* Country ID and League ID are same, could merge those two tables
* In ‘Match’ table, the very last columns are home win odds, draw odds and away win odds computed using different methods. But all these seem to be giving results in the same ratio. We could take the average of these numbers and reduce it to just three columns, “Home win odds”, “Draw Odds”, and “Away win odds”.
* How are the player attributes stored in the tables? Also, what are the relationships of these attributes to the match? Can we use it to train our tree? (183978 rows in player\_Attributes table). It has a lot of attributes, but the question is how to associate the attributes of a particular player to the match? Shall we consider the attributes of the team as the aggregation of the attributes of the individual player playing that match?
* Also, there is only so much information given. For instance, the values of particular attributes are given but it is not provided how to interpret that information!
* Given an attribute, potential, it not provided how the attribute is measured. What is the scale of the attribute? Is it evaluated on a scale of 100 or it is indefinite? Is a higher number better or worse? How to compare two numbers of this attribute? Is it linear? Can we just subtract the number to compare which player has how much more potential compare to a particular player?
* Also, while making a complex Decision Tree, should we give different weight to different attributes? How do we decide how much weight to give to which attribute? One way is to just make different trees with different weights assigned to different attributes, and pick the one that gives the best results.
* One problem with this approach is we may never find the best decision tree also there are exponentially many trees possible with different weight attributes, we can’t compute them all.
* Hit and miss without anything also gives 50% ! :P