

# Analiza poziomu meczu w grze League of Legends

Mikołaj Machalski  
Informatyka stosowana  
Politechnika Wrocławska  
MSiD Lab 15:15 TN  
272672@student.pwr.edu.pl

## I Wstęp

League of Legends to jedna z najpopularniejszych gier komputerowych na rynku od wielu lat. W meczu League of Legends spotykają się dwie drużyny, składające się z 5 graczy każda. Celem eksperymentu jest analiza czynników wpływających na ogólny poziom meczu w grze League of Legends. Analizie zostały poddane mecze w formacie Solo Queue, na serwerze EUNE. Do określenia poziomu meczu niezbędne jest przedstawienie systemu rankingowego.

W grze jest 10 dywizji:

- Iron
- Bronze
- Silver
- Gold
- Platinum
- Emerald
- Diamond
- Master
- Grandmaster
- Challenger

Każda dywizja do *Diamond* włącznie, posiada 4 podrangii: IV, III, II, I. Pomiędzy każdą rangą jest 100 punktów ligowych. Powyżej rangi *Diamond*, rangi *Master*, *Grandmaster*, *Challenger* są jedną rangą. Każda ze wspomnianych trzech rang mają ograniczoną liczbę miejsc na serwerze. Poziom meczu rozumiany jest jako średni poziom każdego z dziesięciu graczy w meczu. Poziom gracza oznacza sumę punktów ligowych, które składają się na jego rangę. Gracz z rangą *Emerald II 60 lp* ma punktów  $5 \times 4 \times 100 + 2 \times 100 + 60 = 2260$ . Jednakże poziom gry potocznie oceniany jest za pomocą jedynie **Dywizji**, tj. (*Iron*, *Bronze*, *Silver*...). bez specyfikacji poszczególnej **Rangi**, tj. (*Iron III*, *Bronze II*, *Silver I*...). Z tego powodu, błąd oceny meczu na poziomie  $\pm 200$  lp (*punktów rankingowych*), jest czymś zadowalającym. Taki błąd oznacza margines na poziomie Dywizji, co tak jak wspomniano wyżej, można postrzegać za bardzo poprawny wynik.

## II Zbieranie i przetwarzanie danych

### II-A Metodyka zbierania danych

Do eksperymentu zostało wybranych 100 graczy, ze strony op.gg [1]. Wybrano 10 graczy na każdą z 10-ciu dywizji. Dla każdego gracza zanalizowano 100 ostatnich zagranych meczy. Dane dotyczące meczy i graczy zostały zebrane za pomocą Riot Api [2]. Mecze nie zostały uwzględniane, jeżeli trwały mniej niż 3 minuty. Taka długość meczu oznacza mecz anulowany. Zbiór danych zawiera około dane dotyczące około 9800 meczów.

### II-B Analiza meczu

Z meczu można wyczytać wiele statystyk. Cechy zawarte w tabeli I, zostały wybrane w następujące sposób:

- 1) Intuicja autora. Cechy zostały wybrane na podstawie przemyśleń autora, mające podstawę w rozgrywce i obserwacji meczów w League of Legends
- 2) Konsultacja z graczami. Wybór cech został skonsultowany z kilkunastoma graczami. Konsultacja przebiegała zarówno z profesjonalnymi, jak i przeciętnymi graczami.

Z meczu zostały wybrane i przetworzone następujące cechy:

Tabela I  
CECHY MECZU

Cecha	Znaczenie
lvl	Średni poziom gry
time	Czas trwania gry w sekundach
solokills_t	Stosunek solo-zabójstw do czasu
kills_t	Stosunek zabójstw do czasu
kp	Średni udział w zabójstwach
deaths_t	Stosunek śmierci do czasu
assists_t	Stosunek asyst do czasu
csscore_t	Stosunek średniego wyniku w stworach do czasu
teamgold_t	Stosunek złota do czasu
vision_score_t	Stosunek poziomu wizji do czasu
wards_placed_t	Stosunek postawionych totemów wizji do czasu
gold_graph_diffs_t	Miara dynamiki zmiany złota w czasie

Statystyki, które według graczy są uważane jako wyznacznik poziomu gry, to przede wszystkim wynik w stworach (*cssscore\_t*) oraz wizja (*vision\_score\_t*).

Każda ze statystyk jest agregacją wszystkich statystyk wszystkich dziesięciu graczy w meczu. Przykładowo, *Stosunek śmierci do czasu* oznacza sumę śmierci wszystkich graczy w meczu podzieloną na czas. Szczegóły dotyczące ekstrakcji cech znajdują się w pliku *match.py*

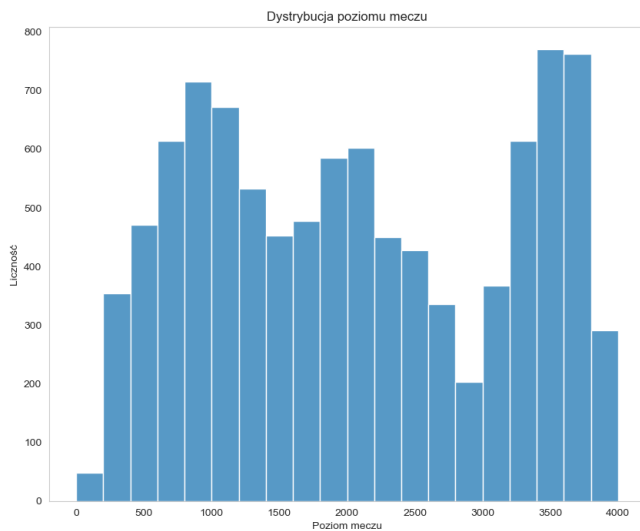
Tabela II  
PRZYKŁADOWE WARTOŚCI

Cecha	Przykładowa wartość
lvl	3800
time	1811
solokills_t	0.008282716731087797
kills_t	0.040861402540033134
kp	0.4744047619047619
deaths_t	0.040861402540033134
assists_t	0.054665930425179456
csscore_t	0.8674765323025952
teamgold_t	68.92932081722805
vision_score_t	0.9559816633728934
wards_placed_t	0.06294864715626726
gold_graph_diffs_t	12.279955825510768

## II-C Przetwarzanie wstępne

Trudno mówić o przetwarzaniu wstępnym. Dane zostały wyselekcjonowane z wielkiej liczby statystyk. Sposób selekcji i opracowania danych nie wymaga oczyszczeń. Jednakże po wstępnej eksploracji zauważono, że należy usunąć cechę *deaths\_t*, z powodu niemal idealnej korelacji z cechą *kills\_t*. Liczba zabójstw w grze jest niemal identyczna co liczba śmierci.

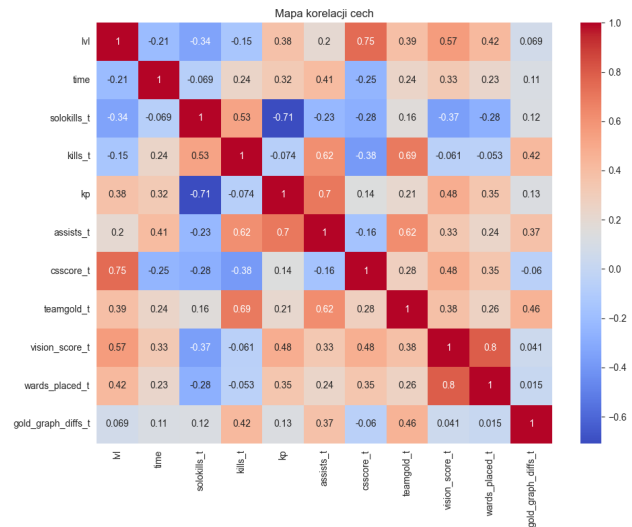
## II-D Analiza eksploracyjna



Rysunek 1. Rozkład poziomu meczu

Jak widać na rysunku 1, poziom meczu nie jest równo rozdystrybuowany. Równa selekcja graczy do testów nie zapewniłażądanego efektu. Może być to spowodowane tym, że ranga **nie jest** wyznacznikiem poziomu gracza, a ukryte punkty *MMR*, *Match-Making Ranking*, niedostępne dla graczy. Te punkty są używane do dobierania graczy.

Mapa korelacji cech pomaga wybrać bardziej znaczące cechy.



Rysunek 2. Mapa korelacji cech

Zgodnie z oczekiwaniami, największą korelację z poziomem meczu mają cechy *csscore\_t* oraz *vision\_score\_t*. Widać oczywiste zależności, jak ujemna korelacja *solo-zabójstw* do *kp* (*drużynowego udziału w zabójstwach*) jak i dodatnia korelacja *wards\_placed\_t* do *vision\_score\_t*. Punkty wizji zdobywamy za pomocą totemów (*wardów*).

## III Eksperymenty

### III-A Użyte modele

- 1) *Linear Regression*
- 2) *Random Forrest Regressor* - model polegający na tworzeniu wielu drzew decyzyjnych i łączeniu ich wyników.

- 3) *Gradient Boost Regressor* - model polegający na tworzeniu lasu losowego stopniowo dodając kolejne drzewa. Z każdą iteracją trenowany model optymalizuje funkcję straty poprzez technikę *gradient descent*.
- 4) *Support Vector Regressor*

### III-B Opis eksperymentów

Dla każdego eksperymentu wykonano osobne dopasowanie parametrów modelu za pomocą narzędzia *GridSearch*

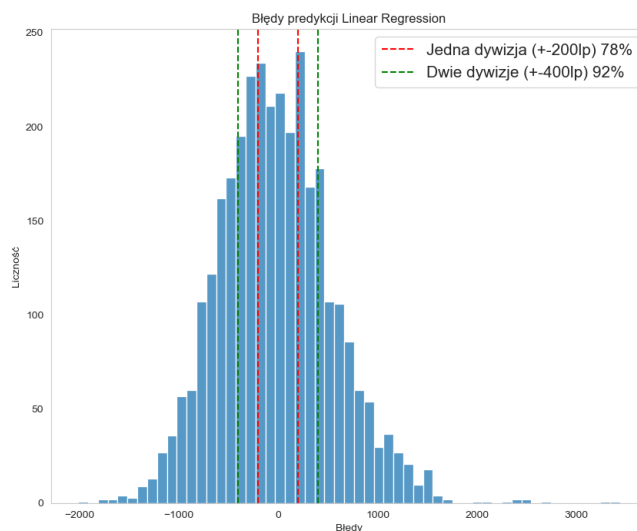
- 1) Eksperyment 1 - Użycie wszystkich cech.
- 2) Eksperyment 2 - Użycie cech bardzo wysoko lub bardzo nisko skorelowanych z poziomem meczu.
- 3) Eksperyment 3 - Użycie cech wysoko lub nisko skorelowanych z poziomem meczu.
- 4) Eksperyment 4 - Zastosowanie PCA (*Principal Component Analysis*), zredukowanie wymiarowości do 5 wymiarów.
- 5) Eksperyment 5 - Zastosowanie PCA (*Principal Component Analysis*), zredukowanie wymiarowości do 2 wymiarów.

### III-C Eksperymenty

- 1) Eksperyment 1:

Tabela III  
WARTOŚCI RMSE DLA EKSPERYMENTU 1

Model	RMSE
Linear Regression	574,87
Gradient Boost	517,99
Random Forrest	527,68
SVR	526,44



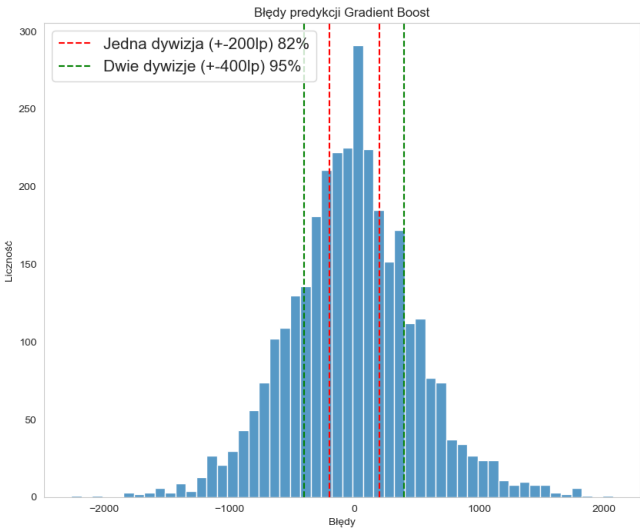
Rysunek 3. Błędy Linear Regression dla eksperymentu 1

Tabela IV  
WARTOŚCI WSPÓŁCZYNNIKÓW REGRESJI I WYRAZ WOLNY

Parametr	Wartość
a	-331.77867
b	88.41569
c	-290.30211
d	-41.40792
e	519.75335
f	592.44395
g	99.64856
h	299.0952
i	-15.48315
j	12.03253
k	2021.66653

Tabela VI  
PARAMETRY MODELU RANDOM FORREST DLA EKSPERYMENTU 1

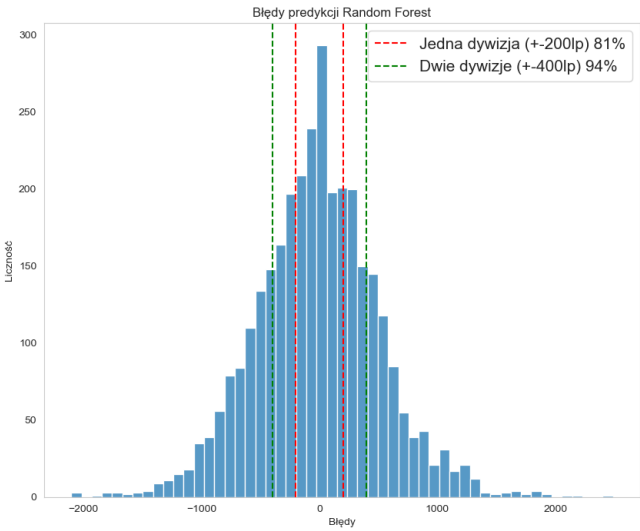
Parametr	Wartość
max_depth	10
min_samples_leaf	4
min_samples_split	10
n_estimators	50



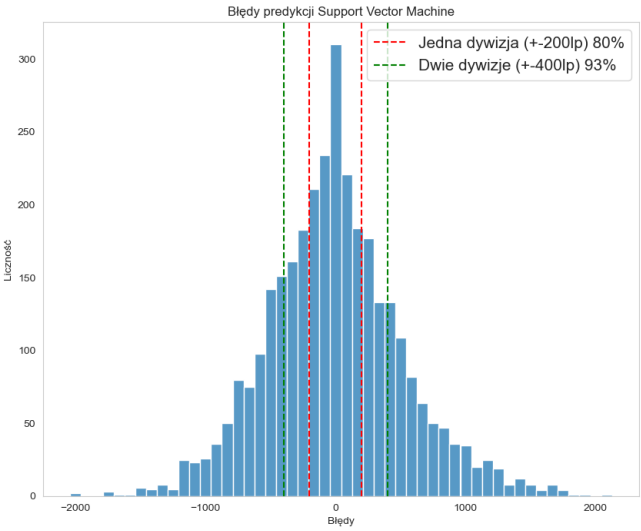
Rysunek 4. Błędy Gradient Boost dla eksperymentu 1

Tabela V  
PARAMETRY MODELU GRADIENT BOOST DLA EKSPERYMENTU 1

Parametr	Wartość
learning_rate	0.085
max_depth	5
n_estimators	100



Rysunek 5. Błędy Random Forrest dla eksperymentu 1



Rysunek 6. Błędy SVR dla eksperymentu 1

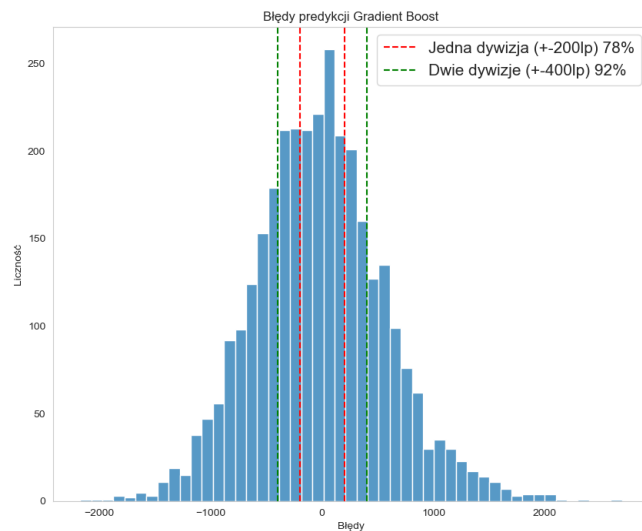
Tabela VII  
PARAMETRY MODELU SVR DLA EKSPERYMENTU 1

Parametr	Wartość
C	100
epsilon	1
gamma	0.1
kernel	rbf

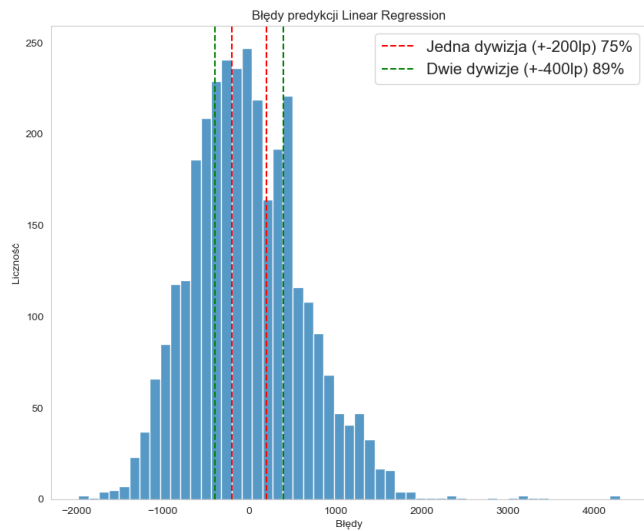
## 2) Eksperyment 2:

Tabela VIII  
WARTOŚCI RMSE DLA EKSPERYMENTU 2

Model	RMSE
Linear Regression	656,71
Gradient Boost	589,26
Random Forrest	590,17
SVR	592,72



Rysunek 8. Błędy Gradient Boost dla eksperymentu 2



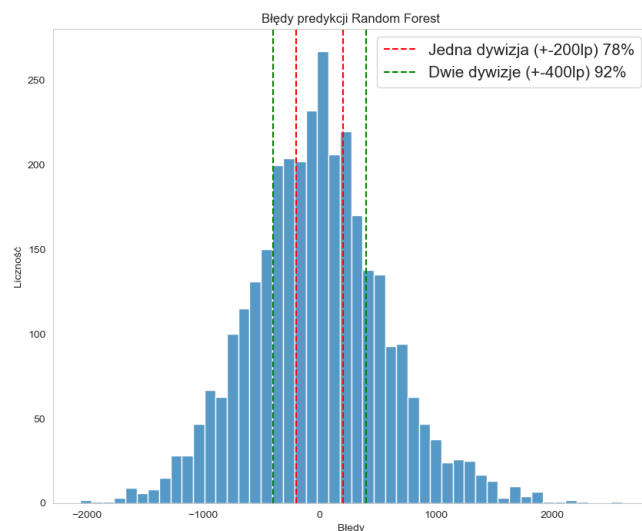
Rysunek 7. Błędy Linear Regression dla eksperymentu 2

Tabela X  
PARAMETRY MODELU GRADIENT BOOST DLA EKSPERYMENTU 2

Parametr	Wartość
learning_rate	0.095
max_depth	5
n_estimators	100

Tabela IX  
WARTOŚCI WSPÓŁCZYNNIKÓW REGRESJI I WYRAZ WOLNY

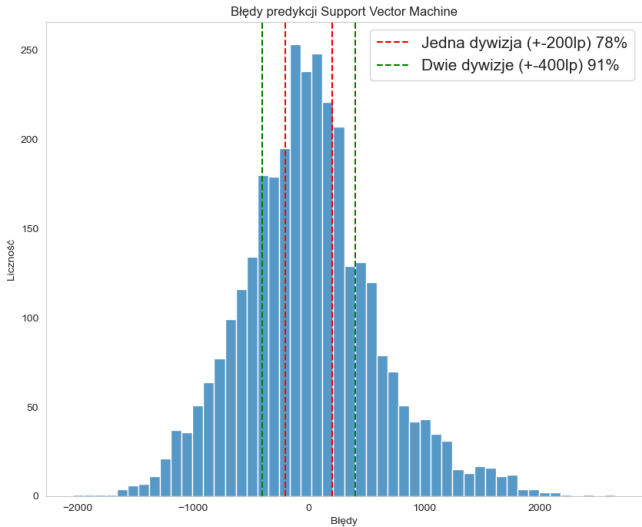
Parametr	Wartość
a	151.62496
b	698.83197
c	213.72177
d	136.1411
e	2021.12878



Rysunek 9. Błędy Random Forrest dla eksperymentu 2

Tabela XI  
PARAMETRY MODELU RANDOM FORREST DLA EKSPERYMENTU 2

Parametr	Wartość
max_depth	10
min_samples_leaf	4
min_samples_split	10
n_estimators	50



Rysunek 10. Błędy SVR dla eksperymentu 2

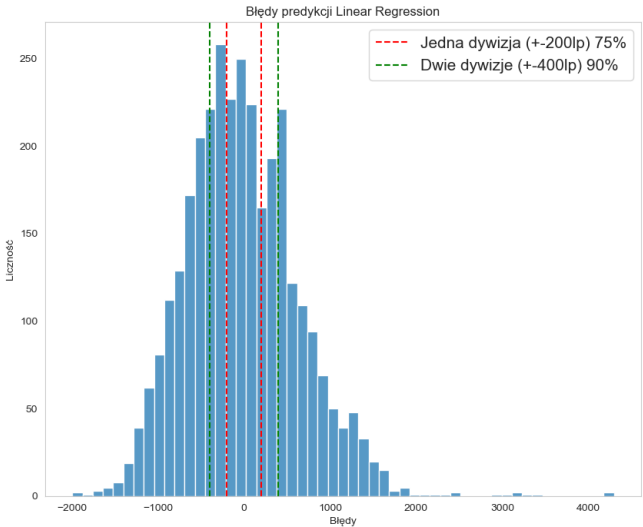
Tabela XII  
PARAMETRY MODELU SVR DLA EKSPERYMENTU 2

Parametr	Wartość
C	100
epsilon	0.01
gamma	0.1
kernel	rbf

### 3) Eksperyment 3:

Tabela XIII  
WARTOŚCI RMSE DLA EKSPERYMENTU 3

Model	RMSE
Linear Regression	657,00
Gradient Boost	587,61
Random Forrest	589,81
SVR	591,87



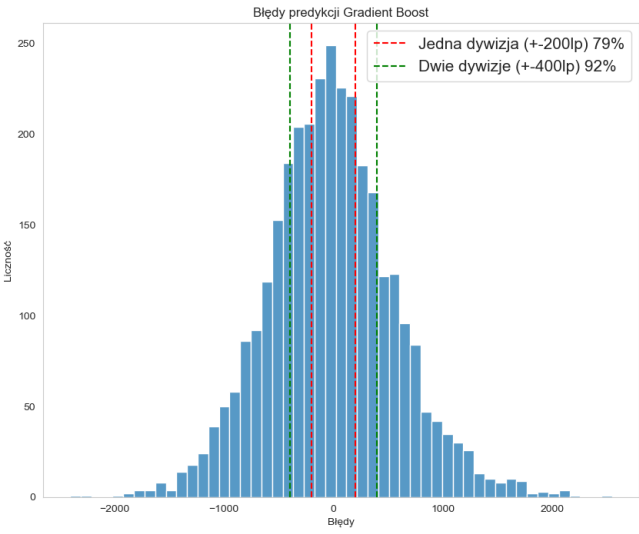
Rysunek 11. Błędy Linear Regression dla eksperymentu 3

Tabela XIV  
WARTOŚCI WSPÓŁCZYNNIKÓW REGRESJI I WYRAZ WOLNY

Parametr	Wartość
a	134.79818
b	700.61733
c	218.13757
d	135.09463
e	19.92032
f	4.53345
g	2021.10075

Tabela XVI  
PARAMETRY MODELU RANDOM FORREST DLA EKSPERYMENTU 3

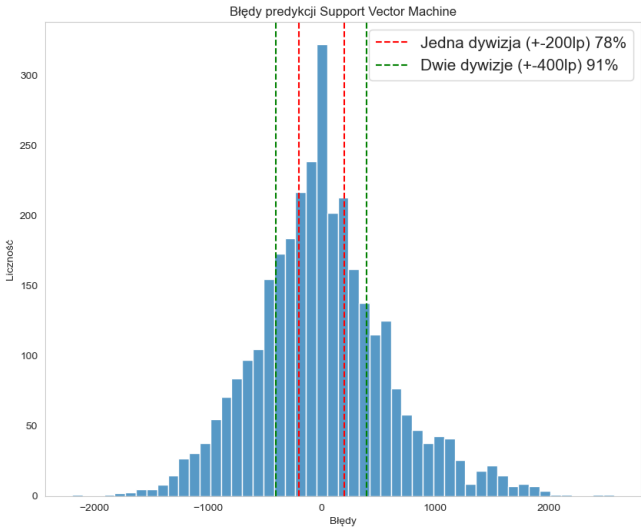
Parametr	Wartość
max_depth	10
min_samples_leaf	4
min_samples_split	10
n_estimators	50



Rysunek 12. Błędy Gradient Boost dla eksperymentu 3

Tabela XV  
PARAMETRY MODELU GRADIENT BOOST DLA EKSPERYMENTU 3

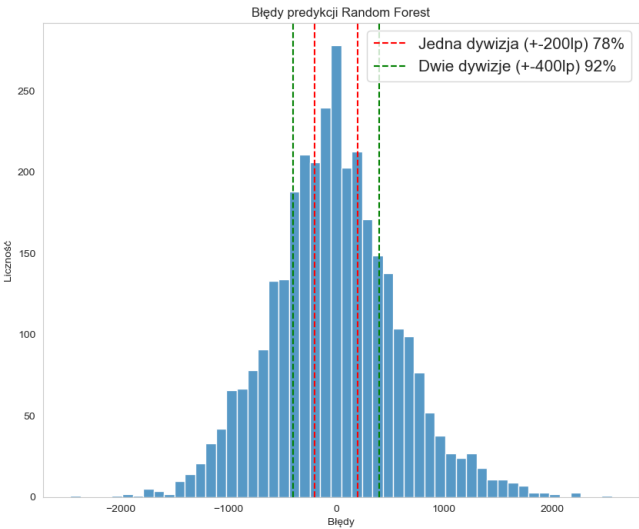
Parametr	Wartość
learning_rate	0.08
max_depth	5
n_estimators	150



Rysunek 14. Błędy SVR dla eksperymentu 3

Tabela XVII  
PARAMETRY MODELU SVR DLA EKSPERYMENTU 3

Parametr	Wartość
C	100
epsilon	0.01
gamma	0.1
kernel	rbf

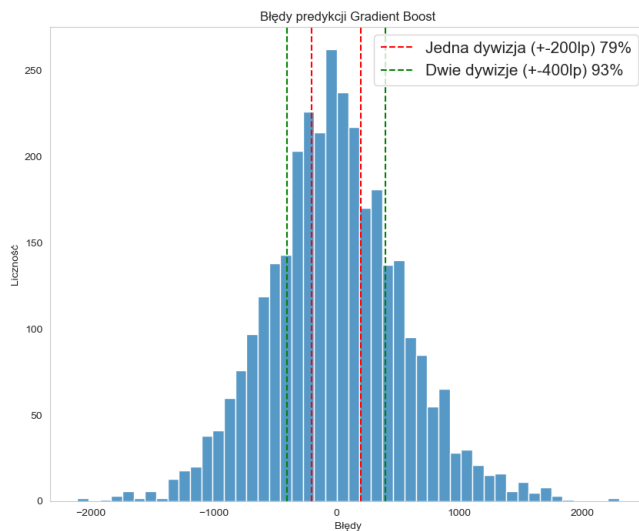


Rysunek 13. Błędy Random Forrest dla eksperymentu 3

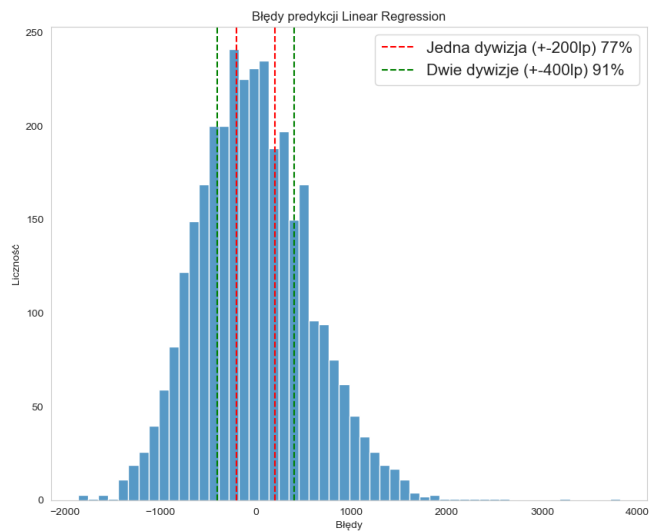
#### 4) Eksperyment 4:

Tabela XVIII  
WARTOŚCI RMSE DLA EKSPERYMENTU 4

Model	RMSE
Linear Regression	591,93
Gradient Boost	550,92
Random Forrest	562,07
SVR	545,67



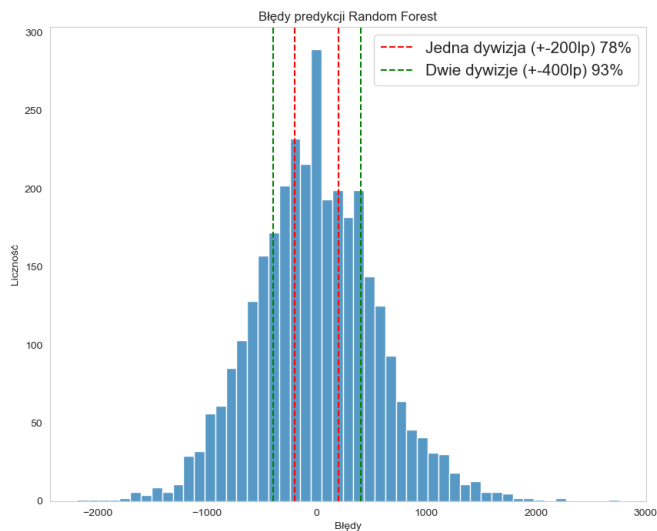
Rysunek 16. Błędy Gradient Boost dla eksperymentu 4



Rysunek 15. Błędy Linear Regression dla eksperymentu 4

Tabela XX  
PARAMETRY MODELU GRADIENT BOOST DLA EKSPERYMENTU 4

Parametr	Wartość
learning_rate	0.075
max_depth	5
n_estimators	100



Rysunek 17. Błędy Random Forrest dla eksperymentu 4

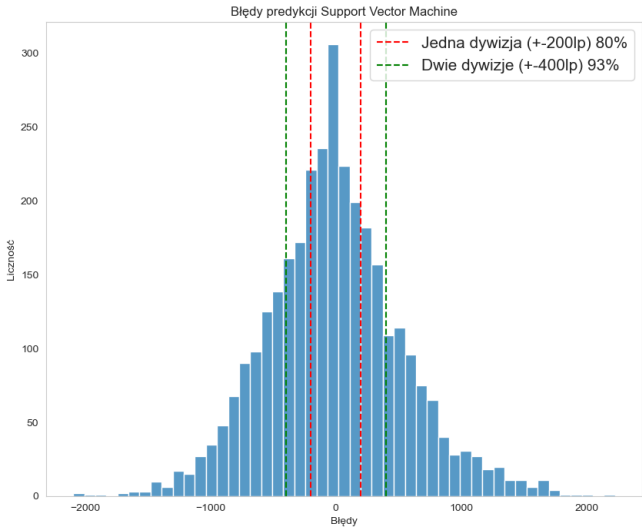
Tabela XIX  
WARTOŚCI WSPÓŁCZYNNIKÓW REGRESJI I WYRAZ WOLNY

Parametr	Wartość
a	284.42272
b	280.1109
c	429.10637
d	-341.35779
e	-203.35292
f	2020.85741



Tabela XXI  
PARAMETRY MODELU RANDOM FORREST DLA EKSPERYMENTU 4

Parametr	Wartość
max_depth	10
min_samples_leaf	4
min_samples_split	2
n_estimators	50



Rysunek 18. Błędy SVR dla eksperymentu 4

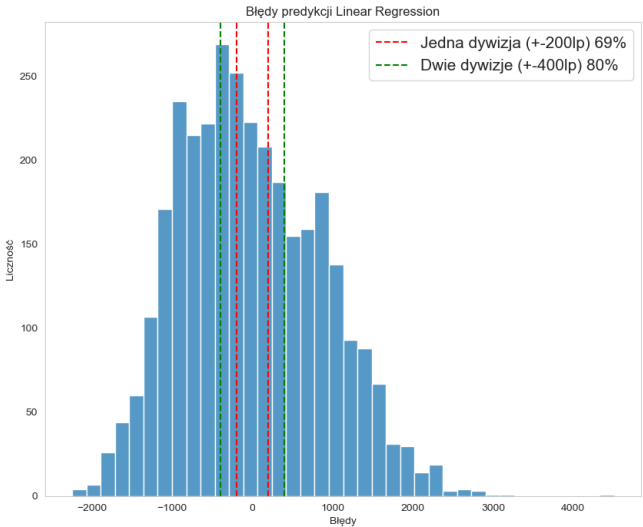
Tabela XXII  
PARAMETRY MODELU SVR DLA EKSPERYMENTU 4

Parametr	Wartość
C	100
epsilon	1
gamma	0.1
kernel	rbf

5) Eksperyment 5:

Tabela XXIII  
WARTOŚCI RMSE DLA EKSPERYMENTU 5

Model	RMSE
Linear Regression	890,27
Gradient Boost	865,04
Random Forrest	869,94
SVR	876,64



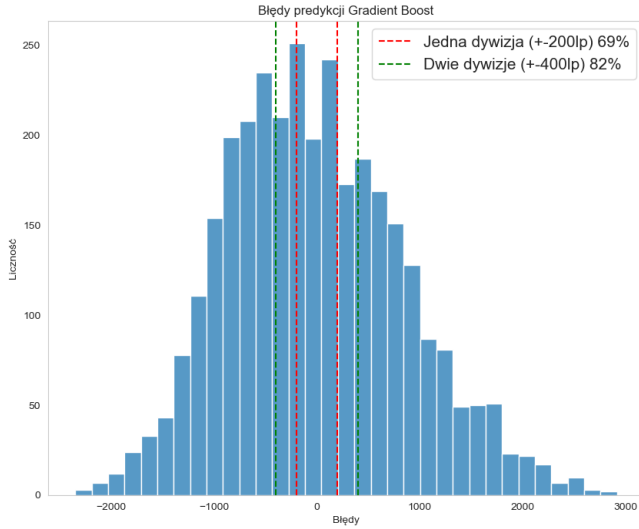
Rysunek 19. Błędy Linear Regression dla eksperymentu 5

Tabela XXIV  
WARTOŚCI WSPÓLCZYNNIKÓW REGRESJI I WYRAZ WOLNY

Parametr	Wartość
a	283.98397
b	277.41878
c	2021.14254

Tabela XXVI  
PARAMETRY MODELU RANDOM FORREST DLA EKSPERYMENTU 5

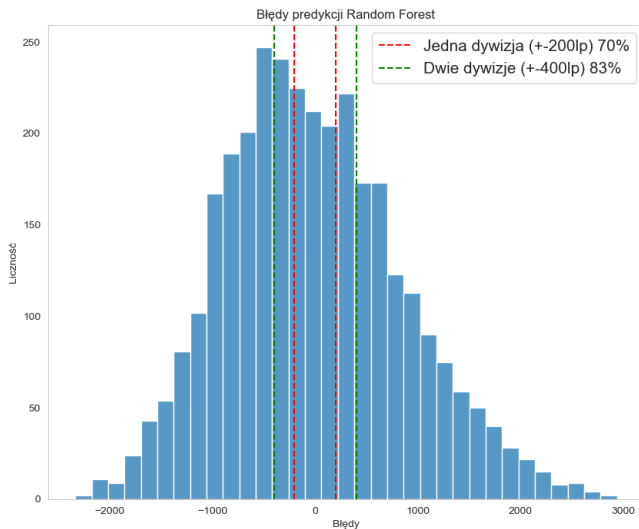
Parametr	Wartość
max_depth	5
min_samples_leaf	2
min_samples_split	2
n_estimators	40



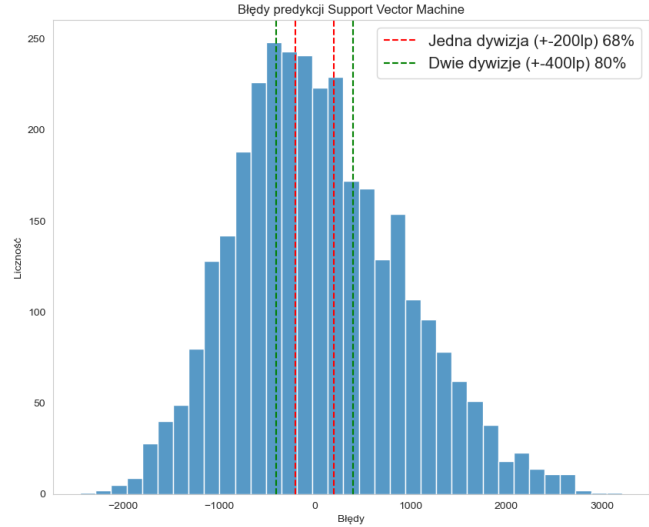
Rysunek 20. Błędy Gradient Boost dla eksperymentu 4

Tabela XXV  
PARAMETRY MODELU GRADIENT BOOST DLA EKSPERYMENTU 5

Parametr	Wartość
learning_rate	0.075
max_depth	5
n_estimators	50



Rysunek 21. Błędy Random Forrest dla eksperymentu 5



Rysunek 22. Błędy SVR dla eksperymentu 5

Tabela XXVII  
PARAMETRY MODELU SVR DLA EKSPERYMENTU 5

Parametr	Wartość
C	90
epsilon	0.01
gamma	0.01
kernel	rbf

### III-D Analiza Eksperymentów

- 1) Najlepsze wyniki uzyskano dla eksperymentu 1. Ogólny trend można opisać jako *więcej cech - lepszy wynik*.
- 2) Dla eksperymentu 1 najlepszy wynik osiągnął *Gradient Boost* - **517,99 lp** błędu RMSE (rysunek 4).
- 3) Równie istotnym wyznacznikiem jakości jest pułap dywizji, zaznaczony odpowiednio zielonymi i czerwonymi liniami przerywanymi. W eksperymencie 1 model Gradient Boost uzyskał **82 %** błędów w pułapie 1 dywizji, oraz **95 %** w pułapie dwóch. (rysunek 4)

## IV Wnioski

Zebrane dane pozwalają na zadowalającą estymację poziomu meczu. Błąd RMSE na poziomie **517,99 lp** dla eksperymentu 1 (tabela III) oznacza średni odchył  $\pm 1$  dywizje i 1 rangę. Innymi słowy, dla predykcji **1000 lp** oznaczającego *Silver II*, średnio wynik mieści się w przedziale **[482,01,1517,99]** lp, lub słownie [*Bronze III*, *Gold I*]. Takie wahania są jak najbardziej spotykane w świecie League of Legends. Gracz raz zagra jak **Bronze** a raz jak **Gold** w zależności od formy. Wahania formy graczy przekładają się na poziom całego meczu. Warto wspomnieć, że mecze w League of Legends potrafią być bardzo niemiarodajne pod względem oceny poziomu. Szczególnie na wysokich dywizjach, gdzie oczekiwany poziom jest wysoki, a mecz jest rozgrywany jak na niższych rangach. Jak widać na rysunku 4 trafiają się błędy predykcji na poziomie 1000 lp i więcej. Takie odstające przypadki zawyżają błąd RMSE. Dlatego równie ważne jest kryterium okienka jednej i dwóch dywizji. Najlepszy wynik pod tym względem osiągnął model *Gradient Boost* w eksperymencie 1 (rysunek 4). **82 %** błędów mieszczących się w zasięgu jednej dywizji można uznać za bardzo dobry wynik.

### IV-A Perspektywy

W celu poprawy wyniku należałoby dogłębniej zanalizować mecz. Aspekty, które warto zawrzeć to:

- 1) Analiza heatmapy graczy
- 2) Analiza heatmapy graczy podczas walk drużynowych
- 3) Analiza zachowań graczy w czasie ważnych celów na mapie (smoków i Barona Nashora)
- 4) Analiza mikro-zachowań gracza.

Te statystyki wymagają bardzo dogłębnego spojrzenia na mecz, szczególnie przygotowanie heatmapy graczy. W tej analizie poddane badaniom zostały tylko proste w przetworzeniu statystyki. Mimo to, można uznać, że na bazie tych statystyk da się ocenić poziom meczu.

## Literatura

- [1] "Ranking graczy serwera eune," <https://www.op.gg/leaderboards/tier?page=2region=eune>.  
[2] "Riot api," <https://developer.riotgames.com/apis>.