Model Card - Basketball Statistics Forecast for Young NBA Players

Model

- Developed by students at the University of Toronto for a project course under supervision of Prof. Lisa Zhang
- Transformer model for sequence-to-sequence statistics prediction through time-series forecasting

Intended Use

- Particularly intended for recreational purposes to predict and compare different NBA players' careers eg. fantasy basketball
- Particularly intended for audience interested in NBA youth
- Possibly used by coaches and team scouts to forecast growth potential in terms of raw statistics for talented rookies
- Not suitable for determining player value or salary based on statistics
- Not suitable for betting purposes

Factors

- Aside from talent, other relevant factors affecting player statistics include overall team ranking, injuries, position, salary and other personal factors affecting physical health and training.
- All the statistics the model will predict are also heavily correlated and may act as factors for all other statistics eg. minutes played affects points scored.

Metrics

- Average mean squared error for all 12 statistics to measure months or statistics for which accuracy is higher (i.e. month 1 predictions more accurate than month 12, points scored predictions more accurate than minutes played)
- All data points are standardized to make the evaluation metric comparable across the 12 NBA statistics.
- Baseline accuracy reported as an average MSE of 0.6 using last season statistics.

Training Data

- nba_api[1], training data split
- Tuning hyperparameters with validation data split

Evaluation Data

- nba_api[1], testing data split
- Chosen as a basic proof-of-concept
- Includes more younger players with less experience in NBA to reflect intended use case

Ethical Considerations

- Model is meant to be used as a tool to help predict player performance statistics rather than a replacement for scouts in the professional industry.
- Model is not meant to provide sports betters an incentive to gamble nor to provide an unfair advantage to teams as it will be publicly available and free to use for everyone.
- Player data and statistics collected from companies pursuing studies in other areas such as player lifestyle and diet without full approval of certain players.

Caveats and Recommendations

- Model performs best with players who have "good" statistics as the training data has biases present towards players with longer careers who post "good" statistics throughout their career. Also ideal for 1 season predictions as compared to 2.
- Only takes into account personal statistics. Further work required to incorporate team performance and ranking as features in the model.
- Possible research could explore correlation between model performance and player predictions in order to optimize or create multiple models to target different positions in the lineup.

Quantitative Analyses

- The transformer model yields a MSE loss of approximately 0.5 under the training set and 0.53 under the validation set, lower than the baseline model.
- The first season has more accurate predictions than the second season, with MSE loss of approximately 0.46 vs. 0.55, respectively.