Beyond Propensity Scores: Investigating College Athletic Success with Random Forest Techniques

PSM Group: Ty McClure, Natalie Mikosz, Anshuman Puppala, Ethan Tabiang, and Juan Rodriguez
Texas A&M University - Dr. M. Jahangir Alam - ECON470:500 - Spring 2024
May 7, 2024

Abstract

In this study, we focused on refining Anderson's analysis, which assessed the impact of successful athletic programs on institutional success. Anderson originally employed bookmaker data to identify college football favorites and used a propensity score design to measure the effects of collegiate athletic success specifically success in football programs. We advanced this approach by implementing Random Forest regression, which enhanced the decision trees for more precise predictions and more accurate and robust propensity scores. Our enhanced Random Forest analysis showed that an additional football win resulted in a much smaller, more modest effect on alumni athletic operating donations (increase of 10,790) and an equally notable decrease in the magnitude of the effect on non-athletic donations (decrease of -23,970). Importantly, it led to a large negative effect on total alumni donations which contrasted to the originally large positive effect (267,400 to -228,100). These results underscore the value of Random Forest regression in providing a nuanced understanding of the relationships between athletic success and institutional health.

Introduction

- **Anderson's Objective:** Examine the impact of football wins on university health metrics like donations and acceptance rates.
- **Methodology:** Used propensity score matching based on team win likelihood.
- **Enhancement Strategy:** Implement Random Forest for more accurate propensity scores, reduced overfitting, and decreased endogeneity.
- **Hypothesis:** Random Forest will yield more precise outcomes and uncover deeper variable relationships.

Effects of Football Wins on Outcomes

Table: Effects of Football Wins on Various Outcomes

	STE				Replication				ML			
	Coeff.	P-val	Conf. Int.		Coeff.	P-val	Conf. Int.		Coeff.	P-val	Conf. Int.	
Alumni Athletic Operating Donations	191,200 (65,000)	0.001	63,800	-318,600	191,240 (65,035)	0.004	63,700	-318,700	10,790 (13,100)	0.412	-15,000	-36,600
Alumni Nonathletic Operating Donations	-137,400 (96,100)	0.210	-325,756	-50,956	-137,410 (96,077)	0.156	-325,700	-50,900	-23,970 (66,600)	0.719	-155,000	-107,000
Total Alumni Donations	267,400 (266,900)	0.450	-255,744	-790,544	267,380 (266,945)	0.319	-255,800	-790,600	-228,100 (173,000)	0.187	-567,000	-111,000

Methodology

Methodology:

- Utilized **Machine Learning Techniques**: Random Forest and Logistic Regression.
- Calculated propensity scores to adjust for potential confounders.
- Employed nearest neighbors matching to compare schools with similar profiles but different levels of athletic success.

Results:

- More robust analysis due to inclusion of ML techniques.
- Revealed nuanced insights into how changes in sports performance influence donation behavior.

Findings

- **Propensity Score Matching (PSM) Results:**
  - Additional football win led to a non-significant increase in athletic donations.
  - Non-significant decrease of -137,412 observed in non-athletic donations.
  - Substantial growth in overall alumni donations of 267,379 despite the above.
- **Advanced Machine Learning (ML) Techniques Results:**
  - Smaller, yet more precise adjustments in donation patterns.
  - Athletic donations showed a non-significant change of 10,790 (p-value = 0.412).
  - Significant decrease in non-athletic donations from STE with -137,412 to -23,970.
  - Total alumni donations significantly positively affected, albeit with a smaller impact magnitude than PSM indicated.
- **ML Implications:**
  - Nuanced results highlight the value of integrating ML methodologies.
  - Enhanced understanding of the dynamics between collegiate sports success and alumni donation behaviors.
- **ML Statistically Significant Values:**
  - Academic Reputation and Application proved to be significant as seen in the final report as seen in Figure 4 in the final report.
  - The inclusion of alumni athletic donations, alumni nonathletic donations, and total donations in this poster is due to their focus in the original paper.

Discussion

- **Methodological Comparison**:
  - **STE**: Large increase in donations post-football wins.
  - **ML**: More conservative, often non-significant results.
- **Alumni Donations**:
  - **Athletic**:
    - **STE**: Coefficient: 191,200 ($p = 0.001$).
    - **ML**: Coefficient: 10,790 ($p = 0.412$).
  - **Nonathletic**:
    - **STE**: Coefficient: -137,400 ($p = 0.210$).
    - **ML**: Coefficient: -23,970 ($p = 0.719$).
- **Total Donations**:
  - **STE**: Coefficient: 267,400 ($p = 0.450$).
  - **ML**: Coefficient: -228,100 ($p = 0.187$).
- **Implications & Strategy**:
  - Possible donation shifts towards athletic programs.
  - Recommend reviewing university financial strategies.

Conclusions

- **Random Forest Application Findings:**
  - Smaller impact of football wins on alumni donations.
  - Athletic donations: Coefficient of 10,790.
  - Nonathletic donations: Coefficient of -23,970.
  - Overall negative relationship in total donations.
- **Implications for University Funding:**
  - Possible reduction in sports funding due to low returns or losses.
  - Potential improvement in educational facilities and institutional health.
- **Future Research Directions:**
  - Explore impacts of different university departments on institutional health.
  - Identify programs yielding the greatest returns for strategic funding.
  - Potential to enhance higher education quality nationwide.

Literature Review, Distribution Graphs, References, Google Colab Links

Anderson, Michael L. 2017. "The Benefits of College Athletic Success: An Application of the Propensity Score Design." The Review of Economics and Statistics 99(1): 119–134

Zhao, Peng, et al. 2016. "Propensity score and proximity matching using random forest." Contemporary Clinical Trials 47(1): 85–92.