Can a Nation's Food Economy Predict World Cup Success?

An Empirical Analysis Using Socio-Economic, Nutritional, and Machine Learning Approaches

Emir Eğilli
Sabancı University – DSA210 (Fall 2025)

---

1. Introduction & Motivation

International football success is often attributed to talent, coaching, and tactical excellence. However, long-term athletic performance is also shaped by broader socio-economic conditions, including nutrition, income levels, and access to health resources. National food economies provide a useful proxy for understanding living standards and physical well-being at a population level.

This project investigates whether food supply and related socio-economic indicators can help explain or predict international football success, measured by a country's ability to reach the Top-8 stage in the FIFA World Cup.

The central research question is:

Can a nation's food economy, combined with economic controls, help predict World Cup success?

---

2. Data Sources

Multiple publicly available datasets were combined to construct a panel dataset at the country–World Cup level.

Food & Nutrition Indicators (FAOSTAT):

- Daily calorie supply (kcal/person/day)
- Protein supply (g/person/day)
- Meat, sugar, dairy, fruit, and vegetable supply per capita

Economic & Demographic Indicators (World Bank / WHO):

- GDP per capita (USD)
- Total population
- Urban population share
- Unemployment rate

- Health expenditure (% of GDP)
- Obesity prevalence among adults

Football Performance Data:

- FIFA World Cup final stage outcomes by country and year
- Countries reaching the quarter-finals or better were labeled as **Top-8**

To ensure temporal relevance, all explanatory variables were averaged over the **1–3 years preceding each World Cup edition**.

---

3. Data Preparation

Country names were standardized across all datasets to enable consistent merging. Observations with insufficient data coverage were excluded. The final dataset is a structured panel where each observation represents a country's socio-economic and nutritional profile prior to a given World Cup.

The target variable is binary:

- **Top8 = 1** if a country reached the Top-8
- **Top8 = 0** otherwise

This setup allows both statistical comparison and supervised machine learning analysis.

---

4. Exploratory Data Analysis

Exploratory analysis was conducted to compare Top-8 and Non-Top-8 countries across key variables.

(Insert Figure 1: Boxplot of GDP per capita by World Cup success)
(Insert Figure 2: Boxplot of average food indicators by World Cup success)

Top-8 countries tend to exhibit higher GDP per capita and slightly stronger food supply indicators on average. However, substantial overlap exists between successful and unsuccessful teams across all variables. Correlation analysis further shows that while economic and nutritional variables are related, none alone provides a clear separation between outcomes.

These observations suggest that World Cup success cannot be explained by a single indicator and motivate the use of multivariate and non-linear methods.

---

5. Statistical Testing

To formally test differences between Top-8 and Non-Top-8 groups, Welch's t-tests were applied to selected variables, including calorie supply, protein supply, meat consumption, and GDP per capita.

The results indicate that most food-related variables do not exhibit statistically significant mean differences when considered individually. GDP per capita shows marginal explanatory power, but its effect is not strong enough to fully explain World Cup success.

This reinforces the hypothesis that success is driven by the interaction of multiple factors rather than isolated variables.

---

6. Machine Learning Methodology

The problem is formulated as a binary classification task predicting Top-8 success.

Two models were implemented:

- **Logistic Regression** as a baseline linear classifier
- **Random Forest** to capture non-linear relationships and interactions

Given the strong class imbalance (few Top-8 observations), stratified cross-validation was used. Model performance was evaluated using AUC and F1-score, which are more informative than accuracy in imbalanced settings.

---

7. Model Results

The logistic regression model captures limited predictive signal and struggles to generalize. In contrast, the Random Forest model achieves improved performance by learning non-linear patterns across variables.

(Insert Table or Figure: Model performance comparison)

Feature importance analysis highlights GDP per capita as the dominant predictor, followed by food supply indicators. Population size plays a secondary role.

These results suggest that while economic capacity is crucial, nutritional and food-related factors contribute meaningful additional information.

---

8. Discussion

The findings indicate that food economy indicators alone are insufficient to explain international football success. However, when combined with economic and demographic controls, they enhance predictive performance.

This supports a holistic interpretation of World Cup success as the outcome of long-term socio-economic conditions rather than short-term sporting factors alone.

---

9. Conclusion & Future Work

This study demonstrates that socio-economic and nutritional indicators, when analyzed jointly using machine learning methods, provide valuable insight into World Cup performance. Success emerges from a combination of economic strength, population characteristics, and food availability rather than any single determinant.

Future research could incorporate football-specific variables such as FIFA rankings, player export data, or youth development indicators. Causal inference techniques could also be explored to move beyond prediction toward explanation.