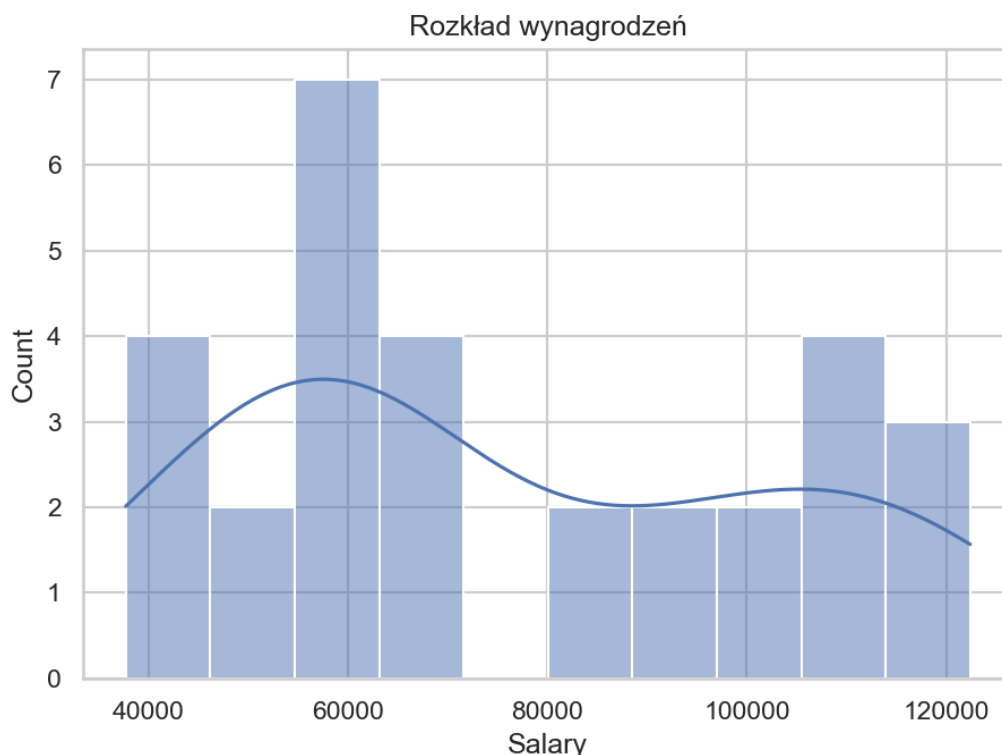


Regresja liniowa – raport

Wprowadzenie

Do wykonania zadania wykorzystaliśmy dane dotyczące wynagrodzeń i lat doświadczenia zawodowego dostępne na platformie Kaggle. Do analizy danych użyliśmy bibliotek takich jak Pandas, NumPy, Matplotlib, Seaborn oraz Sklearn.

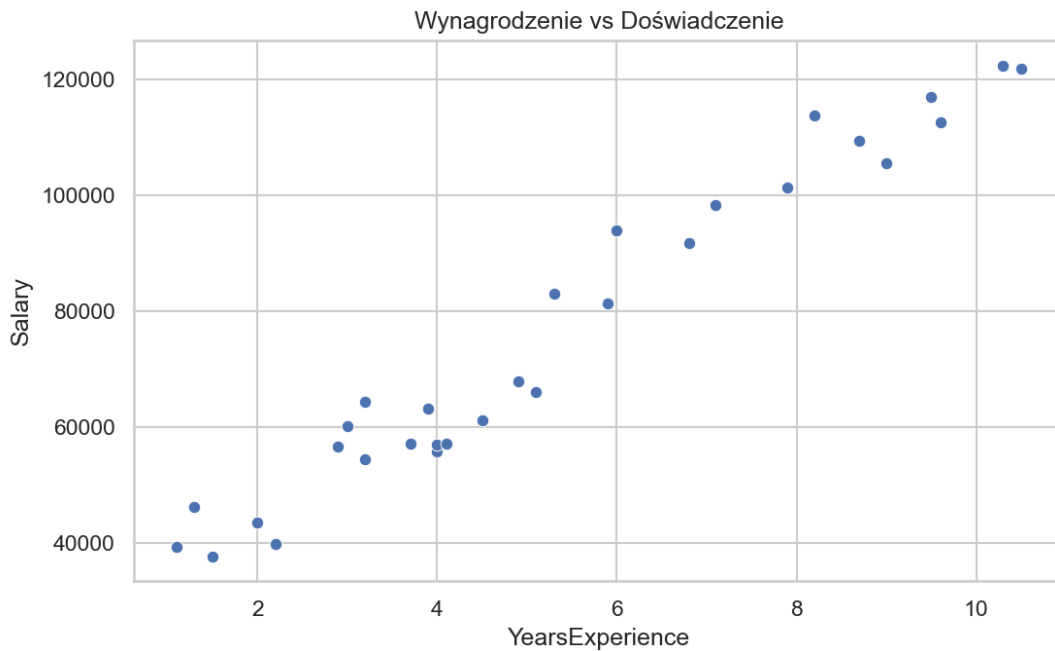
Analiza Rozkładu Danych



Rozkład wynagrodzeń został przeanalizowany za pomocą wykresu Seaborn distplot. Główne obserwacje:

- Wynagrodzenia mieszczą się w zakresie od około 35,000 do 125,000.
- Najwięcej wynagrodzeń zaobserwowano w przedziale 60,000 – 70,000 oraz 100,000 – 110,000.
- Najmniej wynagrodzeń zaobserwowano w przedziale 70,000 – 80,000 oraz powyżej 120,000.
- Rozkład wynagrodzeń jest asymetryczny z wyraźnym szczytem w niższych przedziałach, sugerującym prawostronną skośność (skośność dodatnia).
- Krzywa gęstości wskazuje na dwa lokalne maksimum, czyli przedziały wynagrodzeń z największą gęstością obserwacji.

Zależność między Wynagrodzeniem a Doświadczeniem

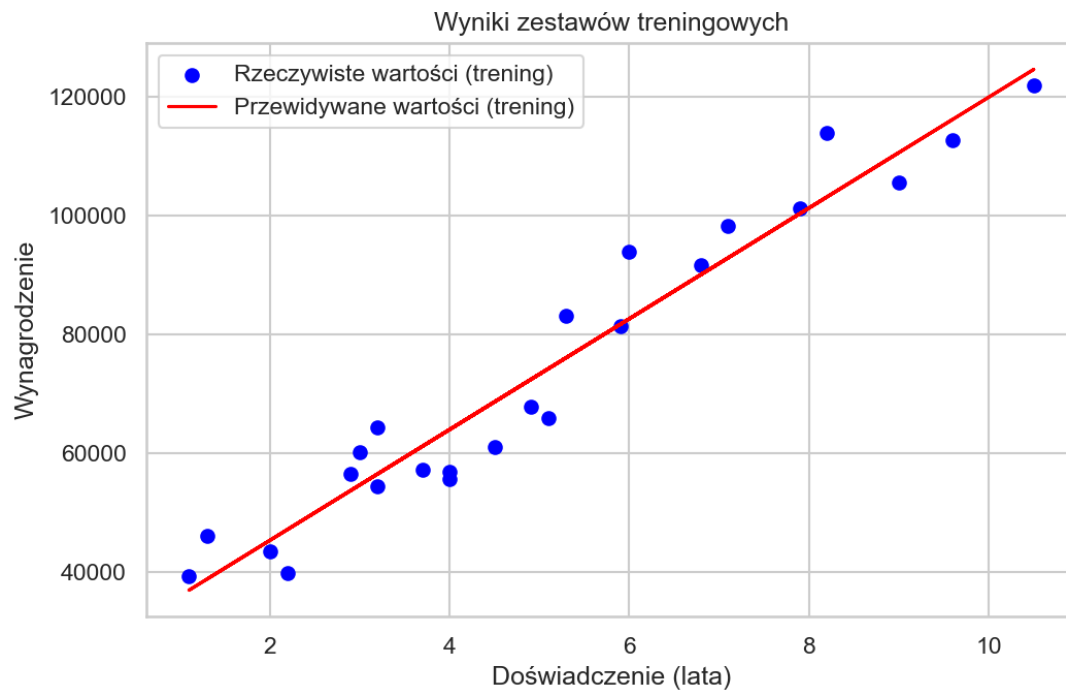


- Wykres zależności między latami doświadczenia a wynagrodzeniem pokazuje, że wraz ze wzrostem doświadczenia wzrasta wynagrodzenie.
- Zależność ta jest w przybliżeniu liniowa.
- Dla osób z doświadczeniem poniżej 2 lat wynagrodzenia są bardziej zróżnicowane. W miarę wzrostu doświadczenia wynagrodzenia stają się bardziej przewidywalne.

Modelowanie Regresji Liniowej

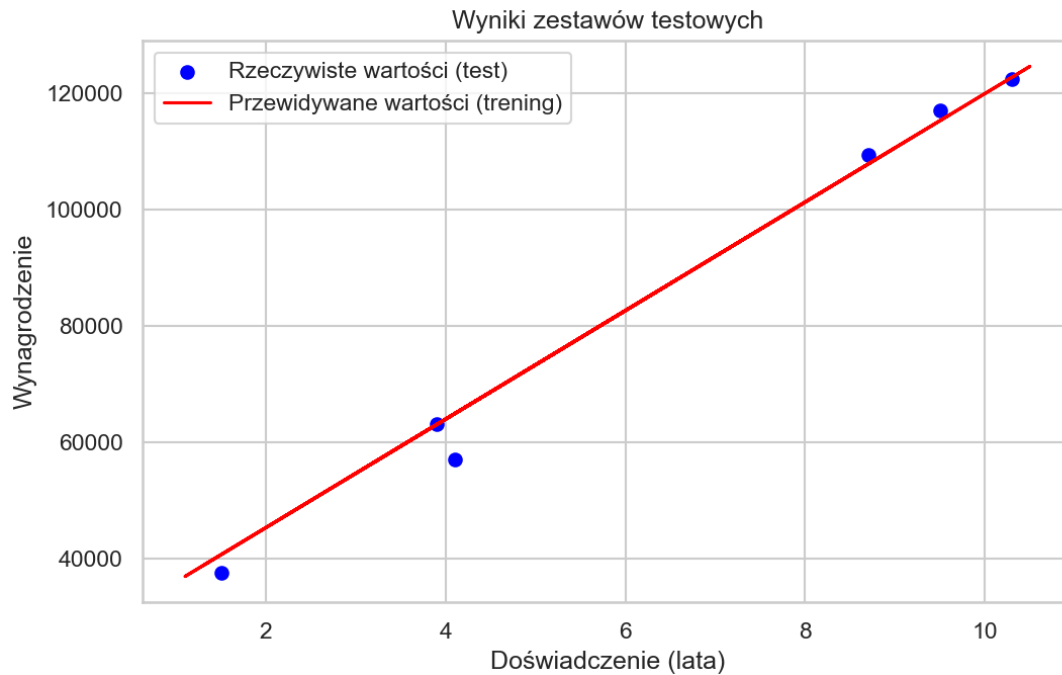
- Dane zostały podzielone na zmienne niezależne (X: doświadczenie) i zależne (y: wynagrodzenie).
- Podział danych na zestawy treningowe (80%) i testowe (20%) został przeprowadzony za pomocą funkcji `train_test_split`.
- Model regresji liniowej został wytrenowany na zestawie treningowym (`X_train`, `y_train`) przy użyciu metody `regressor.fit`.
- Przewidywania wynagrodzenia dla zestawu testowego (`X_test`) zostały dokonane za pomocą `regressor.predict`.

Wyniki Modelu



Zbiór Treningowy

- Model dobrze odwzorował zależność między doświadczeniem a wynagrodzeniem dla zbioru treningowego.
- Większość punktów znajduje się blisko linii najlepszego dopasowania, jednak najbardziej odchylone są dane dla 6 i 8 lat doświadczenia.
- Największe wynagrodzenia są bardziej rozproszone od linii regresji, co może świadczyć o trudności modelu w dokładnym przewidywaniu wynagrodzeń dla tych poziomów doświadczenia.



Zbiór Testowy

- Większość punktów testowych jest blisko linii co sugeruje, że model poprawnie przewiduje wynagrodzenia na podstawie lat doświadczenia.
- Błędy przewidywalne są nieznaczne, jedyne zauważalne odchylenie wystąpiło dla doświadczenia około 4 lat.
- Jak w zestawie treningowym, wynagrodzenia rosną wraz ze wzrostem doświadczenia.

Wnioski

Model regresji liniowej skutecznie odwzorowuje zależność między latami doświadczenia a wynagrodzeniem, chociaż ma pewne trudności z dokładnym przewidywaniem wynagrodzeń dla najwyższych poziomów doświadczenia.