



Metody Systemowe i Decyzyjne L

Raport zaliczeniowy

Przewidywanie czasu trwania gry w League of Legends
na podstawie analizy wydarzeń z gry

Marcin Wawer
kierunek: **Informatyka Stosowana**

7 marca 2025

1 Wstęp - sformułowanie problemu

League of Legends to jedna z najpopularniejszych gier online na świecie i najpopularniejsza gra typu MOBA w historii, przyciągająca miliony graczy codziennie od wielu lat.

Celem tego raportu jest zbadanie, jak wydarzenia w trakcie gry, takie jak m.in. liczba zabójstw, zdobywanie celów (np. wież, smoków, Barona Nashora) czy zdobyte złoto przez drużyny, wpływają na długość meczu. Przeanalizuje dane meczowe graczy, aby zidentyfikować kluczowe zmienne i stworzyć modele prognozujące czas trwania gry.

Raport zawiera omówienie znaczenia analizy czasu trwania gry, metodologię zbierania i przetwarzania danych, techniki analizy i modelowania, wyniki analiz oraz wnioski.

2 Zbiór danych i jego przetwarzanie

2.1 Zbiór danych

Listę graczy, których gry zostały użyte do stworzenia zbioru danych uzyskane zostały ze strony op.gg. Gracze wykorzystani do stworzenia zbioru danych mają rangę 'Diamant' oraz ich konta są zarejestrowane w regionie EUNE (Europa Północ i Wschód). Następnie na podstawie nazw graczy uzyskane zostały statystyki z ich gier korzystając z API Riot Games developer.riotgames.com.

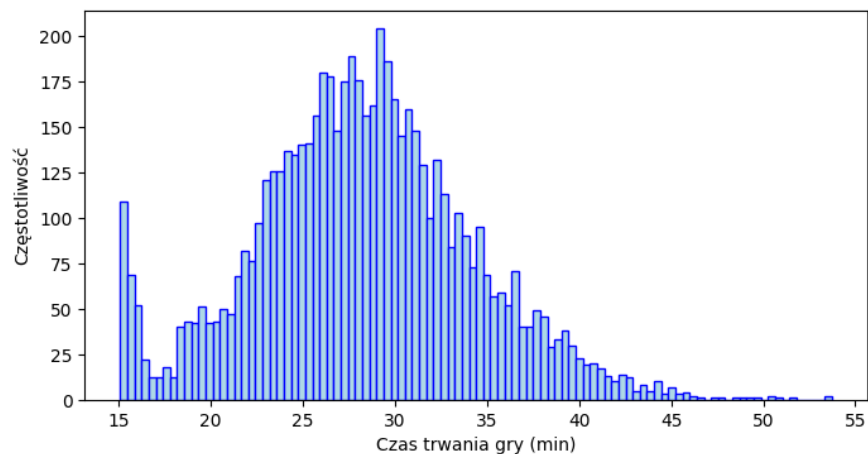
2.2 Przetwarzanie wstępne

Ze zbioru zostały usunięte powtarzające się wiersze oraz gry, które trwały mniej niż 15 minut, ze względu na to, że zdecydowana większość takich gier, to mecze, które skończyły się nienaturalnie, np. poprzez przerwanie gry, ze względu na niedołączenie wszystkich graczy. Następnie ze zbioru usunięte zostały zbędne kolumny, nieprzydatne do analizy. W zbiorze danych nie znajdowały się żadne błędne wiersze lub komórki. Następnie zsumowane zostały dane z kolumn, które były podzielone na statystyki zespołów, aby uzyskać statystyki całej gry. Ostatecznie zbiór został podzielony na taki zawierający gry poddane i niepoddane oraz taki zawierający tylko gry niepoddane.

2.3 Analiza eksploracyjna

Tabela 1 prezentuje dane ze zbioru z grami poddanymi i niepoddanymi po wstępnym przetwarzaniu. Tabela 2 prezentuje dane ze zbioru z grami tylko niepoddanymi. Rysunek 1 prezentuje histogram czasu trwania gier w minutach dla zbioru z grami poddanymi i niepoddanymi. Rysunek 2 prezentuje histogram czasu trwania gier w minutach dla zbioru z grami tylko niepoddanymi.

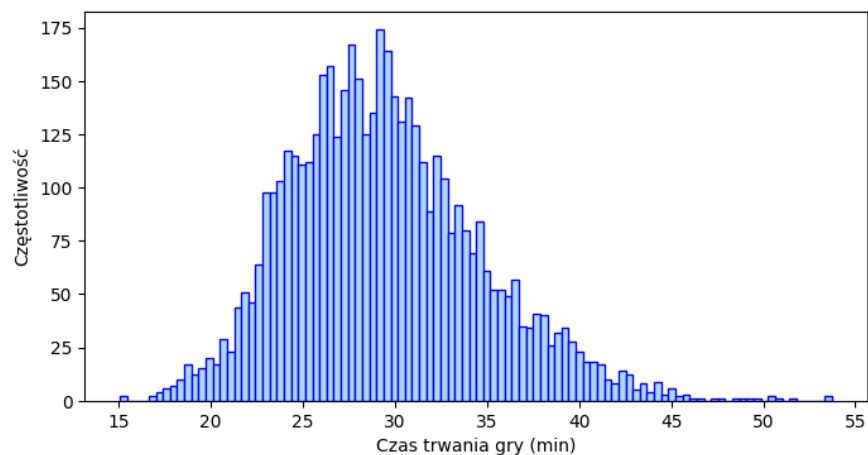
Tak duża różnica liczby gier o czasie trwania około 15 minut w tych zbiorach występuje, ponieważ w League of Legends jest możliwość poddawania gier dopiero od 15 minuty. Co skutkuje częstą praktyką poddawania gier od razu, gdy występuje taka możliwość, np. ze względu na to, że jedna drużyna uzyskała bardzo dużą przewagę na początku gry.



Rysunek 1: Histogram czasu trwania gier w minutach dla gier poddanych i niepoddanych

Minimum	Maksimum	Średnia	Mediana
15.08	53.73	28.27	28.13

Tabela 1: Statystyki czasu trwania gier w minutach dla gier poddanych i niepoddanych



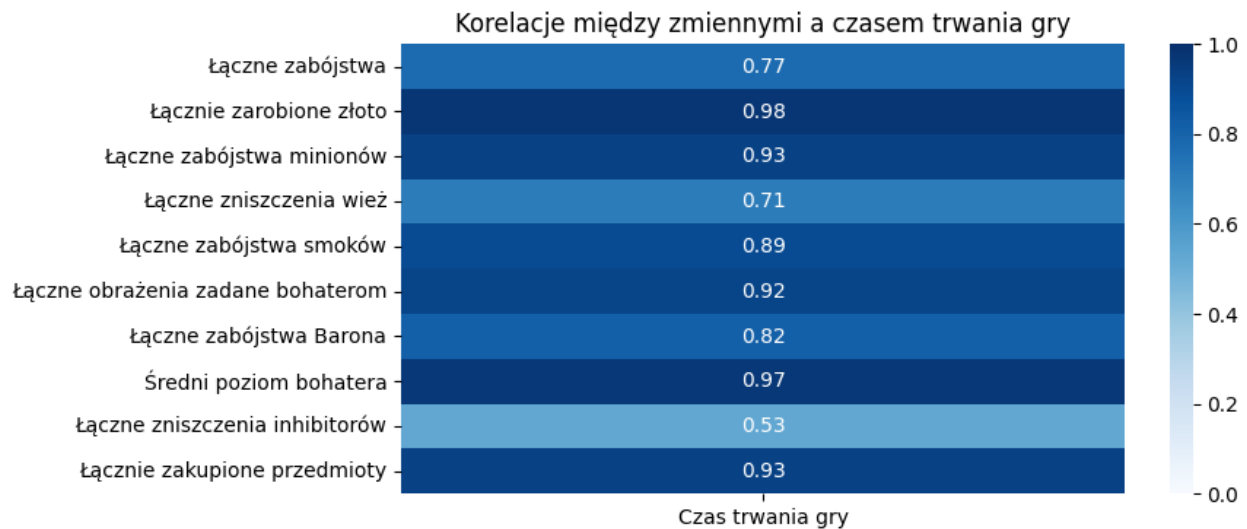
Rysunek 2: Histogram czasu trwania gier w minutach dla gier niepoddanych

Minimum	Maksimum	Średnia	Mediana
15.08	53.73	29.42	29.00

Tabela 2: Statystyki czasu trwania gier w minutach dla gier niepoddanych

- **Zależność między zmiennymi, a czasem trwania gry**

Rysunek 3 przedstawia mapę korelacji między zmiennymi a czasem trwania gry. Mapa wykazała bardzo wysoką korelację niektórych zmiennych z czasem trwania gry m.in. łącznie zarobione złoto, średni poziom bohatera. Tak wysoka korelacja między danymi pozwoli dokładnie przewidywać czas trwania gry.

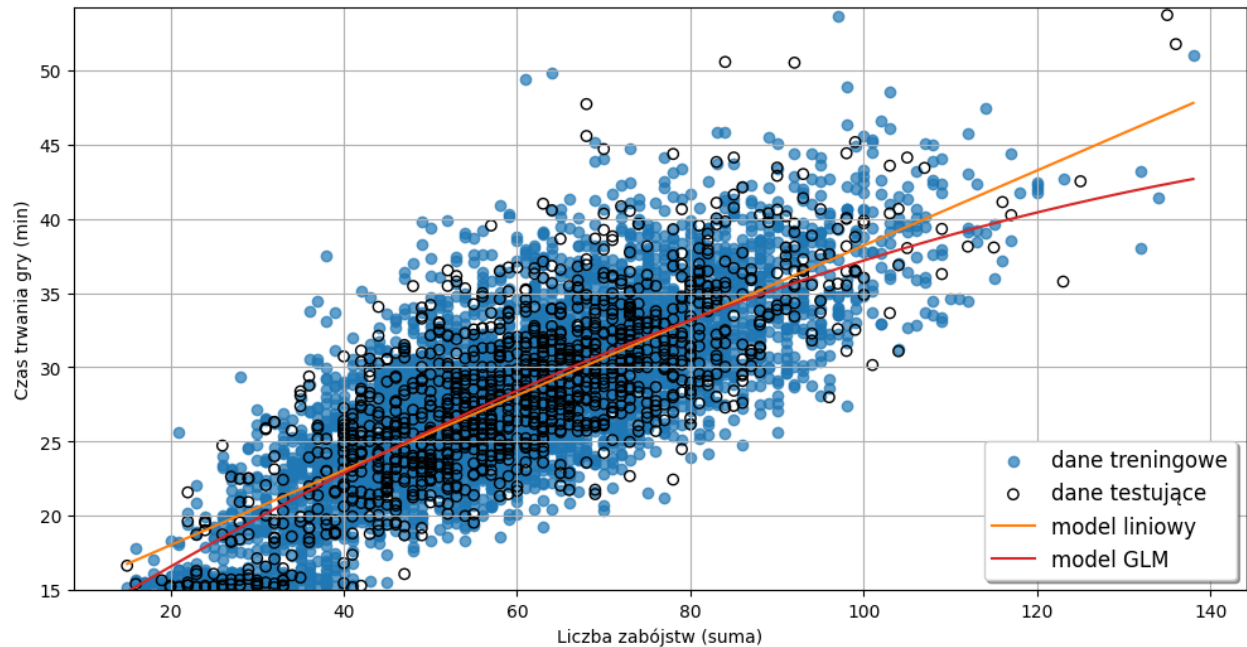


Rysunek 3: Mapa korelacji zmiennych

3 Eksperymenty

W kontekście tego badania, kluczowym aspektem było rozważenie zarówno gier, które zostały poddane, jak i tych, które nie zostały poddane. Poddanie gry jest istotnym zdarzeniem, które może znacząco wpływać na czas trwania meczu. Porównanie wyników błędu średniokwadratowego dla obu zbiorów wykazało, że różnice między nimi były nieznaczne. Wyniki eksperymentów wskazywały, że modele predykcyjne oparte na obu zbiorach danych osiągały bardzo podobne wartości błędu średniokwadratowego. To sugeruje, że uwzględnienie gier poddanych nie wprowadza znaczących różnic w dokładności przewidywań, a jednocześnie pozwala na bardziej wszechstronne zrozumienie mechaniki gry i lepsze modelowanie różnych scenariuszy, dlatego do eksperymentów użyty został zbiór zawierający wszystkie gry.

1. Przewidywanie czasu trwania gry na podstawie liczby zabójstw w grze.



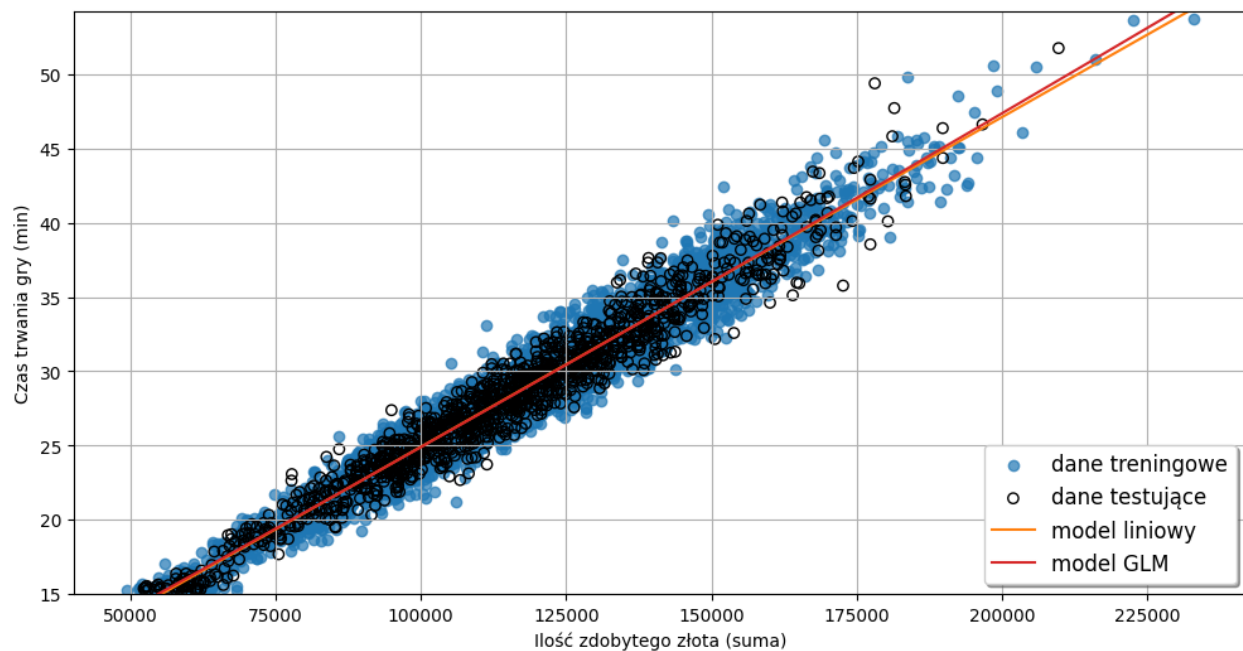
Rysunek 4: Porównanie modelu regresji liniowej oraz modelu GLM w przewidywaniu czasu trwania gry na podstawie liczby zabójstw.

Model	Mean Squared Error (MSE)
Regresja liniowa	14.60
Model GLM	14.50

Tabela 3: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Oba modele osiągnęły zbliżoną dokładność w przewidywaniu czasu trwania gry na podstawie liczby zabójstw. Model GLM nieznacznie przewyższa model regresji liniowej, co sugeruje, że może lepiej uchwycić zależności między zmiennymi w tym kontekście.

2. Przewidywanie czasu trwania gry na podstawie ilości zdobytego złota.



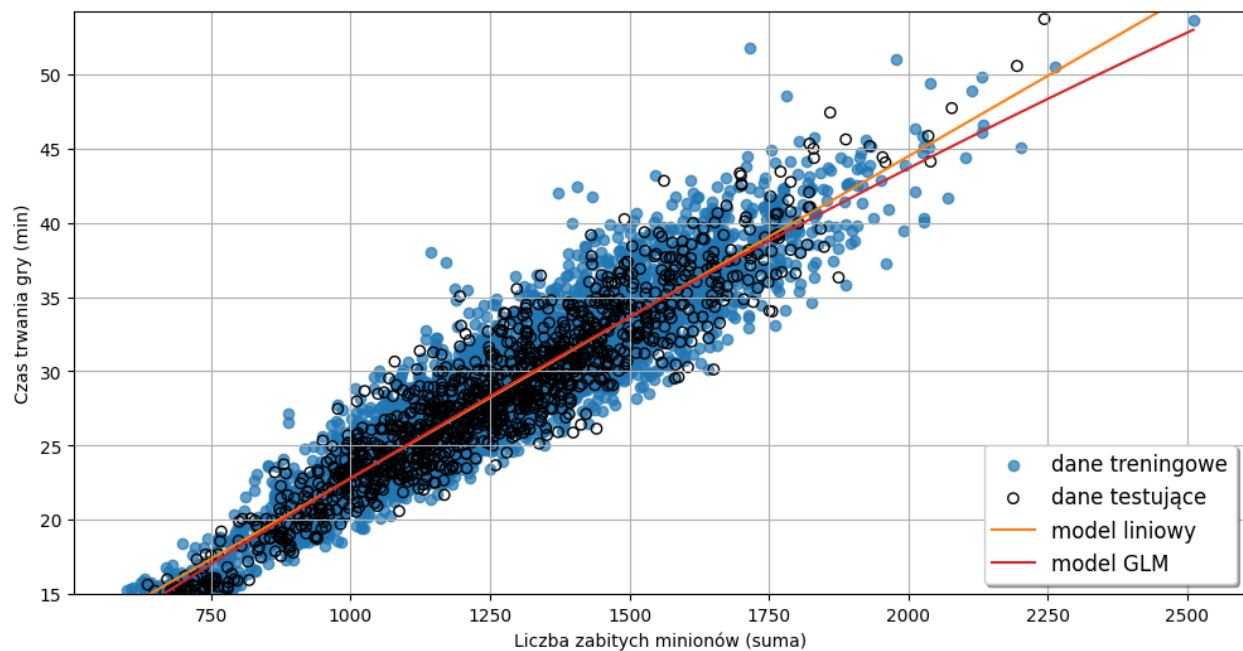
Rysunek 5: Porównanie modelu regresji liniowej oraz modelu GLM w przewidywaniu czasu trwania gry na podstawie ilości zdobytego złota.

Model	Mean Squared Error (MSE)
Regresja liniowa	1.68
Model GLM	1.68

Tabela 4: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Oba modele uzyskały identyczny błąd średniokwadratowy, co wskazuje na równie dobrą zdolność przewidywania czasu trwania gry na podstawie ilości zdobytego złota przez obie drużyny.

3. Przewidywanie czasu trwania gry na podstawie liczby zabitych minionów.



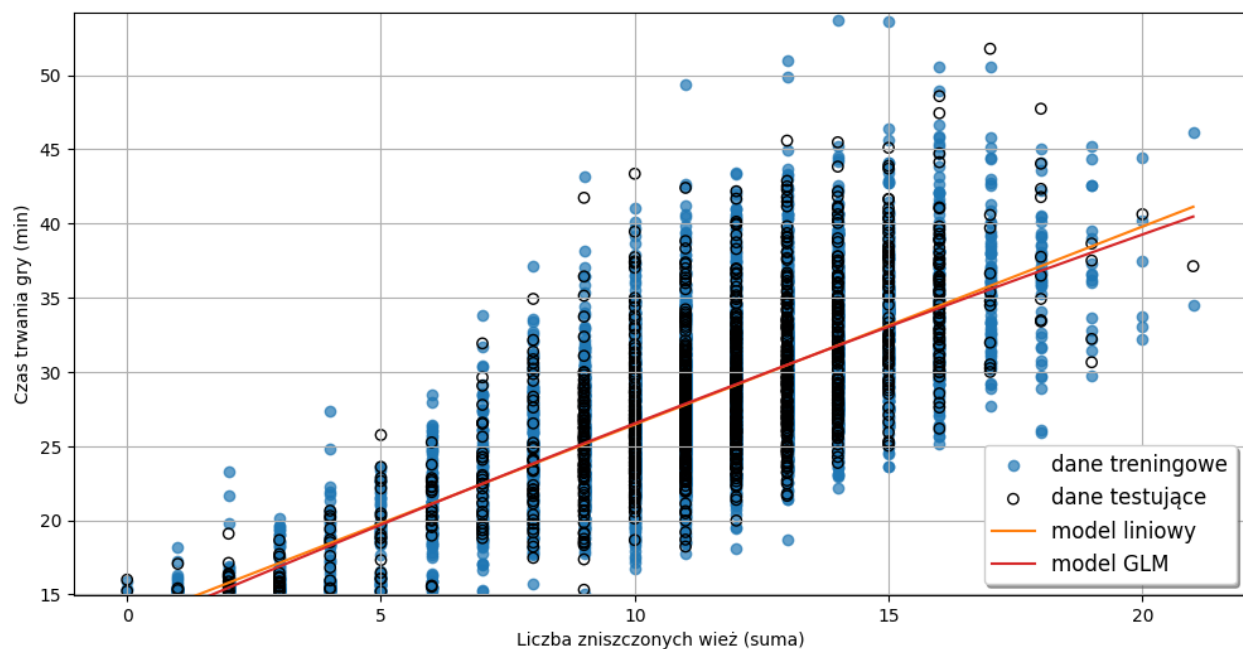
Rysunek 6: Porównanie modelu regresji liniowej oraz modelu GLM w przewidywaniu czasu trwania gry na podstawie liczby zabitych minionów.

Model	Mean Squared Error (MSE)
Regresja liniowa	4.48
Model GLM	4.48

Tabela 5: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Oba modele osiągnęły identyczne wyniki, co sugeruje, że liczba zabitych minionów jest równie dobrze modelowana przez regresję liniową i GLM.

4. Przewidywanie czasu trwania gry na podstawie liczby zniszczonych wież.



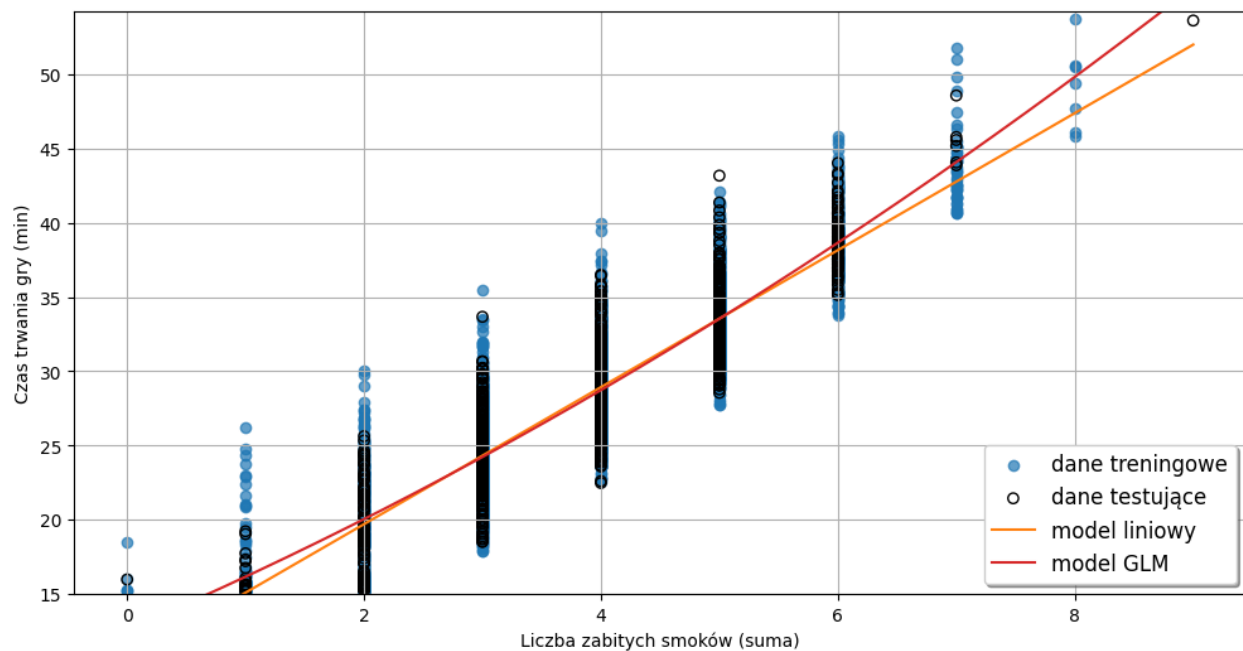
Rysunek 7: Porównanie modelu regresji liniowej oraz modelu GLM w przewidywaniu czasu trwania gry na podstawie liczby zniszczonych wież.

Model	Mean Squared Error (MSE)
Regresja liniowa	18.30
Model GLM	18.30

Tabela 6: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Błędy średniokwadratowe dla obu modeli są identyczne, co sugeruje, że liczba zniszczonych wież jest równie dobrze modelowana przez oba typy modeli.

5. Przewidywanie czasu trwania gry na podstawie liczby zabitych smoków.



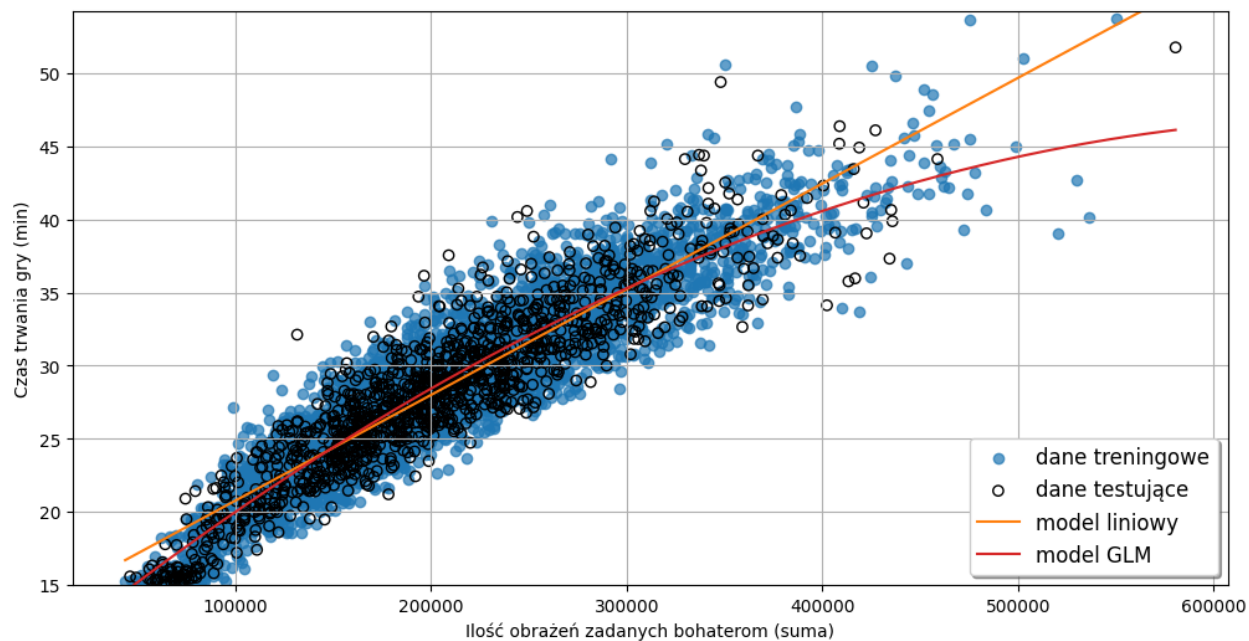
Rysunek 8: Porównanie modelu regresji liniowej oraz modelu GLM w przewidywaniu czasu trwania gry na podstawie liczby zabitych smoków.

Model	Mean Squared Error (MSE)
Regresja liniowa	6.51
Model GLM	6.48

Tabela 7: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Model GLM osiągnął nieco lepszy wynik niż regresja liniowa, co sugeruje, że może lepiej uchwycić wpływ liczby zabitych smoków na czas trwania gry.

6. Przewidywanie czasu trwania gry na podstawie ilości obrażeń zadanych bohaterom.



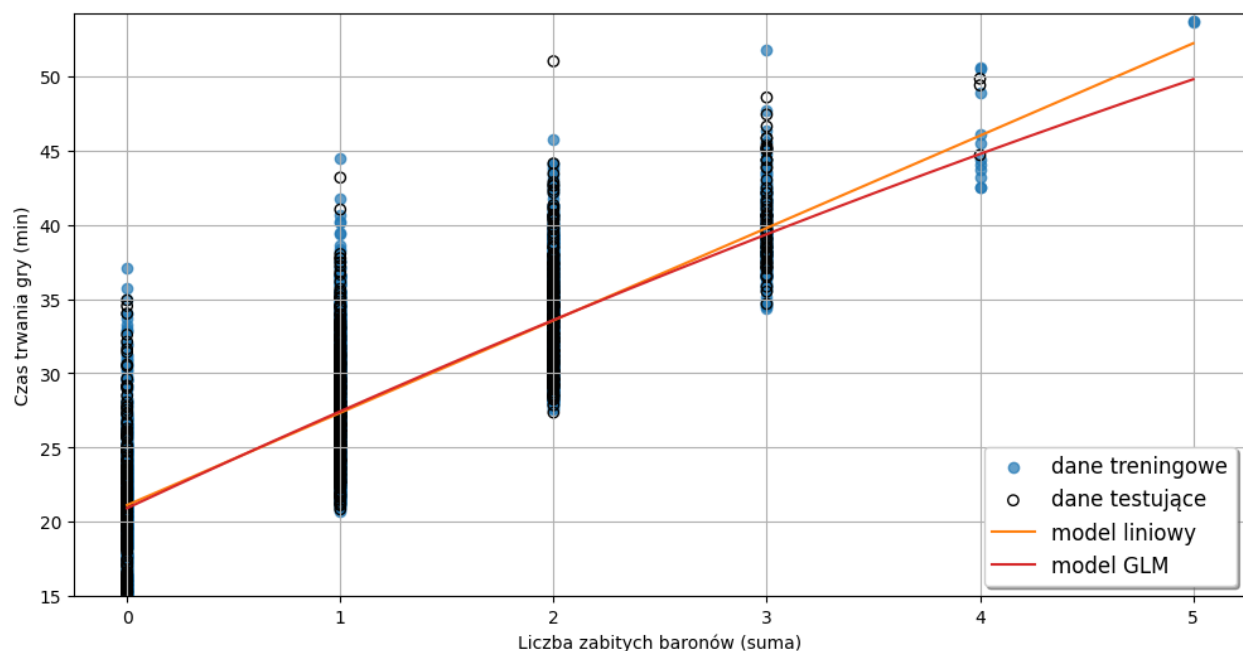
Rysunek 9: Porównanie modelu regresji liniowej oraz modelu GLM w przewidywaniu czasu trwania gry na podstawie ilości obrażeń zadanych bohaterom.

Model	Mean Squared Error (MSE)
Regresja liniowa	5.59
Model GLM	5.15

Tabela 8: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Model GLM osiągnął lepszy wynik niż regresja liniowa, co sugeruje, że lepiej uchwycił zależności między obrażeniami zadanymi bohaterom a czasem trwania gry.

7. Przewidywanie czasu trwania gry na podstawie liczby zabitych baronów.



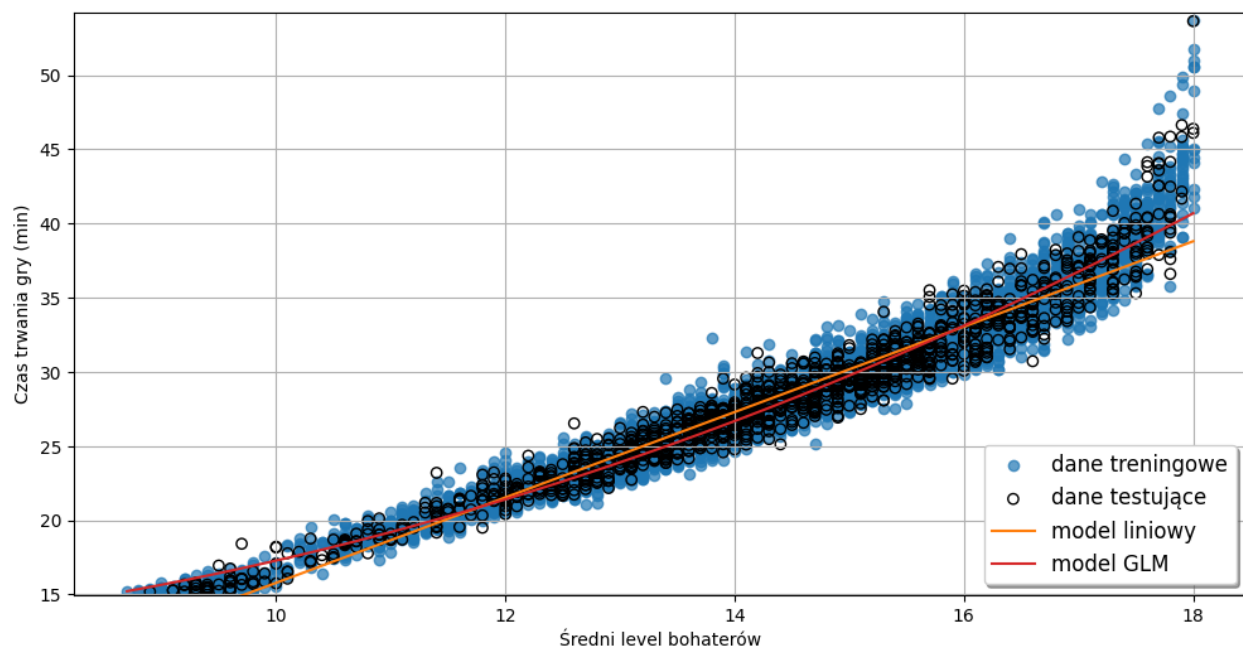
Rysunek 10: Porównanie modelu regresji liniowej oraz modelu GLM w przewidywaniu czasu trwania gry na podstawie liczby zabitych baronów.

Model	Mean Squared Error (MSE)
Regresja liniowa	13.00
Model GLM	13.00

Tabela 9: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Oba modele osiągnęły identyczny wynik, co wskazuje, że liczba zabitych baronów jest równie dobrze modelowana przez oba typy modeli.

8. Przewidywanie czasu trwania gry na podstawie średniego levelu bohaterów.



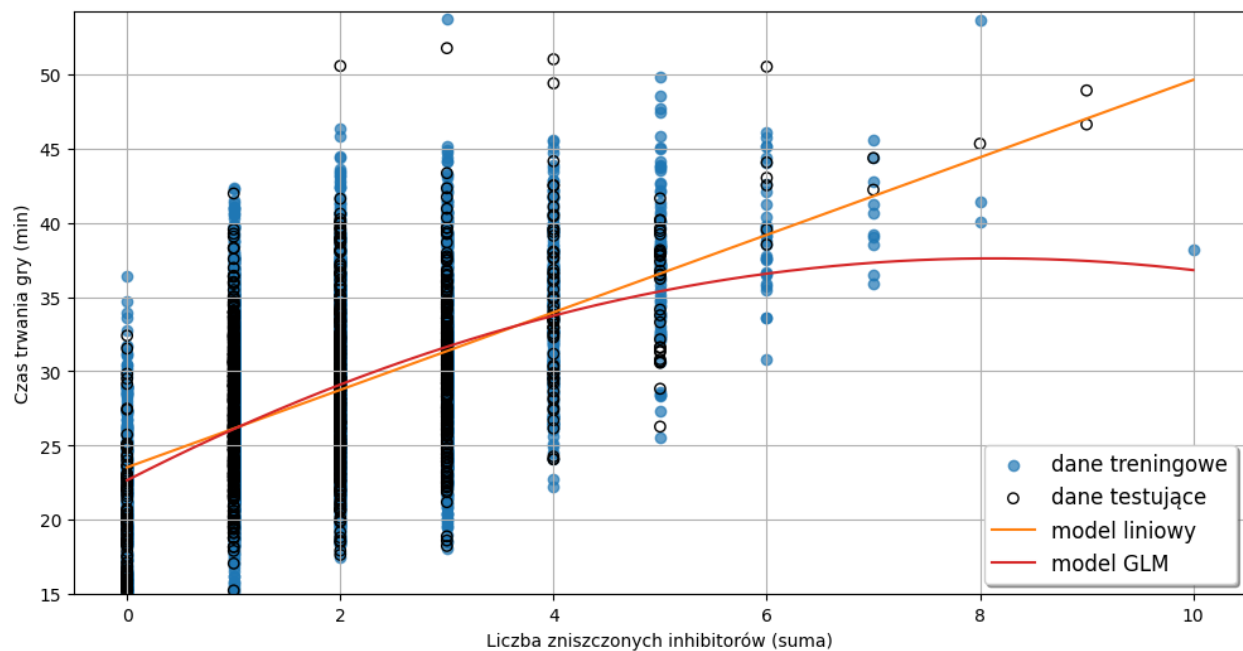
Rysunek 11: Porównanie modelu regresji liniowej oraz modelu GLM w przewidywaniu czasu trwania gry na podstawie średniego levelu bohaterów.

Model	Mean Squared Error (MSE)
Regresja liniowa	2.18
Model GLM	1.73

Tabela 10: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Model GLM osiągnął lepszy wynik niż regresja liniowa, co sugeruje, że lepiej uchwycił zależności między średnim poziomem bohaterów a czasem trwania gry.

9. Przewidywanie czasu trwania gry na podstawie liczby zniszczonych inhibitorów.



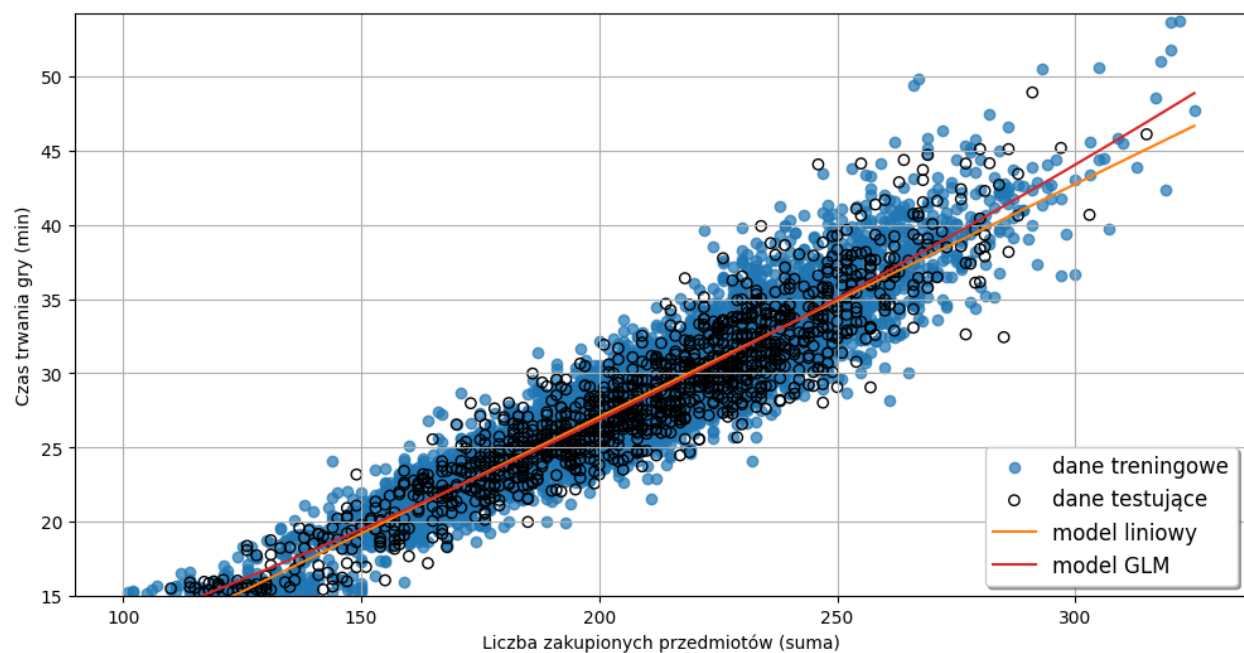
Rysunek 12: Porównanie modelu regresji liniowej oraz modelu GLM w przewidywaniu czasu trwania gry na podstawie liczby zniszczonych inhibitorów.

Model	Mean Squared Error (MSE)
Regresja liniowa	24.70
Model GLM	24.60

Tabela 11: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Model GLM osiągnął nieco lepszy wynik niż regresja liniowa, co sugeruje, że lepiej uchwycił wpływ liczby zniszczonych inhibitorów na czas trwania gry.

10. Przewidywanie czasu trwania gry na podstawie liczby zakupionych przedmiotów.



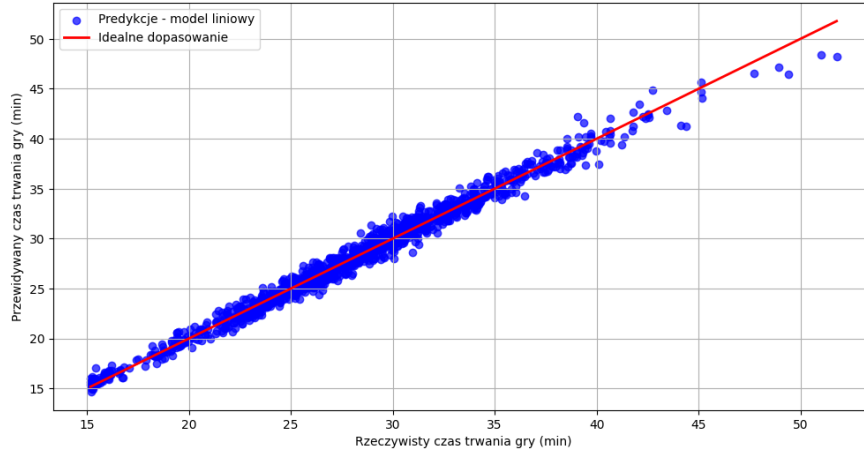
Rysunek 13: Porównanie modelu regresji liniowej oraz modelu GLM w przewidywaniu czasu trwania gry na podstawie liczby zakupionych przedmiotów.

Model	Mean Squared Error (MSE)
Regresja liniowa	4.64
Model GLM	4.55

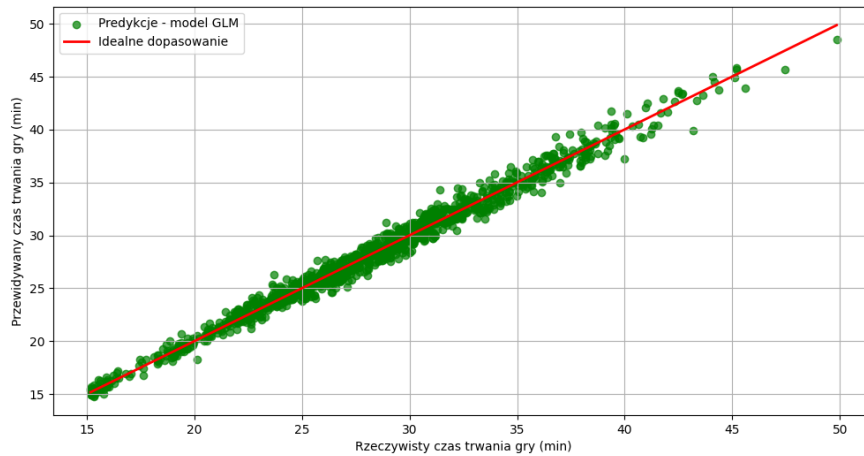
Tabela 12: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Model GLM osiągnął nieco lepszy wynik niż regresja liniowa, co sugeruje, że lepiej uchwycił wpływ liczby zakupionych przedmiotów na czas trwania gry.

11. Celem tego eksperymentu było zbudowanie modelu regresji liniowej i modelu GLM w celu przewidywania czasu trwania gry w League of Legends na podstawie wszystkich zmiennych ze zbioru danych. Wykres przedstawia wyniki predykcji modeli w porównaniu do rzeczywistych wartości. W eksperymencie użyto następujących zmiennych niezależnych: liczba zabójstw (suma), całkowite zdobyte złoto, całkowita liczba zabitych minionów, całkowita liczba zniszczonych wież, całkowita liczba zabitych smoków, całkowite zadane obrażenia bohaterom, całkowita liczba zabitych baronów, średni poziom bohaterów, całkowita liczba zniszczonych inhibitorów, całkowita liczba zakupionych przedmiotów.



Rysunek 14: Wyniki predykcji modelu regresji liniowej w porównaniu do rzeczywistych wartości.



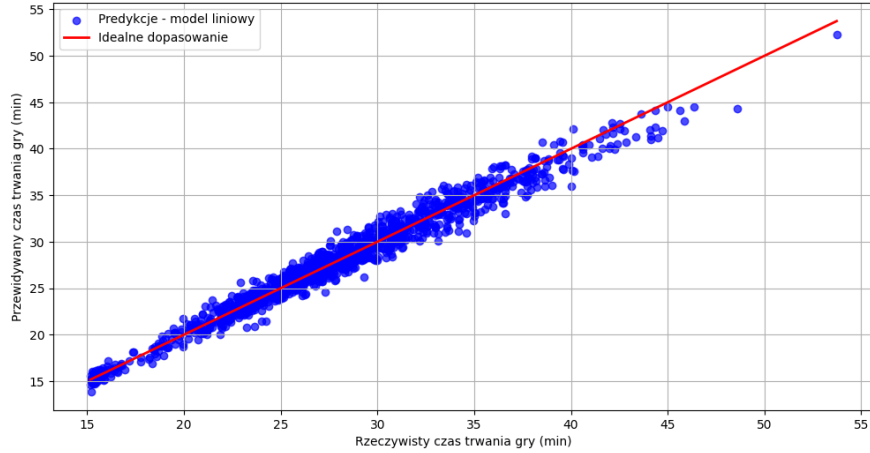
Rysunek 15: Wyniki predykcji modelu GLM w porównaniu do rzeczywistych wartości.

Model	Mean Squared Error (MSE)
Regresja liniowa	0.526
Model GLM	0.507

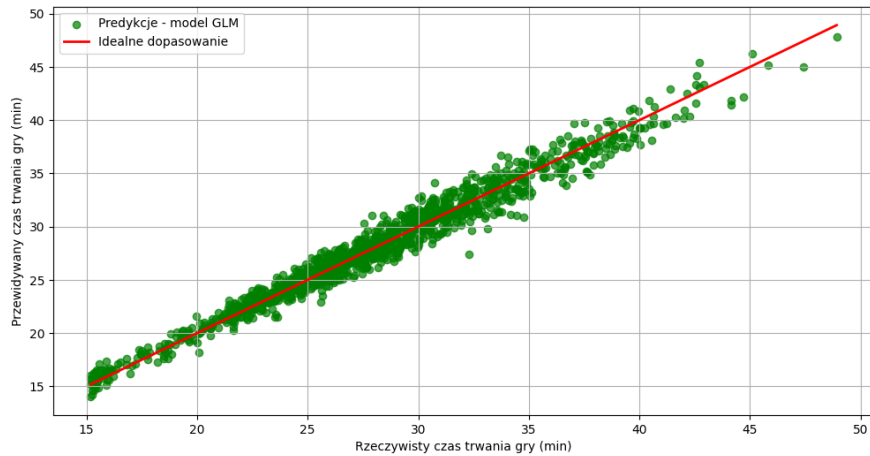
Tabela 13: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Model GLM osiągnął lepszy wynik niż regresja liniowa, co sugeruje, że lepiej uchwycił zależności między wszystkimi zmiennymi a czasem trwania gry.

12. Celem tego eksperymentu było zbudowanie modelu regresji liniowej i modelu GLM w celu przewidywania czasu trwania gry w League of Legends na podstawie zmiennych ze zbioru danych, które mają wartość korelacji powyżej 0.92 względem czasu trwania gry. Wykres przedstawia wyniki predykcji modeli w porównaniu do rzeczywistych wartości. W eksperymencie użyto następujących zmiennych niezależnych: całkowite zdobyte złoto, całkowita liczba zabitych minionów, całkowite zadane obrażenia bohaterom, całkowita liczba zabitych baronów, średni poziom bohaterów, całkowita liczba zakupionych przedmiotów.



Rysunek 16: Wyniki predykcji modelu regresji liniowej w porównaniu do rzeczywistych wartości.



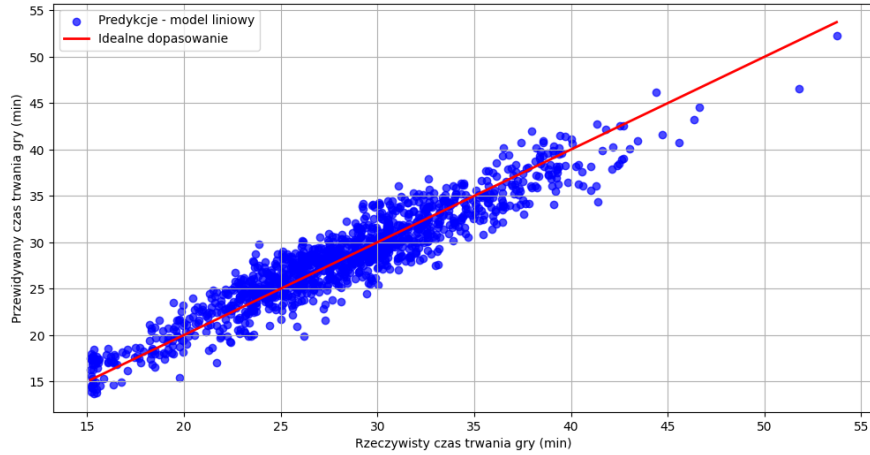
Rysunek 17: Wyniki predykcji modelu GLM w porównaniu do rzeczywistych wartości.

Model	Mean Squared Error (MSE)
Regresja liniowa	1.10
Model GLM	1.03

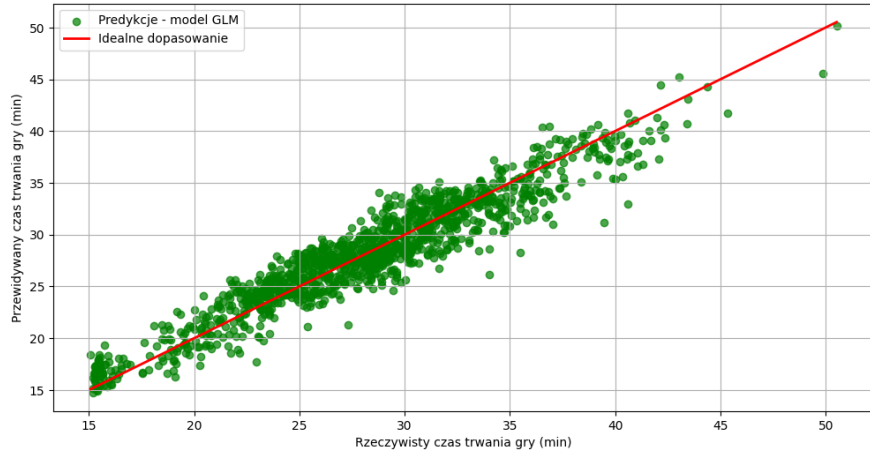
Tabela 14: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Model GLM osiągnął lepszy wynik niż regresja liniowa, co sugeruje, że lepiej uchwycił zależności między zmiennymi o wysokiej korelacji z czasem trwania gry.

13. Celem tego eksperymentu było zbudowanie modelu regresji liniowej i modelu GLM w celu przewidywania czasu trwania gry w League of Legends na podstawie zmiennych ze zbioru danych, które mają wartość korelacji poniżej 0.92 względem czasu trwania gry. Wykres przedstawia wyniki predykcji modeli w porównaniu do rzeczywistych wartości. W eksperymencie użyto następujących zmiennych niezależnych: liczba zabójstw (suma), całkowita liczba zniszczonych wież, całkowita liczba zabitych smoków, całkowita liczba zabitych baronów, całkowita liczba zniszczonych inhibitorów.



Rysunek 18: Wyniki predykcji modelu regresji liniowej w porównaniu do rzeczywistych wartości.



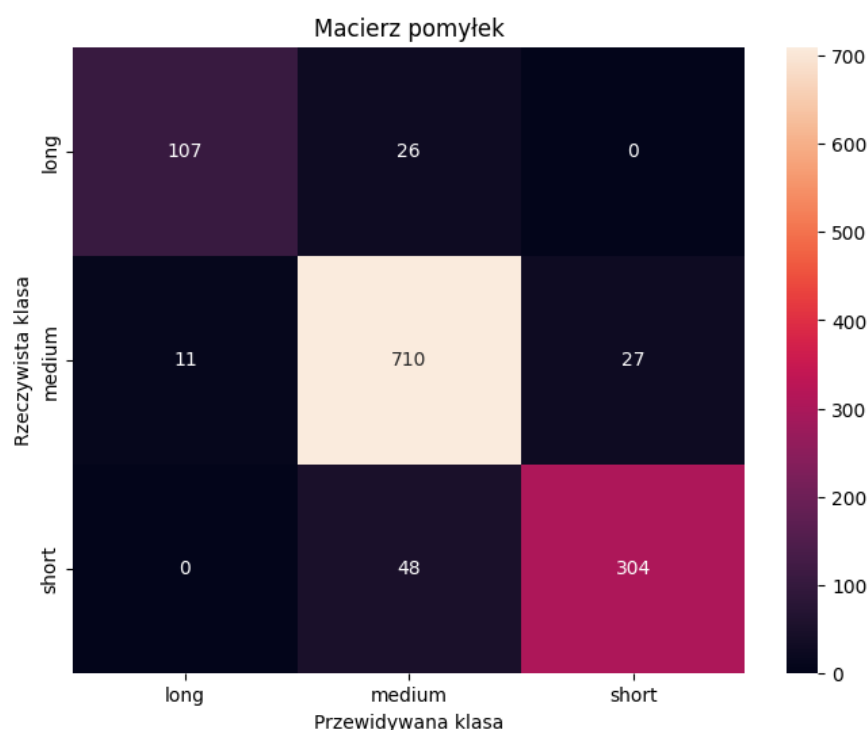
Rysunek 19: Wyniki predykcji modelu GLM w porównaniu do rzeczywistych wartości.

Model	Mean Squared Error (MSE)
Regresja liniowa	3.75
Model GLM	3.48

Tabela 15: Porównanie błędu średniokwadratowego (MSE) dla modelu regresji liniowej i modelu GLM.

- **Obserwacje:** Model GLM osiągnął lepszy wynik niż regresja liniowa, co sugeruje, że lepiej uchwycił zależności między zmiennymi o niskiej korelacji z czasem trwania gry.

14. Celem tego eksperymentu było sklasyfikowanie czasu trwania gry w League of Legends na podstawie wszystkich zmiennych ze zbioru danych. Klasyfikacja miała na celu przypisanie czasu trwania gry do jednej z trzech kategorii: krótki (short) - ≤ 25 minut, średni (medium) - 26-35 minut, długi (long) - > 35 minut. Wyniki klasyfikacji zostały przedstawione za pomocą macierzy pomyłek. Użyto klasyfikatora SVC (Support Vector Classifier). W eksperymencie użyto następujących zmiennych niezależnych: liczba zabójstw (suma), całkowite zdobyte złoto, całkowita liczba zabitych minionów, całkowita liczba zniszczonych wież, całkowita liczba zabitych smoków, całkowite zadane obrażenia bohaterom, całkowita liczba zabitych baronów, średni poziom bohaterów, całkowita liczba zniszczonych inhibitorów, całkowita liczba zakupionych przedmiotów.



Rysunek 20: Wyniki klasyfikacji przedstawione za pomocą macierzy pomyłek.

Klasa	Precyzja	Czułość (recall)	F1-score
long	0.91	0.80	0.85
medium	0.91	0.95	0.93
short	0.92	0.86	0.89

Tabela 16: Raport klasyfikacyjny modelu SVC z jądrem RBF dla przewidywania kategorii czasu trwania gry

- **Obserwacje:** Model SVC z jądrem RBF osiągnął wysoką dokładność (91% i 92%) w klasyfikacji czasu trwania gry do trzech kategorii. Najlepsze wyniki osiągnął dla kategorii "medium", co wskazuje na jego skuteczność w rozróżnianiu gier o średnim czasie trwania.

4 Podsumowanie

Raport dotyczący przewidywania czasu trwania gry w League of Legends na podstawie analizy wydarzeń z gry dostarcza cennych spostrzeżeń na temat związków między różnymi zmiennymi a czasem trwania meczu. Zbiór danych pochodził z meczów graczy z rangą 'Diamant' w regionie EUNE, uzyskany przy użyciu API Riot Games. Gry trwające krócej niż 15 minut zostały usunięte, ponieważ mogły być wynikiem nienaturalnych zakończeń. Histogramy czasu trwania gier wykazały znaczącą różnicę w liczbie gier trwających około 15 minut między zbiorami danych zawierającymi gry poddane i niepoddane, co odzwierciedla praktykę poddawania gier od razu po 15 minutach.

W większości eksperymentów modele GLM (Generalized Linear Model) osiągały lepsze wyniki lub były równorzędne z modelami regresji liniowej pod względem błędu średniokwadratowego (MSE). GLM często lepiej uchwycił złożone zależności między zmiennymi.

Wykonana na początku raportu mapa korelacji dostarcza wartościowych informacji przydatnych podczas eksperymentów. Wszystkie zmienne użyte podczas eksperymentów pokazują wyraźną korelację z czasem trwania meczu. Najwyżej skorelowana zmienna to 'łącznie zarobione złoto' z korelacją na poziomie 0.98. Już modele wykonane na podstawie tylko tej jednej zmiennej dają zadowalające wyniki z błędem średniokwadratowym na poziomie 1.68 dla obu wykorzystanych modeli. Najniższą korelację z wybranymi cechami osiąga zmienna 'łącznie zniszczenia inhibitorów' - 0.53. Wynika to z faktu, że zniszczenia inhibitorów pojawiają się najczęściej w końcówce gry, co niepozwala modelowi na efektywne przewidywanie czasu trwania średnio długiej i krótkiej gry. Model regresji liniowej osiąga błąd średniokwadratowy na poziomie 24.7, natomiast model GLM - 24.6.

Następnie wykonane zostały trzy eksperymenty:

1. Przewidywanie czasu trwania gry na podstawie wszystkich dostępnych zmiennych. Modele użyte w tym eksperymencie, zgodnie z przewidywaniami, dają bardzo zadowalające wyniki z błędem średniokwadratowym na poziomie 0.526 dla regresji liniowej i 0.507 dla modelu GLM.
2. Przewidywanie czasu trwania gry na podstawie zmiennych z korelacją z czasem trwania gry na poziomie co najmniej 0.92. W tym eksperymencie zostało użyte tylko 6 zmiennych, co znacznie upraszcza modele względem poprzednich, zachowując ich skuteczność nadal na wysokim, zadowalającym poziomie z błędem średniokwadratowym 1.10 dla regresji liniowej i 1.03 dla modelu GLM.
3. W ostatnim z tych trzech eksperymentów wykorzystane zostały tylko te cechy, których wartość korelacji z czasem trwania gry była na poziomie poniżej 0.92. Zgodnie z przewidywaniami modele użyte w tym eksperymencie cechowały się znacznie wyższym błędem średniokwadratowym, co za tym idzie znacznie niższą skutecznością. Regresja liniowa - 3.75, GLM - 3.48.

Ostatnim eksperymentem było sklasyfikowanie czasu trwania gry na podstawie wszystkich dostępnych zmiennych, aby osiągnąć jak najwyższą dokładność stworzonego modelu. Użyto SVC, który osiągnął wyniki na bardzo zadowalającym poziomie z precyzją 0.91 dla długich gier, 0.91 dla średnich gier i 0.92 dla krótkich gier. Precyzja na takim poziomie pozwala efektywnie przewidywać kategorie czasu trwania gry na podstawie wybranych zmiennych, z niewielką ilością błędów popełnionych przez model.

Z wykonanego raportu jasno wynika wysoka dokładność dobranych modeli w kwestii przewidywania czasu trwania gry na podstawie dostępnych zmiennych. Najwyższą dokładnością oczywiście cechują się modele, wykorzystujące całe spektrum dostępnych zmiennych, natomiast przy znacznym uproszczeniu modelu, ignorując niektóre z mniej istotnych cech, nadal osiągnięta jest zadowalająca dokładność. Pozwala to na odpowiednie dobranie skomplikowania modelu, względem jego skuteczności, zgodnie z naszymi oczekiwaniami. Raport zgodnie ze wstępnie postawionymi założeniami pokazał, że można skutecznie przewidywać czas trwania gry w League of Legends na podstawie zmiennych dostępnych podczas rozgrywki.