**MÔ TẢ ĐỒ ÁN**

1. **Mô tả bài toán**

* Input: Thông tin trận đấu
* Ouput: Kết quả thắng, thua , hòa của đội nhà

1. **Mô tả bộ dữ liệu**

* Cách thức xây dựng bộ dữ liệu:
* Lấy dữ liệu về chỉ số cầu thủ từ trang web: <https://www.fifaindex.com/>
* Lấy dữ liệu về trận đấu (tên 2 đội, ngày giờ thi đấu, đội hình của 2 đội) từ trang web: <https://www.betstudy.com/>
* Số lượng:
* Số trận đấu qua 10 mùa giải: 3800
* Số cầu thủ trong 1 mùa khoảng 660 cầu thủ. Tuy nhiên mỗi mùa các cầu thủ sẽ khác nhau và khác về chỉ số nên 10 mùa có khoảng 6600 cầu thủ
* Các thao tác tiền xử lý dữ liệu:
* Phân chia dữ liệu 75% train, 10% dev, 15% test.

1. Mô tả về đặc trưng

* 22 chỉ số của 22 cầu thủ. Từ thông tin về đội hình mỗi trận đấu, đồng bộ với chỉ số của cầu thủ của mùa giải đó.
* Lịch sử đấu của hai đội trong 5 trận gần nhất: số nguyên a1, a2, b1, b2 tương ứng số trận thắng, hòa của đội nhà, đội khách.
* Lịch sử đối đầu của 2 đội trong các mùa trước đó: số nguyên a nếu đội 1 thắng a trận, số nguyên dương b nếu 2 đội hòa b trận.
* Input là một mảng số nguyên 42 phần tử. Trong đó 36 phần tử đầu chia đều cho 2 đội (vì đội hình các trận đấu của các đội khác nhau nên ta sẽ dành trong đó 1 phần tử cho chỉ số thủ môn, 6 phần tử tiếp theo cho chỉ số hậu vệ, 7 phần tử tiếp theo cho các tiền vệ, và 4 cho các tiền đạo). Các phần tử tiếp theo đại diện các các chỉ số lịch sử đấu.
* Ouput: 0 – thắng, 1 – hòa, 2 – thua.

1. Thuật toán máy học: Random forest