

Professional Golf Broadcast Viewer Companion Dashboard

Using PGA TOUR ShotLink® Data

P O S	AUTHOR	TODAY	THRU	TOTAL
1	DMITRIY BLAGNIN	-6	F	-10
T2	SANTOSH GOTETI	-4	F	-8
T2	AMIT KESARWANI	+1	F	-8
4	CHRISTOPHER NELSON	-3	F	-6
5	KYLE WEST	E	F	-5

MOTIVATION AND OBJECTIVE

Current State of Golf Visualizations



shot tracing



leaderboard

- Golf has an abundance of player, course, and even shot data available, but little variation in analytical models or visualizations to show for it.
- Consumers interact with (and occasionally pay for) data-driven spectator companion applications in a variety of different sports, e.g. MLB At Bat, Fantasy Football.
- A golf companion app creates a new market for broadcast networks and tournament sponsors to reach consumer viewers, both in-person and on television.

Proposal: Golf Companion Dashboard

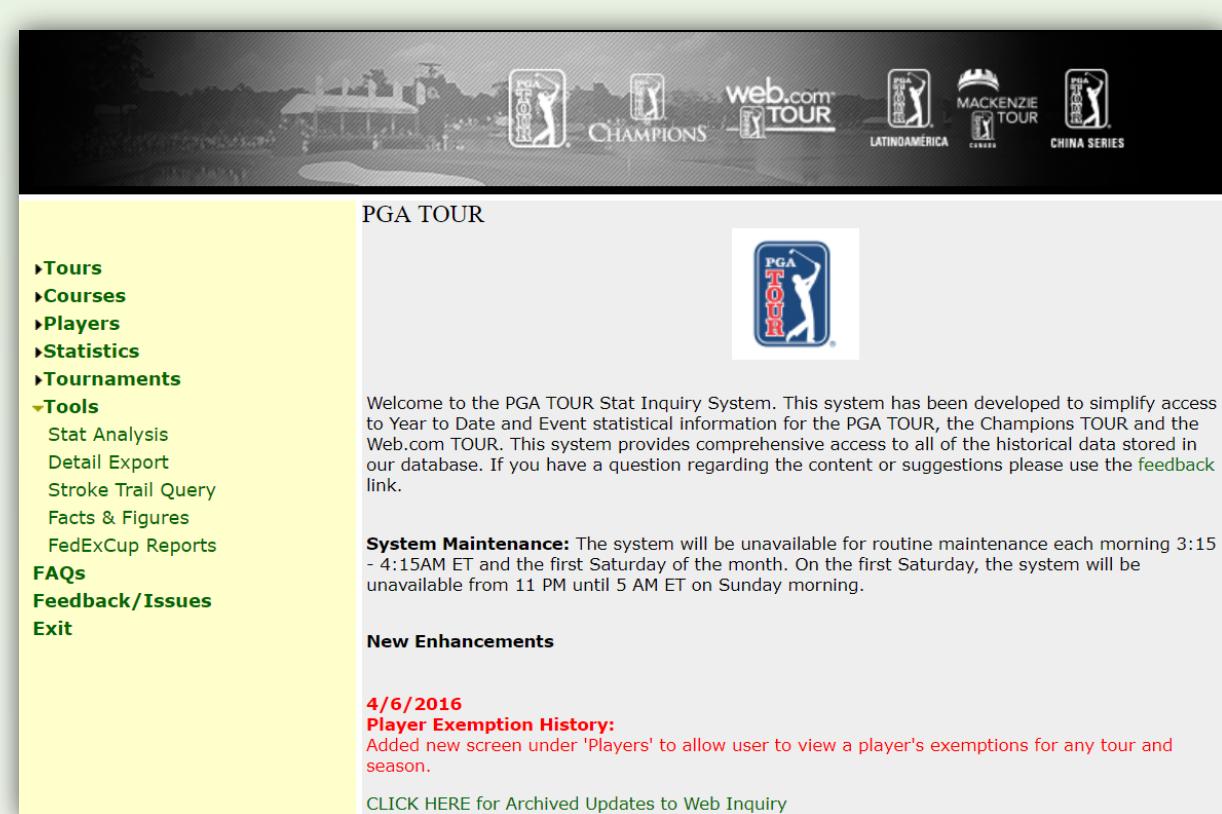
- Use ShotLink® data to model projected outcomes and set odds of live PGA TOUR events.
- Highlight statistically anomalous player performances to increase viewer engagement.
- Develop web-hosted dashboard of shot-by-shot results of PGA TOUR tournaments with real-time updates to predicted final standings.

SOURCE DATA

- Source data was requested and obtained from PGA TOUR ShotLink® proprietary database.
- PGA web portal provided access to standard and custom queries.

Team utilized data from 2004 – 2017 at the:

- Shot level: **16.4 MM** shots, **2045** unique players with **43** attributes.
- Hole level: **10098** holes, **4.1 MM** player-holes with **55** attributes.
- Round level: **2034** rounds with **171** attributes.
- Tournament level: **561** tournaments with **190** attributes.
- Course level: **110** courses with **33** attributes.



PREDICTIVE MODELING APPROACH

Top Innovations:

- Combines historical and real-time data to dynamically predict tournament outcomes and player odds of winning.
- Identifies statistically anomalous shots with the greatest impact on tournament outcome.
- Updates odds after each shot is processed and updates player projected final score after each hole is completed.

3-Step Modeling Approach to Determine Each Player's Final Projected Score:

Model 1: Baseline Score Prediction

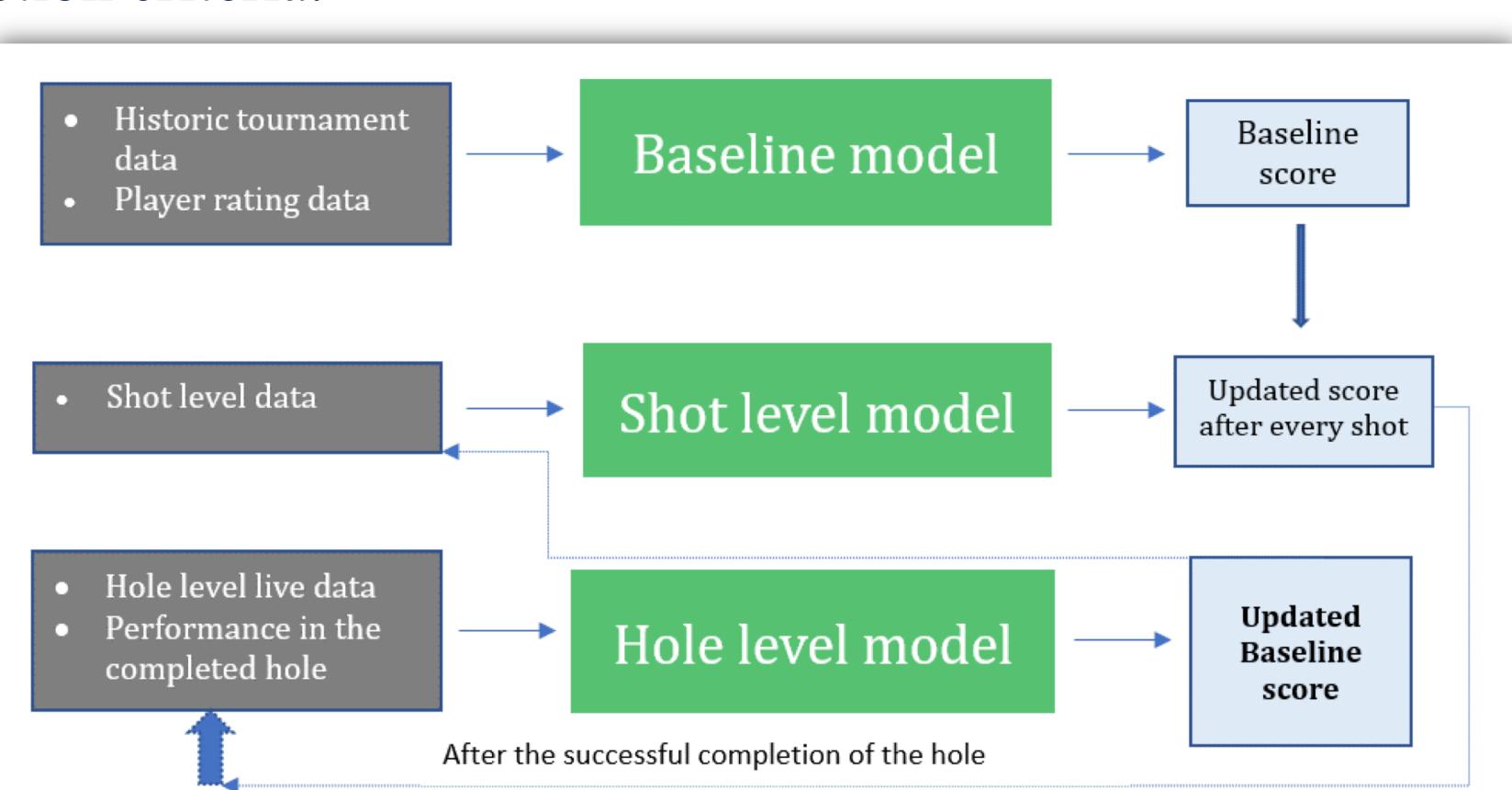
- Create baseline prediction for each player-hole within a tournament using historical data.
- Linear regression with 10-fold cross validation.

Model 2: Shot Level

- Update current player-hole prediction after each shot by calculating probability of finishing the hole in n shots.
- Multiclass logistic regression, Random Forest, gradient-boosting algorithms with 10-fold cross-validation and AUC selection criteria.

Model 3: Hole Level

- Update all remaining player-hole predictions upon completion of any single player-hole to account for variability due to course setup, weather, or any other day-to-day factors.



$$P(\text{Score on hole } H \mid \text{'Competitors perf. on } H \& \text{ Current player's perf. on holes 1,2,3 ... } H')$$

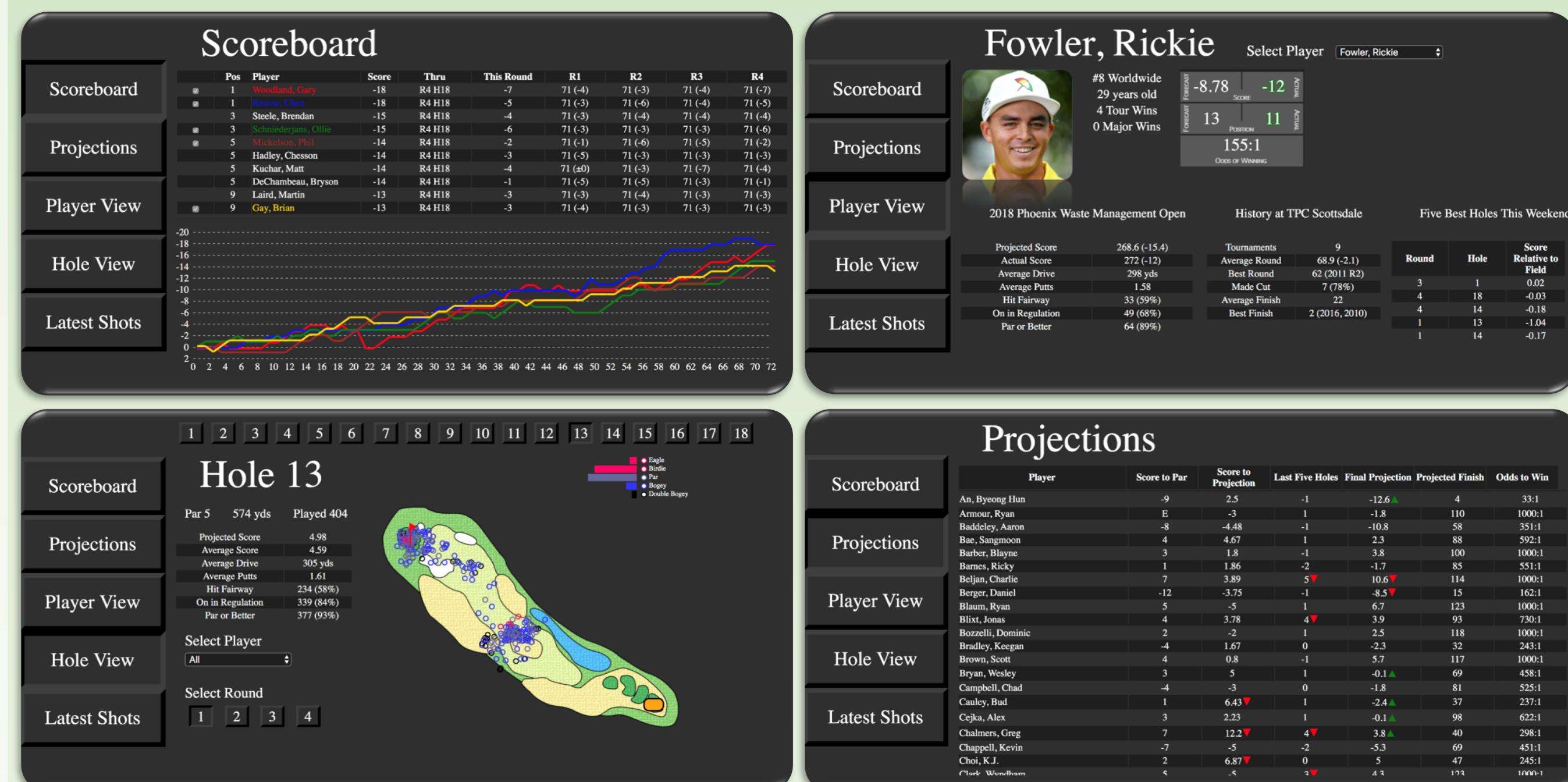
VISUALIZATION APPROACH

Top Innovations:

- Give consumers web-based access to real-time tournament predictions through an interactive UI allowing them to dive into specific shots or holes which have greatest impact on tournament outcome.
- Immediate access to the latest shots allowing consumers to “see” all holes on the course simultaneously.

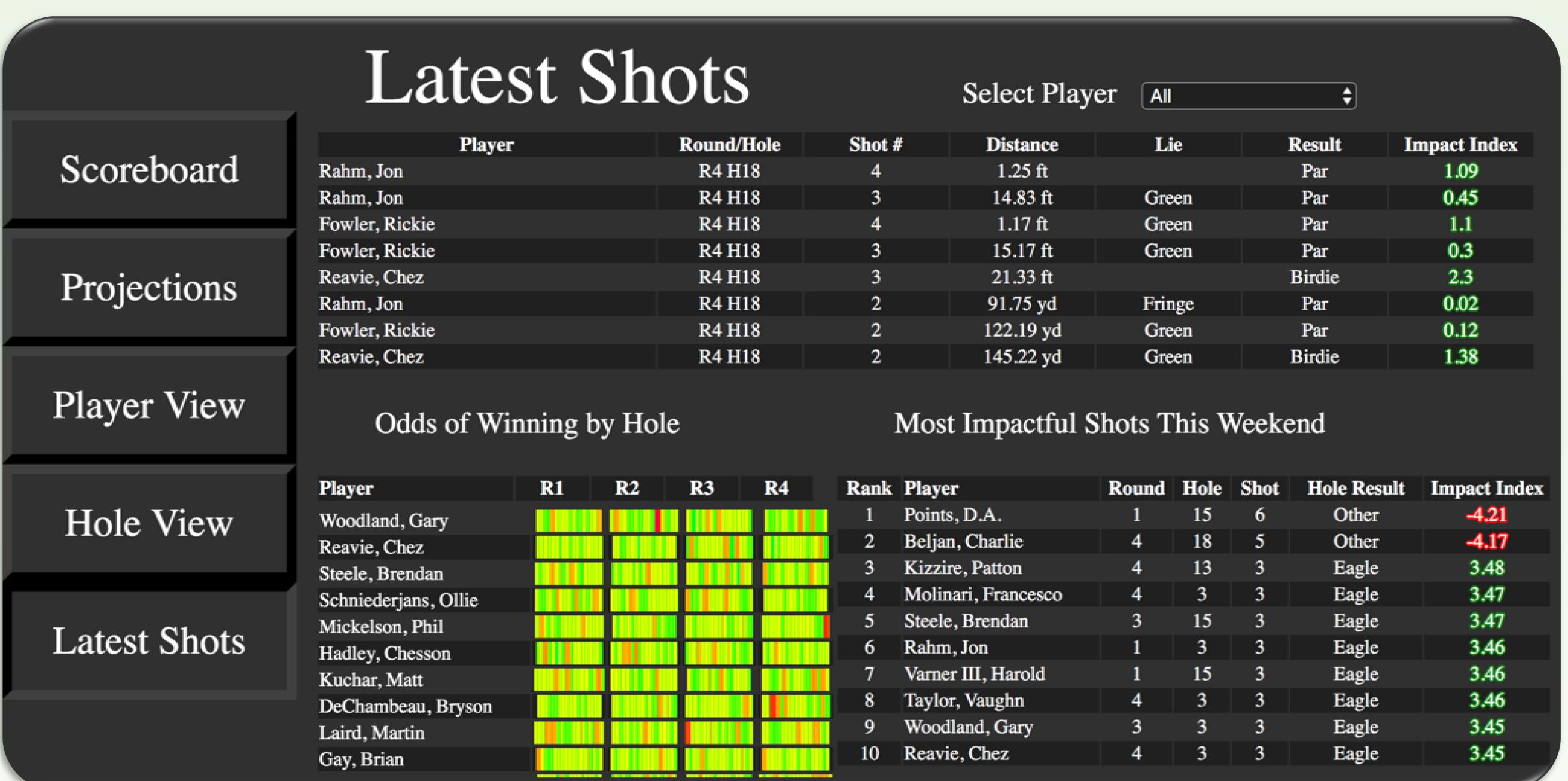
Consumer Experience – Select Views:

- Traditional leaderboard for consumer comfort fully integrated with predictive model.
- Custom selection to compare specific players' tournament predictions.



Latest Shot Update & Impact Index

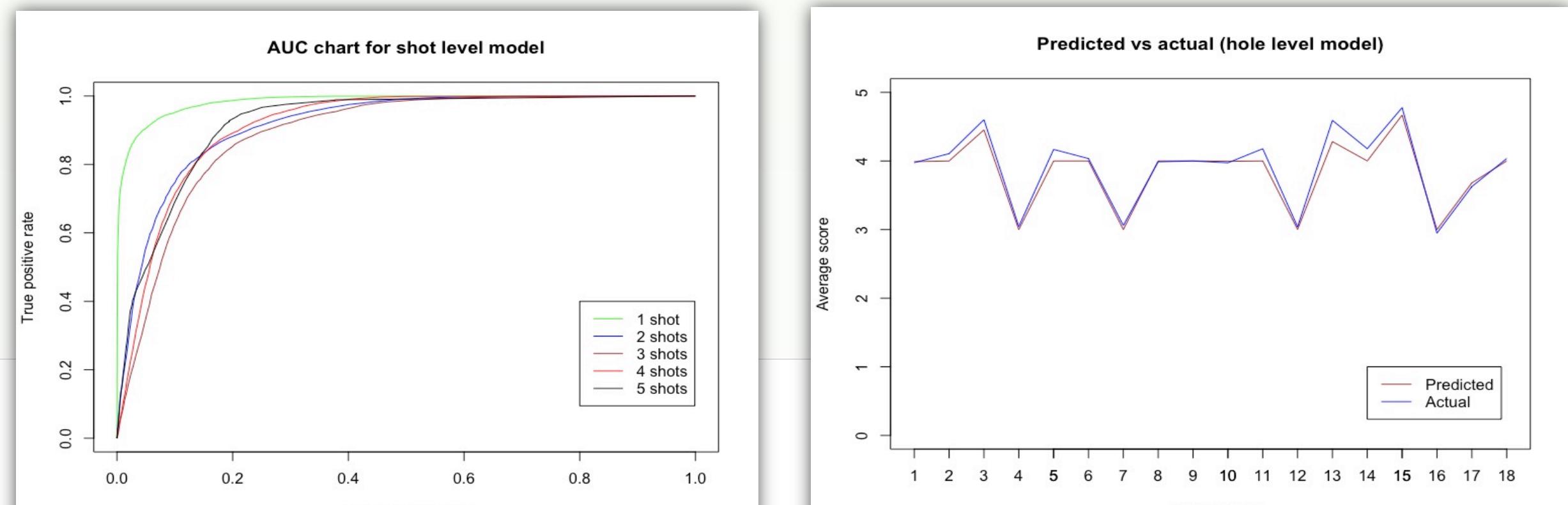
- Real-time access to all tournament shots, displayed with an ‘Impact Index’ to highlight shots that most significantly impact tournament outcome.



SUCCESS MEASURES AND EXPERIMENTATION RESULTS

Evaluating Success:

Determining Model Accuracy on a Test Set from 2018 Phoenix Waste Management Open



Shot Level

- The Shot Level model, using a Random Forest method, enjoyed 72% accuracy predicting remaining shots at each step of shot sequence for a given player-hole.
- The Hole Level Model, using Multi-Class Logistic Regression, showed only 2.2% absolute mean error for projected average score against actual average over a given round.

Other Success Indicators:

- The model-visualization pair gives consumers unprecedented access to familiar, although previously inaccessible data.
- The model combines multiple levels of aggregated performance data and runs ‘behind-the-scenes’ to output a simple score prediction for each player.
- Visualization interactivity allows consumers to keep an eye on the entire tournament field while focusing on players of their choice.