

Report Title*

(COMP3125 Individual Project)

*Note: Do not used sub-title

Matthew Ralon
dept. name of organization

Abstract—This project explores the intersection of data science, sports analytics, and betting strategy by examining historical NBA data to evaluate predictive models and financial outcomes. Using team metrics, injury reports, betting odds, this study investigates whether statistical features can reliably predict game results against “the spread”. By using a quantitative model to assess predictive accuracy, comparing analyses measure the impact of injuries on line reliability. Simulated betting strategies are tested to determine potential ROI across various wagers a person can take.

Keywords—example1, example2, example3, example 4, example 5 (provide 3-5 keywords)

I. INTRODUCTION (HEADING I)

Sports betting has emerged as one of the fastest-growing industries in the United States, with the legalization of wagering in almost every state fueling both economic expansion and academic interest in predictive analytics. The NBA provides a compelling case study because of the sheer amount of statistical data that gets generated game by game across the entire season and the popularity of basketball among bettors. The abundance of data creates a perfect opportunity to examine how quantitative models can be used to forecast outcomes and evaluate if there are viable betting strategies that can be applied. This project focuses on the relationship between team performance metrics, player availability, and betting market behavior. By analyzing both offensive and defensive ratings, turnovers, and other statistics, the study investigates whether it is possible to reliably predict game results with relation to the spread. At the same time, by considering how player injuries affect both sides of the coin, the betting lines, and the team performance, the project considers important factors that get overlooked in simple number analysis. Finally, by simulating betting strategies—Moneyline, spread, over/under, etc. we can evaluate the ROI values and connect predictive modeling to financial decision making.

II. DATASETS

A. Source of dataset (Heading 2)

The datasets that were used in the project were obtained from publicly available and credible providers that focus on sports statistics and betting information. Historical NBA game data and statistics were collected from [Basketball Reference](#); they are a widely recognized source for a variety of basketball statistics. The dataset includes box scores, team ratings, player contributions that are generated and still maintained by the site's authors game by game.

The betting line information was gathered from [Sportsbook Reviews](#) and Kaggle sports [betting datasets](#) which combine betting markets across multiple seasons. The datasets

were generated by scraping sportsbook odds and placing them into tables that make it easier to analyze.

Injury reports and player availability were retrieved from the official [NBA website](#) as well as [third party](#) roster trackers that post real time updates regarding player injuries, likelihood to play, etc. These sources are updated daily as teams must post them before their games.

B. Character of the datasets

Dataset	Format	Size	Key Features	Cleaning/Merging
Basketball Reference Injury Report	HTML Tables	1k+	Player name, team, start/end of injury, games missed	Convert to a CSV
Basketball betting Dataset	CSV	Roughly 50k rows	Game ID, date, teams, spread, moneyline, over/under	Clean missing odds, normalize odds format
NBA Stats and Betting Data	CSV	Roughly 100k rows	Team stats, offensive/defensive rating, game outcome	Remove duplicate games, align with the other datasets using dates
Sports book Review NBA Odds	HTML/CSV	Thousands of games	Bookmaker odd tracker, spread, totals	N/A

III. METHODOLOGY

In this part, you should give an introduction of the methods/model. First, what's the method/model. What's the assumption of this method/model. What's the advantage/disadvantage of this method/model. Why did you choose it. What Python module or function do you apply to apply this method/model. Any optional input/extrawork did you adjust to make the results better. If you have multiple methods, feel free to use subsection A., B. to separate them.

A. Method A

Method/Model: Logistic Regression, Random Forest, and XGBoost classifiers were applied to predict win/loss outcomes against the spread. Linear Regression and Gradient Boosting regressors were used to estimate the point differentials and totals

Assumptions: Team performance metrics are predictive of game outcomes, historical relationships between features and outcomes remain consistent across seasons

Advantages/Disadvantages: Logistic Regressions provides the ability to interpret the data by using coefficients to show importance. Random Forest and XGBoost show nonlinear relationships and interactions. However, logistic regression can underperform, and models risk overfitting if they are not tuned correctly

Why: These models balance the ability to interpret while still having predictive power, which is exactly what we need/want when studying sports analytics

Python Modules: pandas, scikit learn, XGBoost

Adjustment: Parameter tuning, feature scaling

B. Method B

Method/Model: Statistical tests such as t-tests and chi-squared tests are applied to measure the effect of player absence on line accuracy. An Impact score was created to weight absences by player and minutes played

Assumptions: Player availability SIGNIFICANTLY alters team performance and the reliability of betting lines. Historical injury data represents the impact on outcomes.

Advantages/Disadvantages: Statistical tests give clear evidence of the factors beyond just performance. However, injury data can be incomplete, or teams “lie” which can reduce accuracy.

Why: Injuries are a critical factor that often gets overlooked in other predictive models. By incorporating them we can improve the realism of the model

Python Modules: pandas, numpy, scipy

Adjustments: Starter vs bench to give better accuracy/weight to affect

C. Method C

Method/Model: Simulated betting strategies (moneyline,spread,over/under) were applied to outcomes using model predictions. ROI was calculated as cumulative profit/loss over time

Assumptions: Historical odds and outcomes reflect betting conditions and simulated strategies reflect real bettor behavior/bets

Advantages/Disadvantages: ROI analysis connects predictive modeling to financial outcomes, which in turn shows practical limitations. However, simulations cannot account for external factors such as books limiting a user, live market shifts, bettor psychology.

Python Modules: pandas, numpy, matplotlib, seaborn

Adjustments: Normalized odds formats, applied some bankroll management (betting certain units, mega bets), and compared strategies across seasons

IV. RESULTS

In this section, present your findings using an appropriate method, such as equations, numerical summaries, or visualizations like charts and graphs. Clearly explain all results and provide guidance on how to interpret them. If any unexpected results arise, discuss possible reasons or contributing factors. To improve clarity and organization, consider using subsections (e.g., A, B) to separate different aspects of your results.

Example: After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. Result A

Example: XXX

1) *For papers with more than six authors:* Add author names horizontally, moving to a third row if needed for more than 8 authors.

2) *For papers with less than six authors:* To change the default, adjust the template as follows.

a) *Selection:* Highlight all author and affiliation lines.

b) *Change number of columns:* Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

c) *Deletion:* Delete the author and affiliation lines for the extra authors.

B. Results B

Example: Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

C. Results C

a) *Positioning Figures and Tables:* Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

TABLE I. TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
copy	More table copy ^a		

^a. Sample of a Table footnote. (Table footnote)

Fig. 1. Example of a figure caption. (figure caption)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes

with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

V. DISCUSSION

Every method/project has its shortage or weakness. Please discuss the unsatisfied results in your project. And discuss the feasible suggestions of future work to revise/improve your result.

Example: xxx

VI. CONCLUSION

In this part, you should summarize your project. What important results did you find for your topic and what's the effect of this result on the real-world?

Example: xxx

ACKNOWLEDGMENT (*Heading 5*)

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Use the IEEE format for the citation. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...” Unless there are six authors or

more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

- [1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (*references*)
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] K. Elissa, “Title of paper if known,” unpublished.
- [5] R. Nicole, “Title of paper with only first word capitalized,” *J. Name Stand. Abbrev.*, in press.
- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [*Digests 9th Annual Conf. Magnetics Japan*, p. 301, 1982].
- [7] M. Young, *The Technical Writer’s Handbook*. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord “Format” pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.