

NCAA 4th and 1 Decision Tree Classifier

1. My tree first split on the direction of the pass on the 4th and 1 play. For the 5304 plays where the pass direction was to the left or the center (≤ 0.5), the tree then split on whether or not the play was an intentional quarterback sneak. For the 136 plays where the pass direction was to the right (> 0.5), the tree split on whether or not the play was an intended quarterback scramble.
2. The decision tree split on the direction of the pass play because this variable provided the highest information gain about the first down conversion status of the play. For plays to the left or the middle of the field, 3733 resulted in a first down and 1571 did not. Of the plays to the right, just 48 were first downs and 88 were not. This is a very sharp difference, and there might be an underlying variable or data leakage here. The decision tree then split on the binary QB sneak and QB scramble variables, which have the next highest information gain values on the left and right sides of the tree.
3. This tree might be somewhat helpful to a head coach, depending on the true validity of the interpretations above. If the coach wants to set up a pass play, it may be beneficial to direct the pass to the left or the middle of the field. If the coach wants to run the ball on the left or middle, a quarterback sneak is very effective. If the coach wants to run the ball on the right, a designed quarterback scramble is effective.