Introduction to Math for DS Group Project

Predicting the Premier League Winner

IMDS Group 24

Zehao Qian, Mohammad Jamshaid Iqbal, Chloe Mendez

Department of Natural Sciences
Durham University, England, UK

November 24, 2023

Outline

The Premier League

- Premier League: Top tier of English Football League System.
- 20 teams play 38 home and away matches.
- Globally renowned and challenging to predict outcomes.

Background

1. Outcome predictions involve expert analysis.
2. Factors include team performance, player form, and tactics.
3. Growing data, e.g., player touches, team running stats, manager experience.

Outline

Overview of Entropy Weight Method in Football

1. Introduction
   - The Entropy Weight Method is a powerful analytical technique used in football team evaluation. It goes beyond traditional methods by considering the inherent information entropy within various performance attributes.
2. Key Characteristics
   - Entropy: Reflects the degree of uncertainty or randomness within a dataset.
   - Weight Assignment: Assigns weights to attributes based on their information entropy.
3. Objective
   - The method aims to provide a nuanced evaluation, giving higher importance to attributes that contribute more to understanding a team's performance.

Key Steps in Entropy Method

1. **Data Collection and Attribute Selection**
2. **Entropy Calculation:**
   - Utilize mathematical formulas to calculate the entropy of each selected attribute.
   - Entropy = $-\sum(p_i \cdot \log_2(p_i))$, where p_i is the probability of each attribute value.
3. **Weight Assignment:**
   - Assign weights to attributes based on their calculated entropy.
   - Attributes with higher entropy receive lower weights, and vice versa.
   - The sum of weights equals 1 for normalization.
4. **Outcome:**
   - The result is a set of weights that reflect the relative importance of each attribute in evaluating a football team's performance.

Entropy Method in Our Model

Step 1: Get Data Set from FootyStats with web crawler

Football Stats								
Team	MP	Win	Draw	Loss	GF	GA	GD	Pts
MU	38	25	10	3	80	22	58	87
...	...	...	...	...	...	...	...	...

Step 2: Attribute Selection

In order to reduce the complexity of data processing, the model input is simplified.

An Example:

$$\left.\begin{aligned} Loss = MP - Min - Draw \\ GA = GF - GD \end{aligned}\right\} \Rightarrow$$

They are negative and can be represented by other data

Won't take these coloums in to consideration

Step 3: Normalization the Matrix

Implement it with MinMaxScaler

$$x_{ij} = \frac{x_{ij} - x_{min}}{x_{max} - x_{min}}$$

Step 4: Calulate the Information Entropy

$$E_i = -\sum_{j=1}^{n} p_{ij} \log(p_{ij})$$

Step 5: The entropy weight for criterion *i*

$$W_i = \frac{1 - E_i}{n - \sum_{i=1}^{n} E_i}$$

Step 6: Apply weights

Multiply the weight by the value of the corresponding criterion to obtain a weighted sum.

Team Rank for Reference

Manchester City FC

Liverpool FC

...

Outline

Gradient Ascend for Prediction

Result Analysis

Future Work