

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**GIẢI ĐẤU LCK LMHT:**  
**THU THẬP DỮ LIỆU, PHÂN TÍCH**  
**VÀ DỰ ĐOÁN TỈ LỆ THẮNG**

<b>Nhóm 4</b>			
Sinh viên thực hiện:			
STT	Họ tên	MSSV	Ngành
1	Nguyễn Hoàng Duy	22520327	CNTT
2	Trần Thanh Hải	22520392	CNTT
3	Nguyễn Minh Trí	22521522	CNTT
4	Cao Trường Vũ	22521686	CNTT

## 1. GIỚI THIỆU

Đề tài của chúng tôi nhằm phân tích bộ dữ liệu của các trận đấu trong khuôn khổ giải Vô địch Liên Minh Huyền Thoại Hàn Quốc (LCK) để dự đoán tỷ lệ thắng giữa các đội tuyển trong một ván đấu hoặc đội tuyển vô địch trong giải đó. Các công cụ và thư viện được sử dụng cho đề án này bao gồm: pandas và numpy để xử lý dữ liệu; scikit-learn cho quy trình học máy và chuẩn bị dữ liệu; XGBoost, LightGBM và Random Forest cho mô hình dự đoán; Streamlit để xây dựng giao diện và dashboard; Matplotlib, Seaborn, Plotly để trực quan hóa dữ liệu. Dữ liệu đầu vào là thông tin của hai đội thi đấu và các tướng họ lựa chọn, đầu ra sẽ là tỷ lệ thắng của mỗi đội cùng các thông số liên quan. Bộ dữ liệu tham khảo từ **Oracle Elixir**, đã được chúng tôi tự thu thập và điều chỉnh để phù hợp với mục tiêu thực hiện đề án.

## 2. MÔ TẢ BỘ DỮ LIỆU

### 2.1 Chi tiết về bộ dữ liệu

Bộ dữ liệu phân tích có nguồn gốc từ **League of Legends esports stats – Oracle Elixir** [1]. Trang web này chuyên thống kê về các thông số trong các trận đấu chuyên nghiệp của tựa game **League of Legends**. Tác giả của trang web này – ông **Tim Sevenhuysen**, đã cung cấp công khai bộ dữ liệu được sử dụng trên trang web do ông tự thu thập và tổng hợp.

Bộ dữ liệu chúng tôi xây dựng được trích lọc từ tập dữ liệu sau: **2024\_LoL\_esports\_match\_data\_from\_OraclesElixir.csv** [2]. Bộ dữ liệu này hiện nay vẫn đang được cập nhật hằng ngày, với dự định sẽ chốt lại bộ dữ liệu (không cập nhật thêm dữ liệu nữa) vào ngày **26/10/2024**.

Bộ dữ liệu gốc thống kê các thông số trong mỗi ván đấu và được thống kê theo 2 nhóm đối tượng: theo từng tuyển thủ và theo từng đội tuyển. Chúng tôi **dự định** sẽ điều chỉnh, trích lọc lại bộ dữ liệu gốc theo hướng **thống kê theo các đội tuyển trong giải đấu LCK** sau khi cân nhắc qua một số lý do như giới hạn đề án, thời gian thực hiện.

Thống kê ban đầu của bộ dữ liệu kế thừa: (26/10/2024)

- Tên bộ dữ liệu: **LCK\_Tournament.csv**
- Số cột: 147 cột.

- Số dòng: 964 dòng.

	Tên cột	Kiểu dữ liệu	Loại biến	Mô tả chi tiết
1	gameid	object	Định danh	ID của trận đấu
2	datacompleteness	object	Phân loại	Độ hoàn thiện dòng dữ liệu
3	league	object	Phân loại	Tên giải đấu
4	year	int64	Phân loại	Năm diễn ra
5	split	object	Phân loại	Mùa giải
6	playoffs	int64	Phân loại	Vòng loại trực tiếp
7	date	object	Thời gian	Ngày thi đấu
8	game	int64	Phân loại	Game đấu thứ n trong cặp trận.
9	patch	float64	Phân loại	Phiên bản game thi đấu
10	participantid	int64	Định danh	ID của tuyển thủ tham gia ván đấu
11	side	object	Phân loại	Đội xanh/Đội đỏ
12	position	object	Phân loại	Vị trí của tuyển thủ trong team
13	teamname	object	Phân loại	Tên đội tuyển
14	teamid	object	Định danh	ID đội tuyển
15	ban1	object	Phân loại	Cắm tướng lượt 1
16	ban2	object	Phân loại	Cắm tướng lượt 2
17	ban3	object	Phân loại	Cắm tướng lượt 3
18	ban4	object	Phân loại	Cắm tướng lượt 4
19	ban5	object	Phân loại	Cắm tướng lượt 5

20	pick1	object	Phân loại	Chọn tướng lượt 1
21	pick2	object	Phân loại	Chọn tướng lượt 2
22	pick3	object	Phân loại	Chọn tướng lượt 3
23	pick4	object	Phân loại	Chọn tướng lượt 4
24	pick5	object	Phân loại	Chọn tướng lượt 5
25	gamelength	int64	Số liệu	Độ dài ván đấu (giây)
26	result	int64	Phân loại	Kết quả ván đấu
27	kills	int64	Số liệu	Số mạng hạ gục
28	deaths	int64	Số liệu	Số mạng nằm xuống
29	assists	int64	Số liệu	Số mạng hỗ trợ
30	doublekills	float64	Số liệu	Số double kills trong ván đấu
31	triplekills	float64	Số liệu	Số triple kills trong ván đấu
32	quadrakills	float64	Số liệu	Số quadra kills trong ván đấu
33	pentakills	float64	Số liệu	Số penta kills trong ván đấu
34	firstblood	float64	Phân loại	Chiến công đầu
35	team kpm	float64	Số liệu	Số mạng hạ gục trung bình mỗi phút của đội
36	ckpm	float64	Số liệu	Số mạng hạ gục trung bình mỗi phút của ván đấu
37	firstdragon	float64	Phân loại	Rồng đầu
38	dragons	float64	Số liệu	Tổng số rồng nguyên tổ đội hạ gục trong ván đấu

<b>39</b>	opp_dragons	float64	Số liệu	Tổng số rồng đội địch hạ gục trong ván đấu
<b>40</b>	elementaldrakes	float64	Số liệu	Số rồng nguyên tố đội hạ gục
<b>41</b>	opp_elementaldrakes	float64	Số liệu	Số rồng nguyên tố đội địch hạ gục
<b>42</b>	infernals	float64	Số liệu	Số rồng lửa đội hạ gục
<b>43</b>	mountains	float64	Số liệu	Số rồng đất đội hạ gục
<b>44</b>	clouds	float64	Số liệu	Số rồng gió đội hạ gục
<b>45</b>	oceans	float64	Số liệu	Số rồng nước đội hạ gục
<b>46</b>	chemtechs	float64	Số liệu	Số rồng hoá kĩ đội hạ gục
<b>47</b>	hextechs	float64	Số liệu	Số rồng công nghệ đội hạ gục
<b>48</b>	elders	float64	Số liệu	Tổng số rồng ngàn tuổi đội hạ gục
<b>49</b>	opp_elders	float64	Số liệu	Tổng số rồng ngàn tuổi đội địch hạ gục
<b>50</b>	firstherald	float64	Phân loại	Sứ giả đầu
<b>51</b>	heralds	float64	Số liệu	Sứ giả đội hạ gục
<b>52</b>	opp_heralds	float64	Số liệu	Sứ giả đội địch hạ gục
<b>53</b>	void_grubs	float64	Số liệu	Tổng số sâu hư không đội hạ gục
<b>54</b>	opp_void_grubs	float64	Số liệu	Tổng số sâu hư không đội địch hạ gục
<b>55</b>	firstbaron	float64	Phân loại	Baron đầu
<b>56</b>	barons	float64	Số liệu	Tổng số baron đội hạ gục

<b>57</b>	opp_barons	float64	Số liệu	Tổng số baron đội địch hạ gục
<b>58</b>	first_tower	float64	Phân loại	Hạ trụ đầu
<b>59</b>	towers	float64	Số liệu	Tổng số trụ đội đánh hạ
<b>60</b>	opp_towers	float64	Số liệu	Tổng số trụ đội địch đánh hạ
<b>61</b>	firstmidtower	float64	Phân loại	Hạ trụ 1 đường giữa đầu tiên
<b>62</b>	firsttothreetowers	float64	Phân loại	Đội đánh hạ 3 trụ trước tiên
<b>63</b>	turretplates	float64	Số liệu	Số lớp giáp trụ đội lấy được
<b>64</b>	opp_turretplates	float64	Số liệu	Số lớp giáp trụ đội địch lấy được
<b>65</b>	inhibitors	float64	Số liệu	Số nhà lính đội phá hủy
<b>66</b>	opp_inhibitors	float64	Số liệu	Số nhà lính đội địch phá hủy
<b>67</b>	damagetochampions	int64	Số liệu	Sát thương gây ra lên tướng địch
<b>68</b>	dpm	float64	Số liệu	Sát thương trung bình gây ra mỗi phút lên tướng
<b>69</b>	damagetakenperminute	float64	Số liệu	Sát thương gánh chịu trung bình mỗi phút
<b>70</b>	damagemitigatedperminute	float64	Số liệu	Sát thương tự giảm trung bình mỗi phút
<b>71</b>	wardsplaced	int64	Số liệu	Số mắt đã cắm
<b>72</b>	wpm	float64	Số liệu	Số mắt cắm mỗi phút

<b>73</b>	wardskilled	int64	Số liệu	Số mất tiêu diệt
<b>74</b>	wcpm	float64	Số liệu	Số mất tiêu diệt mỗi phút
<b>75</b>	controlswardsbought	int64	Số liệu	Số mất kiểm soát đã mua
<b>76</b>	visionscore	int64	Số liệu	Điểm tầm nhìn
<b>77</b>	vspm	float64	Số liệu	Điểm tầm nhìn mỗi phút
<b>78</b>	totalgold	int64	Số liệu	Vàng sở hữu
<b>79</b>	earnedgold	int64	Số liệu	Vàng kiếm được
<b>80</b>	earned gpm	float64	Số liệu	Vàng kiếm được mỗi phút
<b>81</b>	goldspent	int64	Số liệu	Vàng sử dụng
<b>82</b>	gspd	float64	Số liệu	Tỉ lệ chênh lệch lượng vàng sử dụng giữa đội mình với đội địch
<b>83</b>	gpr	float64	Số liệu	Tỉ lệ vàng dẫn
<b>84</b>	total cs	float64	Số liệu	Chỉ số quái/lính
<b>85</b>	minionkills	float64	Số liệu	Số lính tiêu diệt
<b>86</b>	monsterkills	int64	Số liệu	Số quái tiêu diệt
<b>87</b>	cspm	float64	Số liệu	Chỉ số quái/lính mỗi phút
<b>88</b>	goldat10	float64	Số liệu	Vàng sở hữu tại thời điểm phút thứ 10
<b>89</b>	xpat10	float64	Số liệu	Tổng kinh nghiệm có được tại thời điểm phút thứ 10
<b>90</b>	csat10	float64	Số liệu	Chỉ số quái/lính tại thời điểm phút thứ 10
<b>91</b>	opp_goldat10	float64	Số liệu	Vàng đối phương sở hữu tại thời điểm phút thứ 10

<b>92</b>	opp_xpat10	float64	Số liệu	Kinh nghiệm đối phương có được tại thời điểm phút thứ 10
<b>93</b>	opp_csat10	float64	Số liệu	Chỉ số quái/lính của đối phương tại thời điểm phút thứ 10
<b>94</b>	golddiffat10	float64	Số liệu	Vàng chênh lệch tại thời điểm phút thứ 10
<b>95</b>	xpdiffat10	float64	Số liệu	Kinh nghiệm chênh lệch tại thời điểm phút thứ 10
<b>96</b>	csdiffat10	float64	Số liệu	Chỉ số quái/lính chênh lệch tại thời điểm phút thứ 10
<b>97</b>	killat10	float64	Số liệu	Số mạng hạ gục tại thời điểm phút thứ 10
<b>98</b>	assistat10	float64	Số liệu	Số mạng hỗ trợ tại thời điểm phút thứ 10
<b>99</b>	deathat10	float64	Số liệu	Số mạng nằm xuống tại thời điểm phút thứ 10
<b>100</b>	opp_killsat10	float64	Số liệu	Số mạng hạ gục của đối phương tại thời điểm phút thứ 10
<b>101</b>	opp_assistat10	float64	Số liệu	Số mạng hỗ trợ của đối phương tại thời điểm phút thứ 10
<b>102</b>	opp_deathat10	float64	Số liệu	Số mạng nằm xuống của đối phương tại thời điểm phút thứ 10



<b>103</b>	goldat15	float64	Số liệu	Vàng sở hữu tại thời điểm phút thứ 15
<b>104</b>	xpat15	float64	Số liệu	Tổng kinh nghiệm có được tại thời điểm phút thứ 15
<b>105</b>	csat15	float64	Số liệu	Chỉ số quái/lính tại thời điểm phút thứ 15
<b>106</b>	opp_goldat15	float64	Số liệu	Vàng đối phương sở hữu tại thời điểm phút thứ 15
<b>107</b>	opp_xpat15	float64	Số liệu	Kinh nghiệm đối phương có được tại thời điểm phút thứ 15
<b>108</b>	opp_csat15	float64	Số liệu	Chỉ số quái/lính của đối phương tại thời điểm phút thứ 15
<b>109</b>	golddiffat15	float64	Số liệu	Vàng chênh lệch tại thời điểm phút thứ 15
<b>110</b>	xpdiffat15	float64	Số liệu	Kinh nghiệm chênh lệch tại thời điểm phút thứ 15
<b>111</b>	csdiffat15	float64	Số liệu	Chỉ số quái/lính chênh lệch tại thời điểm phút thứ 15
<b>112</b>	killat15	float64	Số liệu	Số mạng hạ gục tại thời điểm phút thứ 15
<b>113</b>	assitat15	float64	Số liệu	Số mạng hỗ trợ tại thời điểm phút thứ 15
<b>114</b>	deathat15	float64	Số liệu	Số mạng nằm xuống tại thời điểm phút thứ 15

<b>115</b>	opp_killsat15	float64	Số liệu	Số mạng hạ gục của đối phương tại thời điểm phút thứ 15
<b>116</b>	opp_assistat15	float64	Số liệu	Số mạng hỗ trợ của đối phương tại thời điểm phút thứ 15
<b>117</b>	opp_deathat15	float64	Số liệu	Số mạng nằm xuống của đối phương tại thời điểm phút thứ 15
<b>118</b>	goldat20	float64	Số liệu	Vàng sở hữu tại thời điểm phút thứ 20
<b>119</b>	xpat20	float64	Số liệu	Tổng kinh nghiệm có được tại thời điểm phút thứ 20
<b>120</b>	csat20	float64	Số liệu	Chỉ số quái/lính tại thời điểm phút thứ 20
<b>121</b>	opp_goldat20	float64	Số liệu	Vàng đối phương sở hữu tại thời điểm phút thứ 20
<b>122</b>	opp_xpat20	float64	Số liệu	Kinh nghiệm đối phương có được tại thời điểm phút thứ 20
<b>123</b>	opp_csat20	float64	Số liệu	Chỉ số quái/lính của đối phương tại thời điểm phút thứ 20
<b>124</b>	golddiffat20	float64	Số liệu	Vàng chênh lệch tại thời điểm phút thứ 20
<b>125</b>	xpdiffat20	float64	Số liệu	Kinh nghiệm chênh lệch tại thời điểm phút thứ 20

<b>126</b>	csdiffat20	float64	Số liệu	Chỉ số quái/lính chênh lệch tại thời điểm phút thứ 20
<b>127</b>	killat20	float64	Số liệu	Số mạng hạ gục tại thời điểm phút thứ 20
<b>128</b>	assistat20	float64	Số liệu	Số mạng hỗ trợ tại thời điểm phút thứ 20
<b>129</b>	deathat20	float64	Số liệu	Số mạng nằm xuống tại thời điểm phút thứ 20
<b>130</b>	opp_killsat20	float64	Số liệu	Số mạng hạ gục của đối phương tại thời điểm phút thứ 20
<b>131</b>	opp_assistat20	float64	Số liệu	Số mạng hỗ trợ của đối phương tại thời điểm phút thứ 20
<b>132</b>	opp_deathat20	float64	Số liệu	Số mạng nằm xuống của đối phương tại thời điểm phút thứ 20
<b>133</b>	goldat25	float64	Số liệu	Vàng sở hữu tại thời điểm phút thứ 25
<b>134</b>	xpat25	float64	Số liệu	Tổng kinh nghiệm có được tại thời điểm phút thứ 25
<b>135</b>	csat25	float64	Số liệu	Chỉ số quái/lính tại thời điểm phút thứ 25
<b>136</b>	opp_goldat25	float64	Số liệu	Vàng đối phương sở hữu tại thời điểm phút thứ 25

<b>137</b>	opp_xpat25	float64	Số liệu	Kinh nghiệm đối phương có được tại thời điểm phút thứ 25
<b>138</b>	opp_csat25	float64	Số liệu	Chỉ số quái/lính của đối phương tại thời điểm phút thứ 25
<b>139</b>	golddiffat25	float64	Số liệu	Vàng chênh lệch tại thời điểm phút thứ 25
<b>140</b>	xpdiffat25	float64	Số liệu	Kinh nghiệm chênh lệch tại thời điểm phút thứ 25
<b>141</b>	csdiffat25	float64	Số liệu	Chỉ số quái/lính chênh lệch tại thời điểm phút thứ 25
<b>142</b>	killat25	float64	Số liệu	Số mạng hạ gục tại thời điểm phút thứ 25
<b>143</b>	assistat25	float64	Số liệu	Số mạng hỗ trợ tại thời điểm phút thứ 25
<b>144</b>	deathat25	float64	Số liệu	Số mạng nằm xuống tại thời điểm phút thứ 25
<b>145</b>	opp_killsat25	float64	Số liệu	Số mạng hạ gục của đối phương tại thời điểm phút thứ 25
<b>146</b>	opp_assistat25	float64	Số liệu	Số mạng hỗ trợ của đối phương tại thời điểm phút thứ 25
<b>147</b>	opp_deathat25	float64	Số liệu	Số mạng nằm xuống của đối phương tại thời điểm phút thứ 25

Bảng 2.1 Mô tả chi tiết các đặc trưng của bộ dữ liệu.

## 2.2 Thống kê điền khuyết và làm sạch dữ liệu.

Chúng tôi quyết định sẽ loại bỏ các cột có số lượng khuyết trên 70%, một phần vì số lượng khuyết lớn, phần vì các cột này cũng không mang ý nghĩa gì đối với bộ dữ liệu chúng tôi phân tích (đa phần đến từ việc các cột này thống kê các thông số mà tuyển thủ mới có).

Cột	Dữ liệu khuyết	Phần trăm khuyết (%)
damageshare	5876	100
earnedgoldshare	5876	100
playername	5876	100
playerid	5876	100
champion	5876	100
total cs	5876	100
firstbloodvictim	5876	100
firstbloodassist	5876	100
firstbloodkill	5876	100
dragons (type unknown)	4260	72.49829816
monsterkillsownjungle	4260	72.49829816
monsterkillsenemyjungle	4260	72.49829816
url	4260	72.49829816

Bảng 2.2 Các cột có tỉ lệ khuyết trên 70%.

Tiếp theo chúng tôi tiến hành điền khuyết ở một số cột có giá trị null nhưng không phải là null, bao gồm các trường hợp sau:

\* **Các cột thống kê theo thời gian thực trong ván đấu:** Nếu ván đấu nào kết thúc sớm hơn mốc thời gian này thì sẽ điền 0, còn các trường hợp còn lại giữ nguyên khuyết.

\* **Các cột banX, pickX:** Do số lượng giá trị khuyết của các cột này rất nhỏ nên chúng tôi sẽ fill thủ công. Ở các cột ban do có một số trường hợp các tuyển thủ thực sự không cầm lượt đó, dẫn đến thiếu dữ liệu nên những cột này sẽ được giữ nguyên.

Sau đó sẽ tiến hành lọc theo 'league' để chỉ lấy các dữ liệu ở LCK.

Bộ dữ liệu sau khi trích lọc vốn đã rất sạch sẽ cho nên chỉ cần thống kê dữ liệu khuyết để kiểm tra và tiến hành điền khuyết (nếu có).

Đặc trưng	Số lượng khuyết	Phần trăm khuyết (%)
ban5	1	0.103724

Bảng 2.3 Thống kê dữ liệu khuyết cuối cùng.

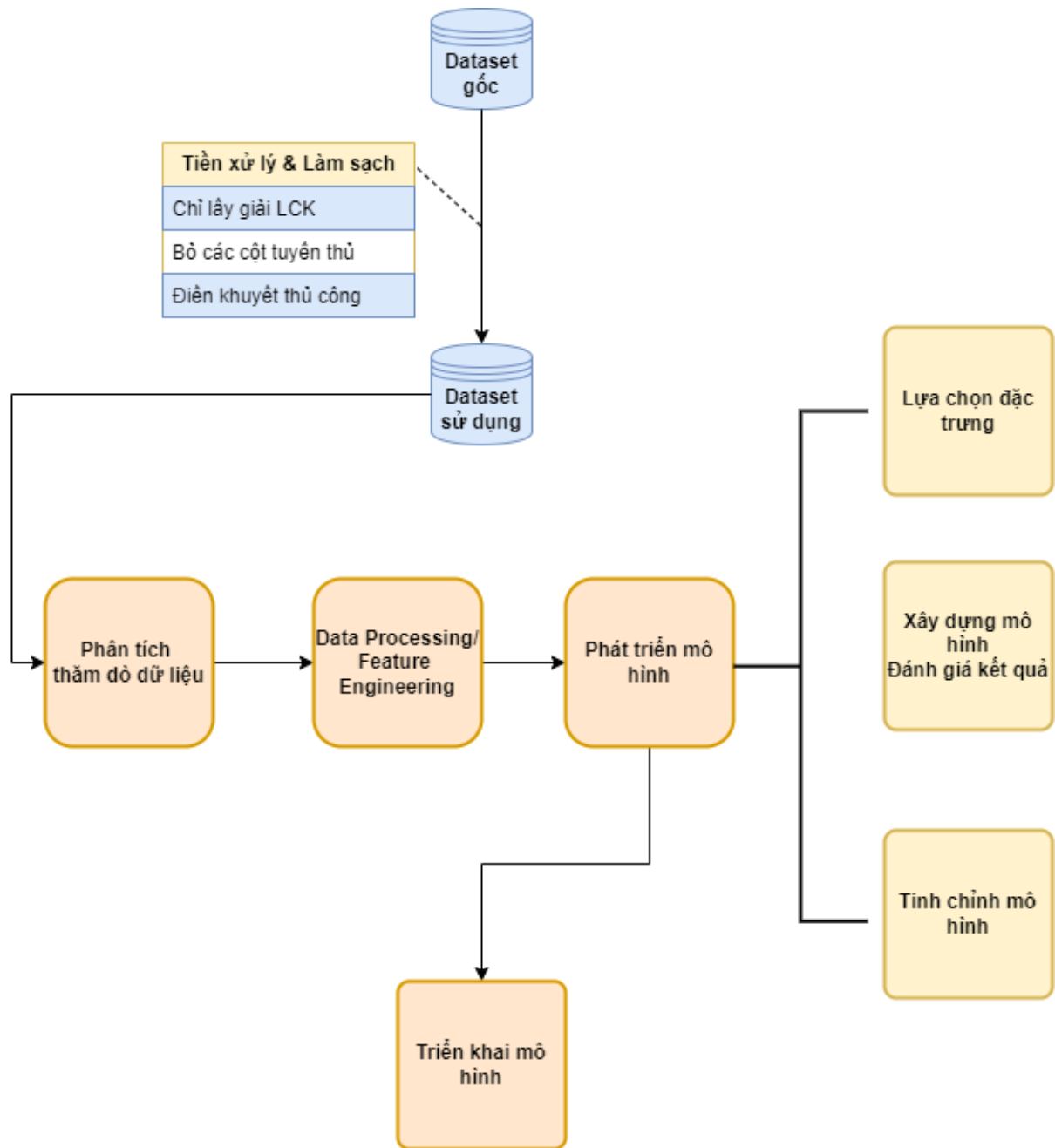
Bộ dữ liệu sau cùng có thống kê như sau:

Metrics	Value	Kiểu dữ liệu	Số lượng cột
Số dòng	964	float64	109
Số cột	147	int64	19
Memory Usage	1.08 MB	object	19

Bảng 2.4, 2.5 Thống kê bộ dữ liệu hậu tiền xử lý.

### 3. THỰC HIỆN PHÂN TÍCH THĂM DÒ & HUẤN LUYỆN MÔ HÌNH

Quy trình thực hiện đồ án của nhóm có thể tóm gọn lại như sau:



Hình 3.1 Quy trình phân tích dữ liệu.

### 3.1 Phân tích thăm dò theo ván đấu/trận đấu

Hình 9.1 cho thấy sự biến động về số lượng ván đấu theo từng phiên bản (patch) của giải đấu LCK trong 2 mùa giải của năm 2024. Sự khác biệt về số lượng ván đấu có thể xuất phát từ thời gian mỗi phiên bản được áp dụng (Xem bảng Bảng 8.1) và việc các trận đấu diễn ra trong ván đó có thường xuyên có ván 3 hay không. (Hình 9.2) Thông qua sự tương quan giữa 2 biểu đồ, có thể dễ dàng chỉ ra rằng tổng số lượng ván đấu có **mối quan hệ tuyến tính** với tổng số lượng ván 3.

Việc một số phiên bản có số lượng ván đấu ít hơn rõ rệt so với trung bình cũng có thể được giải thích thông qua Hình 9.3. Có thể thấy rằng, phiên bản **14.06** và **14.16** chỉ thi đấu các trận thuộc vòng loại trực tiếp cho nên số lượng ván đấu/trận đấu ít hơn đáng kể so với vòng bảng. Trường hợp của patch **14.02** thì sau khi đối chiếu thực tế, chúng tôi biết rằng khoảng thời gian của patch này trùng khớp với thời gian nghỉ Tết của giải, cho nên các quan sát trên đều hoàn toàn hợp lý. Ngoài ra, sự thiếu hụt các trận đấu từ patch **14.07** đến **14.10** là do khoảng nghỉ giữa 2 mùa giải trong năm của LCK.

Từ Hình 9.4, ta có thể thấy biểu đồ phân phối thời lượng ván đấu có xu hướng hơi lệch phải, với phần lớn các ván đấu tập trung trong khoảng từ **27 đến 34 phút**. Đỉnh của phân phối nằm ở khoảng 31-32 phút (80 ván), phản ánh khoảng thời gian kết thúc ván đấu phổ biến. Tần suất các ván đấu kết thúc sớm trước 25 phút hoặc kéo dài quá 45 phút là rất ít, phản ánh việc những ván đấu “một chiều” hay giằng co căng thẳng kéo dài rất ít khi xuất hiện và cho thấy sự cân bằng và cạnh tranh cao giữa các đội tuyển trong giải đấu.

Các đội thường duy trì thế trận ổn định, tập trung vào kiểm soát mục tiêu lớn và tận dụng các cơ hội để giành lợi thế, thay vì áp đảo hoàn toàn hoặc kéo dài trận đấu một cách không cần thiết. Tất nhiên trong số đó vẫn có những ván đấu “huỷ diệt” giữa đội mạnh với đội yếu, hay những ván đấu “ngang tài ngang sức” giữa các đội cùng trình độ. Điều này góp phần tạo nên sức hấp dẫn và tính cạnh tranh đặc trưng của giải đấu LCK.



### 3.2 Phân tích thăm dò theo tỉ lệ cấm/chọn (ban/pick)

**Hình 9.5** và **Hình 9.6** cho ta thấy rõ rệt sự khác biệt của xu hướng cấm/chọn tướng giữa hai mùa Xuân và Hè trong cùng một năm tại LCK.

Trong mùa Xuân, **Azir**, **K'Sante** và **Aatrox** là ba vị tướng được chọn nhiều nhất, trong đó Azir và K'Sante hoà nhau với số lần chọn là 109. Các vị trí còn lại không có sự chênh lệch nhiều, cho thấy sự đa dạng trong việc lựa chọn tướng với các vai trò khác nhau như đấu sĩ, hỗ trợ, và xạ thủ.

Về phần cấm, có thể thấy các đội trong mùa Xuân thường xuyên đưa ra các lượt cấm nhắm vào các vị tướng xạ thủ khi mà trong danh sách có tới 5 vị tướng xạ thủ. **Senna** là tướng bị cấm nhiều nhất với 150 lần, cho thấy các đội thật sự đánh giá rất cao sức mạnh của con bài này.

Trong mùa Hè, ta được chứng kiến sự chuyển dịch meta khi **Rumble** và **Corki** lần lượt leo lên dẫn đầu danh sách cấm/chọn và bỏ xa các vị trí còn lại, cho thấy sức mạnh áp đảo của 2 vị tướng này. Điều này càng được chứng minh khi ta xem qua **Hình 9.7**, Rumble với tỉ lệ xuất hiện lên tới **98,8%** gần như tuyệt đối, tức là đây là một vị tướng “không cấm thì chọn”.

Tỉ lệ cấm/chọn của một vị tướng phụ thuộc rất nhiều vào meta, và meta thì sẽ bị ảnh hưởng, điều chỉnh bởi các bản cập nhật. Để có cái nhìn chính xác hơn, chúng tôi có thực hiện 1 biểu đồ nhiệt thể hiện tỉ lệ xuất hiện của mỗi vị tướng từng được ghi nhận tại LCK trong năm 2024 qua từng bản cập nhật, tuy nhiên vì kích cỡ ảnh quá lớn cho nên không thể đưa vào báo cáo này, thầy và các bạn có thể xem tại [đây](#).

**Hình 9.9** cho biết các cặp tướng thường được chọn cùng nhau trong đội hình. Cặp **Lucian + Nami** dẫn đầu với 57 lần xuất hiện cùng nhau, vượt trội hơn hẳn so với các cặp khác. Điểm đáng chú ý ở đây là cả 3 cặp đứng đầu danh sách đều là các cặp xạ thủ - hỗ trợ, cho thấy xu hướng các đội LCK rất chú trọng và ưu tiên việc xây dựng các cặp đôi đường dưới theo đúng mong muốn, chiến thuật của đội. Lựa chọn **Corki** xuất hiện trong 3 cặp khác nhau, phản ánh đúng thực tế trong giai đoạn mùa hè khi Corki là lựa chọn thường xuyên được ưu tiên cho vị trí đường giữa, có thể kết hợp với nhiều tướng hỗ trợ mở khác nhau, hay các đường trên chống chịu như K'Sante.

### 3.3 Phân tích thăm dò theo các đội tuyển

**Hình 9.10** cung cấp cho ta 2 biểu đồ về số lượng ván đấu và tỉ lệ chọn bên của mỗi đội. Trong đó, **T1** dẫn đầu về số ván đấu với 114 ván, chứng tỏ họ tham gia nhiều trận và thường kéo dài đến các ván quyết định. Các đội mạnh khác như **Hanwha Life Esports** (112 ván) và **Dplus KIA** (111 ván) cũng cho thấy sự cạnh tranh sát sao. Ngược lại, **OKSavingsBank BRION** thi đấu ít nhất với 82 ván, phản ánh sự chênh lệch trình độ giữa nhóm đầu và nhóm cuối.

Tuy vậy, số lượng ván đấu không phải là 1 thước đo tốt để đánh giá đội mạnh và đội yếu, vì nó sẽ không phân biệt được giữa những đội thường xuyên thắng áp đảo 2-0 và những đội thua chóng vánh 0-2, mà cụ thể trong trường hợp này là **Gen.G**. Với 103 ván đấu (chỉ xếp thứ 4), họ vẫn là nhà vô địch của giải mùa Xuân và á quân của giải mùa Hè với thành tích thi đấu áp đảo, khi có tỉ lệ thắng 2-0 tại vòng bảng cao nhất (**83.3%** trong tổng số **36 trận**) và tỉ lệ thắng 3-0/3-1 ở mức **50%** trong **6 trận** thi đấu. (Chi tiết xem tại **Hình 9.11** và **Hình 9.12**)

Về tỷ lệ chọn bên, hầu hết các đội ưu tiên Đội Đỏ hơn Đội Xanh, ngoại trừ **OKSavingsBank BRION** có tỷ lệ chọn Đội Xanh cao nhất (**59.8%**), thể hiện chiến thuật có phần khác biệt so với các đội còn lại, khi họ có lối chơi xoay quanh đường trên và chủ động chọn đội Xanh để có thể chọn tướng cuối cùng nhằm counter-pick cho đường trên của họ. **T1** nổi bật với tỷ lệ chọn hai bên cân bằng (**50.9%** Đội Xanh, **49.1%** Đội Đỏ), cho thấy khả năng thay đổi, thử nghiệm linh hoạt của đội tuyển này.

So sánh % ván thắng tại vòng bảng giữa hai mùa giải (**Hình 9.13**) phản ánh sự thay đổi phong độ của các đội. **Gen.G** vượt trội với tỷ lệ thắng cao nhất cả hai mùa, đặc biệt mùa Hè đạt **92.1%**. Trong khi đó, **T1** cho thấy sự bất ổn khi từ **80%** tại mùa Xuân giảm còn **56.8%** tại mùa Hè. Ngược lại, **OKSavingsBank BRION** và **DRX** có tỷ lệ thắng thấp cả hai mùa, phản ánh sự khó khăn trong việc cạnh tranh với các đội khác. Xu hướng chung là tỷ lệ thắng của các đội tăng từ mùa xuân sang mùa hè, chủ yếu đến từ phong độ sụt giảm nghiêm trọng của **T1**.

**Gen.G** dẫn đầu cả về tỷ lệ thắng trận (**83.5%**) và thắng ván đấu (**88.3%**), cho thấy phong độ ổn định và khả năng áp đảo đối thủ của đội tuyển này. **Hanwha Life Esports** và **T1** xếp sau với tỷ lệ thắng lần lượt là **73.0%** và **66.4%**. Hai đội **Dplus KIA** và **KT**

**Rolster** có tỷ lệ thắng tương đối gần nhau ở cả hai hạng mục, trong khi các đội còn lại có tỷ lệ thấp, thể hiện sự chênh lệch rõ rệt về sức mạnh và phân hoá giữa nhóm các đội mạnh, đội tầm trung và đội yếu, được củng cố thêm qua số trận thi đấu và tỉ lệ thắng tại vòng loại trực tiếp của mỗi đội. (Hình 9.14,Hình 9.15).

**Dplus KIA** dẫn đầu về tỷ lệ ăn chiến công đầu (**63.96%**), thể hiện sự chủ động và hiệu quả trong giao tranh sớm và kỹ năng cá nhân trong giai đoạn đi đường, theo sau là **Hanwha Life Esports (57.14%)**. Trong khi đó, **Kwangdong Freecs (34.44%)**, là đội thường xuyên mất chiến công đầu vào tay đối thủ. **Gen.G** cho thấy sự vượt trội trong việc kiểm soát hầu hết các mục tiêu đầu game như rồng đầu, trụ đầu, và Baron đầu. **Hanwha Life Esports** cũng rất mạnh ở rồng đầu và Sứ Giả, còn **T1** nổi bật ở trụ đầu và Baron đầu, phản ánh chiến thuật đặc trưng của đội tuyển này là việc kiểm soát Baron đầu rất sớm. Các đội khác có mức kiểm soát không đồng đều, với **OKSavingsBank** **BRION** yếu nhất ở tất cả các mục tiêu. Nhìn chung, **Gen.G**, **Dplus KIA** và **Hanwha Life Esports** dẫn đầu trong việc kiểm soát giai đoạn đầu trận. (Hình 9.16,Hình 9.17)

Để có cái nhìn toàn diện hơn về chiến thuật và ảnh hưởng của các mục tiêu trong trận đấu, bạn có thể tiếp tục khám phá các biểu đồ từ Hình 9.20 đến Hình 9.22. Các biểu đồ này cung cấp thông tin về phân bố tỷ lệ loại Rồng, tỷ lệ thắng khi có hoặc không có các mục tiêu đầu game, cũng như mối quan hệ giữa số lượng Sâu Hư Không hoặc Giáp Trụ với tỷ lệ thắng ván đấu đó. Những phân tích này giúp làm rõ cách các yếu tố chiến thuật cụ thể tác động đến kết quả trận đấu, không phân biệt từng đội, mang đến góc nhìn tổng quan về giải đấu.

### 3.4 Huấn luyện mô hình và kết quả thu được

Quá trình huấn luyện mô hình được nhóm chúng tôi thực hiện qua 3 bước:

1. **Chuẩn bị dữ liệu:** Gồm đọc dữ liệu đầu vào, mã hoá các lượt chọn tướng bằng **LabelEncoder**, vào lấy các đặc trưng để đưa vào mô hình. Hiện tại chúng tôi đang sử dụng khoảng 60 đặc trưng thuộc các nhóm khác nhau.
2. **Huấn luyện mô hình:** Gồm chia tập train/test theo tỉ lệ 8:2, huấn luyện thử nghiệm với 3 mô hình khác nhau (**Random Forest, LightGBM và XGBoost**) bằng Pipeline, fine-tune cho mô hình tốt nhất bằng **GridSearchCV**, đánh giá kết quả mô hình sau đó lưu lại mô hình này để sử dụng về sau.
3. **Đưa ra dự đoán:** Gồm đưa ra dự đoán về xác suất chiến thắng từ đầu vào do người dùng nhập (tên đội – các tướng được chọn mỗi lượt), đồng thời xuất ra thêm các thống số chi tiết hơn như tỉ lệ thắng chi tiết của mỗi vị tướng, phong độ gần đây của đội.

Cả 3 mô hình **LightGBM, RandomForest** và **XGBoost** đều cho kết quả rất tốt và sát sao, đặc biệt hai mô hình XGBoost và LightGBM trong bước đầu huấn luyện có độ chính xác ngang nhau (0.9793). Tuy vậy chúng tôi quyết định chọn **XGBoost** để thực hiện fine-tune là vì đặc tính bộ dữ liệu ít mẫu – nhiều đặc trưng, ít gặp tình trạng overfitting hơn so với LightGBM,...

Một số **Hyperparameter** của mô hình XGBoost sau khi tiến hành fine-tune:

- 'classifier\_\_colsample\_bytree': 0.9
- 'classifier\_\_learning\_rate': 0.3
- 'classifier\_\_max\_depth': 4
- 'classifier\_\_min\_child\_weight': 1
- 'classifier\_\_n\_estimators': 300
- 'classifier\_\_subsample': 0.8

Kết quả đánh giá kiểm chéo bằng kiểm định K-Fold với 5 Folds cho ra độ chính xác khoảng **98,44%**. Còn kết quả kiểm nghiệm trên tập kiểm thử là **96,89%**. Có thể thấy sự chênh lệch nằm ở mức chấp nhận được.

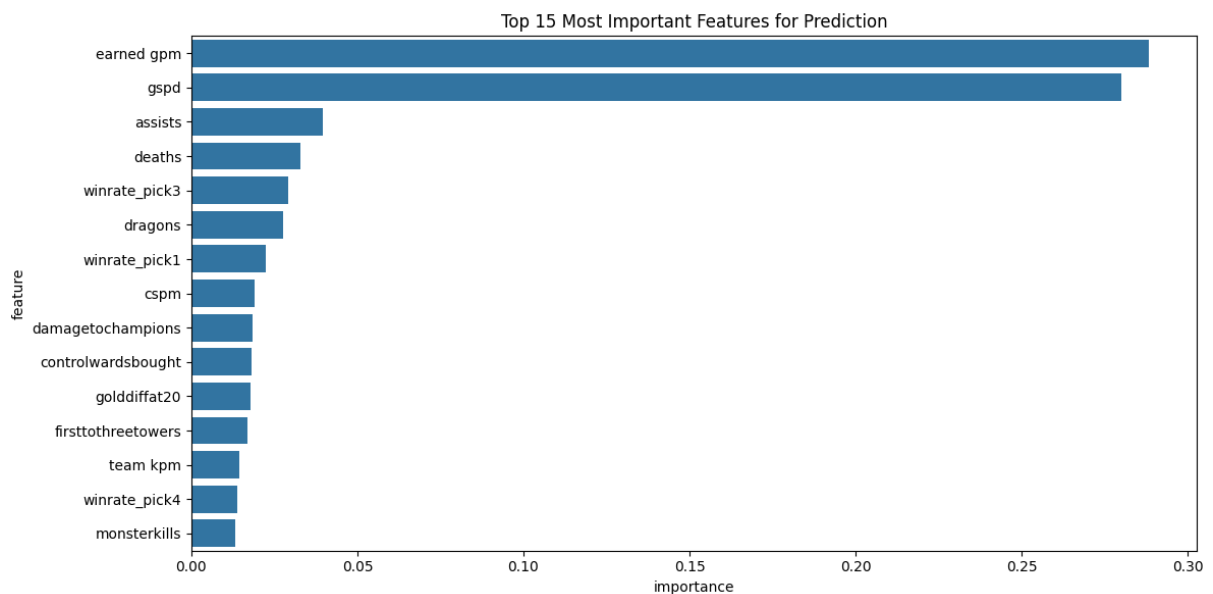
	precision	recall	f1-score	support
<b>0</b>	0.97	0.97	0.97	97
<b>1</b>	0.97	0.97	0.97	96
<b>accuracy</b>			0.97	193
<b>macro avg</b>	0.97	0.97	0.97	193
<b>weighted avg</b>	0.97	0.97	0.97	193

Bảng 3.1 Bảng đánh giá độ hiệu quả của mô hình phân loại.

Ta có thể thấy rằng:

- **Hiệu suất tổng thể của mô hình rất tốt**, với **97%** độ chính xác trên toàn bộ tập kiểm tra.
- Đối với cả hai lớp (**0** và **1**), Precision, Recall, và F1-score đều đạt **0.97**, cho thấy mô hình có khả năng phân biệt rất tốt giữa hai lớp.
- **Số lượng mẫu (Support) được phân bố đều giữa hai lớp (97 và 96 mẫu)**.

Bên dưới là biểu đồ thể hiện **15 đặc trưng quan trọng nhất** trong việc dự đoán kết quả của mô hình. Mỗi đặc trưng (feature) được sắp xếp theo mức độ quan trọng (importance) mà nó đóng góp vào khả năng dự đoán của mô hình.



Hình 3.2 Đánh giá mức độ quan trọng của các đặc trưng.

## 4. KẾT QUẢ PHÂN TÍCH

Qua quá trình phân tích thăm dò và thực hiện huấn luyện mô hình, chúng tôi quyết định vận dụng vào triển khai hai ứng dụng web tương tác – 1 trang thống kê Dashboard thông số của từng đội và 1 trang web GUI để sử dụng mô hình dự đoán trực tiếp. Nền tảng mà chúng tôi sử dụng ở phần này là **streamlit**, một thư viện hỗ trợ xây dựng các ứng dụng web trong Python.

### 4.1 Dashboard phân tích kết quả chung của từng đội.

Điểm mạnh của áp dụng Dashboard là chúng ta có thể tương tác trực tiếp với các biểu đồ hoặc bảng số liệu của từng phần, bộ lọc linh động theo mục đích sử dụng như: năm, mùa giải và đội tuyển... Trong quy mô của tập dữ liệu này chỉ mới của 1 giải và trong 1 năm duy nhất, nhưng tập dữ liệu hoàn toàn có thể được mở rộng quy mô lên nhiều giải khác nhau, thậm chí là qua từng năm.

Dashboard sẽ cung cấp thông tin cơ bản về đội tuyển trong mục **Team Performance Overview** và các phân tích chuyên sâu của đội tuyển đó bao gồm: **GPR Performance Analysis, GPR Performance Categories, GPR Performance Insights, Game Progression Analysis, Objective Control, Game Duration Analysis**.

Chúng tôi sẽ lấy ví dụ với đội tuyển **Dplus KIA**, với các số liệu qua cả 2 mùa giải trong năm 2024.

**Hình 9.23** là ví dụ khi ta lựa chọn các giá trị cho bộ lọc lần lượt là: 2024, All Split và Dplus KIA, khi này web sẽ hiển thị các thống kê của đội tuyển Dplus KIA ở toàn bộ mùa giải 2024. Tại đây ta thấy được các chỉ số cơ bản của đội tuyển bao gồm tỉ lệ thắng, KDA trung bình, lượng vàng mỗi phút, điểm tầm nhìn mỗi phút và điểm **GPR (Gold Percent Rating – Tỉ lệ vàng dẫn/bị dẫn)**. **GPR** sẽ được trực quan và phân tích kỹ hơn thông qua các biểu đồ ở trong **Hình 9.24**. Chỉ số **GPR** ở các biểu đồ này ngoài việc đánh giá bằng giá trị ghi nhận còn thông qua phân loại xếp hạng: Ở mức -4 đổ xuống thì là *Very Poor*, từ -4 đến -2 là *Poor*, từ -2 đến 0 là *Below Average*, từ 0 đến 2 tương đương *Above Average*, 2 đến 4 là *Good* và cuối cùng, từ mức 4+ đổ lên là *Excellent*.

Chỉ số **GPR** là chỉ số được tính toán sau khi kết thúc ván đấu, mang ý nghĩa rằng trong xuyên suốt cả ván đấu, đội thường thi đấu ở thể dẫn vàng hay bị dẫn vàng và % dẫn/bị dẫn là bao nhiêu. Đây là một chỉ số mà nếu hiểu rõ ý nghĩa của nó, có thể giúp

đội hiểu rõ ưu và nhược điểm của mình trong cách tiếp cận trận đấu. Một đội có **GPR cao** nhưng không thể duy trì lợi thế để giành chiến thắng cần xem xét các yếu tố như các tình huống giao tranh tổng, khả năng macro trong giai đoạn cuối trận, hoặc ra quyết định ở những thời điểm quan trọng. Ngược lại, đội có **GPR thấp** nhưng đạt được tỷ lệ thắng cao có thể đã tối ưu hóa lối chơi dựa trên việc phản công hoặc chờ đợi cơ hội lật kèo, chủ động nhường thế trận để đảm bảo điều kiện thắng của đội. Đó cũng là lý do mà chúng tôi dành hẳn 1 mục riêng cho chỉ số này.

**Hình 9.26** thể hiện tương quan về tỉ lệ kiểm soát rồng và trụ của DK so với mặt bằng chung tại LCK. Có thể thấy DK có khả năng kiểm soát công trình khá tốt, tuy nhiên kiểm soát mục tiêu Rồng có phần yếu hơn. Từ đây có thể rút ra kết luận rằng DK là một đội thường tập trung vào yếu tố lợi thế đi đường và chủ động trao đổi rồng để lấy các công trình như giáp trụ và trụ hay sâu hư không, sứ giả.

**Hình 9.27** thể hiện phân phối thời lượng ván đấu của DK trong năm 2024. Quan sát biểu đồ ta có thể rút ra được rằng DK là một đội tuyển thường sẽ kết thúc trận đấu ở khoảng 26-35 phút, giống với hầu hết các đội khác tại LCK.

#### 4.2 GUI áp dụng mô hình dự đoán tỉ lệ thắng dựa vào phần chọn tướng của đội.

Để việc sử dụng mô hình dễ dàng hơn cũng như triển khai thực tế, chúng tôi đã xây dựng một ứng dụng web có tác dụng như một GUI (Graphics User Interface) nhằm sử dụng mô hình dự đoán đã được huấn luyện. Từ đây ta có thể tương tác, lựa chọn giá trị từ đó mô hình sẽ sử dụng dữ liệu có sẵn để đưa ra dự đoán tỉ lệ thắng của đội dựa trên các lượt chọn tướng, đồng thời cũng hiển thị thêm các thông tin liên quan khác giúp người dùng có cái nhìn chi tiết hơn.

**Hình 9.28** là ví dụ thực tế khi tiến hành sử dụng ứng dụng web. Ta sẽ được điền tên đội tuyển và các vị tướng mà đội sẽ lựa chọn **THEO TỪNG LƯỢT**. Sau khi đã nhập đủ đầu vào, nhấn vào nút “Kết quả” sẽ đưa ra kết quả dự đoán trận đấu đầy với thông số hai bên cũng như phân tích chi tiết cho từng đội.

Một điểm đáng chú ý khi sử dụng là vì tính chất và quy mô của mô hình còn hạn chế, cho nên mô hình chỉ có thể dự đoán dựa trên các lượt chọn tướng mà nó **đã được học**, có nghĩa là nếu như bạn đưa vào 1 lượt lựa chọn mà mô hình chưa gặp qua bao giờ,



nó sẽ không thể đưa ra dự đoán và trả về thông báo “Unseen Data”. Ngoài ra, **thứ tự chọn tướng** và **đội tuyển sử dụng** cũng đóng một vai trò rất quan trọng trong việc dự đoán: lấy ví dụ đội A sử dụng đội hình với 5 vị tướng và đội B tương tự 5 vị tướng; với thứ tự chọn tướng khác nhau, kết quả dự đoán sẽ khác nhau; với những đội tuyển khác nhau, tỉ lệ thắng trả về cũng sẽ khác nhau. (cùng một đội hình nhưng Gen.G sử dụng và DRX sử dụng sẽ cho ra kết quả khác nhau khi đối đầu với T1)

Ta sẽ tiến hành phân tích chi tiết dưới trên ví dụ thực tế từ **Hình 9.28** - lựa chọn 2 đội là T1 và DRX và các lựa chọn tướng lần lượt của **T1** là **Lucian, Nami, Oriana, Aatrox và Taliyah** và lựa chọn tướng của **DRX** lần lượt là **Aphelios, Milio, Azir, Rell và Jax**. Kết quả của dự đoán trận đấu trên được hiển thị ở **Hình 9.29** cho ta thấy mô hình đã dự đoán kết quả trận đấu này có **tỉ lệ thắng của T1 cao hơn với 86,3%** và **DRX là 13,7%**. Mô hình còn cho ta thấy số lần các đội tại LCK đã sử dụng khung đội hình này trong mùa giải cũng như tỉ lệ thắng của đội hình đó (24 lần đối với đội hình mà T1 sử dụng và tỉ lệ thắng là 56,8%; 21 lần đối với DRX và tỉ lệ thắng là 63,5%).

Các **Hình 9.30, Hình 9.31, Hình 9.32, Hình 9.33** là phần phân tích chi tiết cho từng đội tuyển.

Đầu tiên là **Hiệu suất của đội tuyển** thông qua Spider Chart (Sơ đồ mạng nhện) với các thông số bao gồm: **Economy** (thể hiện lượng vàng/kinh tế), **Objective Control** (kiểm soát mục tiêu), **Vision Control** (Kiểm soát tầm nhìn), **Early Game** (khả năng chiến đấu đầu trận), **Late Game** (Khả năng chiến đấu về cuối trận) và **Combat Power** (Chỉ số giao tranh). Phần lưới mở rộng về phía nào chứng tỏ khả năng của đội tuyển về chỉ số đấy càng cao. Ở **Hình 9.30** ta thấy rằng T1 có khả năng cân bằng các chỉ số cho ta thấy đây là đội tuyển có khả năng đồng đều giữa lượng kinh tế, khả năng kiểm soát mục tiêu và tầm nhìn cũng như khả năng chiến đấu đầu và cuối game đều ở mức khá. Trong khi đó với DRX có thể thấy đội tuyển này có khả năng kiểm soát tầm nhìn nhìn hơn, tuy nhiên các chỉ số liên quan đến mục tiêu hay giao tranh đều chỉ ở mức trung bình-yếu.

Thứ hai là **Thống kê chi tiết 10 ván gần nhất** của 2 đội. **Hình 9.31** cho ta thấy các thông số của từng đội tuyển lần lượt là: **First Dragon Rate** (Tỉ lệ tiêu diệt rồng đầu), **First Herald Rate** (Tỉ lệ tiêu diệt sứ giả khe nứt đầu), **First Tower Rate** (Tỉ lệ phá hủy



trụ đầu), **GPM** (Lượng vàng mỗi phút), **Vision Score** (Điểm tầm nhìn), **First Blood Rate** (Tỉ lệ có chiến công đầu), **Gold Diff @15** (Lượng vàng cách biệt ở phút thứ 15), **Gold Diff @20** (Lượng vàng cách biệt ở phút 20).

Có thể thấy trên hình các chỉ số của T1 khá đồng đều, cho thấy trong phong độ trong giai đoạn gần đây của T1 không có sự biến động, đặc biệt, nếu xét về lượng vàng cách biệt ở phút thứ 15 hay 20, T1 hoàn toàn vượt trội so với DRX phía bên kia. thấy đội tuyển này đánh đầu game không tốt nên thường để đối thủ dẫn trước về chỉ số mục tiêu hay trụ, từ đó thường bị đối phương dẫn vàng ở phút 15 và phút 20, điểm đáng khen của DRX là điểm tầm nhìn của họ rất tốt, cho ta thấy khả năng kiểm soát bản đồ của họ ở mức tuyệt vời.

Cuối cùng là **Thống kê tỉ lệ thắng của các vị tướng được lựa chọn**. **Hình 9.32** và **Hình 9.33** cho ta thấy lượt pick của từng vị tướng của mỗi đội, ta có thể sử dụng nút mũi tên phía bên phải để xem thêm các chỉ số về Winrate (tỉ lệ thắng) cụ thể của vị tướng đó đi kèm số trận đấu theo các tiêu chí khác nhau như: Tỉ lệ thắng của đội khi pick tướng A ở lượt X; Tỉ lệ thắng chung của tướng A khi pick ở lượt X, v.v...

Cụ thể ở **Hình 9.32** có thể thấy vị tướng Aatrox được T1 chọn ở lượt thứ 4 có các thông số tỉ lệ thắng chi tiết như sau: tỉ lệ thắng của **T1 khi chọn Aatrox ở lượt 4 là 66,7%** (trong tổng số 3 ván đấu), tỉ lệ thắng của **T1 khi chọn Aatrox là 77,8%** (trong tổng số 18 ván đấu), tỉ lệ thắng của **Aatrox khi được chọn ở lượt 4 là 55,2%** (trong tổng số 29 ván đấu), **tỉ lệ thắng tổng của vị tướng Aatrox là 49,5%** (trong tổng số 109 ván đấu).

## 5. KẾT LUẬN

Đồ án đã thực hiện xây dựng hệ thống phân tích và dự đoán kết quả trận đấu trong giải đấu LCK thông qua nhiều giai đoạn quan trọng. Quá trình bắt đầu từ việc thu thập và tiền xử lý dữ liệu, tập trung vào giải đấu LCK và các đội tuyển, sau đó tiến hành phân tích thăm dò dữ liệu (EDA) để tạo ra các biểu đồ và thống kê có ý nghĩa thực tiễn trong trò chơi. Một điểm nổi bật là việc phát triển một mô hình riêng để thực hiện quy trình dự đoán tỷ lệ thắng, bao gồm các bước từ chuẩn bị dữ liệu, chuẩn hóa số liệu, đến việc huấn luyện và so sánh nhiều mô hình khác nhau thông qua Pipeline, cuối cùng đạt được độ chính xác ấn tượng khoảng 97% trên tập kiểm thử. Ngoài ra, đồ án cũng đã phát triển thành công dashboard phân tích bằng Streamlit và hệ thống dự đoán trận đấu trực tiếp. Tuy nhiên, đồ án vẫn còn một số hạn chế đáng chú ý như: tiềm năng chưa được khai thác hết trong quá trình EDA do giới hạn về kiến thức và thời gian, cũng như những bất cập về tính ứng dụng thực tế của mô hình do tập dữ liệu còn nhỏ và chỉ tập trung vào một giải đấu duy nhất, đồng thời cách tương tác và sử dụng mô hình vẫn còn những hạn chế nhất định cần được cải thiện.

## **6. TÀI LIỆU THAM KHẢO**

- [1] League of Legends esports stats - Oracles Elixir. Link: <https://oracleselixir.com> (26/10/2024).
- [2] OE Public Match Data. Link: [OE Public Match Data](#) (26/10/2024).
- [3] Streamlit documentation. Link: <https://docs.streamlit.io> (10/12/2024)

## 7. PHỤ LỤC PHÂN CÔNG NHIỆM VỤ

STT	Thành viên	Nhiệm vụ
1	Nguyễn Hoàng Duy	EDA, xây dựng mô hình và triển khai kết quả, viết báo cáo. Hoàn thiện, chỉnh sửa đồ án, báo cáo (toàn bộ) sau buổi báo cáo.
2	Trần Thanh Hải	Slide thuyết trình, xây dựng mô hình.
3	Nguyễn Minh Trí	Viết báo cáo, xử lý dữ liệu.
4	Cao Trường Vũ	Viết báo cáo, slide thuyết trình Hoàn thiện báo cáo chương 4 sau buổi báo cáo.

8. PHỤ LỤC BẢNG

Bảng 2.1 Mô tả chi tiết các đặc trưng của bộ dữ liệu.....10

Bảng 2.2 Các cột có tỉ lệ khuyết trên 70%.....11

Bảng 2.3 Thống kê dữ liệu khuyết cuối cùng.....12

Bảng 2.4, 2.5 Thống kê bộ dữ liệu hậu tiền xử lý.....12

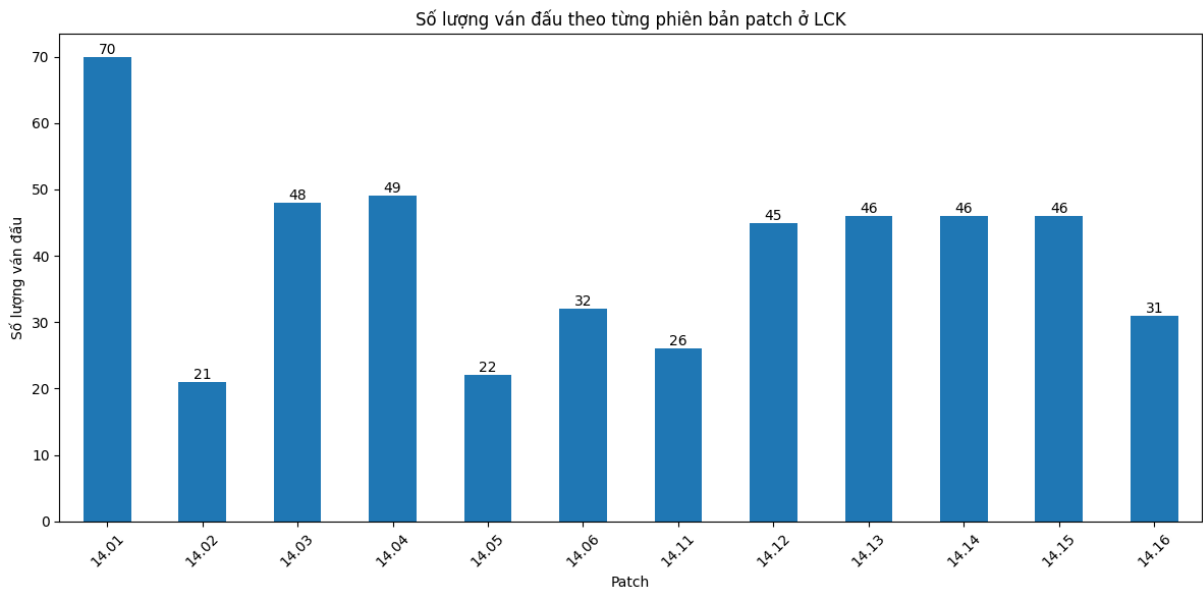
Bảng 3.1 Bảng đánh giá độ hiệu quả của mô hình phân loại.....19

Patch	Ngày cập nhật	Patch	Ngày cập nhật
14.1	10/01/2024	14.11	30/05/2024
14.2	24/01/2024	14.12	12/06/2024
14.3	07/02/2024	14.13	26/06/2024
14.4	22/02/2024	14.14	17/07/2024
14.5	06/03/2024	14.15	31/07/2024
14.6	20/03/2024	14.16	14/08/2024

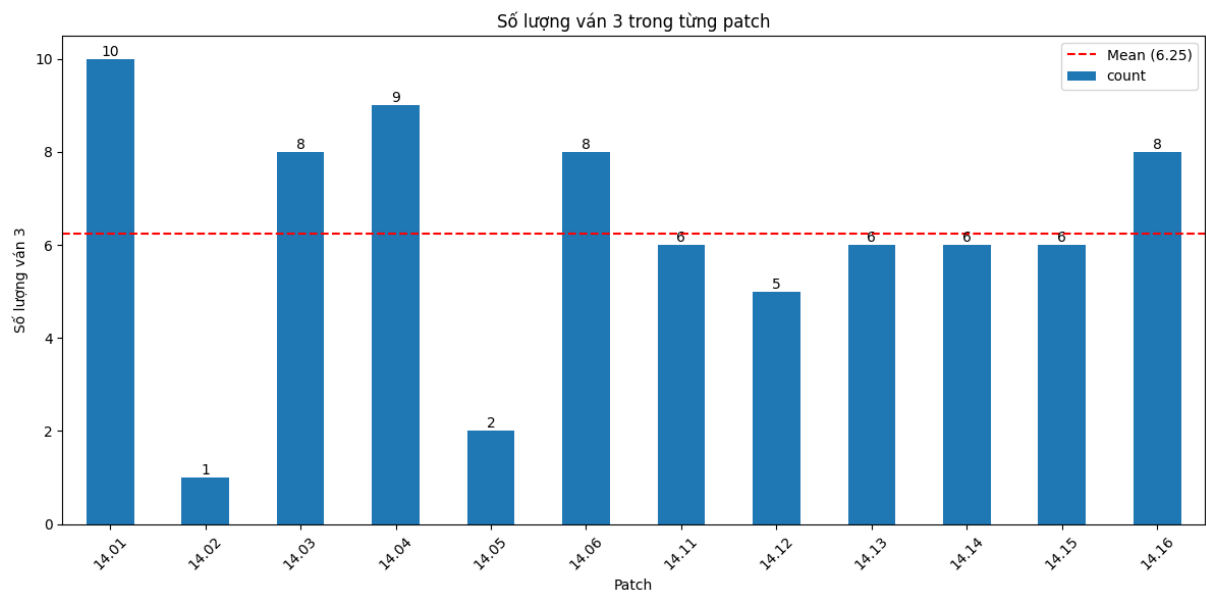
Bảng 8.1 Ngày ra mắt một số phiên bản cập nhật.

9. PHỤ LỤC HÌNH ẢNH

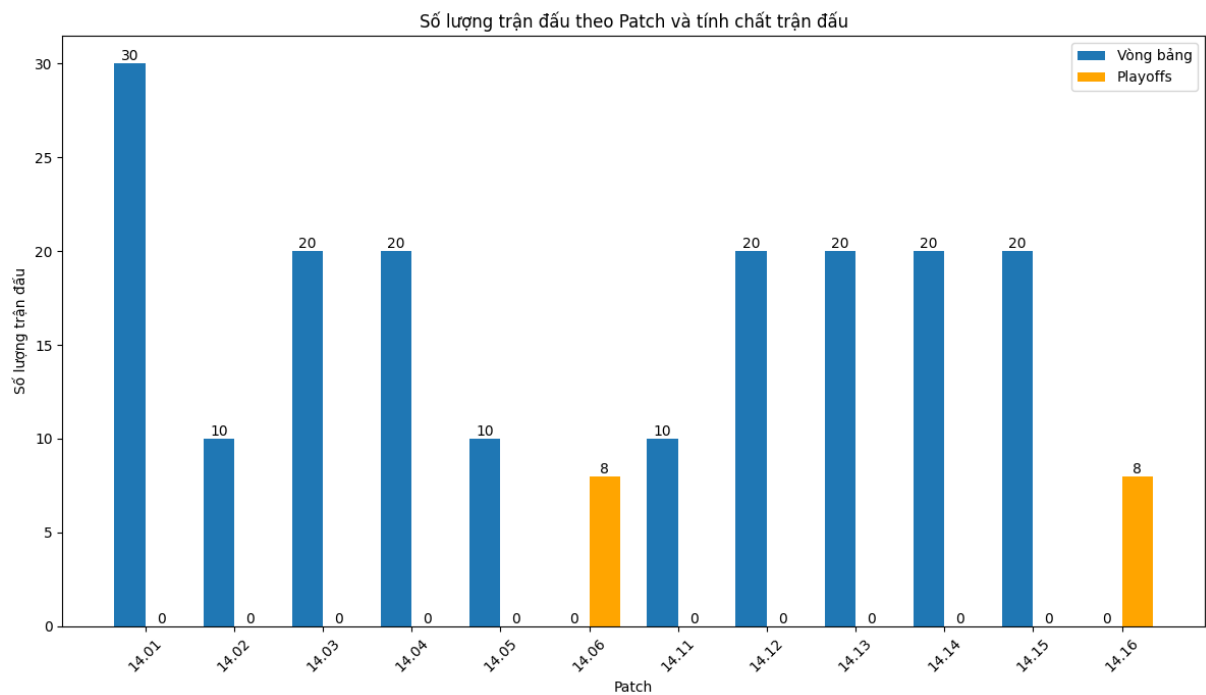
Mục 3.1 Các phân tích theo số lượng ván đấu/trận đấu



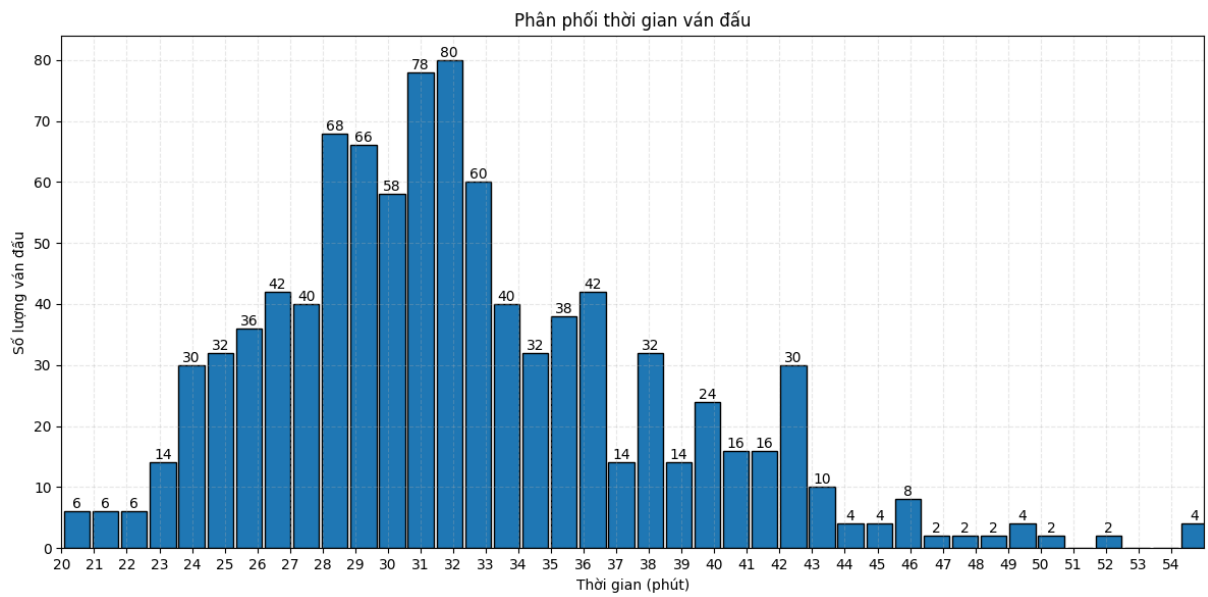
Hình 9.1 Thống kê số lượng ván đấu theo các bản cập nhật.



Hình 9.2 Thống kê số lượng ván 3 trong mỗi phiên bản.

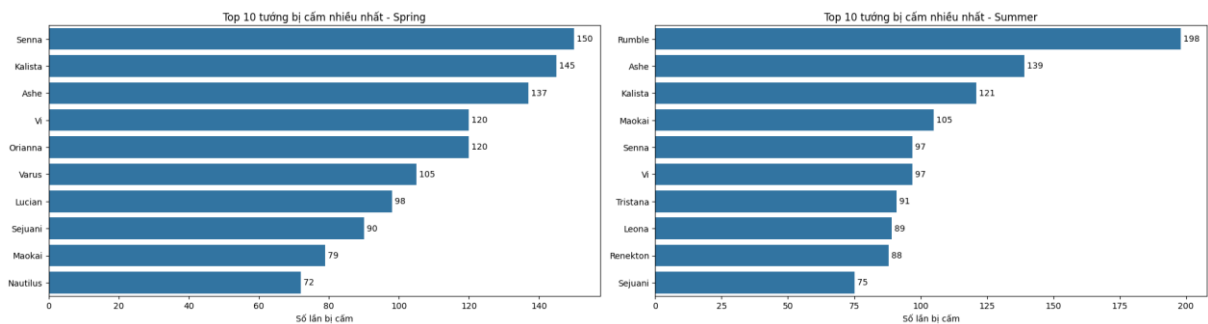


Hình 9.3 Thống kê số lượng trận đấu theo patch và tính chất trận đấu.

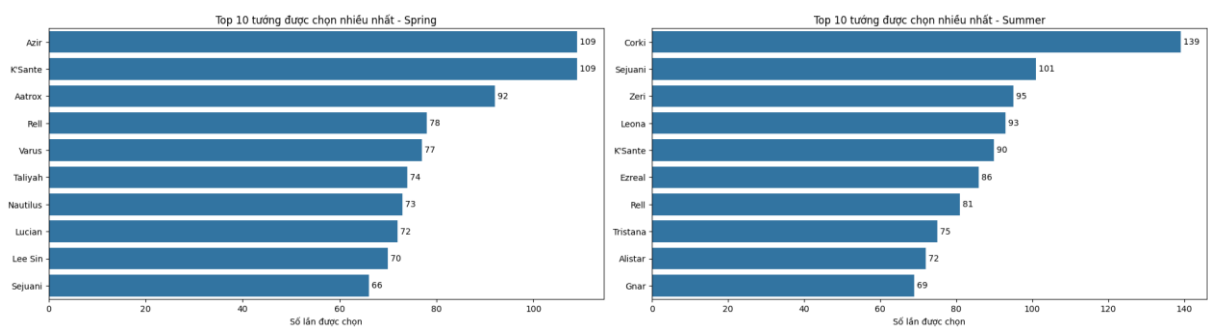


Hình 9.4 Thống kê số lượng ván đấu theo thời gian ván đấu.

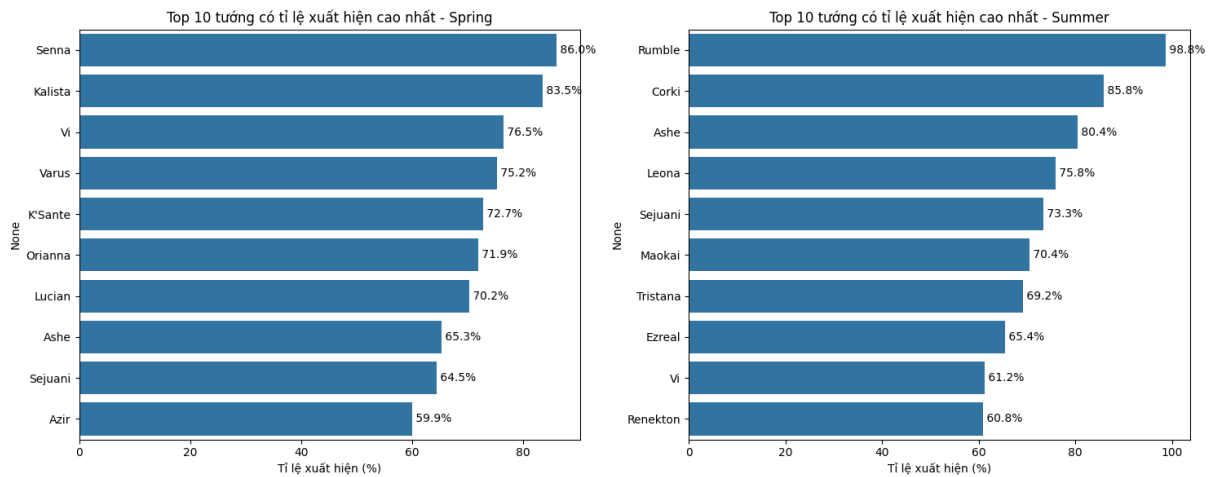
### Mục 3.2 Các phân tích theo tỉ lệ cấm/chọn



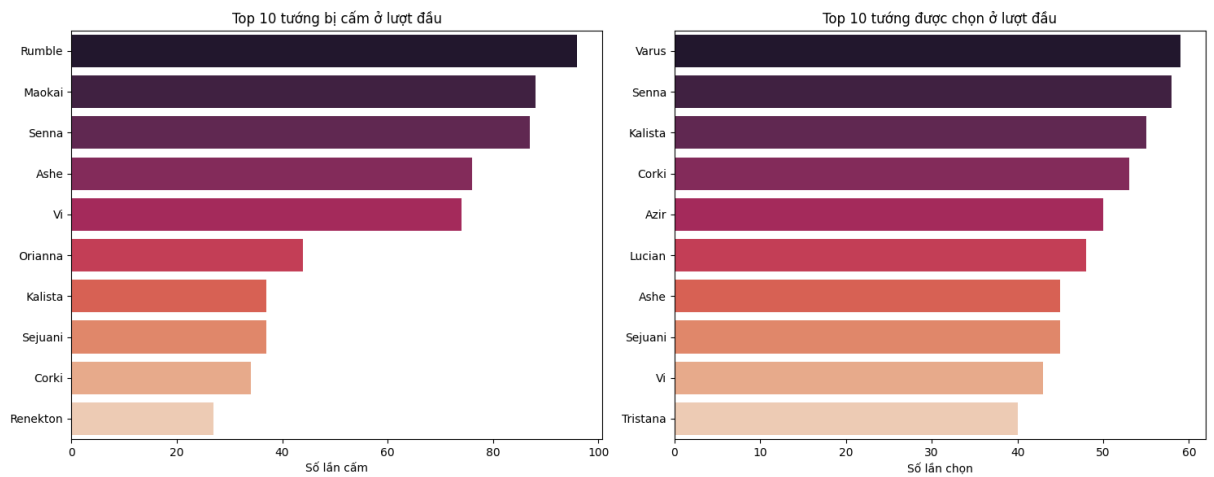
Hình 9.5 Top 10 tướng bị cấm nhiều nhất ở từng mùa giải.



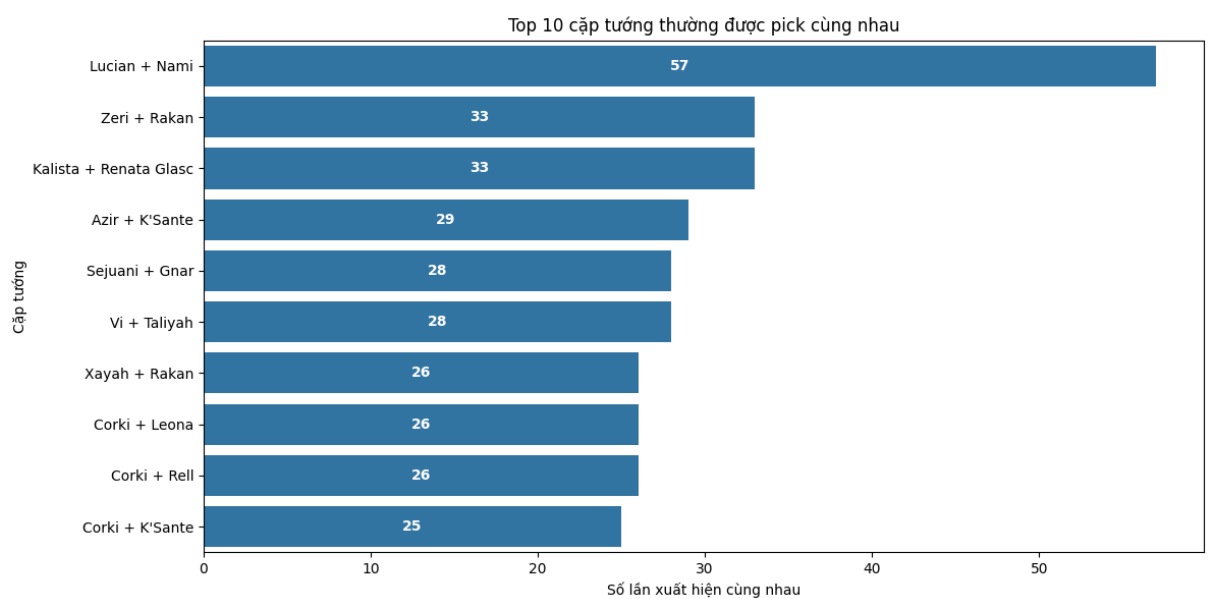
Hình 9.6 Top 10 tướng được chọn nhiều nhất ở từng mùa giải.



Hình 9.7 Top 10 tướng có tỉ lệ xuất hiện cao nhất.



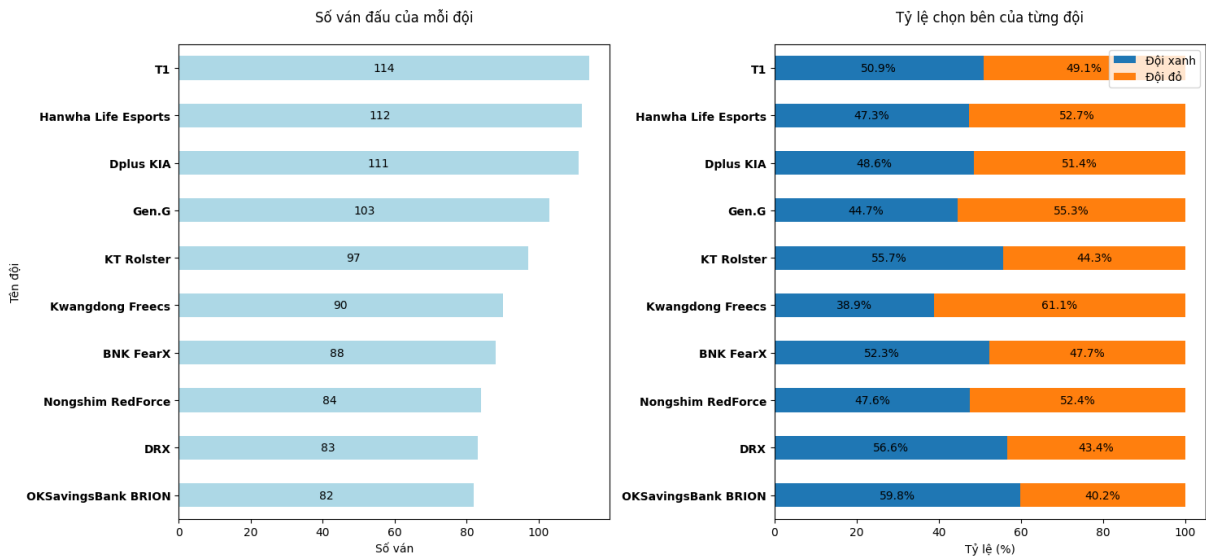
Hình 9.8 Top 10 tướng được cấm/chọn ở lượt đầu.



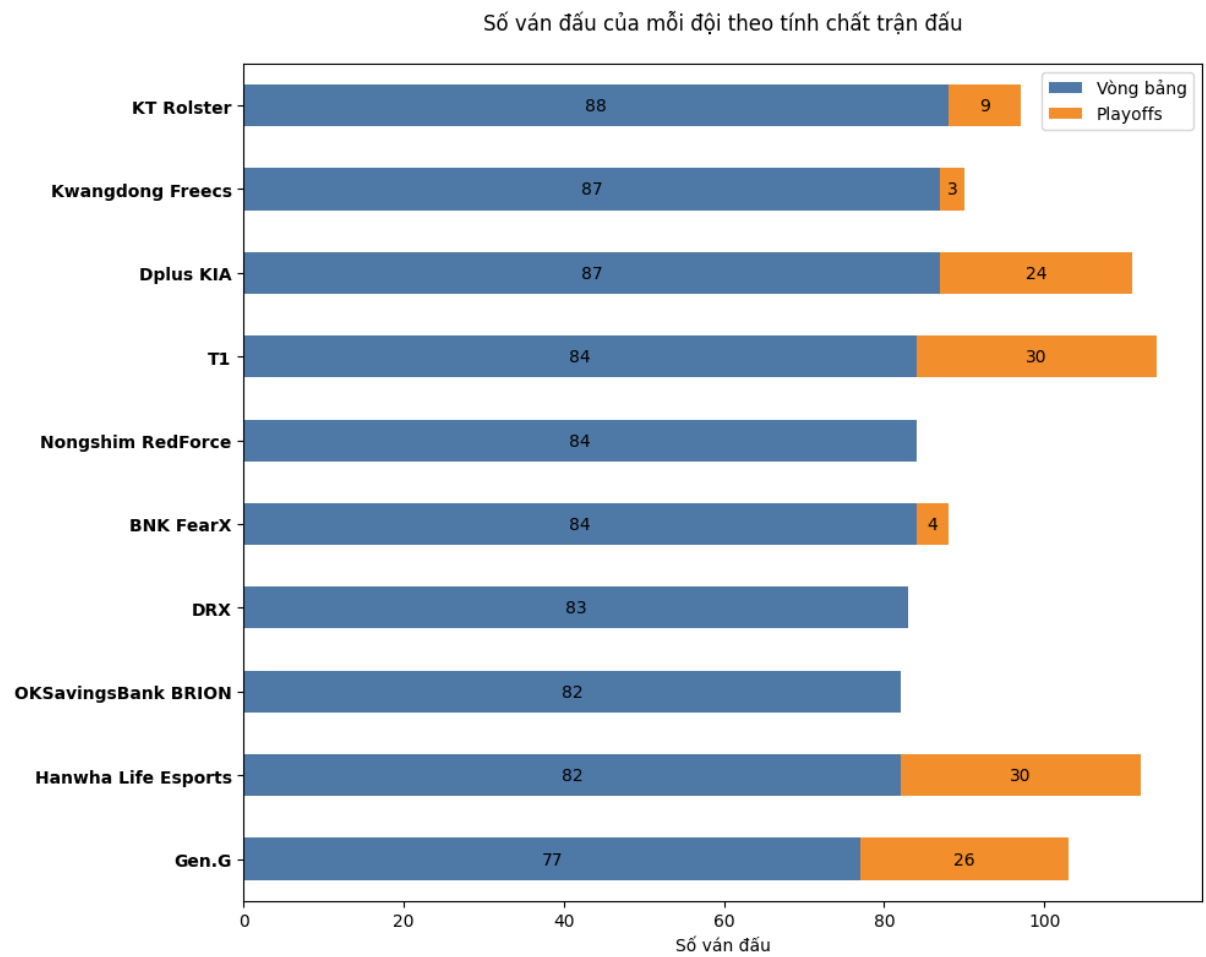
Hình 9.9 Top 10 cặp tướng được pick cùng nhau.



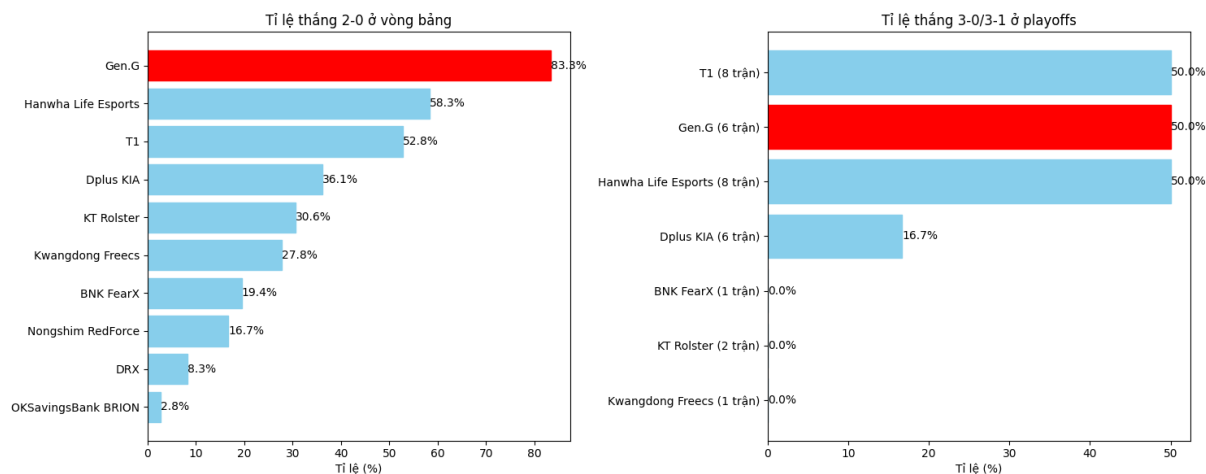
Mục 3.3 Các phân tích theo đội tuyển.



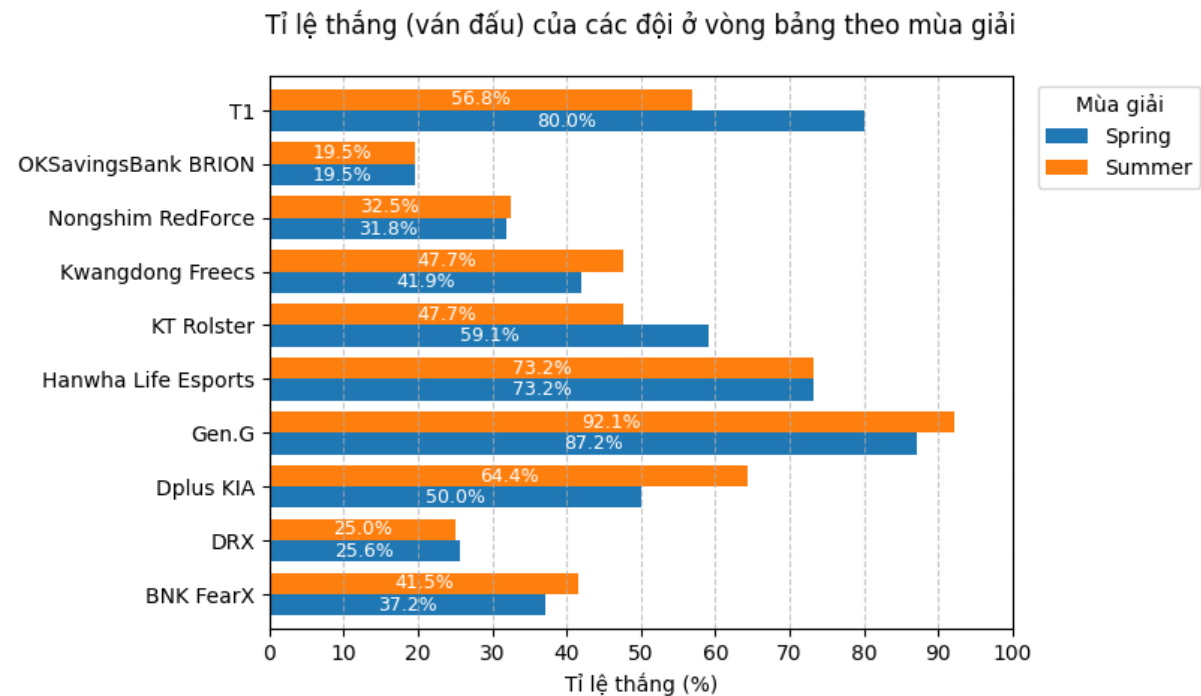
Hình 9.10 Thống kê số lượng ván đấu, tỉ lệ chọn bên của mỗi đội.



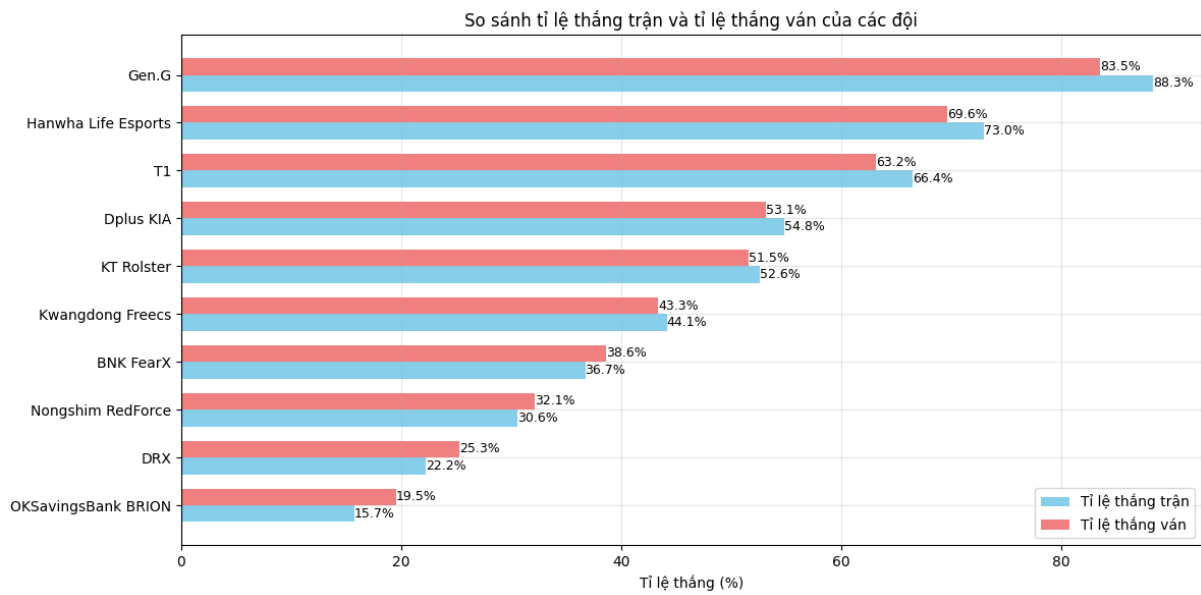
Hình 9.11 Số lượng ván đấu theo tính chất trận đấu.



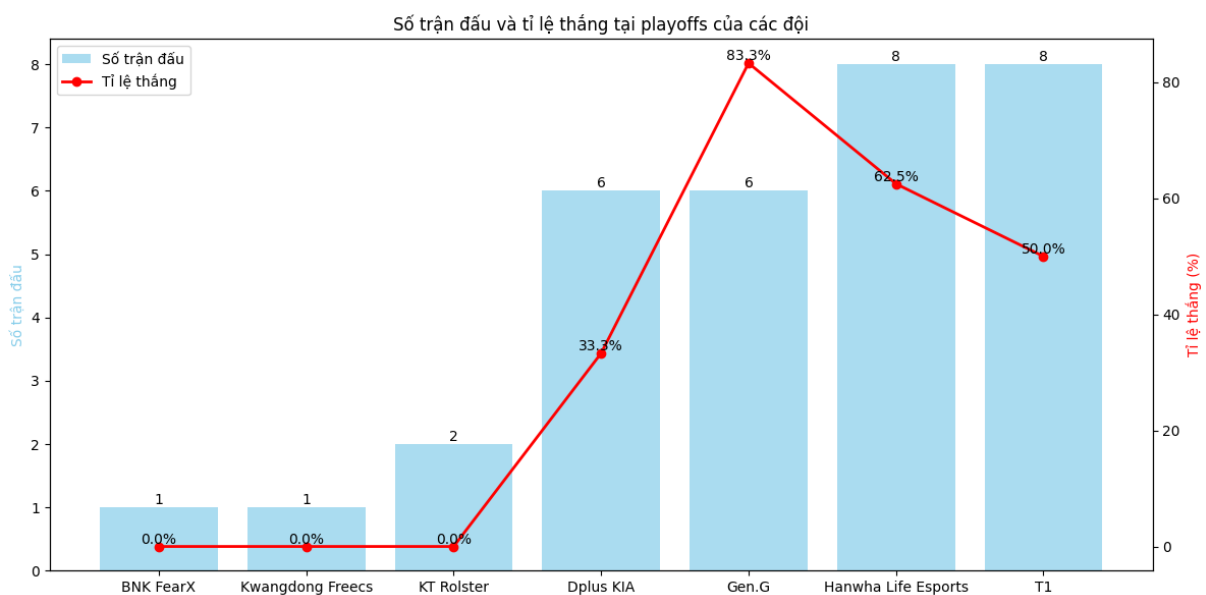
Hình 9.12 Thông số thắng áp đảo của Gen.G so với các đội.



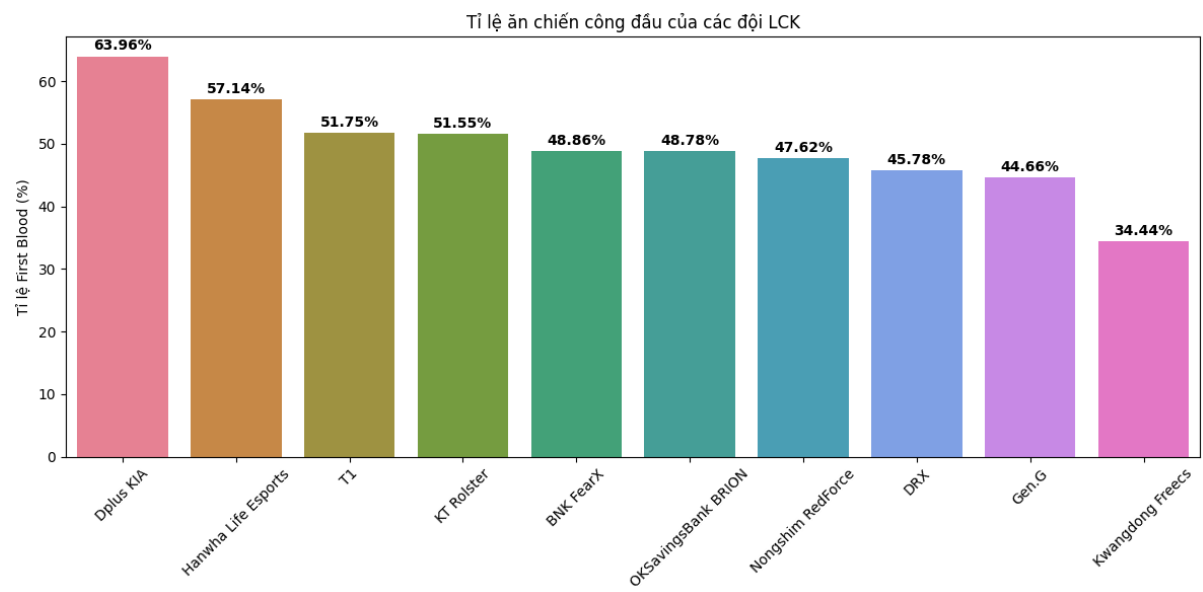
Hình 9.13 Tỉ lệ thắng (ván đấu) của các đội ở vòng bảng theo mùa giải.



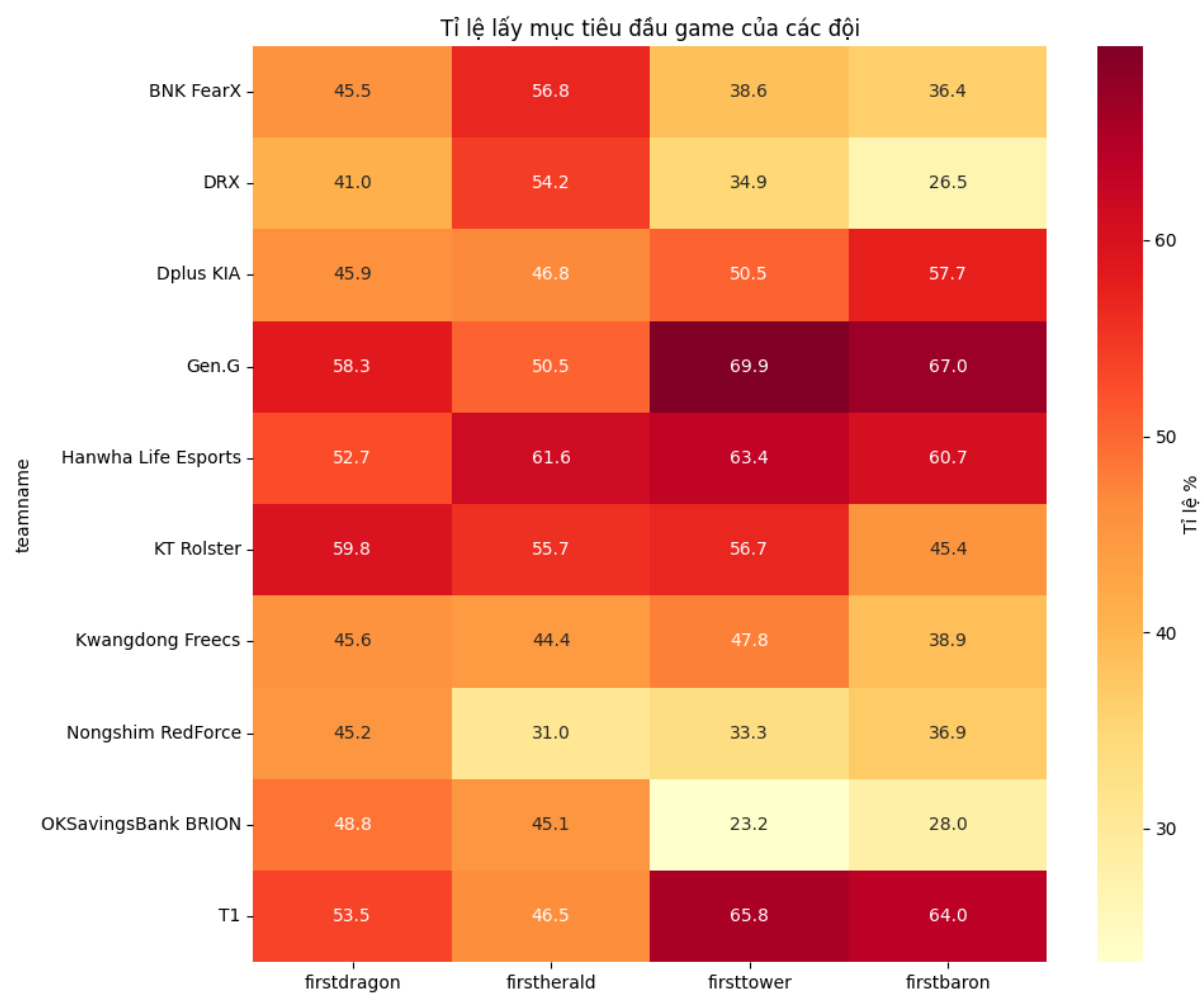
Hình 9.14 So sánh tỉ lệ thắng trận đấu với tỉ lệ thắng ván đấu của các đội.



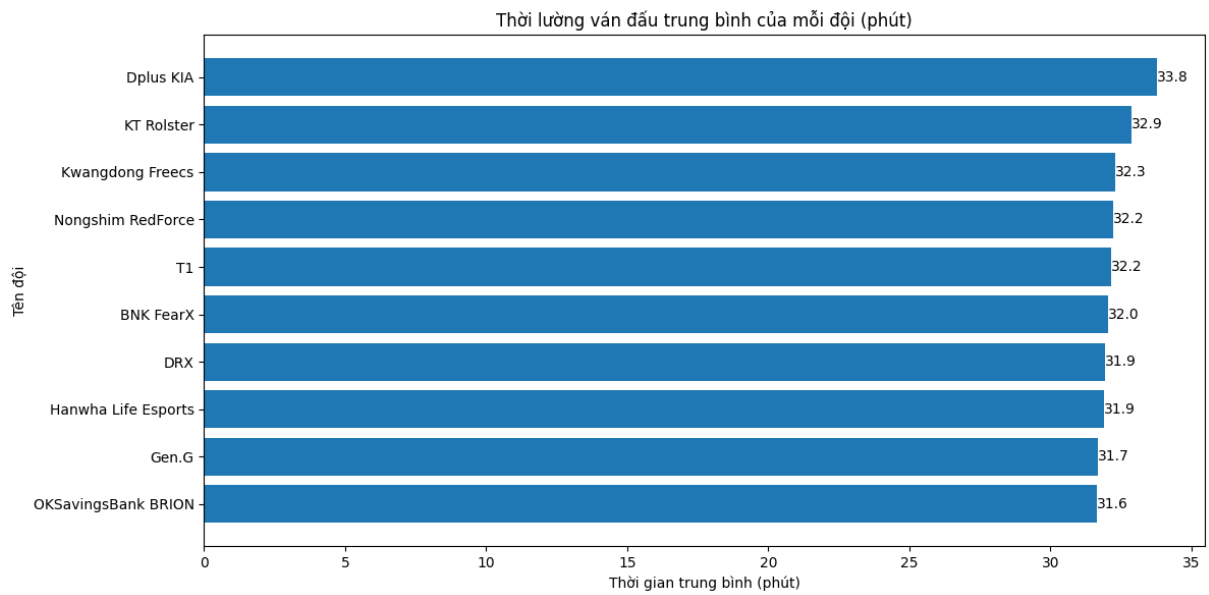
Hình 9.15 Số trận đấu tại playoffs và tỉ lệ thắng của từng đội.



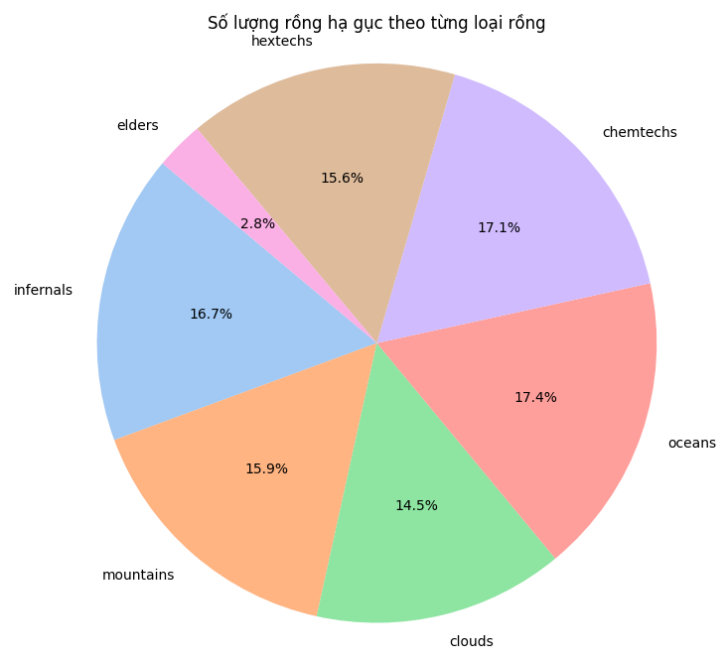
Hình 9.16 Tỉ lệ ăn chiến công đầu mỗi đội.



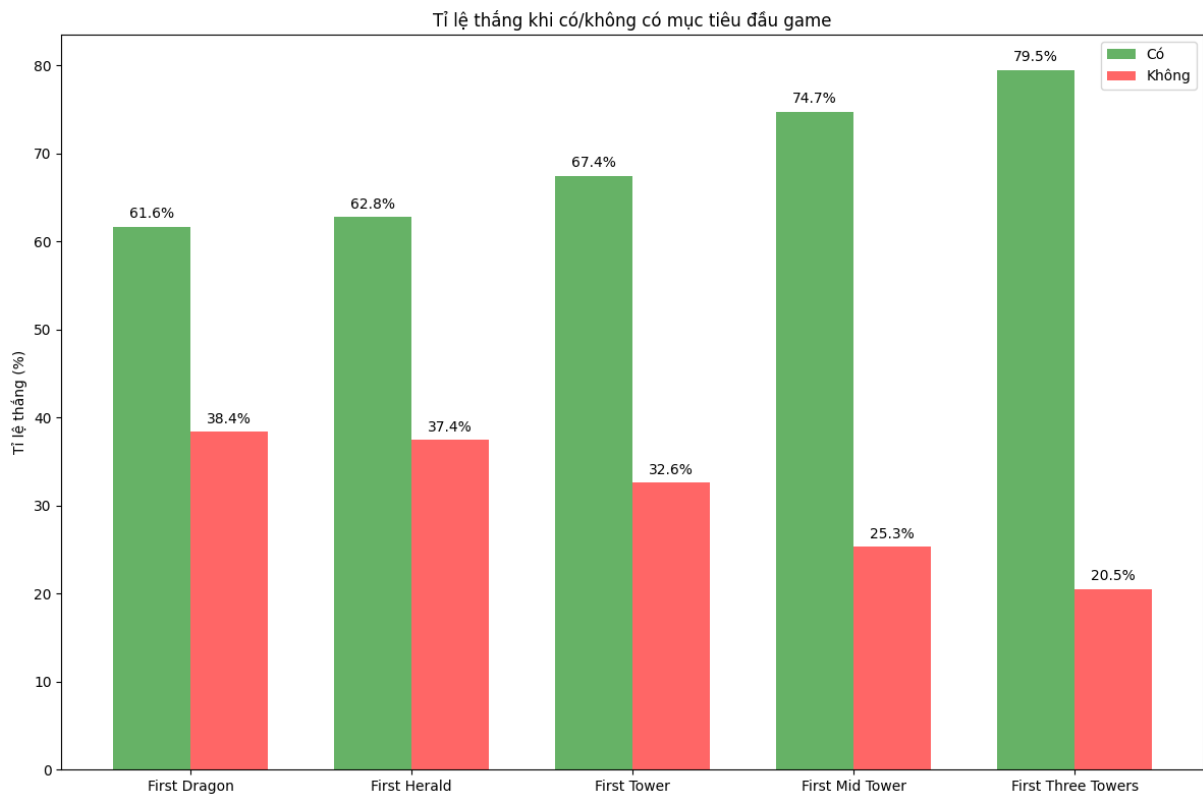
Hình 9.17 Tỉ lệ lấy các mục tiêu đầu game của mỗi đội.



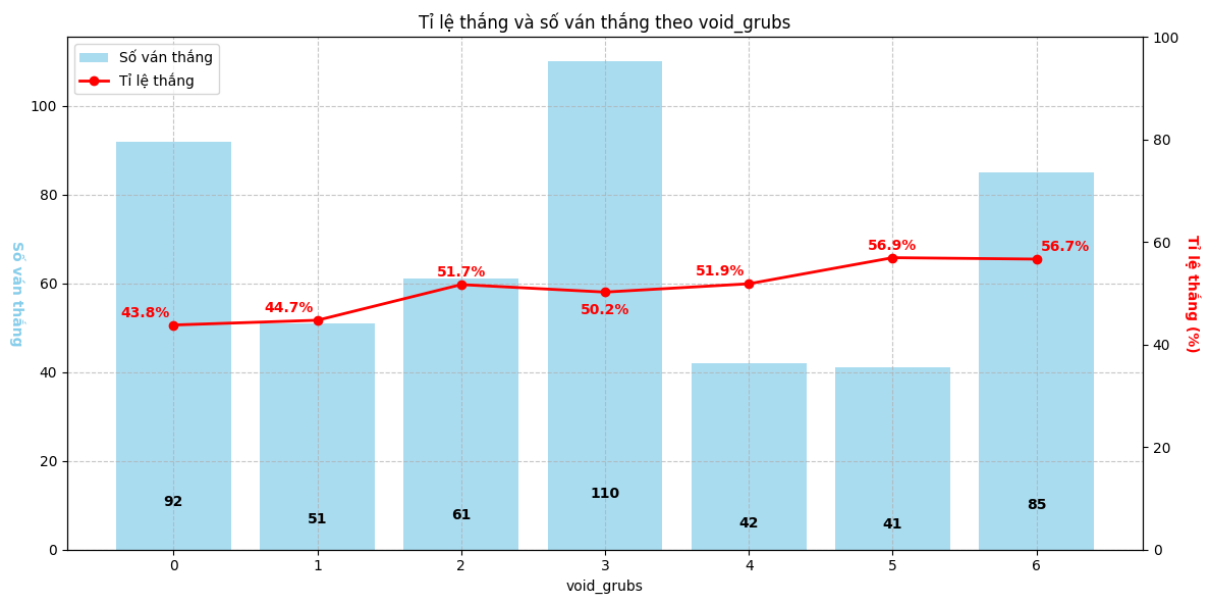
Hình 9.18 Thời lượng ván đấu trung bình theo từng đội.



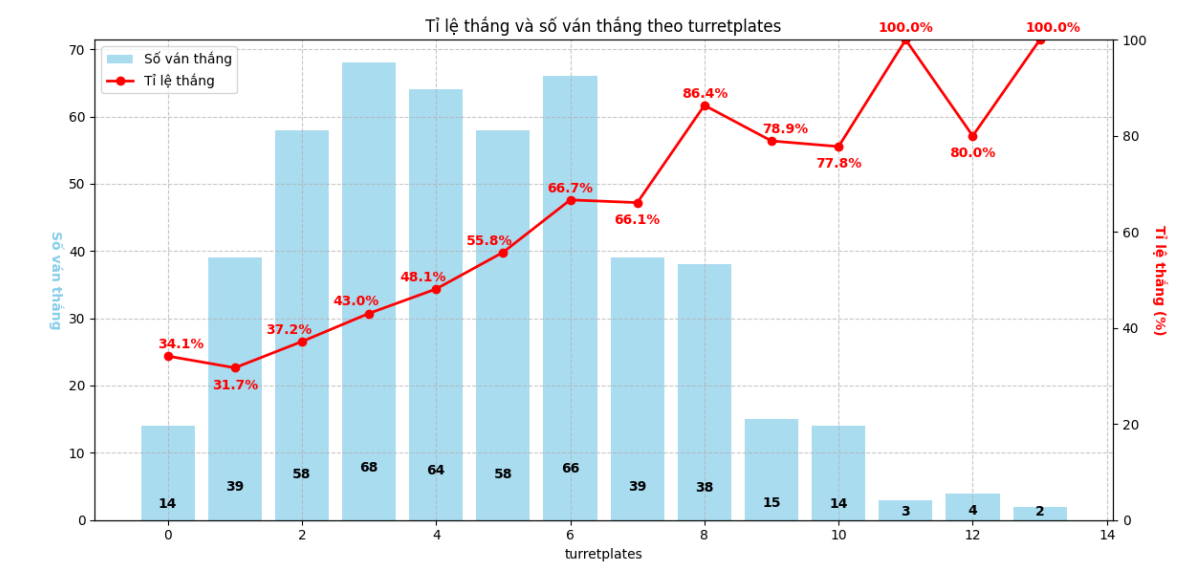
Hình 9.19 Số lượng rồng bị hạ theo từng loại.



Hình 9.20 Tỉ lệ thắng khi có/không có mục tiêu đầu game

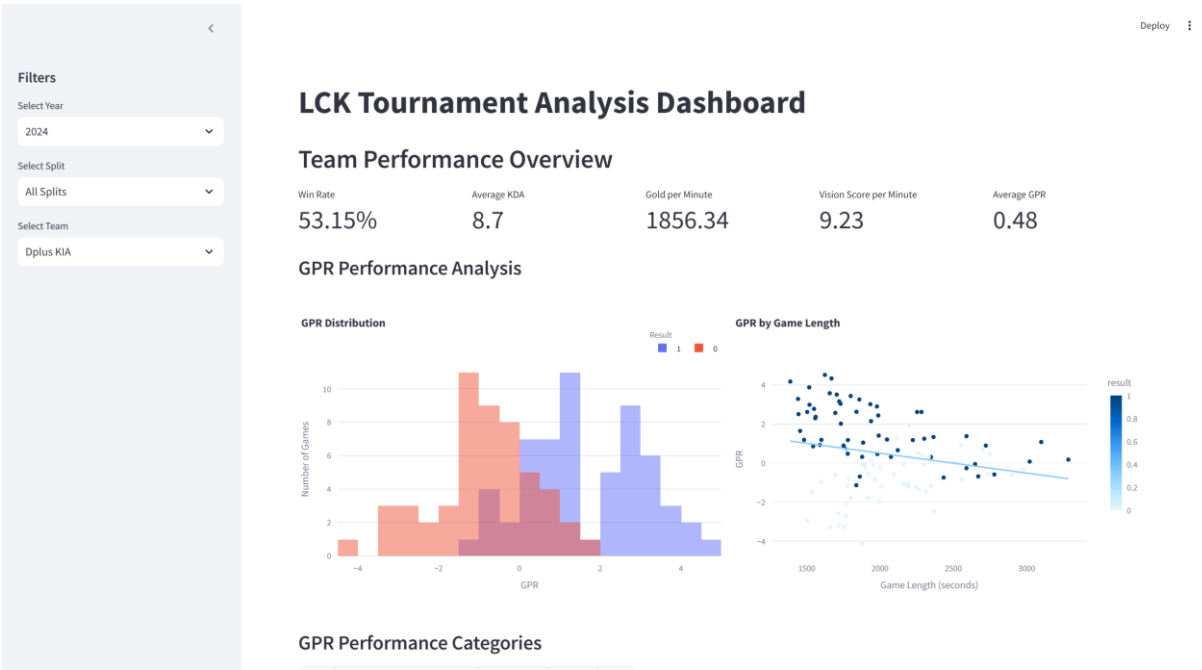


Hình 9.21 Tỉ lệ và số ván thắng với lượng sâu hư không.

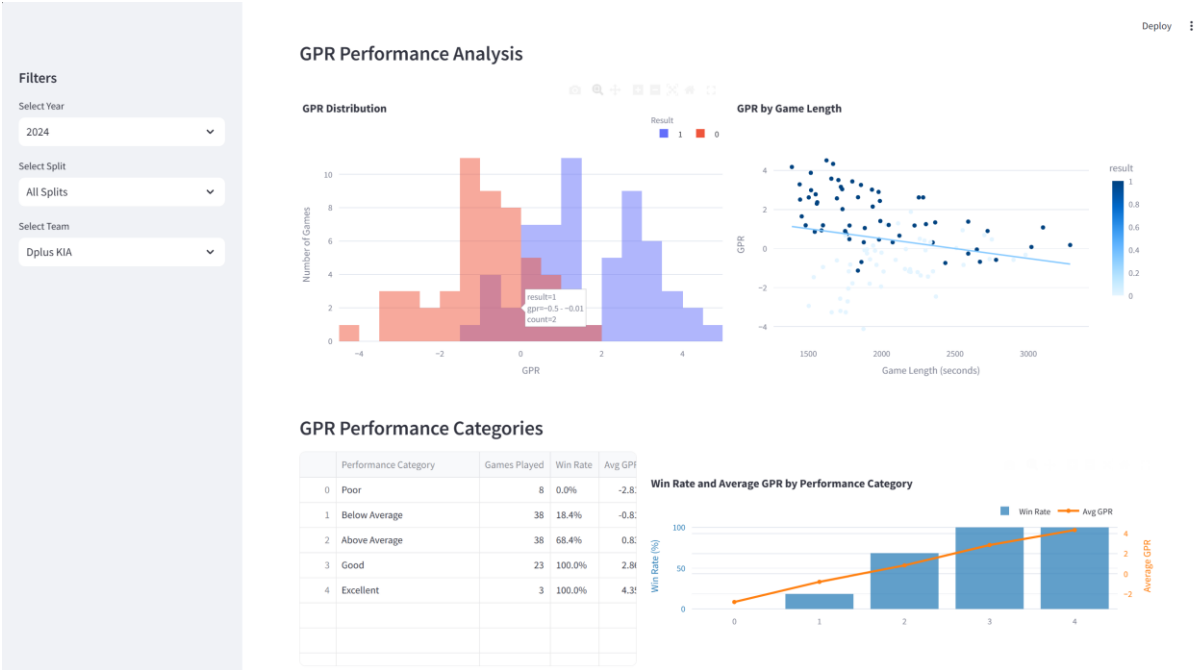


Hình 9.22 Tỉ lệ thắng và số ván thắng theo lượng giáp trụ.

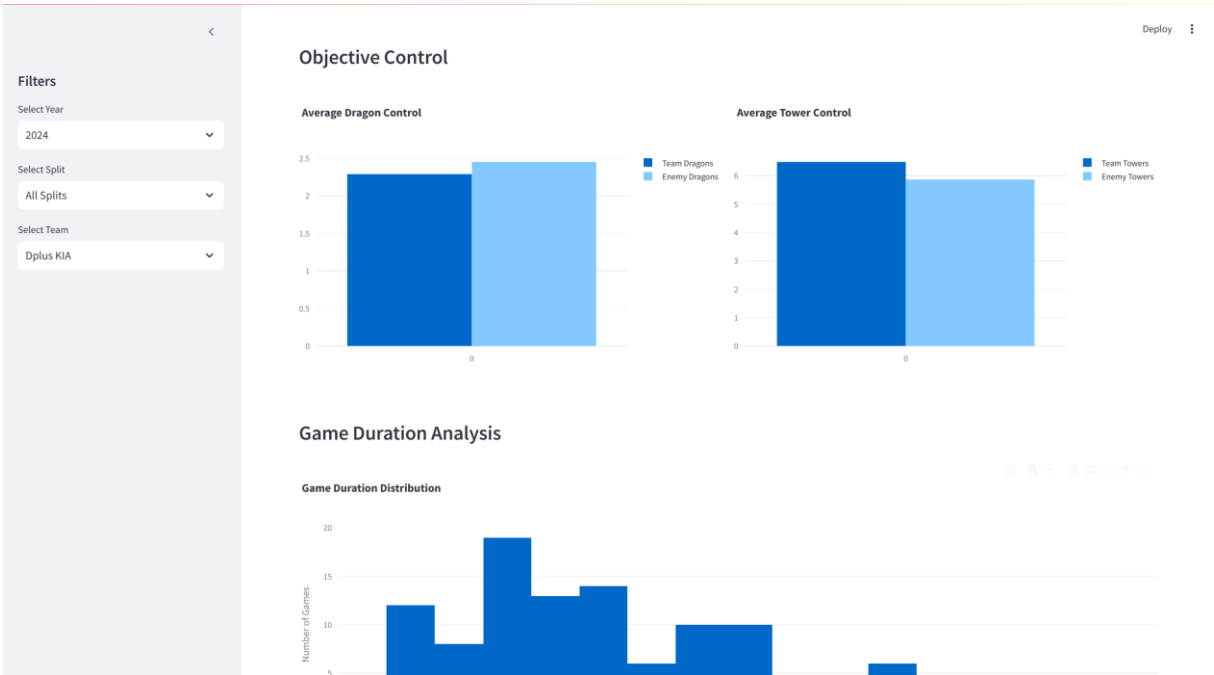
Mục 4.1 Dashboard



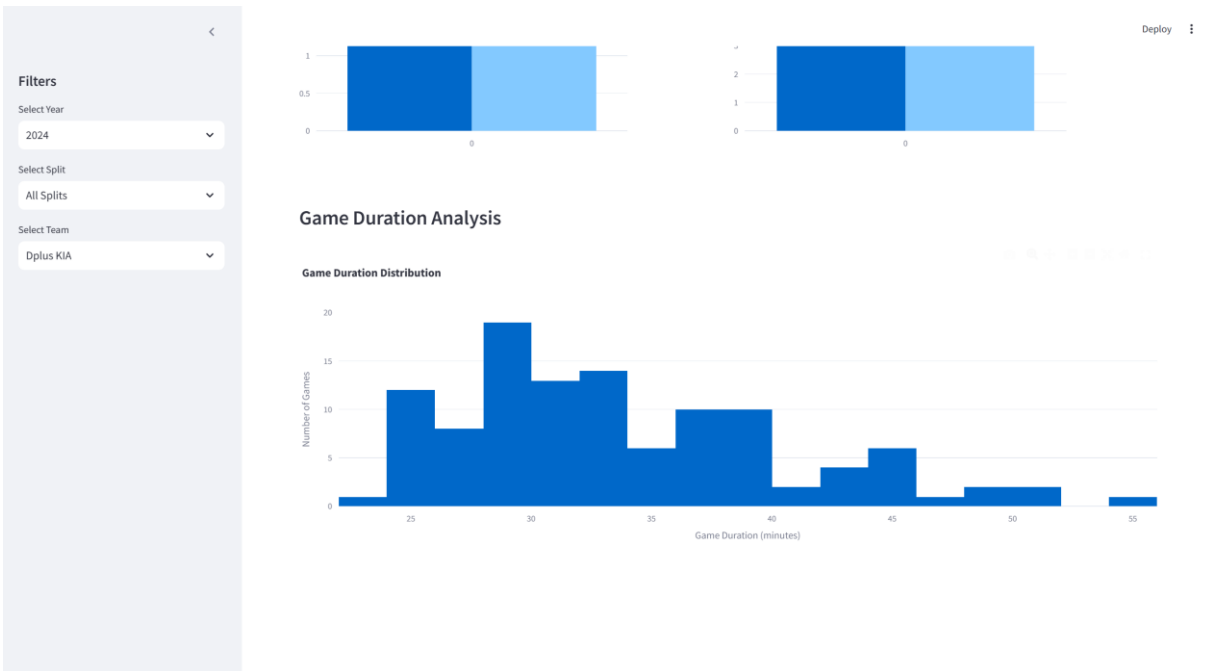
Hình 9.23







Hình 9.26



Hình 9.27

Hình 9.28

21

Deploy ⋮

### Thống kê chi tiết thông số trong 10 ván đấu gần nhất

T1	DRX
First Dragon Rate	First Dragon Rate
40.0%	40.0%
First Herald Rate	First Herald Rate
50.0%	40.0%
First Tower Rate	First Tower Rate
40.0%	20.0%
GPM	GPM
1140.6	1070.5
Vision Score	Vision Score
267.0	282.4
First Blood Rate	First Blood Rate
50.0%	50.0%
Gold Diff @15	Gold Diff @15
-54	-1065
Gold Diff @20	Gold Diff @20
274	-1864

Hình 9.31

Deploy ⋮

### Thông số các tướng lựa chọn

T1 Draft:

Lượt 1 - Lucian	▼
Lượt 2 - Nami	▼
Lượt 3 - Orianna	▼
Lượt 4 - Aatrox Winrate của đội khi pick ở lượt 4: 66.7% (3 games) Winrate của đội khi pick tướng này: 77.8% (18 games) Winrate của tướng này khi pick ở lượt 4: 55.2% (29 games) Winrate tổng của tướng này: 49.5% (109 games)	▲
Lượt 5 - Taliyah	▼

DRX Draft:

Lượt 1 - Aphelios	▼
Lượt 2 - Milio	▼
Lượt 3 - Azir	▼
Lượt 4 - Rell	▼

Hình 9.32

Deploy

Lượt 1 - Lucian	▼
Lượt 2 - Nami	▼
Lượt 3 - Orianna	▼
Lượt 4 - Aatrox	▼
Lượt 5 - Taliyah	▼
DRX Draft:	
Lượt 1 - Aphelios	▼
Lượt 2 - Milio	▼
Lượt 3 - Azir	▼
Lượt 4 - Rell	▼
Lượt 5 - Jax	▲
Winrate của đội khi pick ở lượt 5: 28.6% (7 games)	
Winrate của đội khi pick tướng này: 20.0%(10 games)	
Winrate của tướng này khi pick ở lượt 5: 55.2% (29 games)	
Winrate tổng của tướng này: 44.1% (68 games)	

Hình 9.33