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:** `confidence_score`, `data_quality`, `has_mission`, `has_financial`, `has_impact`, `risk_level`, and `impact_efficiency` (target)

The dataset combines organizational, financial, and qualitative indicators that collectively describe nonprofit performance and risk characteristics.

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.

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

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

Model Results

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

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:

- `confidence_score`
- `data_quality`
- `has_financial`
- `has_mission`
- `risk_level`

These findings suggest that organizational transparency, financial reporting, and clear mission alignment play a critical role in determining grant success.

Key Insights and Recommendations

- Nonprofits with complete and verifiable financial data are more likely to secure grants.
- A well-defined mission statement strongly correlates with higher impact efficiency.
- Managing operational risks (low anomaly scores) enhances donor trust and funding likelihood.

- Consistently high data quality improves the organization's reputation and access to funding opportunities.

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