

Goal Score Dataset

Swapnil Biswas (14.02.04.021)
Shahariar Shibli (14.02.04.027)

February 13, 2018

1 Introduction

This dataset contains information that we have collected to predict whether a football player can score a goal or not in a match. The data were obtained from here. The dataset is small in size with only 102 cases.

2 Data Set Description

The detailed description about the dataset is stated below.

1. **Title:** Goal Score Data Set.
2. **Usage:** This dataset is used in lab project for assessment.
3. **Number of instances:** 102
4. **Relevant Information:**
 - (a) Dataset contains 2 classes which are represented by 0 and 1. 0 indicates the player didn't score any goal and 1 indicates that the player scored in a particular match.
 - (b) Class 0 contains 58 instances and class 1 contains 44 instances.
 - (c) The information of players and teams we have collected to create the dataset is based on the season 2017/2018.
5. **Number of attributes:** Total 9 attributes. 1 object type and 8 numeric, including predictive attributes and the class.
6. **Attribute information:**
 - (a) **Name(object type):** Name of the player.
 - (b) **Player's Current Rating:** Describes the performance factor of the player in the particular match. Ranges between 0 to 10.
 - (c) **Age:** Age of the player.
 - (d) **Recent Form (No. of Goals):** Number of goals the player scored.
 - (e) **Position:** Describes in which position the player plays. Positions are classified into three types that are represented by 1, 2 and 3.
 - i. Forward represented by 1
 - ii. Midfield represented by 2
 - iii. Defense represented by 3
 - (f) **Opponent Team's Rating:** Describes the performance rating of the opponent team in scale of 1 to 10.
 - (g) **Own Team's Rating:** Describes the performance rating of the player's team in scale of 1 to 10..
 - (h) **Venue:** Describes where the match took place. Venue is classified into two types and represented by 1 and 2.
 - i. Home represented by 1
 - ii. Away represented by 2

(i) **Target(predict):**

- i. 0 means the player didn't score any goal.
- ii. 1 means the player scored at least 1 goal.

7. **Missing attribute values:** None.

3 Attribute Correlation

Table 1: Correlation between attributes

	Players Current Rating	Age	Recent Form (No. of Goals)	Position	Opponent Team's Rating	Own Team's Rating	Venue	Target
Players Current Rating	1.000000	-0.022335	0.206547	-0.156921	-0.068704	0.093705	0.024040	0.684175
Age	-0.022335	1.000000	-0.066879	-0.049754	0.003933	0.162185	0.027604	-0.046492
Recent Form (No. of Goals)	0.206547	-0.066879	1.000000	-0.670338	0.014279	0.161060	-0.010021	0.184101
Position	-0.156921	-0.049754	-0.670338	1.000000	-0.044438	-0.001718	-0.015509	-0.103509
Opponent Team's Rating	-0.068704	0.003933	0.014279	-0.044438	1.000000	-0.004469	-0.011725	-0.133720
Own Team's Rating	0.093705	0.162185	0.161060	-0.001718	-0.004469	1.000000	0.022812	0.090394
Venue	0.024040	0.027604	-0.010021	-0.015509	-0.011725	0.022812	1.000000	0.039185
Target	0.684175	-0.046492	0.184101	-0.103509	-0.133720	0.090394	0.039185	1.000000

4 Class Distribution

- 56.86% data from class 0
- 43.14% data from class 1

5 Sources

1. Data Collection Source: whoscored
2. Date: February, 2018