
PrecisionPicksPGA

Agenda

1. What is Golf?
2. What is Sports Betting?
3. Vision & Impact
4. Dataset & EDA
5. Preliminary Model Evaluation
6. Next Steps

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

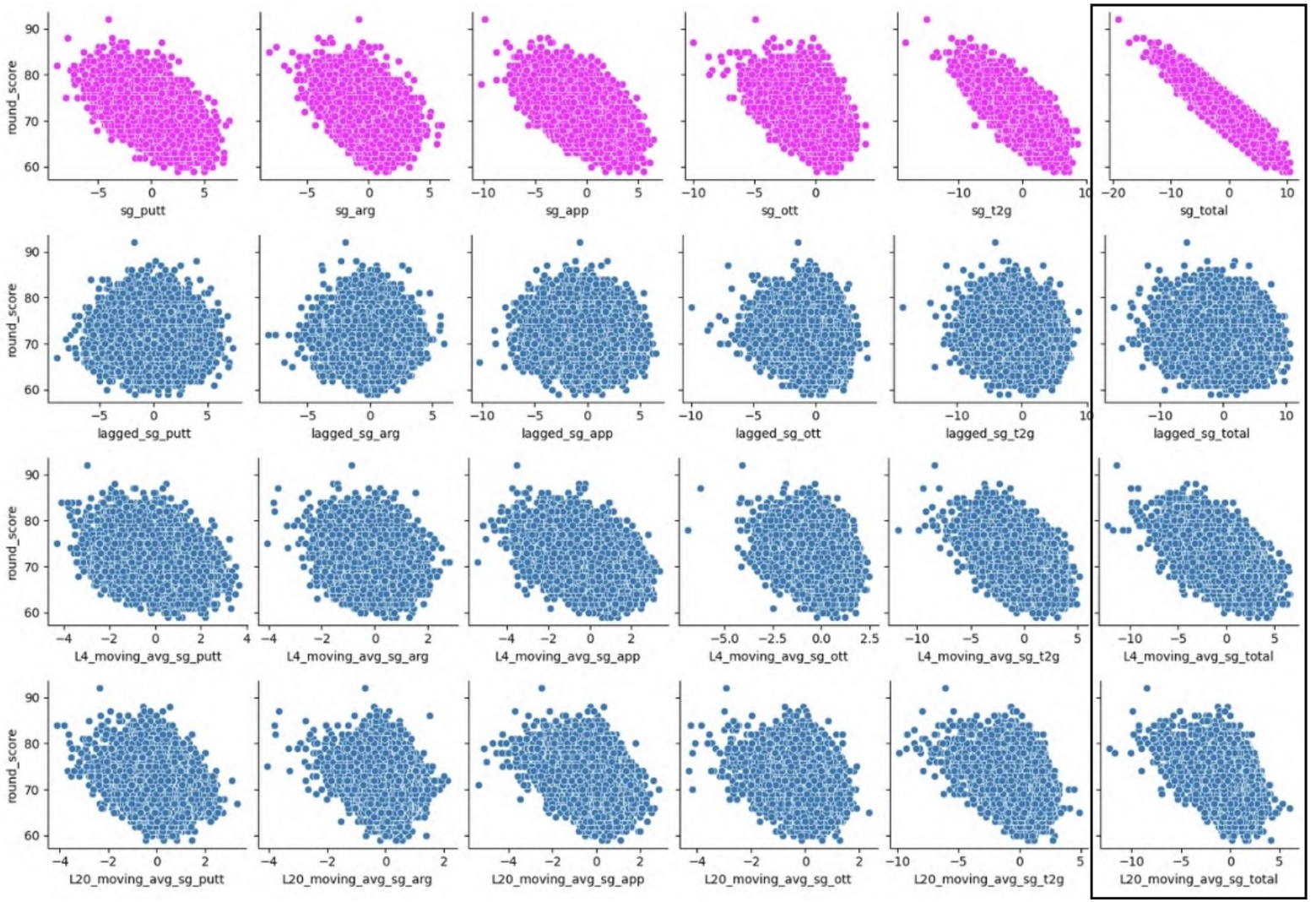
- **Growing Trend:** As the number of sports bettors increases, navigating the complexities of betting becomes challenging.
- **Reality**: Many are unaware that the break-even point for betting odds exceeds a 50% probability, as set by casinos.
- **Objective**: Analyze extensive individual player statistics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

Dataset & EDA

THREE DISTINCT TABLES

1. Round Scoring, Stats & Strokes Gained
   - Shape: (128861, 200)
   - Includes detailed player statistics dating back to 2017
   - Initially 30 features, engineered to 200 features
2. Player List & IDs
   - Includes player data such as ID & Name
3. Field Updates
   - Includes up-to-date data regarding who is playing in the upcoming tournament (Known as the field)

SOURCE: DATAGOLF



Preliminary Model Evaluation

INITIAL LINEAR MODELS

Model	Train R^2	Train MAPE	Test R^2	Test MAPE
Linear Regression	0.4601	0.0269	0.4615	0.0270
Lasso Regression	0.2688	0.0313	0.2760	0.0313
Ridge Regression	0.4600	0.0269	0.4621	0.0270

LINEAR PIPELINE RESULTS

Inputs	Hyperparameters
Scaler	Standard
Dimensionality Reduction	None
Models	Lasso
Alpha	0.002
Test R^2	0.4622

RANDOM FOREST REGRESSOR RESULTS

Inputs	Hyperparameters
Scaler	Standard
Dimensionality Reduction	None
Models	Random Forest Regressor
Max Depth	15
Test R^2	0.5246

NEURAL NETWORK RESULTS

Inputs	Hyperparameters
Activation	Relu
Dropout	0.1
Regularizer	L1
Optimizer	Adam
Losses	Mean Squared Error
Epochs	150
Test R^2	0.6371

Next Steps:

1. Model Optimization
    a. Add newly discovered features
    b. Impute initially dropped records
    c. Tweak model parameters
2. Streamlit Application
3. Re-work code for efficiency
4. Add Logistic Regression Models