Decision tree classifier is used for classifying the given 5 cards into one of the 9 poker hands or to categorize the given input as not recognized hand.

The height of the Decision tree is one of the hyper parameter which is to be figured out. An experiment is conducted to compute the good value of the height of the tree by computing the error rate both on training data and test data using the models with various heights. The training data which is available is divided into data for training and data for validation. As expected, the tree with smaller height gave more training error because of underfitting and as the height increased the training error decreased but the validation error increased due to overfitting of the data. This variation is plotted as shown in the figure 1.

The cards are sorted according to the suit number(as in the data) and then by card number to avoid the complication of the model, as the order of the cards doesn’t matter when calculating the poker hand. For instance ♥6, ♣8, ♦7, ♦4, ♥5 input is same as ♥6, ♦4, ♥5, ♣8, ♦7. During training or predicting the output if these two inputs are considered as different, it will lead to either complications in the model or higher error rate in case of testing.

To decrease the error rate, feature extraction needs to be done. As from the rules of finding the poker hand, it is evident that most of the hands only require the information of whether the cards given are sequential or having the same card number/ suit(except for the Royal Flush which requires the information about the actual cards). Exploiting this, feature is extracted from the given five cards by considering the difference in each of the card numbers and suit numbers.

The feature extracted for the input (Suit1,Card1,Suit2,Card2,Suit3,Card3,Suit4,Card4,Suit5,Card5) is as below:

Feature 1(Input 1) : difference(Card1, Card2)  
Feature 2(Input 2) : difference(Card2, Card3)  
Feature 3(Input 3) : difference(Card3, Card4)  
Feature 4(Input 4) : difference(Card4, Card5)  
Feature 5(Input 5) : Card5(highest ranked card)  
Feature 6(Input 6) : difference(Suit1, Suit2)  
Feature 7(Input 7) : difference(Suit2, Suit3)  
Feature 8(Input 8) : difference(Suit3, Suit4)  
Feature 9(Input 9) : difference(Suit4, Suit5)  
Cards are arranged in the ascending order of their suit number and then by their card number

The features that will be extracted from the input be representing the relative values of the cards instead of their actual values which is what is required in the case of finding the poker hand. As the feature 5 : Card5 is used to account for one of the hands which is Royal Flush, which is the only hand that requires actual value of the cards with the relative values.