**III**

Các độ đo được sử dụng cho mô hình bao gồm:

Precision: Cho biết tỉ lệ dự đoán đúng (true positives) của mô hình so với tổng số dự đoán là positives.

Recall: Cho biết tỉ lệ dự đoán đúng (true positives) trong bộ dữ liệu thực tế.

F1 score: là trung bình điều hoà của precision và recall. F1 được dùng để đánh giá nên chọn mô hình nào là tốt nhất.

Accuracy: cho biết tỉ lệ dự đoán đúng của mô hình (cả true positive và true negative) so với tổng số mẫu.

**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. Kết quả thu thập được ghi ở bảng sau:

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên mô hình | Training Accuracy | Testing Accuracy |
| 1 | XGBoost | 1.0 | 0.9794 |
| 2 | Neural Network(Multilayer Perception) | 1.0 | 0.9716 |
| 3 | Random Forest | 0.9929 | 0.9691 |

Bảng . . . Các mô hình và kết quả sau khi huấn luyện.

Độ chính xác của Mô hình XGBoost cao hơn hai mô hình còn lại là Random Forest và Neural Network (Lần lượt là 0.9876, 0.9742, 0.9665).

Để có thể nâng cao hơn kết quả của ba mô hình nhằm tìm ra mô hình tốt nhất, chúng tôi tiến hành sử dụng GridSearchCV từ thư viện sklearn để nâng cao kết quả của mô hình bằng cách định nghĩa các hyperparameter grid cho mỗi mô hình. Kết quả Test Accuracy trước và sau khi thực hiện fine tuning được ghi nhận ở bảng bên dưới:

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên mô hình | Before | After |
| 1 | XGBoost | 0.9794 | 0.9768 |
| 2 | Neural Network(Multilayer Perception) | 0.9716 | 0.9716 |
| 3 | Random Forest | 0.9691 | 0.9716 |

Bảng . . . Kết quả của các mô hình trước và sau fine-tuning

A graph of blue and orange bars

Description automatically generated

Hình . . . So sánh kết quả của các mô hình trước và sau fine-tuning

Mặc dù sau khi fine-tuning, kết quả của XGBoost có giảm nhẹ so với trước khi thực hiện fine-tuning, Nhưng nó vẫn là một trong ba mô hình có kết quả cao nhất. Vậy nên chúng tôi quyết định sẽ sử dụng mô hình này để tiếp tục huấn luyện.

**V. Kết quả mô hình**

Để có thể thấy rõ được mô hình sau khi huấn luyện hoạt động như thế nào, chúng tôi đã tạo một trang web đơn giản, bao gồm đầu vào là tên hai đội cùng với các lượt chọn và đầu ra là tỉ lệ thắng của hai đội cùng với một số thông số liên quan.

Dữ liệu đầu vào chúng tôi đưa vào lần lượt như sau:

Team 1 (T1): Lucian, Nami, Orianna, Aatrox, Taliyah

Team 2 (DRX): Aphelios, Milio, Azir, Rell, Jax

Kết quả dự đoán được trả về như hình:

A white background with black text

Description automatically generated

Hình . . . Kết quả dự đoán được từ mô hình

Ngoài ra chúng tôi còn có phân tích 10 trận đấu gần nhất của cả hai đội cũng như các thông số liên quan đến tướng.

A white paper with black text

Description automatically generated

Hình . . . Thống kê chi tiết của hai đội