

# US HEALTH INSURANCE APPLICATION

Cel aplikacji :

- Wyznaczenie kwoty potrąceń na podstawie najlepiej dopasowanego modelu uczenia maszynowego na podstawie bazy danych firmy ubezpieczeniowej.

Autorzy:

Joanna Garwacka

Monika Pokojka

Łukasz Wysocki

Grzegorz Piaskowski

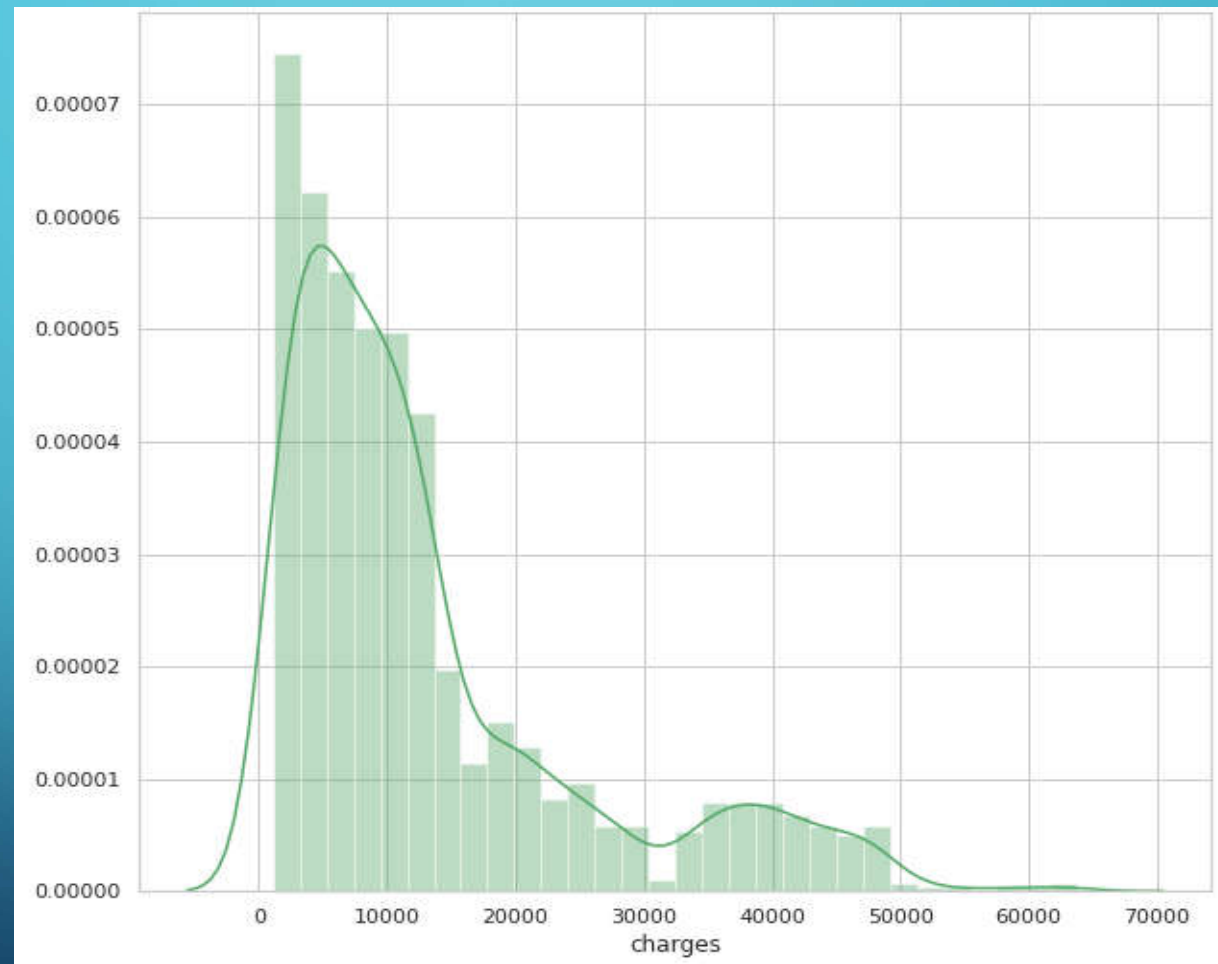
# DANE WEJŚCIOWE:

## 1. ANALIZA DANYCH

### TABELA WYJŚCIOWA :

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
5	31	female	25.740	0	no	southeast	3756.62160

- Zestawienie wielkości kosztów poniesionych na transakcję ubezpieczeniową (charges)



## Potrącenia w zależności o regionu i płci

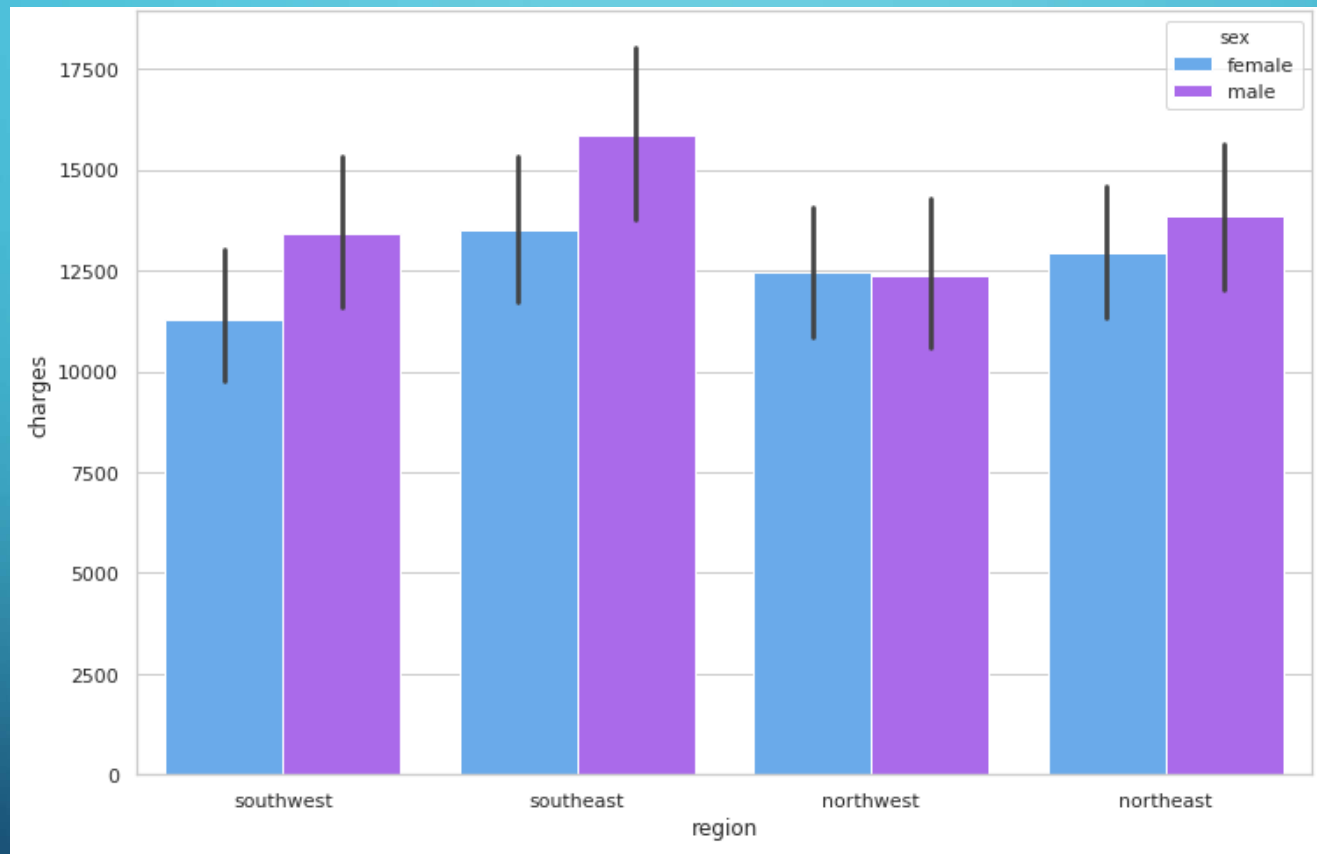
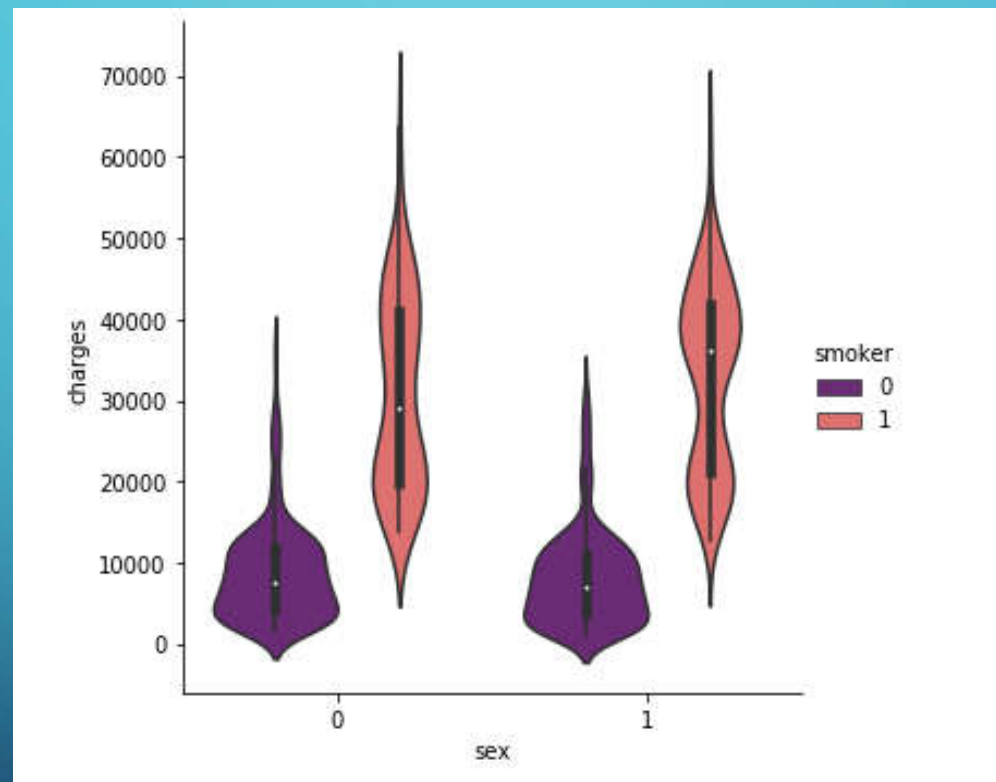
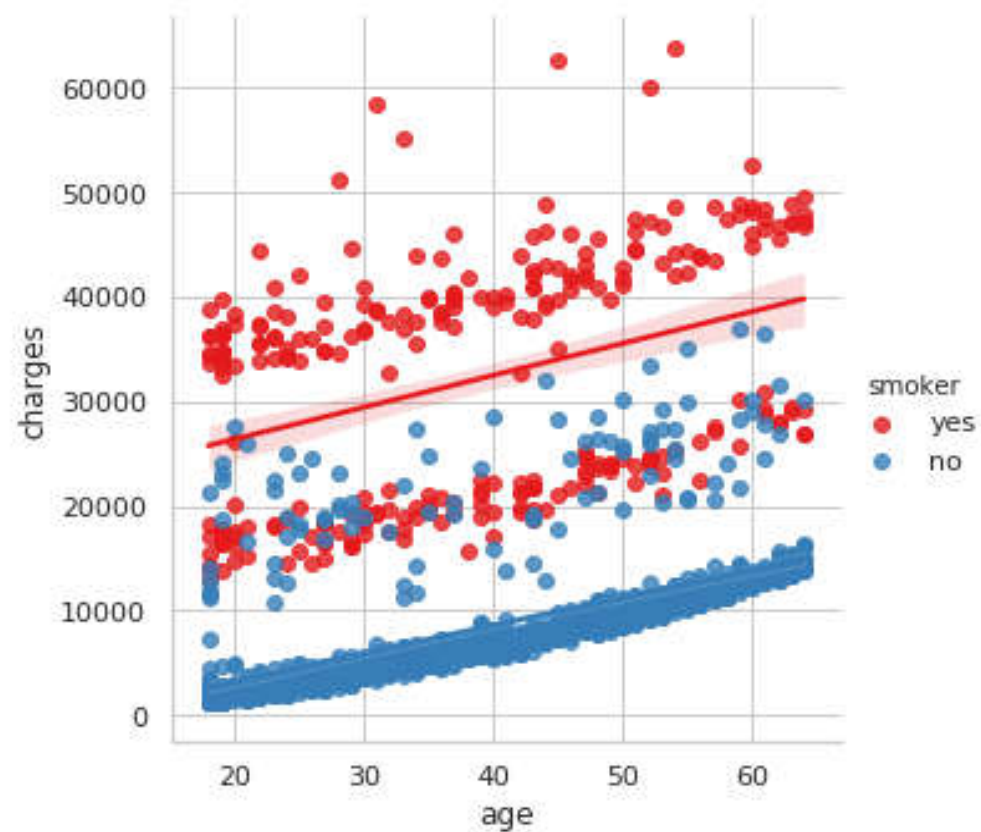


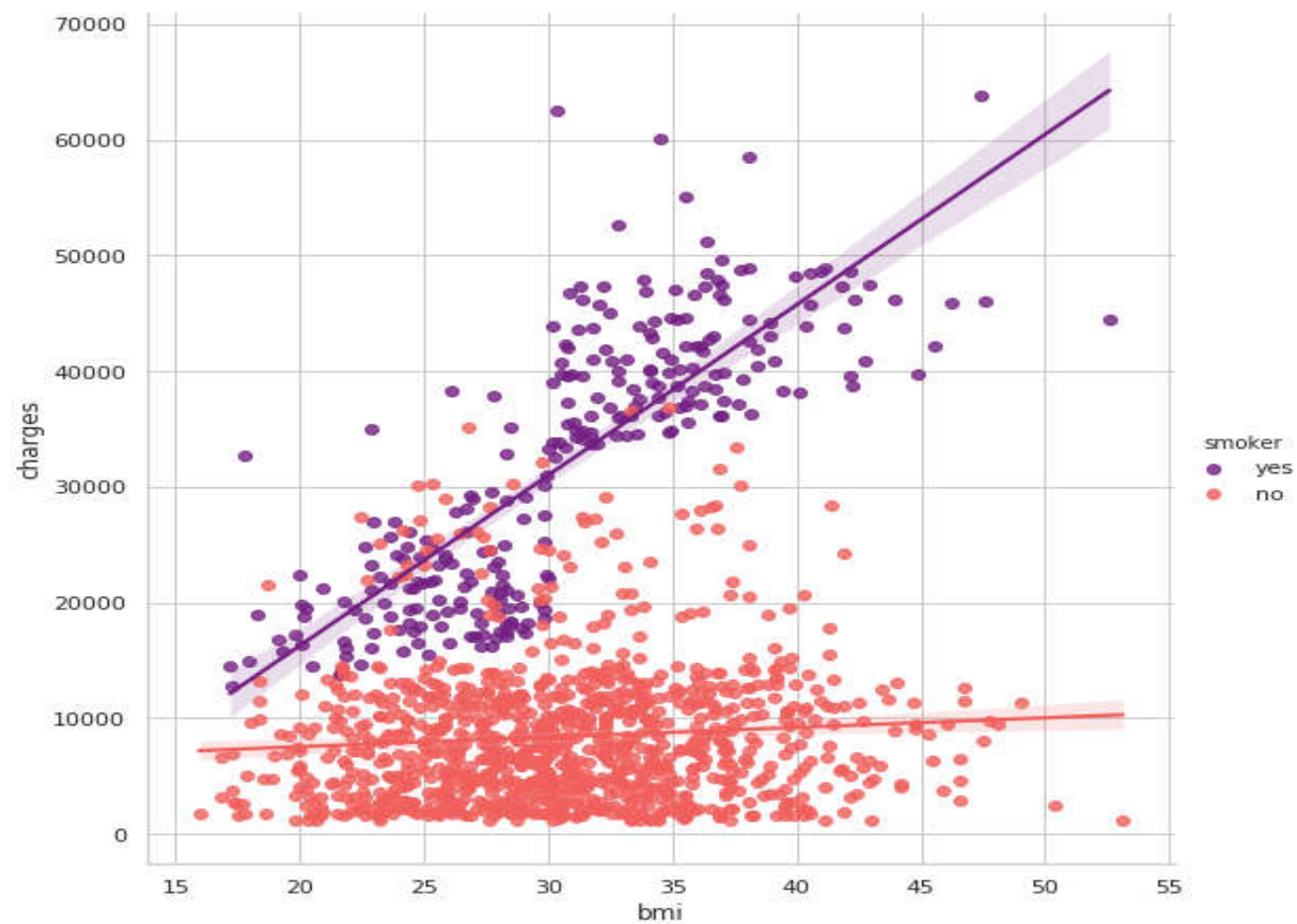
Diagram wielkości potrąceń ze względu na płeć i palaczy



## Potrącenia w zależności od wieku w podziale na palaczy



## Potrącenia w zależności od BMI & palaczy



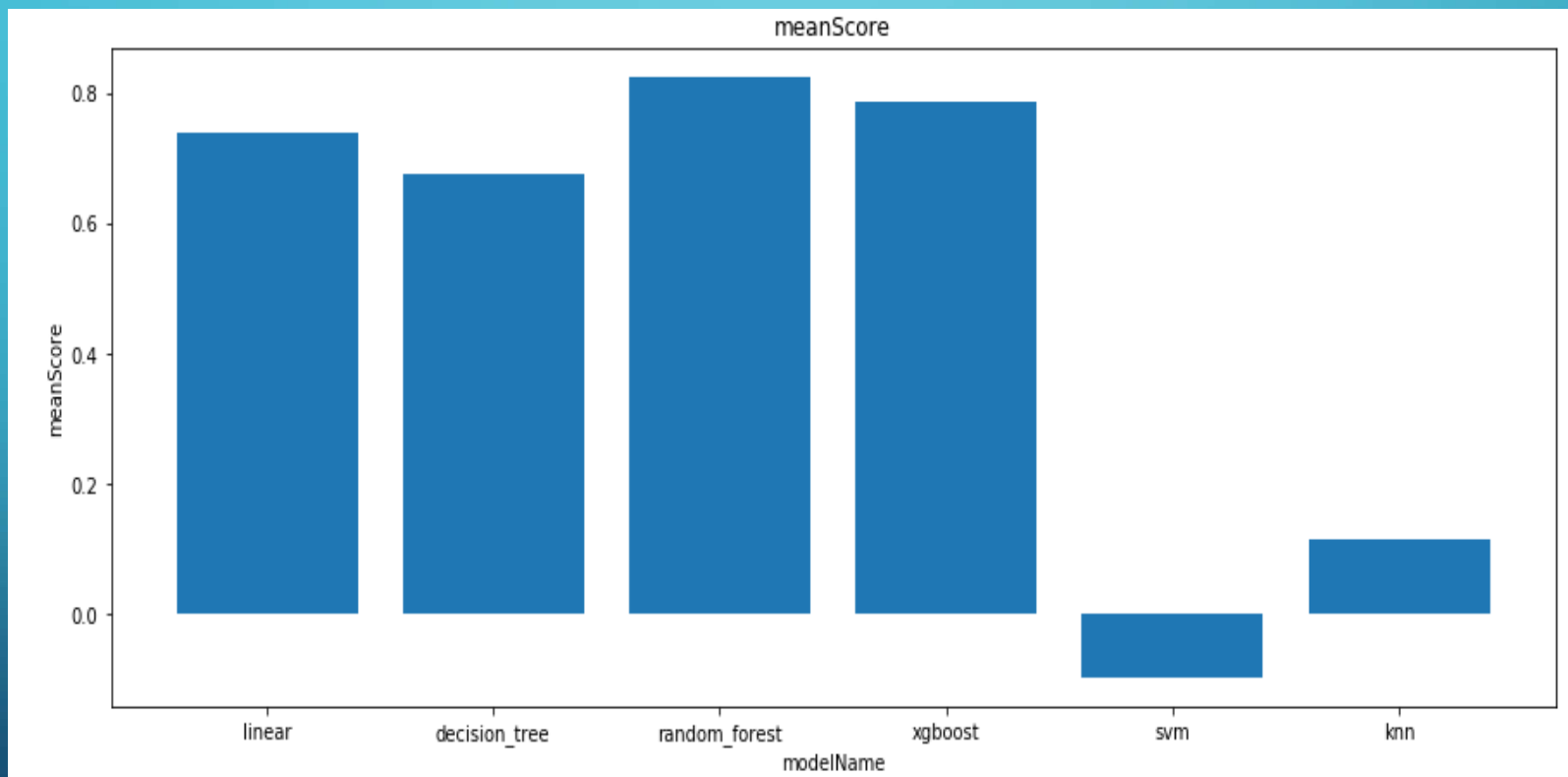
Wybór najlepszego modelu :

Dane przetestowano na modelach:

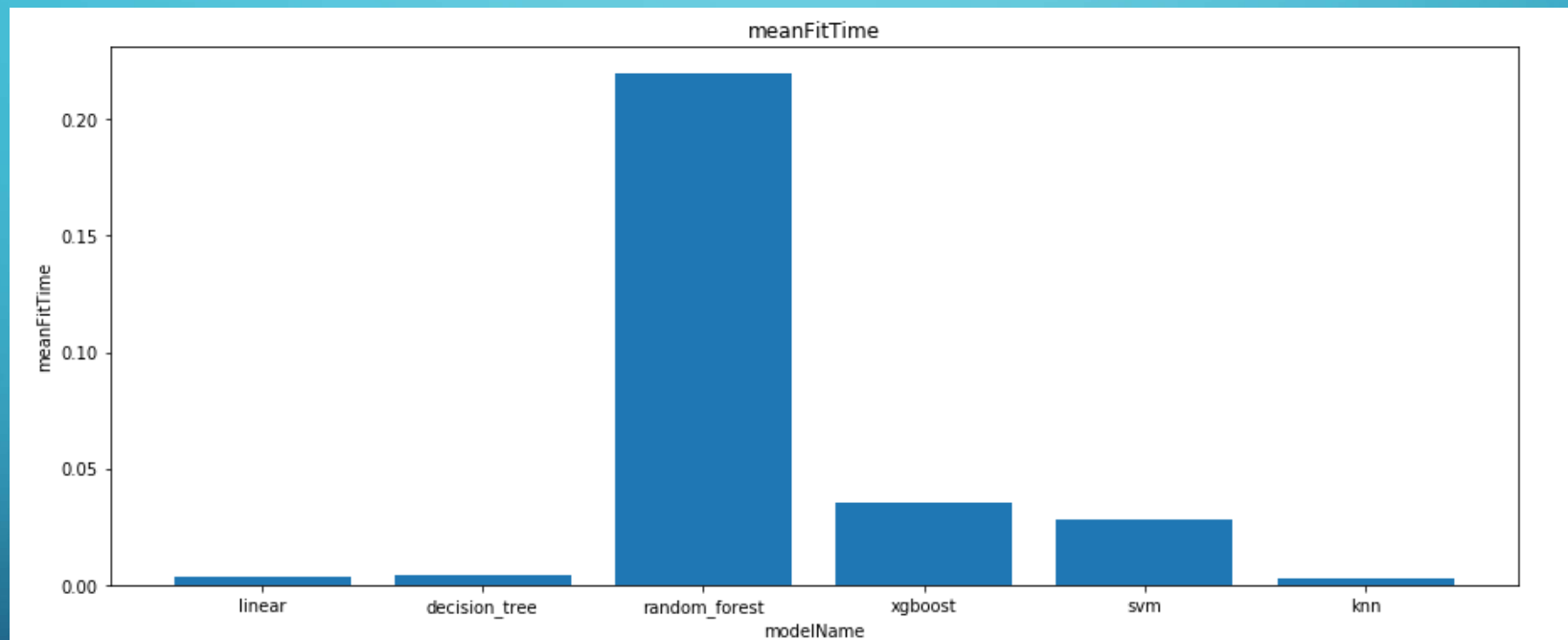
- Linear regresion
- SVM
- Xgboost
- RandomForestRegressor
- DecisionTree
- KNN
- K- MEANS ( metoda nie dala czytelnych rezultatów z uwagi z spory rozrzut danych)



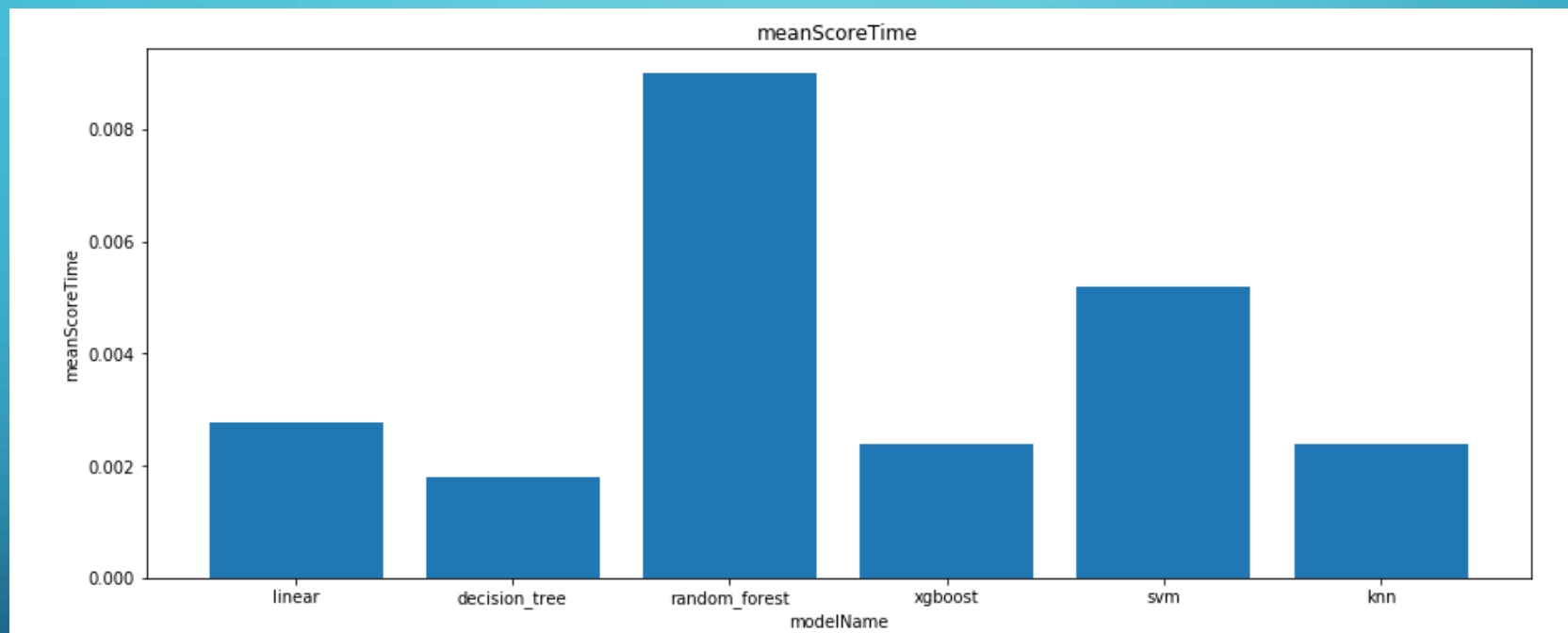
## Wybór najlepszego modelu :



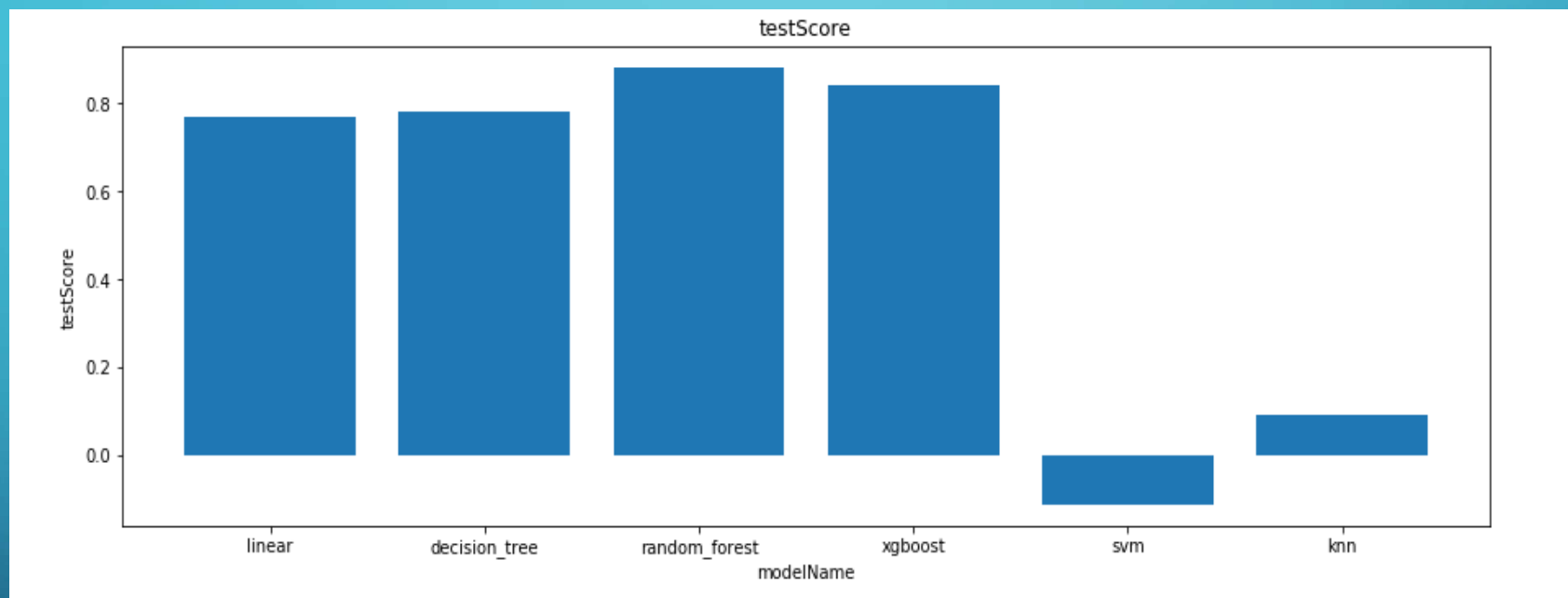
## Wybór najlepszego modelu :



## Wybór najlepszego modelu :

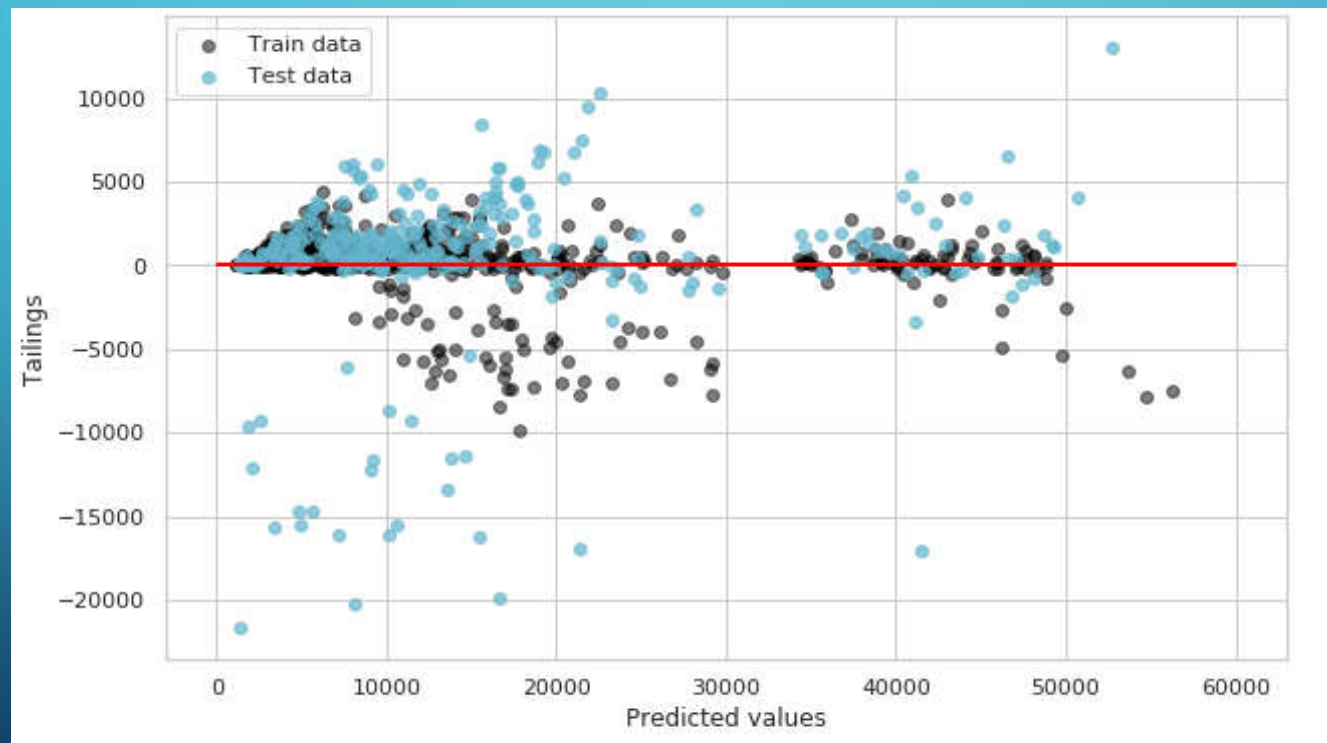


## Wybór najlepszego modelu :



Wybór najlepszego modelu :

Wykres dopasowania danych testowych i treningowych na podstawie metody RandomForestRegressor



# WNIOSKI

- Dlaczego wprowadzić model:
- Model na obecnych obserwacjach prognozuje średni koszt danej transakcji z błędem 0.66% - Czyli z błędem bliskim zero
- Możemy więc używając prezentowanego modelu wygodny sposób zarządzać przychodem zyskiem, który chcemy uzyskać z naszych produktów ubezpieczeniowych .