PREDICTING PROFESSIONAL GOLFER EARNINGS USING MACHINE LEARNING

Supporting code files, reports, and presentations can be found at
https://github.com/nosliwes/dsa5103

1.1.1 SUMMARY

Machine learning was used to predict annual winnings for professional golfers using individual tournament statistics over the years 1992-2020. The project specifically focused on three skills in the game that a golfer can develop to maximize winnings (tee shots, fairway shots, and green shots). Results showed that fairway shots contributed the most to winning, followed by green shots and finally tee shots. Although it is commonly believed that the longest hitters win more frequently, the analysis showed that the most successful players perform better in the mid game and short game.

1.1.2 BACKGROUND

Golf is a sport that has become increasingly popular in the U.S. due to its low impact, availability, and social engagement. Playing golf causes little stress on the body allowing people of all ages and fitness levels to participate. Additionally, the nature of the game equalizes physical differences that exist based on age and gender. These aspects of the game have resulted in a competitive situation where earnings can be related directly to individual performance.

Although many individual statistics are available to analyze, the goal of the project was to determine the most important aspect of a player's game to practice that maximizes performance. As a result, the project aimed to predict a professional golfer's annual tournament winnings based on three specific shots (tee shots, fairway shots, and green shots).

1.1.3 METHOD

The analysis method involved data collection, data preparation, model creation, and model evaluation. Details for each of the phases are included below.

DATA COLLECTION

Thirty years of data (1991-2020) was collected for player performance and prize earnings from the publicly available data at https://www.pgatour.com/stats. The data was extracted to build a player data set spanning multiple years and tournaments.

The statistics used in the analysis are shown in the table below. Individual statistics were summarized with the goal of building an interpretable model that players could use to make informed decisions for

Draft Project Report   DSA 5103 Section 995   Jenny Vance
(O) We're #1   Nic Cejda
Steven Wilson

skill development. In addition, the number of events played was included to capture contribution from player experience in the model.

TABLE 1. STATISTICS USED FOR EACH INDIVIDUAL SKILL

SKILL	STATISTIC
TEE SHOTS	DRIVE DISTANCE
FAIRWAY SHOTS	FAIRWAYS HIT, GREENS HIT, RELATIVE PAR
GREEN SHOTS	SAND SAVES, OVER/UNDER PAR, ONE PUTTS

DATA PREPARATION

MODELING

EVALUATION

1.1.4 MODEL TRAINING AND SELECTION

1.1.5 MODEL EVALUATION

1.1.6 DISCUSSION

TODO – Discussion following model evaluation

1.1.7 CONCLUSIONS

TODO - Conclusions and highlights following completion.

1.1.8 REFERENCES

1. *The Rise of Golf: Increasing Popularity of Golf in the US*. (2020, October 21). Golf365.Com. https://www.golf365.com/news/the-rise-of-golf-increasing-popularity-of-golf-in-the-us/#:%7E:text=Golf%20is%20a%20popular%20pastime,of%20the%20big%20golf%20tournaments.

2. Peters, Andrew (2008). Determinants of Performance on the PGA Tour. *Issues of Political Economy*, *17*(August), 1–10. http://org.elon.edu/ipe/f2%20andrew%20peters%20final.pdf