Model Card

Model Card: Linear Regression Model for NFL Quarterback Performance Prediction

Model Details

Model Name: NFL Quarterback Performance Prediction Model
Version: 1.0
Type: Linear Regression
Developer(s): Created by analyzing historical NFL player game statistics using R's `lm` function.
Date: November 2024

Input Data:
- Dataset: Player-level NFL game statistics from `merged_player_game.parquet`.
- Features: 16 numerical features, including completions, attempts, passing yards, passing touchdowns, interceptions, and rushing statistics.
- Training Data: Filtered observations of Tom Brady's career game statistics.
- Prediction Data: Patrick Mahomes' per-game averages, projected over the next 200 games.

Output: Predicted lifetime performance statistics for Patrick Mahomes over 200 additional games.

Intended Use

This model predicts the lifetime performance metrics of NFL quarterbacks by projecting their per-game averages over a hypothetical number of additional games. It was developed for comparative analysis of quarterbacks' career achievements.

Intended Use Cases:
- Performance projections for historical or active NFL quarterbacks.
- Exploratory analysis of career trajectories.

Not Suitable For:
- Real-time decision-making in NFL game contexts.
- Predicting short-term outcomes such as performance in a single game or season.
- Evaluation of quarterbacks without sufficient historical data.

Performance Characteristics

Evaluation Metrics: The model's fit and prediction accuracy were evaluated using the following metrics:
- Adjusted R^2 values for multivariate regression.
- Mean squared error (MSE) for the fitted model.

Performance on Training Data (Tom Brady):
- The model successfully captured trends in Tom Brady's career statistics, showing high adjusted R^2 values for most features.
- No missing data points were included in the training subset.

Projection Validation:
- Model predictions were cross-referenced with Mahomes' historical averages.
- Assumes statistical continuity in performance.

Considerations for Bias and Fairness

Potential Biases:
- Training data limited to Tom Brady may introduce biases reflecting his unique playing style and career context.
- Excludes contextual variables like team quality, injuries, and opponent defenses.

Fairness:
- Model does not account for demographic, geographic, or era-specific variations in playing conditions.
- Results may not generalize to quarterbacks with different profiles or shorter careers.

Limitations

1. **Contextual Limitations**: The model assumes that Patrick Mahomes' future performance mirrors his historical averages and does not account for potential changes in playing style, team composition, or career disruptions.

2. **Feature Scope**: Only numerical features were included. Non-numerical factors such as play-calling strategies, weather, or team dynamics are excluded.

3. **Era-Specific Differences**: Comparisons between quarterbacks from different NFL eras may not reflect genuine differences in skill due to rule changes or league trends.

4. **Linear Assumptions**: The linear model does not capture non-linear trends or interactions between features.

Ethical Considerations

Transparency:
- All inputs, feature selections, and assumptions are explicitly documented.
- Predictions are saved to a CSV file for reproducibility.

Responsibility:
- This model is designed for exploratory analysis and should not be used for high-stakes decision-making, such as draft picks or contract negotiations.

Accountability:
- Users are encouraged to validate predictions against domain expertise and supplementary data.

Recommendations for Future Work

1. **Data Diversity**: Expand training data to include quarterbacks with varied playing styles and career trajectories.

2. **Feature Enrichment**: Incorporate non-numeric and contextual features for a more holistic analysis.

3. **Model Enhancements**: Explore non-linear models (e.g., random forests, neural networks) to improve accuracy and capture feature interactions.

4. **Bias Mitigation**: Adjust for era-specific and demographic variables to enhance fairness in predictions.

Model Context

This linear regression model was built using the R programming language and the `lm` function. It demonstrates the value of leveraging historical sports data to inform predictions while emphasizing the importance of understanding and addressing model limitations. The results are intended to foster discussions about quarterback performance metrics and their broader implications.

Contact: For questions or feedback, contact the model developer at alexander.guarasci@mail.utoronto.ca.