Nonprofit Grant Success Predictor

AI-powered Analysis of Factors Influencing Nonprofit Grant Success

Prepared by: Ouahabi Abdessamed

October 2025

Data Science Research Report

Executive Summary

This project leverages artificial intelligence (XGBoost) to analyze data from nonprofit organizations and identify the most influential factors that determine how nonprofits win and effectively manage grants. Using a synthetic dataset of over 240,000 nonprofits, the analysis models real-world grant allocation patterns to understand drivers of success and impact efficiency.

Dataset Overview

- **Total records:** 240,585 nonprofits
- **Total features:** 37 columns
- **Source:** Kaggle – Nonprofit Organizations and Grants (Synthetic Data)
- **Key features:** `REVENUE_AMT`, `financial_metric`, `anomaly_score`, `impact_score_numeric`, `PF_FILING_REQ_CD`, `FILING_REQ_CD`, `INCOME_AMT`, `INCOME_CD`, `ICO`, `TAX_PERIOD`, and `impact_efficiency` (target)

The dataset combines financial, operational, and impact indicators that collectively describe nonprofit performance and risk characteristics. The primary determinants of grant success are related to financial performance and risk management.

Methodology

The following data science pipeline was implemented:

1. **Data Cleaning:** Missing values were imputed, and categorical fields were label-encoded.
2. **Feature Selection:** Irrelevant text columns were removed to optimize the model.
3. **Model Training:** An XGBoost Regressor was trained to predict `impact_efficiency`.
4. **Evaluation Metrics:** R^2 and Mean Absolute Error (MAE) were used to measure performance.
5. **Feature Importance:** Model-derived importance scores were used to identify top predictors of success.

Model Results

The XGBoost model achieved outstanding predictive accuracy on the synthetic nonprofit dataset.

Metric	Value
R^2 Score	0.9993
Mean Absolute Error (MAE)	0.00041

Table 1: Updated model performance metrics for predicting nonprofit impact efficiency.

The model explains nearly 99.93% of the variance in the target variable (`impact_efficiency`), with an extremely low mean error. These results confirm the effectiveness of the XGBoost approach for modeling grant success patterns among nonprofits. However, given the synthetic nature of the dataset, the results should be interpreted as indicative of model capability rather than real-world generalization.

Feature Importance Analysis

The analysis revealed that the following features are most predictive of nonprofit grant success, based on model-derived importance scores:

- `REVENUE_AMT` (importance: 0.578) – Total revenue of the organization, the most influential factor.
- `financial_metric` (importance: 0.381) – Additional financial indicators reflecting the organization's financial health.
- `anomaly_score` (importance: 0.032) – Measure of operational risk or unusual patterns.
- `impact_score_numeric` (importance: 0.008) – Quantitative measure of organizational impact.

These findings suggest that financial performance and risk management are the primary determinants of grant success among nonprofits.

Key Insights and Recommendations

- Organizations with higher total revenue and strong financial metrics are more likely to secure grants.

- Managing operational risks (low anomaly scores) enhances donor trust and funding likelihood.
- While impact scores contribute to success, their effect is minor compared to financial factors.

Conclusion

This study demonstrates that grant success among nonprofits can be effectively modeled through data-driven methods. By leveraging XGBoost and structured nonprofit data, stakeholders can gain actionable insights into which organizational characteristics most influence efficiency and funding outcomes. These insights can inform both nonprofit strategy and grant-making decisions.

Acknowledgments

The dataset used in this project originates from Kaggle:

Pooja Yakkala, Nonprofit Organizations and Grants (Synthetic Data), Kaggle, 2025.

Prepared by: Ouahabi Abdessamed