PrecisionPicksPGA

By: Matt Bonadies

Agenda

1. Project Overview
2. Data Summary
3. Insights
4. Model Optimization
5. Streamlit Demo
6. Next Steps
7. Appendix

Project Overview

BACKGROUND

- A passionate golfer who enjoys exploring the influence of data on player performance & sports betting odds

PROBLEM

- Many find golf to be a dull & difficult sport to understand, coupled with an increasing number of people placing uninformed bets across a range of sports

OBJECTIVE

- Leverage machine learning to examine player statistics & predict future scores, thereby informing strategic betting decisions & showcasing the greatness of golf

IMPACT

- In less than 5 years, the legal online sports betting market in the U.S. surpassed **$10 billion** by 2023 with projections to mature to **$45 billion** annually

Data Summary

DATA SOURCE

- All data was retrieved from datagolf.com, a partner of the PGA Tour, APIs

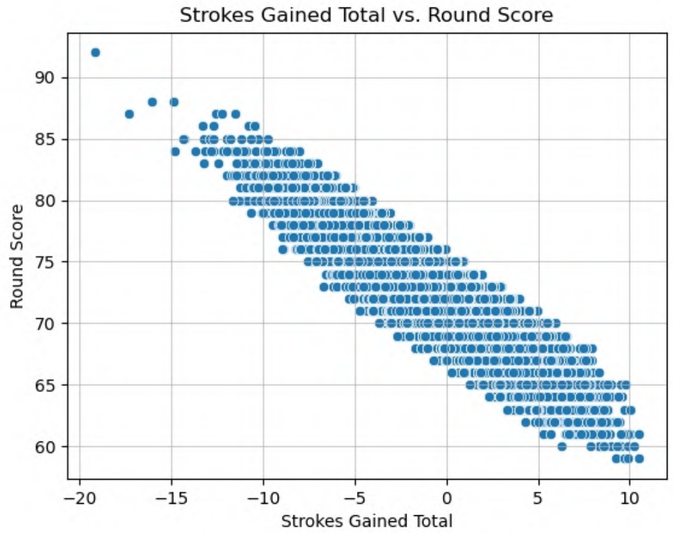
DATASETS UTILIZED

- **Round Scoring & Stats:** Detailed player statistics since 2017
- **Player Information:** Data such as ID, Country, & Name
- **Tournament Field:** Up-to-date data regarding who is playing
- **Player Rankings:** Ranks of the Top 500 Players Globally

GOAL

- Develop robust & detailed data to assess its predictive influence on future outcomes

#	Column	Non-Null Count	Dtype
0	event_completed	131333 non-null	object
1	event_name	131333 non-null	object
2	player_name	131333 non-null	object
3	fin_text	131333 non-null	object
4	course_name	131333 non-null	object
5	teetime	131333 non-null	object
6	sg_putt	104164 non-null	float64
7	sg_arg	104164 non-null	float64
8	sg_app	104164 non-null	float64
9	sg_ott	104164 non-null	float64
10	sg_t2g	104164 non-null	float64
11	sg_total	131333 non-null	float64
12	driving_dist	114688 non-null	float64
13	driving_acc	118779 non-null	float64
14	gir	116134 non-null	float64
15	scrambling	102436 non-null	float64
16	prox_rgh	99588 non-null	float64
17	prox_fw	102480 non-null	float64
18	round_score	130873 non-null	float64



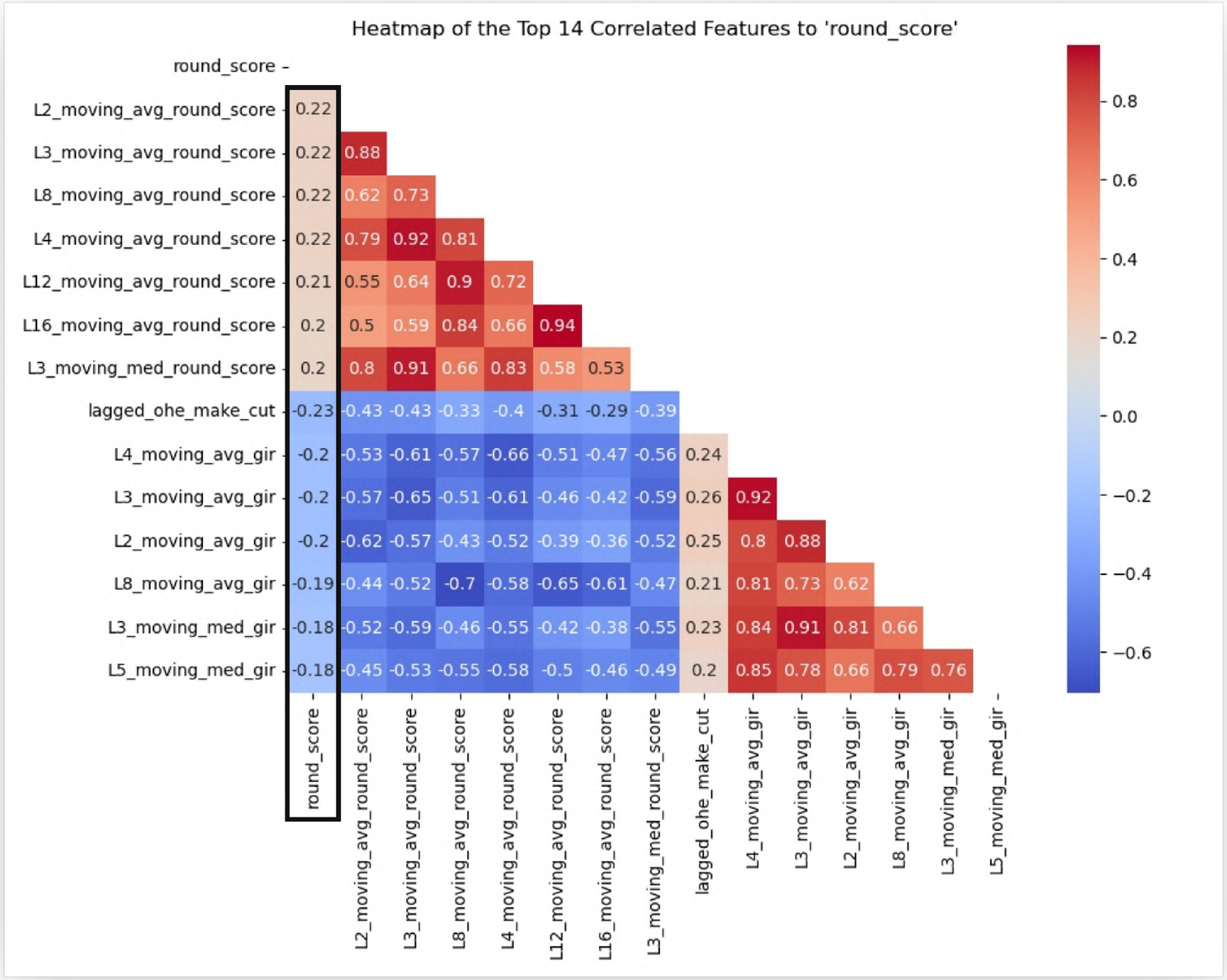
Insights

SOLUTION

- Leveraging the detailed statistics from each round

- Feature engineered 300+ new columns incorporating moving averages, medians, standard deviations, etc.

- Analyzing player trends to evaluate their impact on predicting future performance



Model Optimization

Model Type	Train R^2	Test R^2
Linear Regression	26.29%	26.35%
Lasso Regression	25.81%	26.41%
Neural Netork	42.98%	31.84%

**For additional scoring metrics please refer to Jupyter Notebooks*


Streamlit Demo

Next Steps:

1. Enhance Models with New Features
2. Deploy Streamlit Application
3. License GitHub Repository

Appendix

What is Golf?

SUMMARY OF THE SPORT

- Played on an 18 hole course spanning thousands of yards
- Objective is to hit the ball into each hole in the fewest strokes
- Baseline score for most courses is 72 strokes
- Various different shots to hit such as driving, chipping, etc.

PGA TOUR TOURNAMENT

- 4 rounds of 18 holes each, played over 4 days (Thu-Sun)
- Approx. 120 players competing against each other
- Player with the lowest total score wins
- New competition begins every week in a different location

What is Sports Betting?

SUMMARY OF SPORTS BETTING

- Individuals can bet money on certain outcomes occuring
- Every bet has odds representing the likelihood of an event occurring which determines how much you win
- Bets can be placed online or in-person for most sports

GOLF BETTING

- Golf bets come in many forms:
  - Guessing the winner
  - Predicting an individual to finish in the top 5, 10, or 20
  - Wagering on the winner out of 3 players in a group
  - Betting on one player to finish better than another

Impact & Vision

INDUSTRY IMPACT

- **In 2018**, the Supreme Court reversed a federal law banning sports betting
  - Approx. **38 states** have legalized sports betting, both in-person and online
- In less than 5 years, the legal online sports betting market in the U.S. surpassed **$10 billion** by 2023
- The market is expected to reach **$45 billion** annually at market maturity

VISION

- **Growing Trend:** As the number of sports bettors increases, navigating the complexities of betting becomes challenging.
- **Reality**: Many are unaware that the break-even point for betting odds exceeds a 50% probability, as set by casinos.
- **Objective**: Analyze extensive individual player statistics using machine learning to forecast future scores, thereby guiding the most strategic bets for tournament winners, finishing positions, and head-to-head matchups.