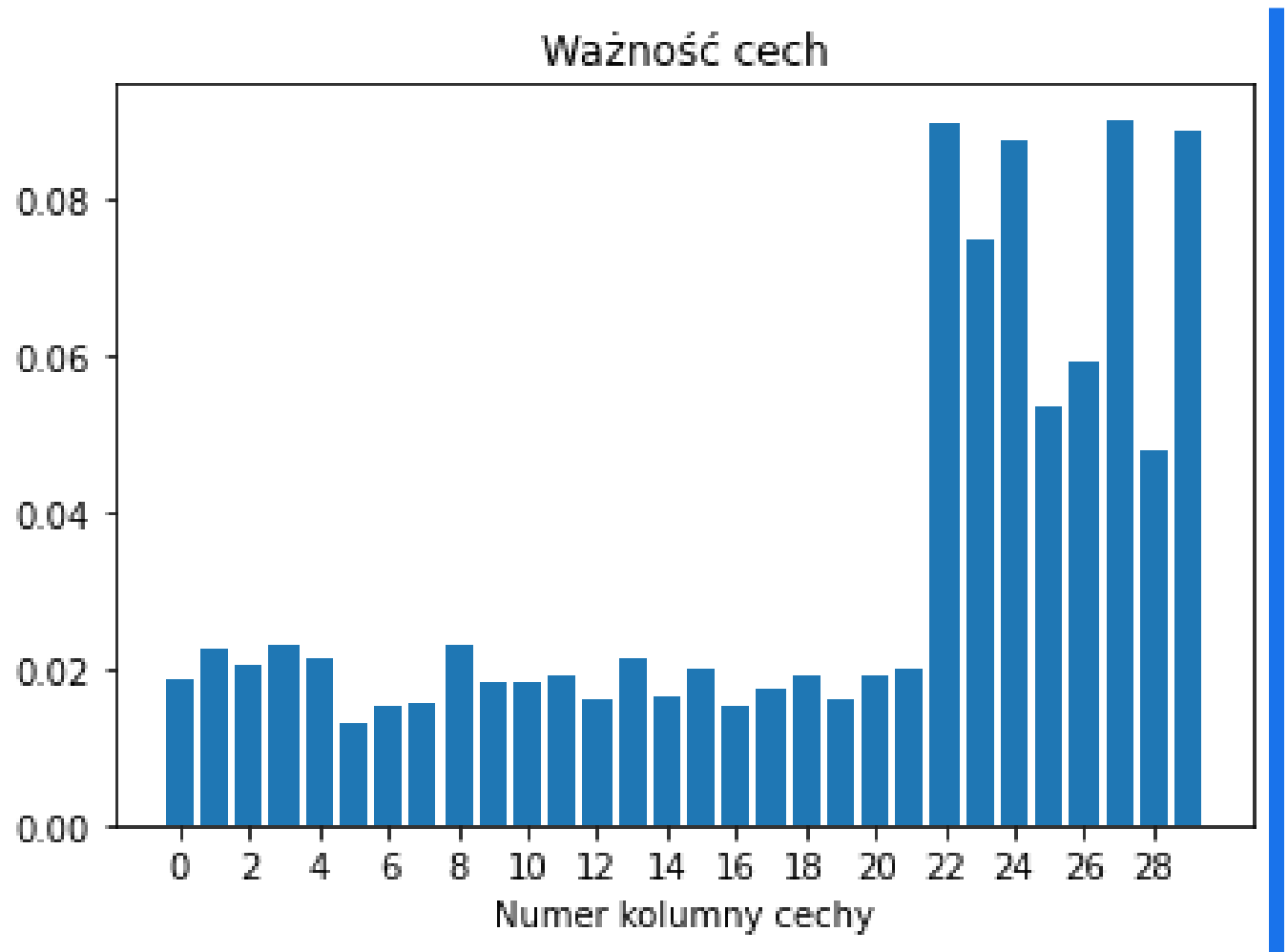


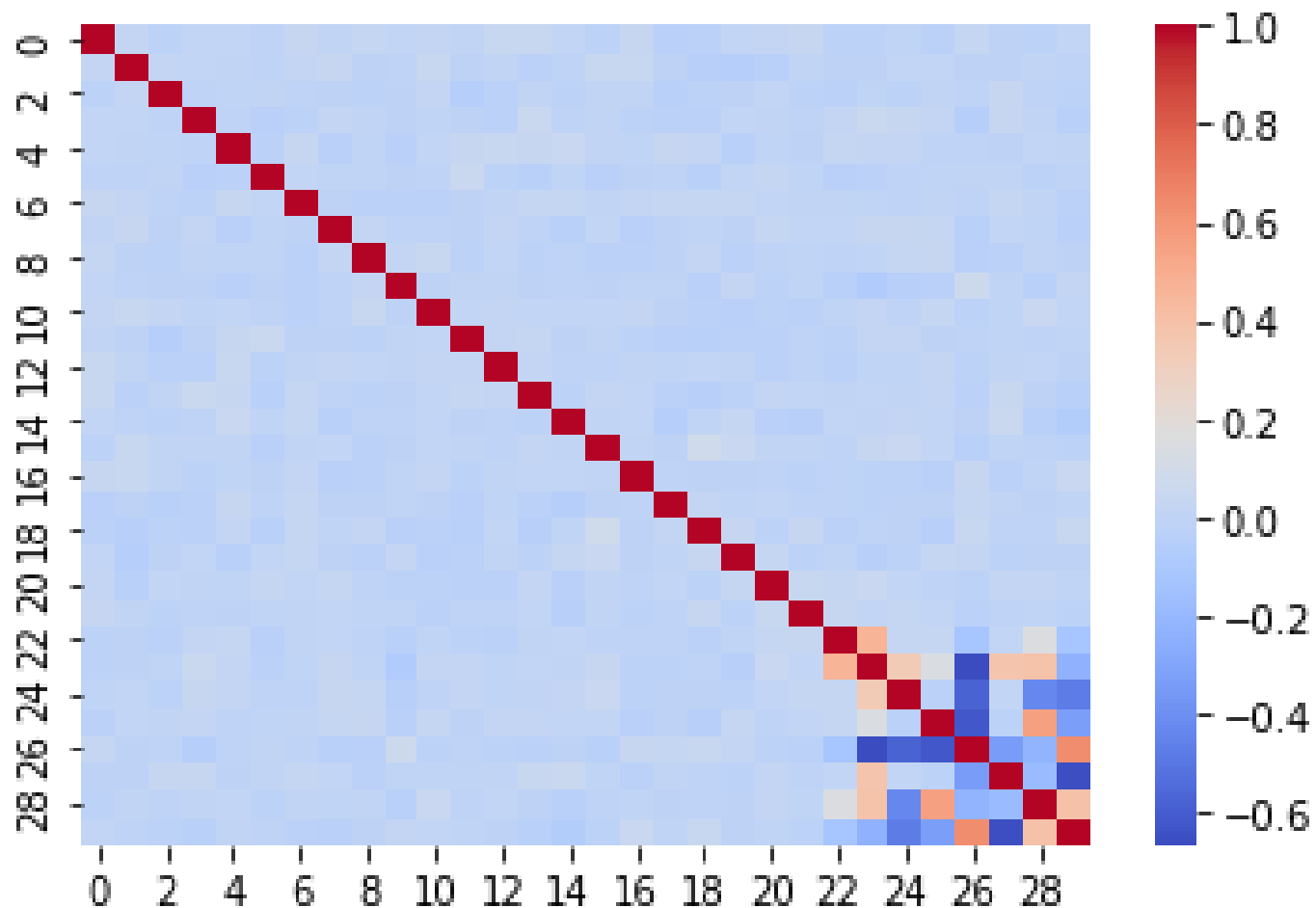
# Wstęp do uczenia maszynowego Projekt

JULIA GIRTLER



Zbadanie  
ważności cech  
za pomocą  
XGBoosta

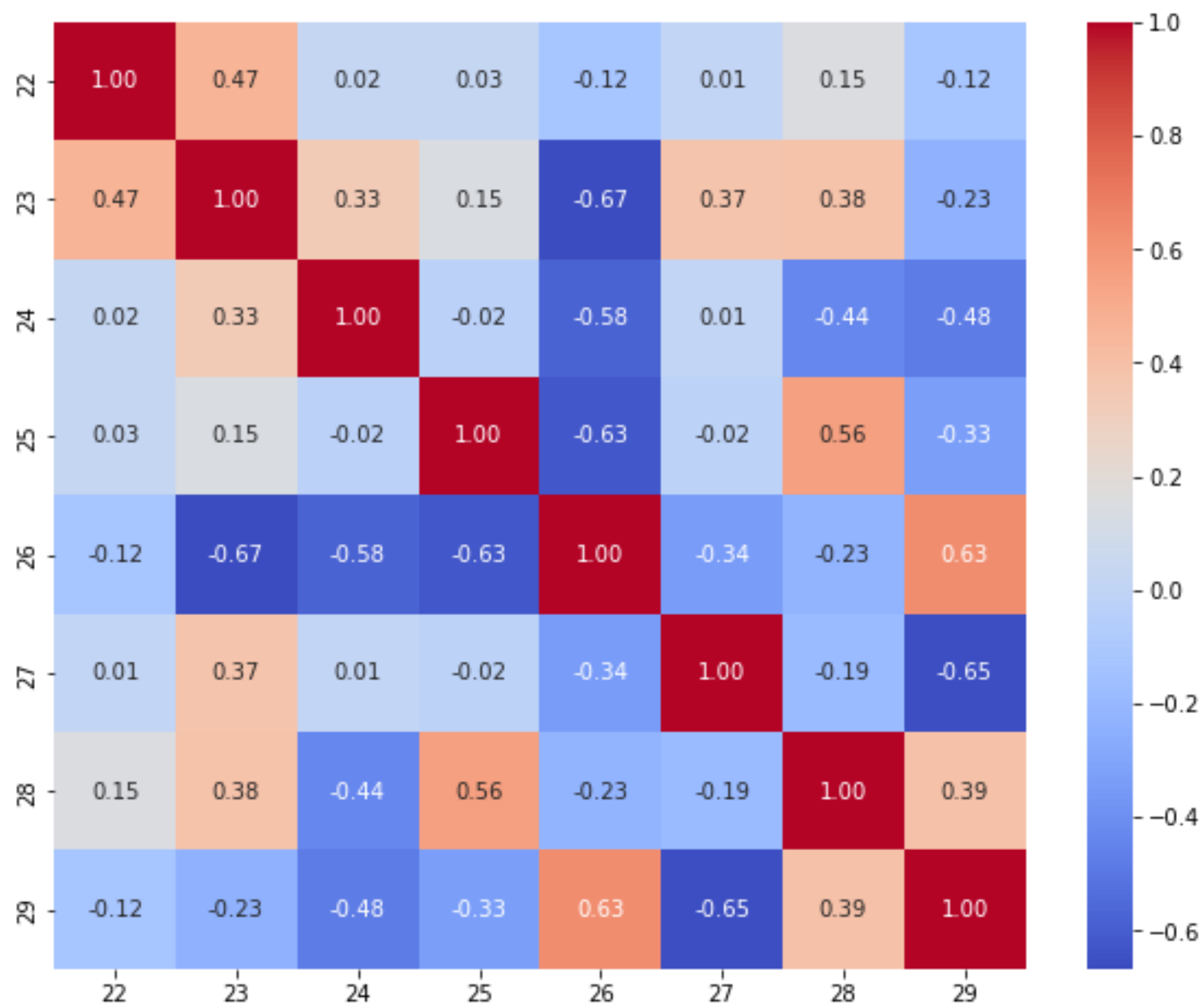
WAŻNOŚCI SUMUJĄ SIĘ DO  
JEDYNKI



## MAPA CIEPŁA

Kolumny od 0 do 21  
nie są ze sobą  
wzajemnie  
skorelowane

—



# DOBÓR NAJLEPSZEGO MODELU

- Testowane modele: DecisionTreeClassifier, LogisticRegression, SVC, KNearestNeighbors, RandomForestClassifier, VotingClassifier, XGBoost, BaggingClassifier

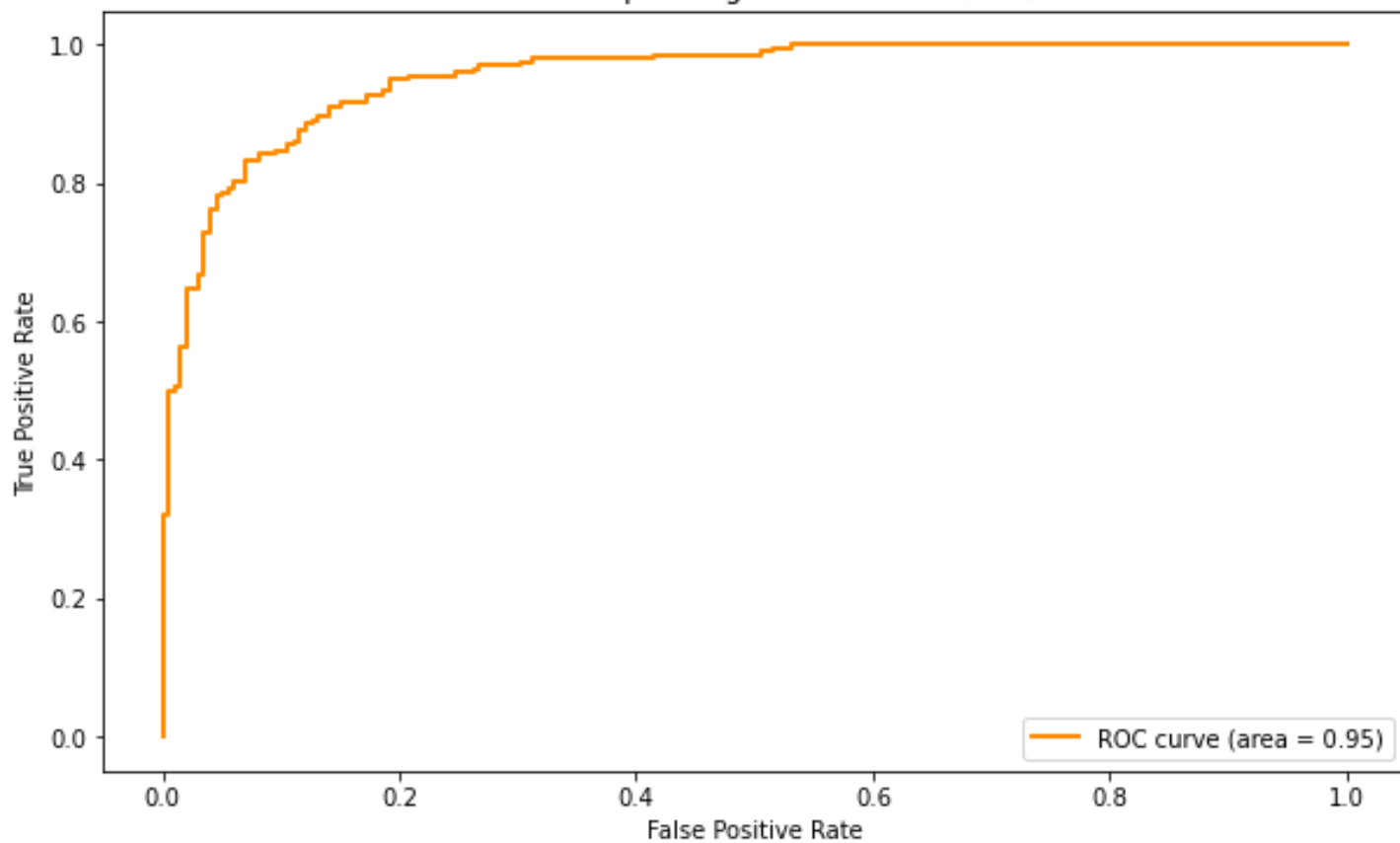
Decision Tree Classifier	Logistic Regression	KNeighbors Classifier	XGBoost	Bagging Classifier	Random Forest Classifier	Voting Classifier	SVC
0.78	0.6275	0.845	0.875	0.855	0.845	0.9025	0.83



# WYBRANY MODEL

- VotingClassifier
- RandomForestClassifier(max\_depth=28, min\_samples\_leaf=1, min\_samples\_split=5, n\_estimators=150), XGBClassifier(max\_depth=20, learning\_rate=0.05, n\_estimators=200, min\_child\_weight=1, subsample=1, colsample\_bytree= 0.8)
- BaggingClassifier(n\_estimators = 200, random\_state = 0, max\_features = 0.5)
- SVC(probability = True, C = 1.25)
- Weights = [3, 2, 3, 1]

Receiver Operating Characteristic (ROC)



Krzywa ROC

# WNIOSKI



- PREPROCESSING
- Z PRZETESTOWANYCH MODELI NAJGORSZA OKAZAŁA SIĘ REGRESJA LOGISTYCZNA, NATOMIAST NAJLEPSZY VOTINGCLASSIFIER
- DOSTOSOWANIE HIPERPARAMETRÓW
- DLA NASZEGO PROBLEMU MAŁO SKUTECZNYM OKAZAŁ SIĘ DOBÓR PARAMETRU VOTING = 'HARD'.

