

TÓM TẮT KHÓA LUẬN

Hiện nay, máy học (Machine learning, gọi tắt là ML) [1] nổi lên như một phương pháp để dự đoán kết quả, phân loại và nhận diện đối tượng một cách chính xác. Một trong những tiềm năng của ML là dự đoán kết quả thi đấu thể thao đặc biệt là bóng đá. Một mô hình dự đoán chính xác sẽ giúp người dùng có thể dễ dàng tham gia dự đoán với độ chính xác cao. Ngoài ra, các huấn luyện viên cũng sẽ được hưởng lợi thông qua việc điều chỉnh chiến lược của mình cho phù hợp.

Chính vì lẽ đó, chúng tôi quyết định xây dựng và chọn ra một mô hình dự đoán kết quả thi đấu bóng đá (thắng/thua/hòa). Đầu tiên, chúng tôi xét các thuộc tính (features) như lịch sử đối đầu của 2 đội, phong độ của từng đội (dựa vào kết quả 10 trận gần nhất), tỉ lệ cược của nhà cái, thứ hạng FIFA của từng đội, việc một đội có phải là đội nhà không và các yếu tố về thời tiết. Sau đó, dùng các phương pháp phân tích dữ liệu khám phá (Exploratory Data Analysis, gọi tắt là EDA) [2] để chọn ra các thuộc tính quan trọng nhất. Các thuộc tính được chọn sẽ được dùng để huấn luyện các mô hình dự đoán (model).

Tiếp theo, chúng tôi sẽ dựa vào độ chính xác trung bình của kết quả dự đoán khi kiểm chứng chéo (Cross-validation) để chọn ra một model tốt nhất cho vấn đề này. Cuối cùng, chúng tôi sẽ xây dựng một trang web đơn giản để trình bày kết quả dự đoán các trận đấu kế tiếp.