

University of Pisa

DEPARTMENT OF COMPUTER SCIENCE

Master's Degree of Computer Science
(Artificial Intelligence)

ARTIFICIAL INTELLIGENCE FUNDAMENTALS

Football-betting Detection System

Project made by:
Andrea Tufo

Professor:
Vincenzo Lomonaco

Academic year 2022/2023

Contents

1	Introduction	2
1.1	Sport-betting	2
1.2	Modus operandi	3
2	Implementation	4

Chapter 1

Introduction

In this document all the specifications of the project can be found, including not only theoretical explanations, but also instances, useful to enucleate the code.

The first chapter is going to introduce how the projects actually works, and why it would be useful for football and more in general for sport. This section is also important to figure out the idea behind the algorithm and how all the development has been organized.

All the main difficulties that I faced during the development and all the most important issues that my algorithm has, are listed in the end of this document, where are explained some possible solutions too, in order to solve them and improve the algorithm.

1.1 Sport-betting

The Sport-betting phenomenon is still nowadays one of the worse side of the sports. It hurts two sports above all: tennis and football. The former because, since there are very few actors involved in the game (for example only two players), it's very easy bribing one of them or both of them and change the flow of events.

The latter because football moves a huge amount of money, thus it's very easy to became millionaire corrupting one or two match per season.

The *modus operandi* is always the same, "bribe and earn", so "sport criminals" used to corrupt players, who have the role to make the match ends as agreed, then criminals will be able to bet and so collect thier money.

1.2 Modus operandi

In order to choose the right way and so understand what kind of data my algorithm needs to work as expected, I studied how criminals operate, trying to find many information as possible on the Internet.

Usually these lawbreakers rig a few matches per division in a season and, furthermore, also few teams are involved, maximum of three or four teams per division. For example in Serie A during the season 2011/2012, only two matches has been faked, for a total earning estimated between 300'000 and 500'000 euros, corrupting players who belong to only two teams. Accordingly, it's quite difficult to detect anomalies observing all the season of a specific team, so it's important to highlight that, especially in this case, technology and human must work together to reach the goal and get rid of false-positive.

Chapter 2

Implementation

For the implementation the main goal was to develop a system that relying and analyzing on matches football data, shows a percentage of "possible unfair match". The basic idea was to retrieve some parameters and values from raw filtered data, collect them, and then looking for some anomalies. In the code has been used as case of study the 2012/2013 season of Serie A, so every match that belongs to the regular season. In my case there is one important metric, which is called "*UGI*" (*Potential Goal Index*) and has the role to measure how a team has been dangerous and offensive during a match, so this value is computed for each football team.

2.1 Dataset and filtering

The first important fase has consisted in select a reliable dataset from the Internet, so in the project my choise is to pick up a dataset found in kaggle, which contains six seasons matches' data of five championships.

2.2 Potential Goal Index

The "UGI" is the main metrics on which the entire project lies on. It's made up by