PODSUMOWANIE RAPORTU

Najlepszy model	Regresja Logistyczna
Wynik walidacji	0.7930
AUROC	0.8437
Gini	0.6875

Raport klasyfikacji

	Precision	Recall	F1-Score	Support
0	0.84	0.89	0.86	1552
1	0.63	0.52	0.57	561
macro avg	0.73	0.70	0.72	2113
weighted avg	0.78	0.79	0.78	2113

Statystyki opisowe danych

Feature	mean	std	min	25%	50%	75%	max
gender	nan	nan	nan	nan	nan	nan	nan
SeniorCitizen	0.16	0.37	0.0	0.0	0.0	0.0	1.0
Partner	nan	nan	nan	nan	nan	nan	nan
Dependents	nan	nan	nan	nan	nan	nan	nan
tenure	32.37	24.56	0.0	9.0	29.0	55.0	72.0
PhoneService	nan	nan	nan	nan	nan	nan	nan
MultipleLines	nan	nan	nan	nan	nan	nan	nan
InternetService	nan	nan	nan	nan	nan	nan	nan
OnlineSecurity	nan	nan	nan	nan	nan	nan	nan
OnlineBackup	nan	nan	nan	nan	nan	nan	nan
DeviceProtection	nan	nan	nan	nan	nan	nan	nan
TechSupport	nan	nan	nan	nan	nan	nan	nan
StreamingTV	nan	nan	nan	nan	nan	nan	nan
StreamingMovies	nan	nan	nan	nan	nan	nan	nan
Contract	nan	nan	nan	nan	nan	nan	nan
PaperlessBilling	nan	nan	nan	nan	nan	nan	nan
PaymentMethod	nan	nan	nan	nan	nan	nan	nan
MonthlyCharges	64.76	30.09	18.25	35.5	70.35	89.85	118.75
TotalCharges	2279.73	2266.79	0.0	398.55	1394.55	3786.6	8684.8
Churn	0.27	0.44	0.0	0.0	0.0	1.0	1.0

Profil Klienta na Podstawie Predykcji Modelu

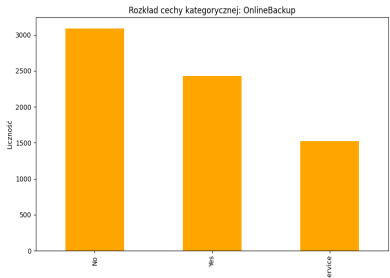| Feature	Overall Average	Predicted High Probability (Target=1)
Contract_Month-to-month	0.55	1.00
Contract_One year	0.20	0.00
Contract_Two year	0.24	0.00
Dependents_Yes	0.29	0.01
DeviceProtection_No	0.44	0.75
DeviceProtection_No internet service	0.22	0.00
DeviceProtection_Yes	0.33	0.25
InternetService_DSL	0.34	0.02
InternetService_Fiber optic	0.43	0.98
InternetService_No	0.22	0.00
MonthlyCharges	64.26	86.43
MultipleLines_No	0.49	0.33
MultipleLines_No phone service	0.10	0.02
MultipleLines_Yes	0.42	0.65
OnlineBackup_No	0.44	0.86
OnlineBackup_No internet service	0.22	0.00
OnlineBackup_Yes	0.33	0.14
OnlineSecurity_No	0.50	1.00
OnlineSecurity_No internet service	0.22	0.00
OnlineSecurity_Yes	0.28	0.00
PaperlessBilling_Yes	0.57	0.94
Partner_Yes	0.48	0.24
PaymentMethod_Bank transfer (automat	0.22	0.00
PaymentMethod_Credit card (automatic	0.21	0.00

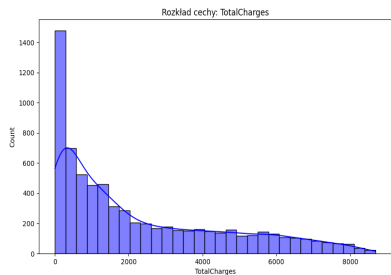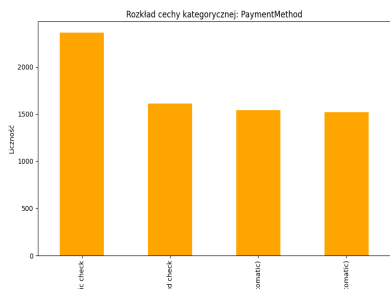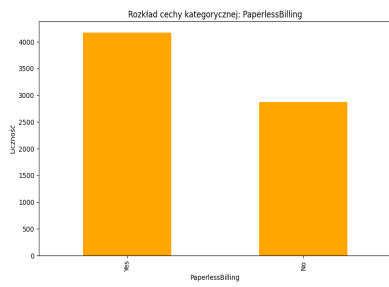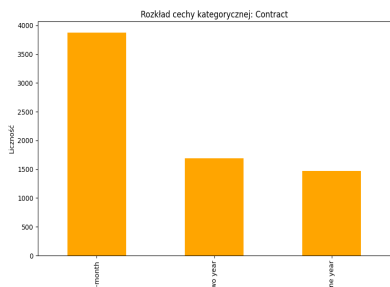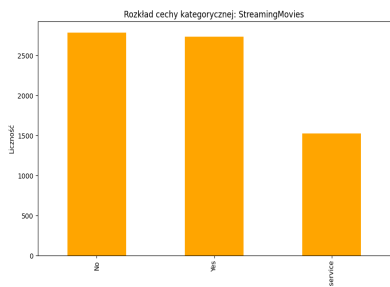| Feature	Overall Average	Predicted High Probability (Target=1)
PaymentMethod_Electronic check	0.33	0.99
PaymentMethod_Mailed check	0.24	0.01
PhoneService_Yes	0.90	0.98
SeniorCitizen	0.16	0.56
StreamingMovies_No	0.39	0.43
StreamingMovies_No internet service	0.22	0.00
StreamingMovies_Yes	0.39	0.57
StreamingTV_No	0.40	0.35
StreamingTV_No internet service	0.22	0.00
StreamingTV_Yes	0.38	0.65
TechSupport_No	0.49	1.00
TechSupport_No internet service	0.22	0.00
TechSupport_Yes	0.28	0.00
TotalCharges	2231.93	537.34
gender_Male	0.50	0.61
tenure	32.02	5.90

Tabela przedstawia znaczenie cech modelu wykorzystywanego do predykcji profilu klienta.

 1. Feature (Cecha): Nazwa cechy.

2. Overall Average (Srednia Ogolna): Srednia wartosc cechy dla wszystkich klientow.

3. Predicted High Probability (Target=1) (Przewidywana Wysoka Prawdopodobienstwo dla Target=1):Przewidywane

prawdopodobienstwo, ze zmienna docelowa wynosi 1 dla danej cechy.

Wiersze z 3 najwazniejszymi cechami sa podswietlone na zielono.
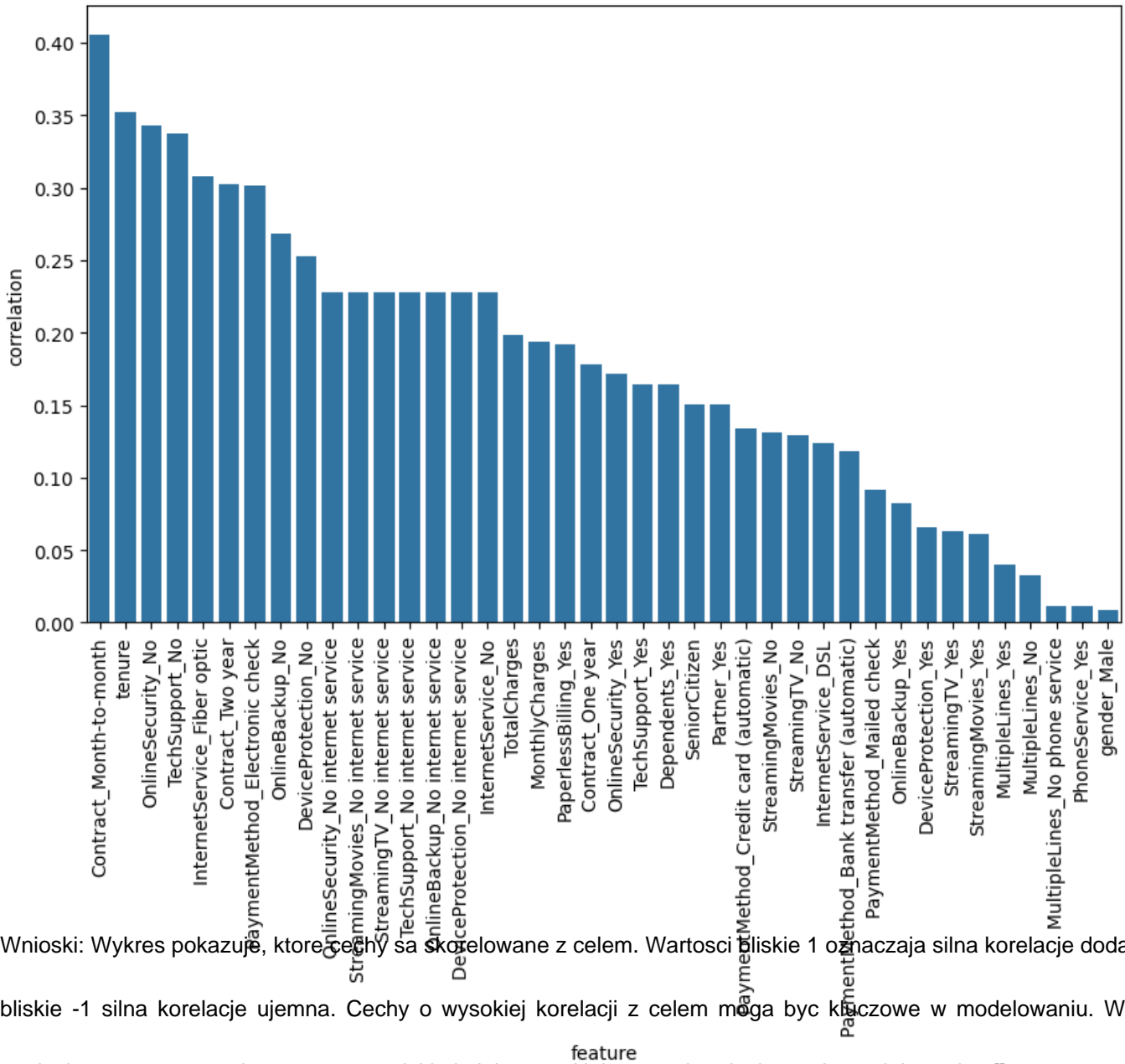
Wizualizacja rozkladu danych

Rozkład cechy kategorycznej: StreamingMovies

Rozkład cechy kategorycznej: Contract

Rozkład cechy kategorycznej: PaperlessBilling

Rozkład cechy kategorycznej: PaymentMethod
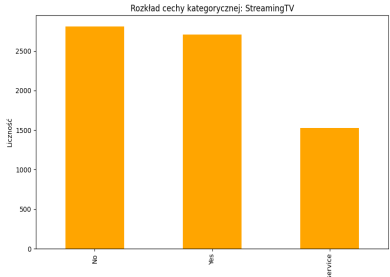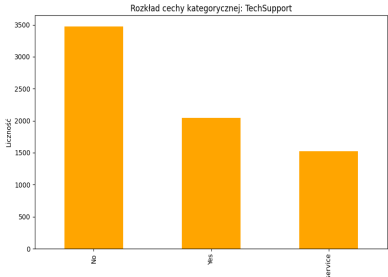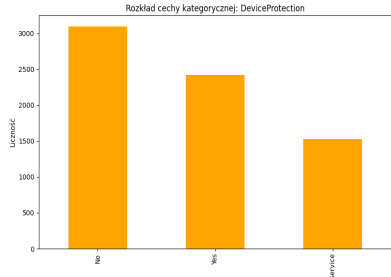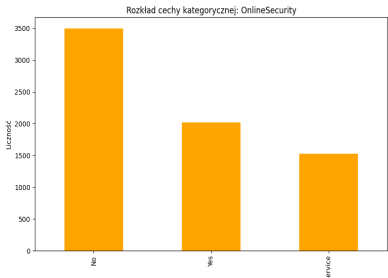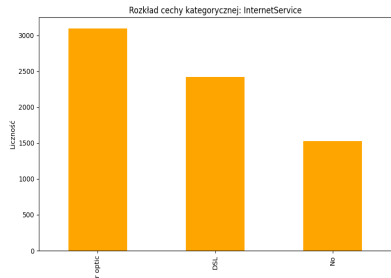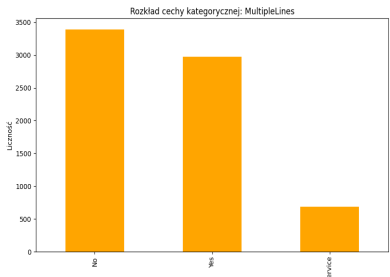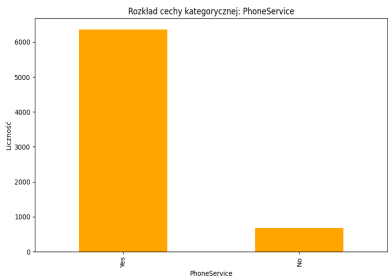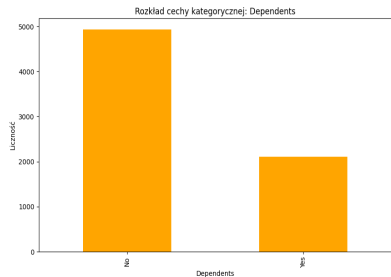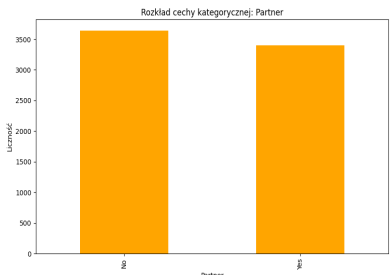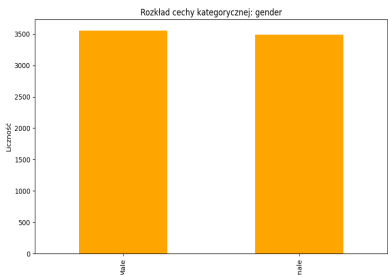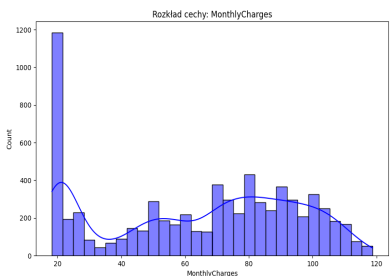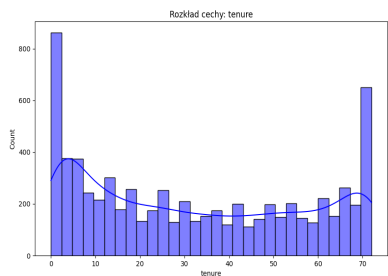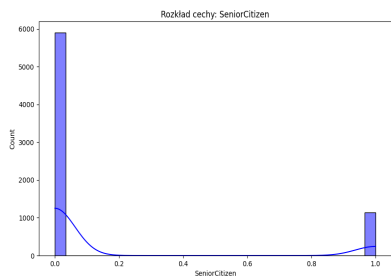
Korelacja cech z celem

ABS Correlation with Churn

Wnioski: Wykres pokazuje, ktore cechy sa skorelowane z celem. Wartosci bliskie 1 oznaczaja silna korelacje dodatnia, a bliskie -1 silna korelacje ujemna. Cechy o wysokiej korelacji z celem moga byc kluczowe w modelowaniu. Wartosci ponizej 0.1 sa zazwyczaj uznawane za niskie i nieistotne. Najwyzsza korelacje z celem mialy cechy: [].

Macierz korelacji cech

Correlation Matrix
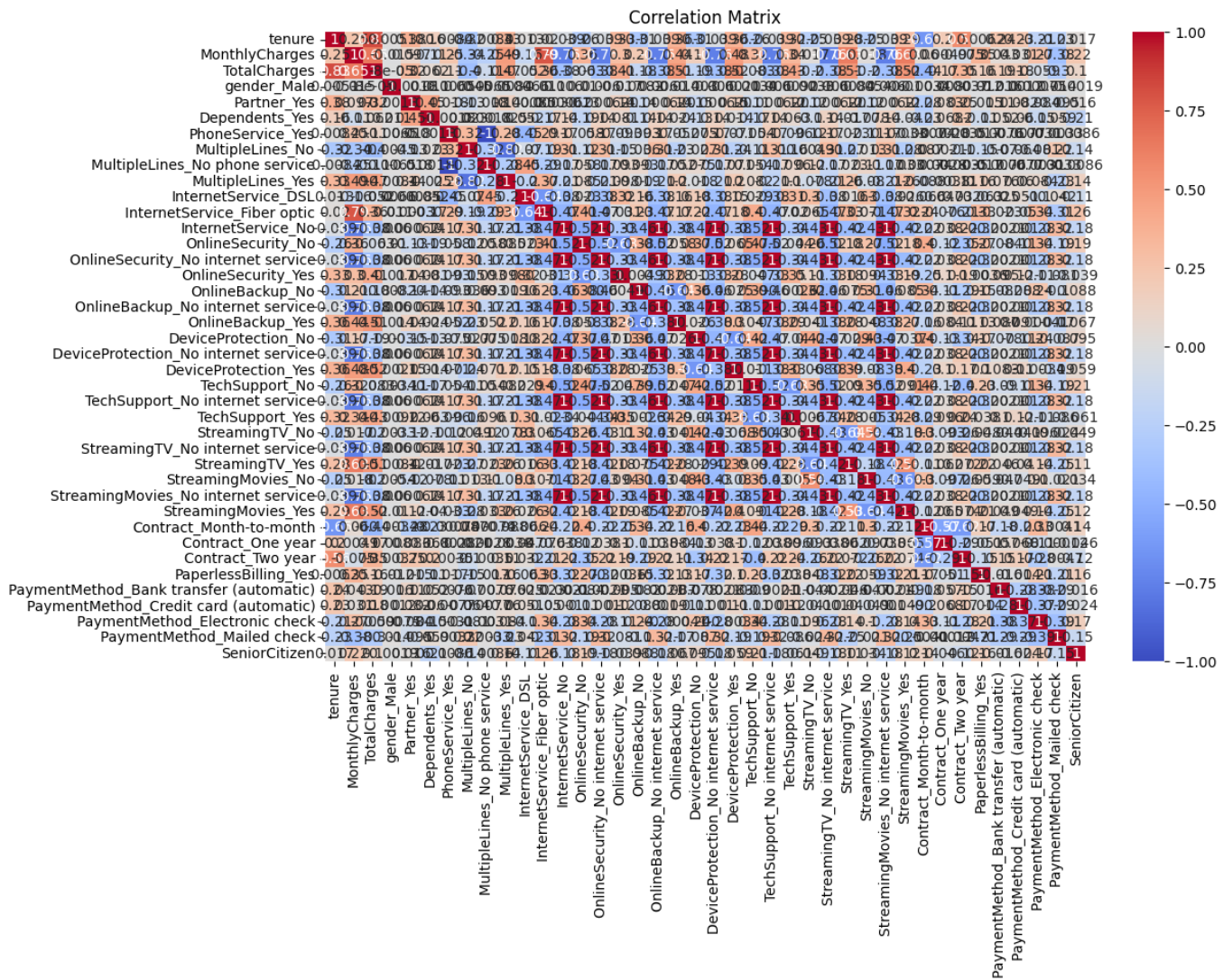
Wnioski: Wykres pokazuje, ktore cechy sa ze soba skorelowane. W wyniku analizy usunieto cechy o wysokiej korelacji miedzy soba: ['TotalCharges', 'Contract_Two year', 'MultipleLines_No', 'TechSupport_No internet service', 'OnlineSecurity_Yes', 'MonthlyCharges', 'InternetService_Fiber optic', 'OnlineSecurity_No', 'DeviceProtection_No', 'StreamingTV_No', 'DeviceProtection_Yes', 'InternetService_No', 'OnlineBackup_Yes', 'StreamingTV_No internet service', 'OnlineBackup_No', 'Contract_Month-to-month', 'StreamingMovies_Yes', 'tenure', 'Contract_One year', 'TechSupport_Yes', 'MultipleLines_Yes', 'OnlineBackup_No internet service', 'InternetService_DSL', 'StreamingMovies_No', 'OnlineSecurity_No internet service', 'StreamingMovies_No internet service', 'DeviceProtection_No internet service', 'TechSupport_No', 'StreamingTV_Yes'].
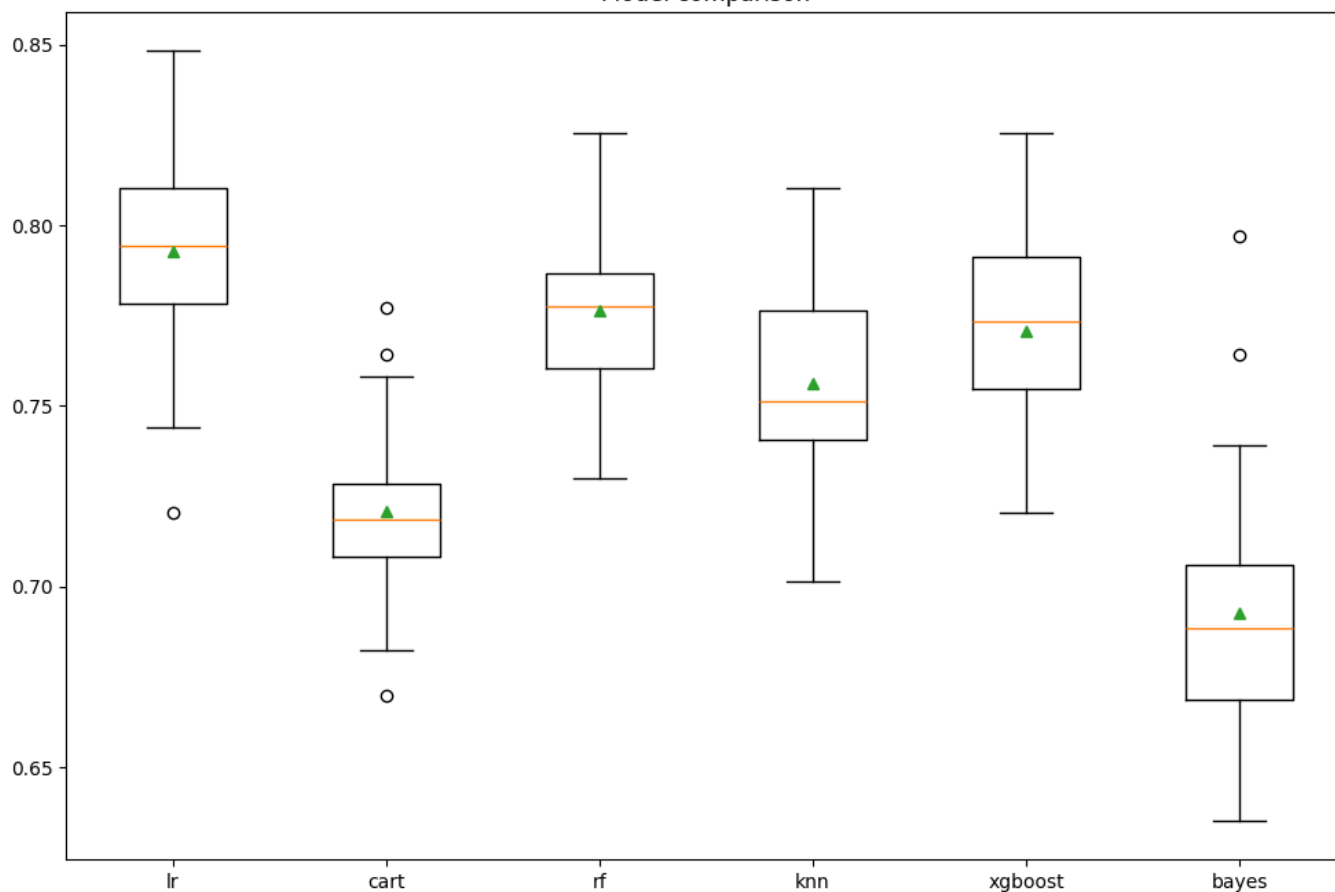
Waznosc cech

Feature Importance
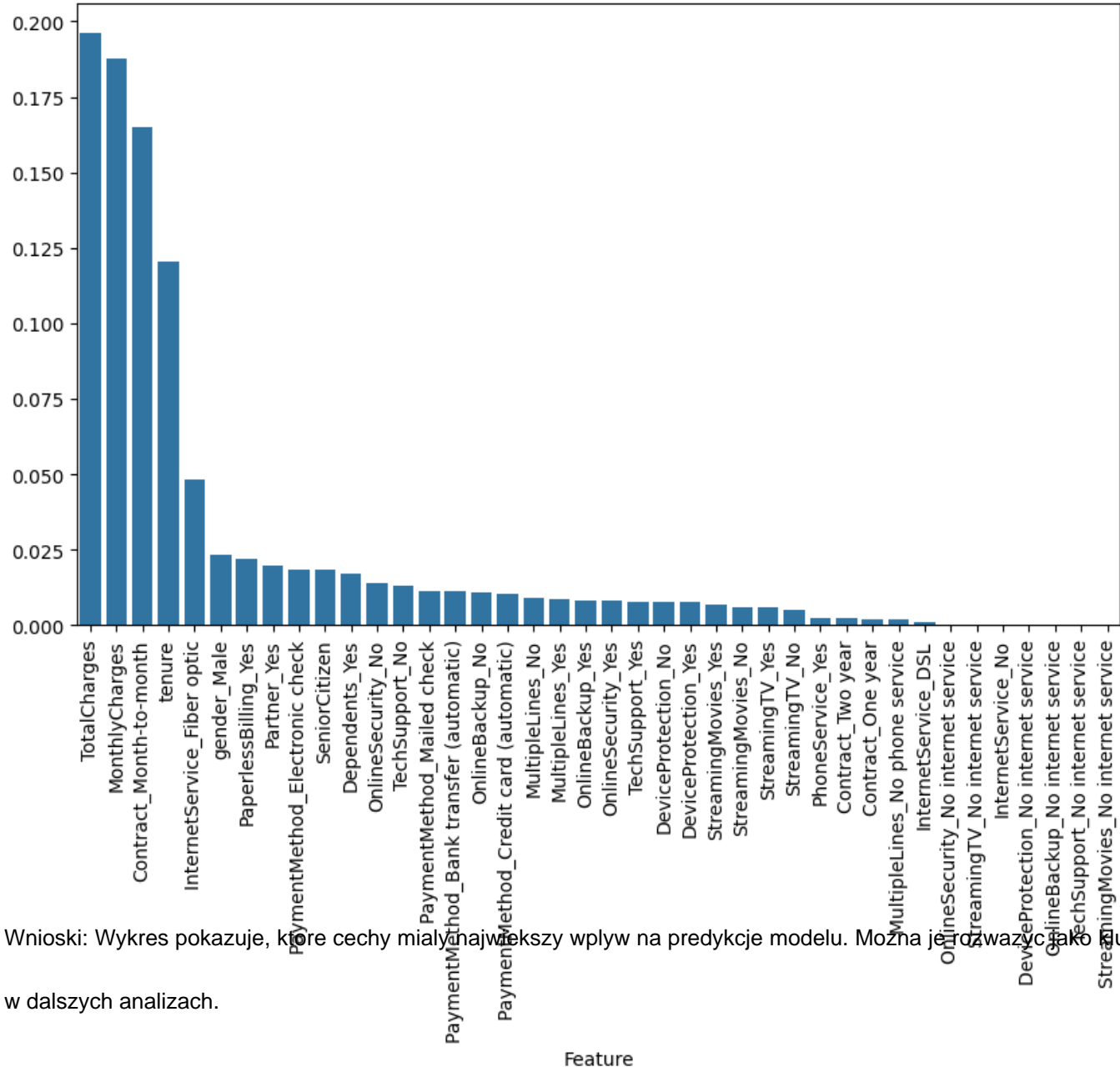
Wnioski: Wykres pokazuje, ktore cechy mialy najwiekszy wplyw na predykcje modelu. Mozna je rozwazyc jako kluczowe w dalszych analizach.
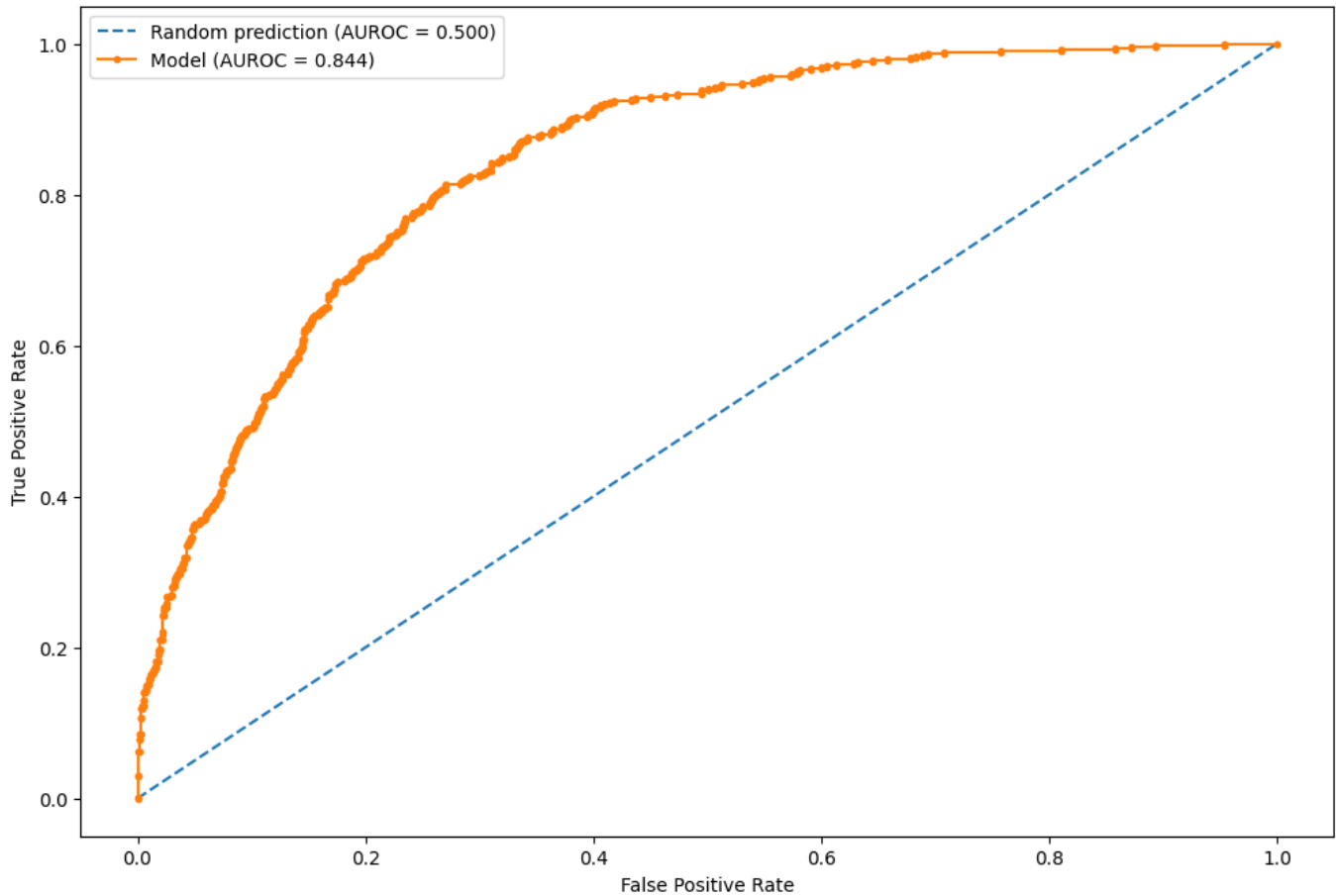
Porownanie modeli

Model comparison

Wnioski: Wykres przedstawia porownanie roznych modeli pod wzgledem ich skutecznosci. Widac, ze model 'lr' osiagnal najwyzsze wyniki w walidacji. Z wynikiem 0.7930 jest to najlepszy model do dalszych analiz.

Krzywa ROC

Wnioski: Wykres przedstawia krzywa ROC dla najlepszego modelu 'lr'. Im wyzsza powierzchnia pod krzywa (AUC), tym

lepszy model. W tym przypadku AUC wynosi 0.8437.