
Application of Classification Models for Predicting Shooting Performance in the NBA

A Case Study of Kobe Bryant

Allen Crane

Brock Friedrich

Examiner: Dr. Anthony Tanaydin

Southern Methodist University

Data Science

Applied Statistics

December 1th, 2018

Writing period

10. 24. 2018 – 12. 1. 2018

Abstract

In a world with continuously increasing technological capabilities, the demand for answers increases alongside it. The domain of professional sports in particular is no exception to this trend. The very essence of competitive sports is rooted in a desire to optimize and gain every advantage over competitors. In this paper, we describe classification methodologies for predicting the performance of professional athletes through a scoped analysis of the long-time NBA allstar, Kobe Bryant. We detail three simulations modelling shooting accuracy over Bryant's 20 year career, using logistic regression and discriminant analyses to yield predictions about Bryant's theoretical future performance. We show that Bryant's potential to score is strongly dependent on his distance from the basket when shooting. We also show that Bryant's scoring performance in the post-season, as compared to that of the regular season, is consistent with his performance at any other point in the season.

Contents

1	Introduction	1
2	Approach	3
2.1	Problem Definition	3
2.2	Exploratory Analysis	3
2.3	Evaluation	3
2.4	Illustration	3
3	Methodology	5
3.1	Establish Notation	5
3.2	Feature Selection	5
3.3	Model Selection	5
3.4	Evaluation	5
3.5	Discussion	5
4	Conclusion	7
5	Limitations and Future Directions	9
6	Appendix	11
.1	Appendix A	11

List of Figures

List of Tables

List of Algorithms

1 Introduction

With modern computers and the popularity of the budding discipline of data science, there are numerous predictive technologies readily available. In professional sports alone, a host of disparate entities, players, organizations, and gamblers alike are hungry for new insights. Each bears an economic imperative to seek advantages and optimizations that will give them an edge over their peers. Technology, however, does not predicate useful and effective models. It is the rigorous application of sound statistical theory that yields actionable conclusions. Unfortunately, the appeal of novel algorithms and the sparkles of big data detract from the importance of a strong statistical foundation. Thus, we seek to right the ship of this rampant malpractice as it applies to this problem domain. In what follows, we explore insights we have garnered from the application of linear discriminant analysis and logistic regression, then leveraging those insights to appropriately tune the final predictive model.

2 Approach

2.1 Problem Definition

This study was birthed out of a need to reliably quantify three key metrics used in assessing a profession basketball player's shooting ability. Those metrics are:

1. Odds of making a shot as distance from the basket increases.
2. Linearity of the decline rate of the probability of making a shot with respect to the distance the shot was taken from the basket.
3. The relationship between the distance from the shooter to the basket and the odds of the shot being made is different when in the regular season verses the post-season.

Rephrasing those metrics as questions helps clarify the focus of the analysis as it relates to Kobe Bryant. Those questions are as follows:

1. Do the odds of Kobe making a shot decrease with respect to distance he is from the hoop?
2. Does the probability of Kobe making a shot decrease linearly with respect to the distance he is from the hoop?
3. Is the relationship between the distance Kobe is from the basket and the odds of him making the shot different if they are in the playoffs.

To appropriately answer the questions of interest above, we must fit a series of classifier models to the training portion of our dataset, iteratively scoring and comparing the various models against one another so that we can tune their parameters and hone in on a final feature set. Below, we set the stage for conducting our analysis through an exploratory analysis of the dataset.

2.2 Exploratory Analysis

2.3 Evaluation

2.4 Illustration

3 Methodology

3.1 Establish Notation

Outline methods and results

3.2 Feature Selection

3.3 Model Selection

3.4 Evaluation

Simulations (validity)

Model Comparison

Illustration

Application (how is this study practically relevant?)

3.5 Discussion

Summarize methodology and "mark territory"

Write last

Potential follow up studies (where you could go with this)

Found interesting results for this population, but

4 Conclusion

Answer these questions with our results:

The odds of Kobe making a shot decrease with respect to the distance he is from the hoop. If there is evidence of this, quantify this relationship.

The probability of Kobe making a shot decreases linearly with respect to the distance he is from the hoop. If there is evidence of this, quantify this relationship.

The relationship between the distance Kobe is from the basket and the odds of him making the shot is different if they are in the playoffs. Quantify your findings with statistical evidence one way or the other.

Include Post hoc analysis:

5 Limitations and Future Directions

Limitations : - Only Kobe's career.

Future Directions: - How much can we limit the dataset without drastically reducing the predictive capability of the model? - Would provide insight into the effectiveness of the model when predicting on non-veteran players.

6 Appendix

.1 Appendix A

Appendix AAA