Table of Contents

[**1.** **Predicting Company Ultimate Status: A Machine Learning Analysis Report**](#_Toc189778557)

[**2.** **Data Overview and Preprocessing** 2](#_Toc189778558)

[**3.** **Methodology** 2](#_Toc189778559)

[**3.1.** **Feature Engineering** 2](#_Toc189778560)

[**3.2.** **Model Selection and Implementation** 3](#_Toc189778561)

[**4.** **Results and Business Insights** 3](#_Toc189778562)

[**4.1.** **Model Performance** 3](#_Toc189778563)

[**4.2.** **Key Findings** 3](#_Toc189778564)

[**5.** **Insights** 3](#_Toc189778565)

[**5.1.** **Corporate Profiles** 3](#_Toc189778566)

[**6.** **Conclusions and Recommendations** 4](#_Toc189778567)

# **Predicting Company Ultimate Status: A Machine Learning Analysis Report**

This analysis aimed to develop a machine learning model to predict whether companies are Domestic or Global Ultimates based on their operational and financial characteristics. Understanding ultimate status is crucial for mapping corporate hierarchies and relationships in the business world.

# **Data Overview and Preprocessing**

The dataset comprised approximately 29,000 companies with various operational metrics including employee counts, sales figures, and company characteristics. A significant challenge in preprocessing was handling the hierarchical nature of company relationships.

We approached data cleaning systematically, first addressing missing values through industry-specific median imputation. This method was chosen because companies within the same industry typically share similar operational structures and scales. For instance, manufacturing companies tend to have comparable employee-to-sales ratios, making industry-specific medians more reliable than overall dataset medians.

Outlier treatment utilized the Interquartile Range (IQR) method rather than z-score normalization. This decision was based on the inherently skewed nature of business data – large corporations legitimately have employee counts and sales figures orders of magnitude larger than smaller companies. The IQR method better preserves these legitimate outliers while removing truly anomalous data points.

# **Methodology**

## **Feature Engineering**

Feature engineering focused on creating meaningful ratios that could capture company relationships regardless of absolute size. Key derived features included:

1. Employee Domestic Ratio: This metric compares single-site employees to domestic total employees. A ratio close to 1 suggests the company operates primarily from one location, often indicating domestic ultimate status.

2. Employee Global Ratio: Comparing domestic total employees to global total employees helps identify global ultimate status. Lower ratios often indicate significant international operations.

3. Sales Global Ratio: This provides a financial perspective on the company's global presence, complementing the employee-based metrics.

## **Model Selection and Implementation**

We selected Random Forest Classification for several compelling reasons. First, business hierarchies often involve non-linear relationships that simpler models like logistic regression might miss. Random Forest excels at capturing these complex patterns through its ensemble of decision trees.

Additionally, Random Forest provides built-in feature importance rankings, crucial for understanding which factors most strongly indicate ultimate status. The model's ability to handle mixed data types (both categorical and numerical) without extensive preprocessing made it particularly suitable for our dataset.

# **Results and Business Insights**

* 1. **Model Performance**

The Random Forest classifier achieved [insert accuracy] accuracy for domestic ultimate prediction and [insert accuracy] for global ultimate prediction. More importantly, the model revealed several key patterns in corporate structures.

## **Key Findings**

Analysis of feature importance revealed that employee ratios are the strongest predictors of ultimate status. Companies with concentrated employee bases (high domestic ratios) are more likely to be domestic ultimates, while those with distributed workforces typically indicate global ultimate status.

Company age emerged as a secondary but significant predictor. Established companies (>15 years) are more likely to be ultimates of either type, possibly reflecting the time needed to develop complex corporate structures.

# **Insights**

## **Corporate Profiles**

Our analysis revealed distinct profiles for different ultimate types:

Domestic Ultimates typically show:

- Employee concentration in one country

- Stable, long-term operational history

- Focus on local or regional markets

- Strong presence in service-based industries

Global Ultimates commonly exhibit:

- Distributed workforce across multiple countries

- Significant variation between domestic and global sales

- Presence in manufacturing or technology sectors

- Complex ownership structures

# **Conclusions and Recommendations**

This analysis provides a robust framework for predicting company ultimate status. The model's insights can inform various business decisions, from market analysis to merger evaluations.

We recommend:

1. Developing industry-specific models for sectors with unique characteristics

2. Incorporating geographic clustering analysis for regional patterns

3. Adding time-series features to capture corporate structure evolution

These enhancements would further refine the model's predictive capabilities and business value.