Basketball Eclipse Analytics Project: Free Throw Pressure Modeling

DOES PRESSURE FEATURES AFFECT FT%?

BY JAY WU AND ALAY ALAY NAWAB

Background

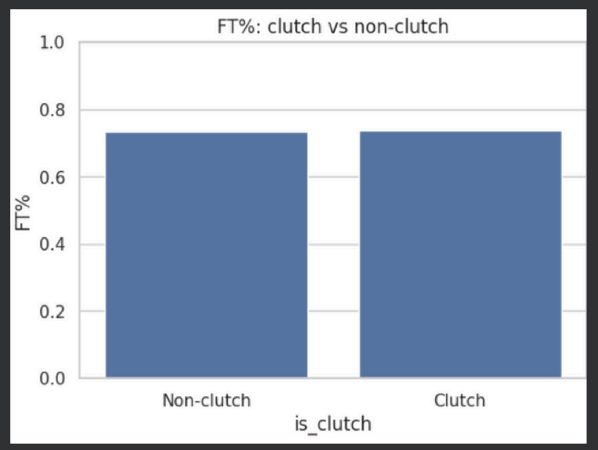
This project investigates whether the outcome of an NBA free throw can be predicted by combining player-level shooting skill with game-context variables that represent in-moment “pressure,” such as time remaining, score margin, and clutch situations. The team chose this topic because free throws often determine the outcome of close games, and there is a long-running debate about whether certain players perform differently under pressure. Free throws offer a controlled, repeatable event where skill and context interact in measurable ways, making them ideal for statistical modeling. Now, this model offers a realistic, high stakes use case for probabilistic classification, and model interpretation, which can be used for storytelling for coaches and fans.

Methodology

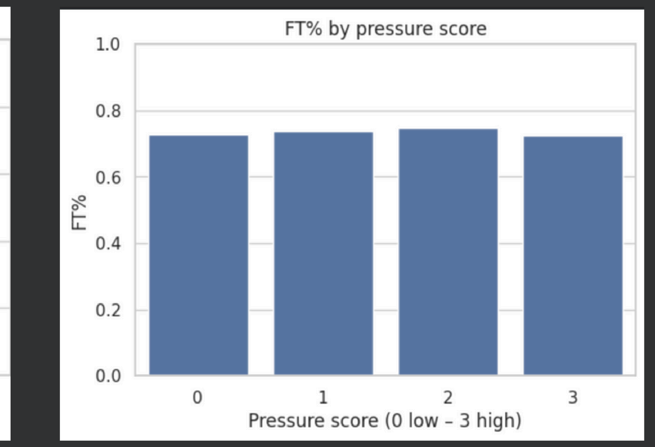
Event-level data were sourced from a public Kaggle NBA play-by-play dataset. From this dataset, all rows describing free-throw attempts were extracted to form the analysis subset. The team then engineered a binary target variable indicating whether each free throw was made, along with several player and context based features. Player skill features included season free throw percentage, career free throw percentage, clutch free throw percentage, and total career attempts. Game context features included period, time remaining, score margin, clutch/close game indicators, and a ‘combined’ pressure score which took all the context features into account to define how intense the in game situation is while the player makes a free throw .

For model preparation, the data were split by game into training and test sets to prevent leakage. Missing values were imputed, numerical variables were standardized, and all engineered features were verified during exploratory data analysis. Three models (Logistic Regression, Random Forest, and XGBoost), were tuned using Optuna, with PR-AUC chosen as the primary evaluation metric due to class balance considerations. The tuned Random Forest model achieved the strongest performance and was selected as the final model.

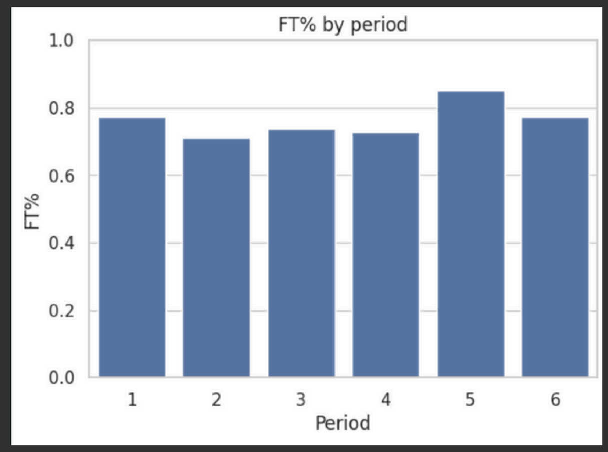
EDA



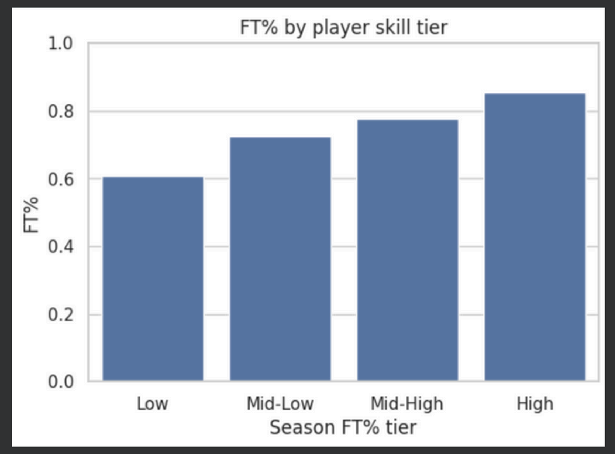
EDA 1: FT%: clutch vs non-clutch



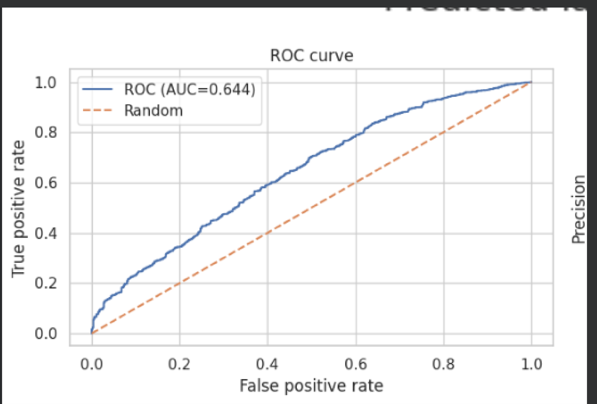
EDA 2: FT%: Pressure Score



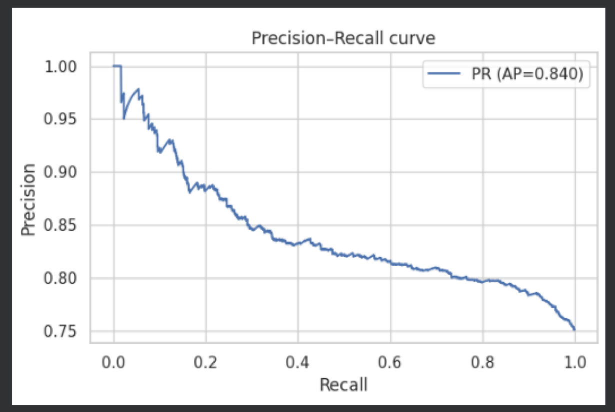
EDA 3: FT%: Period



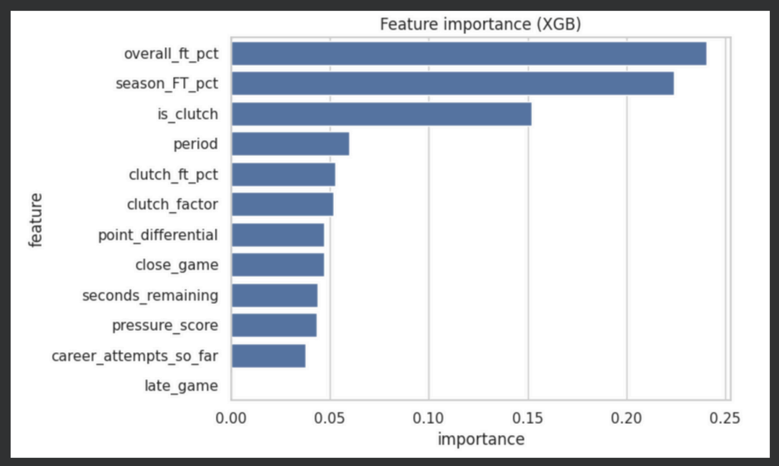
EDA 4: FT%: Player Skill Tier



ROC Curve



Precision-Recall Curve



Feature Importance

Exploratory plots included bar charts comparing free-throw percentage for clutch vs. non-clutch shots, free-throw percentage by game period, and free-throw percentage across player skill tiers defined by season FT%, which helped confirm that both time context and underlying shooting ability meaningfully affect outcomes. Additional EDA visualizations examined the distribution of engineered features such as score margin, seconds remaining, and the composite pressure score, showing that high-pressure situations are relatively rare but exhibit slightly different make rates than low-pressure ones. For the trained Random Forest, the team produced a confusion matrix, ROC and precision-recall curves, a calibration curve with Brier score, and a feature importance bar chart, giving a concise picture of accuracy, ranking quality, calibration, which features the model relies on most, and how these patterns align with the trends observed in the EDA.

Results

The tuned XGB achieved the best validation PR-AUC of the three models and showed strong performance on the held-out test set, with high accuracy, a high PR-AUC, and a low Brier score, indicating that it both ranks shots well and produces reasonably calibrated probabilities. The confusion matrix shows that the model correctly classifies most free throws, with very few cases where it predicts a make for a shot that was actually missed (false positives) and somewhat more cases where it predicts a miss for a shot that was actually made (false negatives), meaning it is conservative about calling makes.

These results broadly match expectations: player-skill features (season and career FT%) and pressure-related context (pressure score, time remaining, score margin) emerge as the most important variables in the XG Boost’s feature-importance ranking, which aligns with basketball intuition that good shooters and low-pressure situations lead to more makes. One mildly unexpected pattern in the EDA and model behavior is that free throw percentage under “high pressure” is not dramatically worse than in low pressure situations; this may reflect that the NBA players who frequently take high pressure shots are already strong free throw shooters, so their baseline skill offsets the added pressure, and that truly extreme clutch moments are relatively rare, limiting statistical power.

The project did not formally test statistical significance for every difference (for example, between each pressure score level), and in some segments(such as very high pressure combined with low volume players), the sample sizes are small enough that apparent differences could easily be due to random variation. Because of this, the team interprets the model’s performance and feature importance patterns as evidence of meaningful relationships rather than definitive causal claims, and treats any subtle differences between similar pressure buckets or player groups as suggestive rather than statistically proven

Conclusion

This project had several clear limitations that affected the final results. The dataset lacked certain variables that would have strengthened the analysis, meaning some patterns were inferred indirectly rather than measured directly. In addition to this, sample size was also relatively small, which limited the reliability of the conclusions. The modeling approach was also simple by design, so it did not capture more complex interactions that may exist in the data. If I were to repeat the project, I would start by expanding the dataset, either by collecting additional attributes or merging external sources to reduce noise and improve prediction stability. I would also use more rigorous validation methods and compare multiple models instead of relying on a single approach. In terms of next steps, the project could be extended by refining the feature set, running stronger baselines, and testing the model on new or unseen data to see if the model can be improved. Overall, the project establishes a working framework, but it can be made significantly more powerful with more data, deeper experimentation, and systematic validation.