Student: 105502506 106502549

Professor: Hung-Hsuan Chen

Course: CE6143-Introduction to Data Science

Date: 2020.1.7

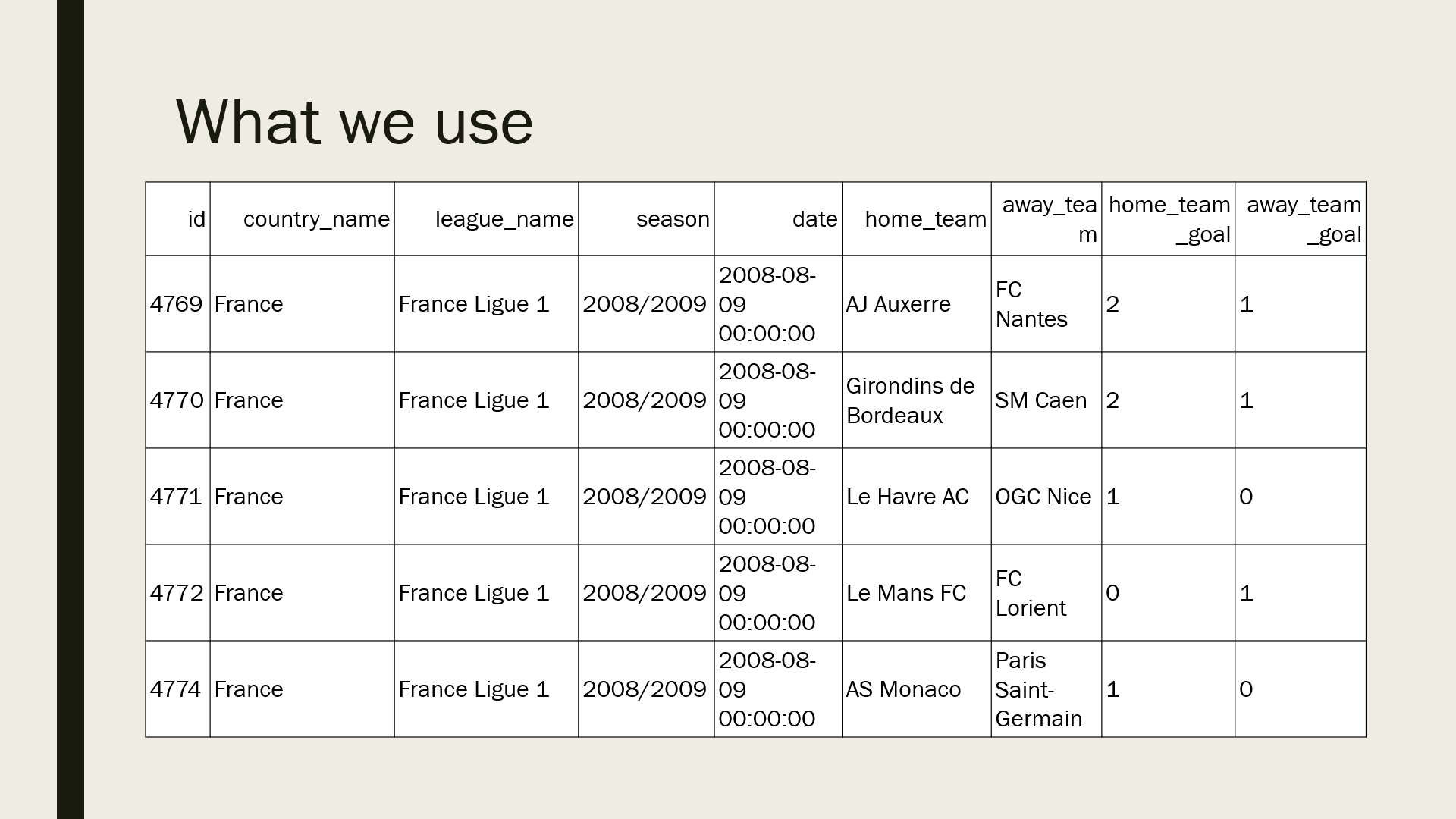
Final Project: Football Game Result Predictor

Being a football fan growing up, we always have that one team we support the most and wish it win every single game. Most of the football-themed video games rate teams, players, coaches and strategies. We would play these games as many times as possible, only looking forward to beat the opponent one time. However, the reality seldom lives up to the expectation.

With the opportunity to develop a Data Science based project, we decided to create a game result predictor. In order to get enough data to train for a model, we choose a nice collected database from Kaggle coming in contents shown below:

* Database: <https://www.kaggle.com/hugomathien/soccer>
* Country 11 x 2
* League 11 x 3
* Match 26.0k x 115
* Player 11.1k x 7
* Player\_Attributes 184k x 42
* Team 299 x 5
* Team\_Attributes 1458 x 25

Data frame Example:



After calculating, we use every team’s score records of which it’s home team or away to predict result. “Home Home Average Goals” means the average home goals of the home team in this match, so on and so forth.



We use the three training methods we learned this semester to train models: KNN, Gaussian and Logistic Regression. The accuracies are 71.48%, 64.27% and 65.29%. But of course, these are just the simple models we trained for final project and practicing. The actual accuracies could be way lower and way less predictable, while there are more factors (like players’ ability, what strategy used, …) need to be considered into the predictions.

References:

https://www.kaggle.com/c/football-data-challenge

http://odds.football-data.co.uk/

https://www.kaggle.com/airback/match-outcome-prediction-in-football

https://towardsdatascience.com/pandas-for-football-analysis-42c23b252995