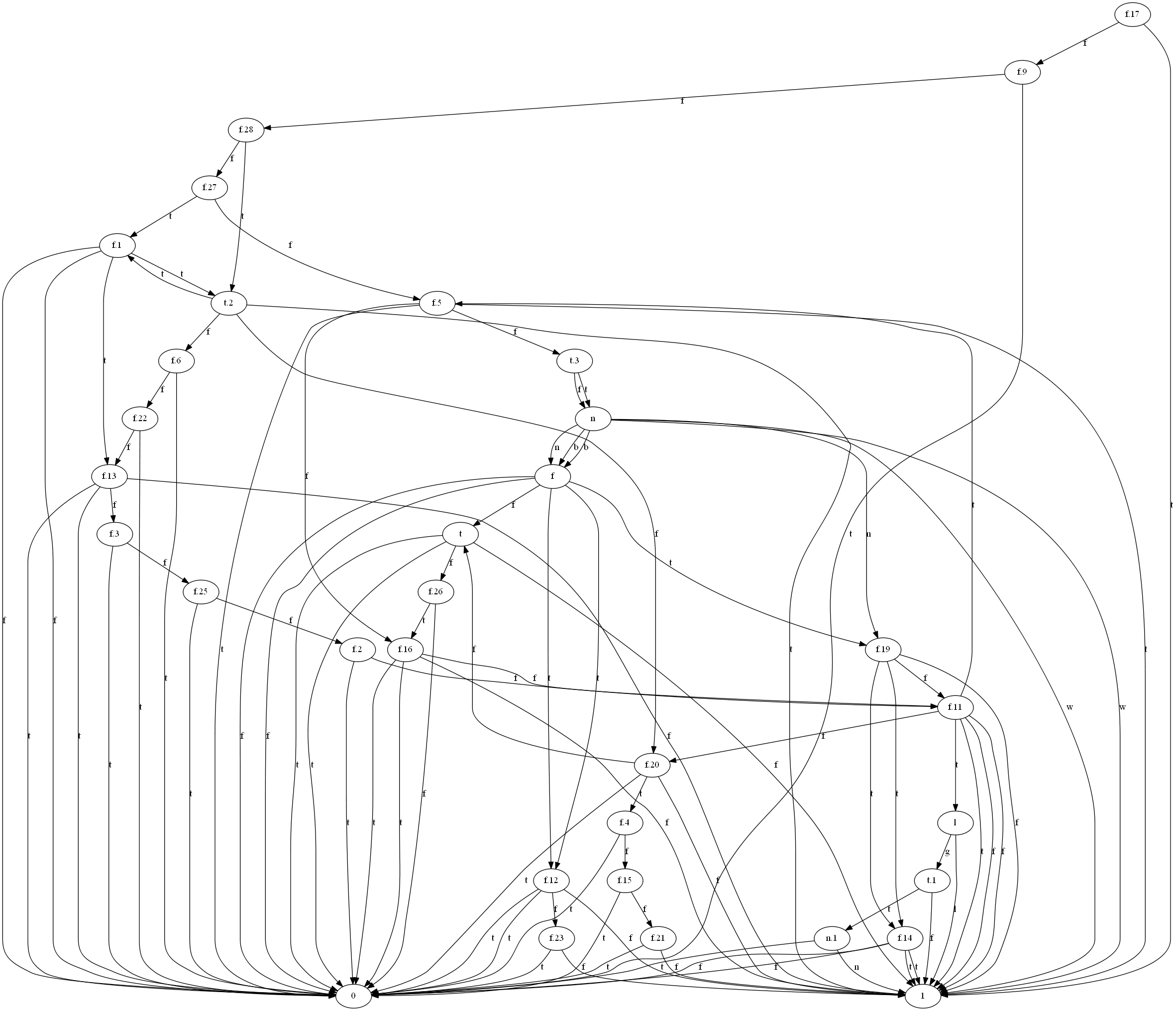
Mushroom Data:



Fishing Data:



**Chess (King-Rook vs. King-Pawn)** Data:



Our model is a binary classification decision tree which is generic. Any binary classification data, categorical or numerical, can be used in our model to build a decision tree. Since the test is giving us a good accuracy and F1 score, our decision tree model can be used in real world application. For example, we have used the hypothetical mushroom data to decide between edible and poisonous mushrooms. We could use real data to figure out if the mushroom is edible or not using our decision tree. Decision trees can be used in application to classify data. For example, application can use decision tree to classify user data according to the information given by the user. Decision trees can be used in many ways by applications to make decision or classify data. For example, decision trees can be used to recognize spam emails.