IV Thử nghiệm

Thử nghiệm được tiến hành dựa trên mô hình tốt nhất với kết quả tốt nhất. Ở đây chúng tôi sử dụng các mô hình XGBoost, Random Forest và Neural Network để huấn luyện mô hình. Trong báo cáo lần này, chúng tôi sử dụng data từ các ván đấu thuộc giaỉ đấu LCK của năm 2023 và 2024. Các tập dữ liệu sẽ được chia làm hai phần là tập train và tập test là tỉ lệ 8:2. Các tập này sẽ được minh hoạ ở hai bảng training\_set và bảng testing\_set. (lát tạo bảng)

A screenshot of a computer

Description automatically generated

Bảng .. . tập dữ liệu training\_set

Trong tập dữ liệu này bao gồm 80% toàn bộ tập dữ liệu và có tổng cộng 60 thuộc tính bao gồm Index, lượt pick từ 1 đến 5 được mã hoá (pick1\_encoded, pick2\_encoded, .. ., pick5\_encoded), tỉ lệ thắng của các từng pick đó (winrate\_pick1 – winrate\_pick5), số lần tướng đó được chọn (count\_pick1 – count\_pick5), và các thuộc tính liên quan đến các thông số trong ván đấn đó bao gồm thông số tổng và thông số ở mỗi 5 phút (tính từ phút 15 trở đi). (Lát tạo bảng)

A screenshot of a computer

Description automatically generated

Bảng … tập dữ liệu testing\_set

Trong tập dữ liệu này bao gồm 20% toàn bộ tập dữ liệu và tương tự với tập training\_set, cũng bao gồm 60 thuộc tính bao gồm index, các lượt pick từ 1 đến 5, tỉ lệ thắng của từng lượt pick đó, số lần tướng đó được chọn ở từng lượt pick và các thuộc tính liên quan đến thông số trong ván đấu đó.

Chúng tôi sử dụng heatmap để trực quan hoá mối liên hệ giữa các thuộc tính trong bộ dữ liệu. Cụ thể, heatmap tính toán và đưa ra các số liệu cùng với màu sắc để thể hiện mức độ tương quan của các thuộc tính trong bộ dữ liệu. Cụ thể dưới đây là ba loại heatmap để minh hoạ các thuộc tính nào có mức độ tương quan nhiều đến thắng thua trận đấu (biến result) A screenshot of a diagram

Description automatically generated

Hình . . . Heatmap cho thấy độ tương quan giữa các thuộc tính thuộc nhóm Overall\_performance

Theo biểu đồ trên ta dễ dàng thấy được rằng biến số mạng hạ gục (kills) cùng với thu nhập kiếm được từ mạng hạ gục (earn kpm) có ảnh hưởng nhiều đến kết quả của toàn trận đấu (độ tương quan đều là 0.72). Chênh lệch vàng tiêu thụ (gspd) và tỉ lệ vàng dẫn/bị dẫn (gpr) nắm vai trò tiên quyết ảnh hưởng nhiều tới kết quả chung cuộc.

A screenshot of a computer screen

Description automatically generated

Hình . . . Heatmap cho thấy độ tương quan giữa các thuộc tính thuộc nhóm Economy

Hoặc ở biểu đồ này, ta dễ dàng thấy được thuộc tính lượng vàng kiếm được mỗi phút (earned gpm) có độ tương quan với result cao nhất và vượt trội hơn các thuộc tính còn lại trong một nhóm. Điều đấy chứng tỏ lượng vàng kiếm được trong ván đấu đó là yếu tố quan trọng quyết định kết quả chung cuộc.

Chúng tôi nhận thấy rằng có một số biến có độ tương quan thấp, không quan trọng trong mô hình này nên chúng tôi quyết định sẽ bỏ đi các biến đó bên cạnh các biến đã bỏ đi ở bước tiền xử lý.

|  |  |
| --- | --- |
| **feature** | **importance** |
| firsttower | 0 |
| firstbaron | 0 |
| firstmidtower | 0.001415 |
| firstdragon | 0.001475 |
| turretplates | 0 |

Bảng. . . các biến không quan trọng đến mô hình

Ngoài ra để dễ quan sát hơn, chúng tôi chọn ra 15 biến đặc trưng quan trọng nhất cho mô hình. Nó được xếp theo thứ tự tăng dần theo mức độ quan trọng.

A graph with a bar graph

Description automatically generated with medium confidence

Hình . . . 15 biến đặc trưng quan trọng ảnh hưởng đến mô hình

Theo như biểu đồ trên ta có thể thấy ngoại trừ ba biến đặc trưng đầu là gspd, kills và earned gpm thì những biến còn lại đều có độ quan trọng được đánh giá là dưới mức 0.05.

Chúng tôi tiếp tục sử dụng bộ dữ liệu này để huấn luyện ba mô hình

Độ chính xác của Mô hình XGBoost cao hơn hai mô hình còn lại là Random Forest và Neural Network (Lần lượt là 0.9768, 0.9742, 0.9665)