
GOLF

Rob Whitaker, Computer Science
rmw2@princeton.edu
Eric Mitchell, Computer Science
eam6@princeton.edu

Abstract

1 Introduction

Motivating ideas in the golf world. What exists in terms of stats? [Describe pga tour stats etc.] What exists in terms of prediction? [betting odds etc.] Machine learning methods are difficult because the data is very intricate, though detailed.

Principle issue is in integrating two distinct datasets: player year-end statistics (henceforth “stats”), and the specifics of a given tournament course. [Discuss this more, give examples, pretend to be a humanities person.]

Things we would like to understand: The best players in the world are interesting to study and we can make predictions about their relative rankings on important tournaments, but the place where automated prediction becomes really important is for players outside of the critical spotlight. Players that are good enough to be on the PGA tour and have data collected about their performance, but not good enough for a golf aficionado to analyze their game enough to make predictions are left out. Simple questions like making the cut etc etc

2 Related Work

3 Methods

3.1 Data

Discuss scraping and pre-pre-processing aka all the crazy fucking json files, milk this for difficulty of the scraping part.

3.2 Exploring Golf Stats

This is our year-end standing predictions from

3.3 Predicting Tournament Results

The logic of our approach was as follows: different courses bring different challenges, and while a naive approach modeling the results based on season stats alone may shed some insights into the importance of different skills for the game in general, general sports wisdom and observation suggest that different courses (and different holes) have their own keys for success. Rather than fitting a new

054 model to every course, or worse for every hole in the game, we attempted to train a single model
055 while correcting for the biases that different courses introduce into the final results.

056
057 As an illustrative example, being able to escape a sand trap is an important skill, but it likely will
058 not enter the calculus on a hole with no sand traps. Our task was then to correct for these biases
059 before player stats were fed to our models. We accomplished this with two similar approaches: bias
060 correction by course and by hole. We first calculated the correlation of each stat with

061 **3.4 Models**

062 **4 Results**

063
064
065 How well did we do?

066 **5 Discussion and Conclusion**

067
068 What does this say about golf in general, how can this be used by people, did we succeed in any of
069 our goals?

070 **6 Further Research**

071
072
073 All the techniques we thought of but couldn't implement.
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161

References

Discussion references:

Code references: