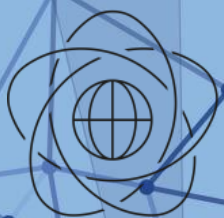


# FEATUREFLEX

Andrii Voznesenskyi

Bartosz Kaczorowski

Auto  
ML  
24/25



Wydział Matematyki  
i Nauk Informatycznych

Politechnika Warszawska

CZYM JEST  
FEATUREFLEX?

# WYBÓR CECH W SYSTEMACH REKOMENDACJI

- powszechny problem w branży e-commerce i streamingowej
- wielkie zbiory danych z mnóstwem etykiet
- potrzeba dynamicznej, wydajnej i precyzyjnej metody personalizacji treści



# ROZWIĄZANIE - FeatureFlex

- **zautomatyzowany wybór najistotniejszych cech** - własne metody selekcji cech, dynamicznie dostosowujące się do danych, trenowalne mechanizmy kontrolne oraz optymalizacja oparta na gradientach
- **optymalizacji modeli** - zaawansowane mechanizmy dostrajania hiperparametrów, dynamiczna zmiana metody optymalizacji w zależności od rozmiaru i charakteru danych (skalowanie rozwiązania do dużych zbiorów)
- **ocena modeli** - pełen zestaw metryk (AUC, dokładność, precyzja, czułość i F1-score), a także raportowanie (macierze konfuzji, krzywe ROC oraz Precision-Recall)
- **wsparcie dla systemów rekomendacji** - usprawnienie procesu selekcji cech oraz optymalizacji modeli dla głębokich systemów rekomendacji



# DLA KOGO JEST PRZEZNACZONY?

**FeatureFlex** skierowany jest do firm z branż e-commerce lub streamingowych, które chcą poprawić wydajność rekomendacji i zmniejszyć koszty obliczeniowe. W szczególności pakiet jest dedykowany:

- **badaczom i inżynierom** rozwijającym systemy rekomendacji oparte na głębokim uczeniu, którzy potrzebują efektywniejszego zarządzania cechami wejściowymi,
- **specjalistom od AutoML i optymalizacji**, którzy pracują nad automatyzacją procesów uczenia maszynowego, w tym wyboru cech w dużych zestawach danych.

# PODOBNE ROZWIĄZANIA

Pakiet inspirowany był pracą [\*AutoField: Automating Feature Selection in Deep Recommender Systems\*](#) panów Yejing Wang, Xiangyu Zhao, Tong Xu i Xian Wu.

Oprócz tego na rynku dostępne są pakiety realizujące podobne funkcjonalności: [Boruta](#), [SHAP](#), [SelectKBest](#) i [ReliefF](#).

**FeatureFlex** wyróżnia się jednak wszechstronnością, wykorzystaniem własnych metod selekcji cech oraz możliwością dynamicznej optymalizacji modeli, co czyni go bardziej elastycznym i skalowalnym rozwiązaniem.

# KOMPONENTY TECHNICZNE

# PRZETWARZANIE DANYCH

- imputacja brakujących wartości dla zmiennych kategorycznych i numerycznych
- skalowanie zmiennych numerycznych
- kodowanie One-Hot dla zmiennych kategorycznych



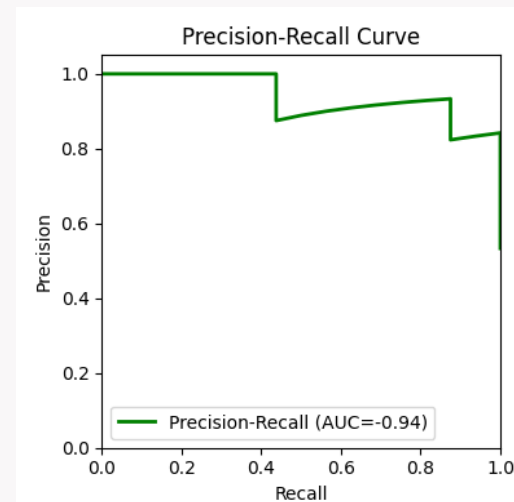
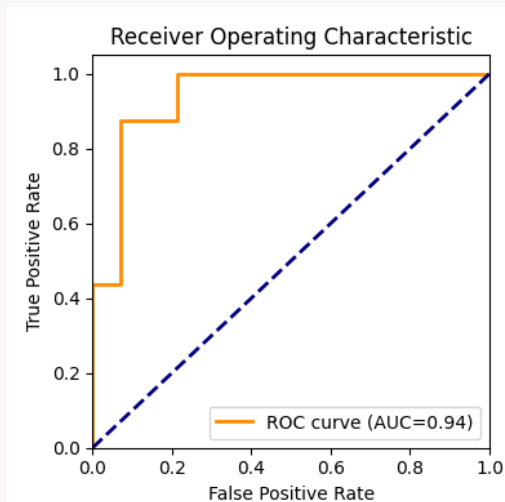
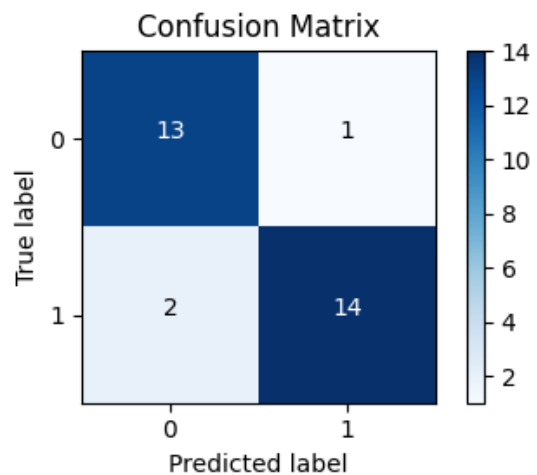
# WYBÓR CECH

- **oparty na modelach głębokiego uczenia** – trenowalny parametr alpha
- **dynamiczny wybór cech** - dostosowanie do różnych typów danych i rozmiarów zbiorów
- **dwie metody określania najistotniejszych cech:**
  - wybór poprzez wykorzystanie SHAP oraz RandomForestClassifier
  - dynamiczny wybór poprzez własny zautomatyzowany proces wykorzystujący zoptymalizowany model ML

# OPTYMALIZACJA MODELI

- **dynamiczne dostrajanie hiperparametrów** - wykorzystanie metod takich jak Grid Search, Random Search oraz Optymalizacja Bayesowska
- **dynamiczna zmiana metody optymalizacji w zależności od rozmiaru i charakteru danych** - skalowanie rozwiązania dla dużych zbiorów danych

Obsługiwane modele: RandomForest, GradientBoosting, LogisticRegression, SVM, XGBoost, KNN.



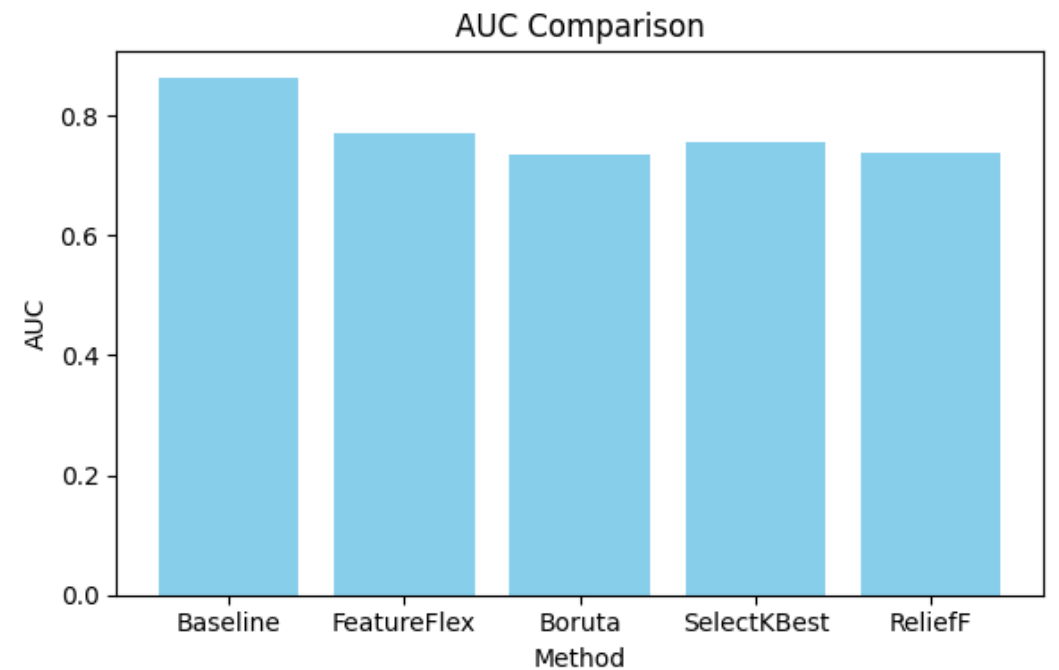
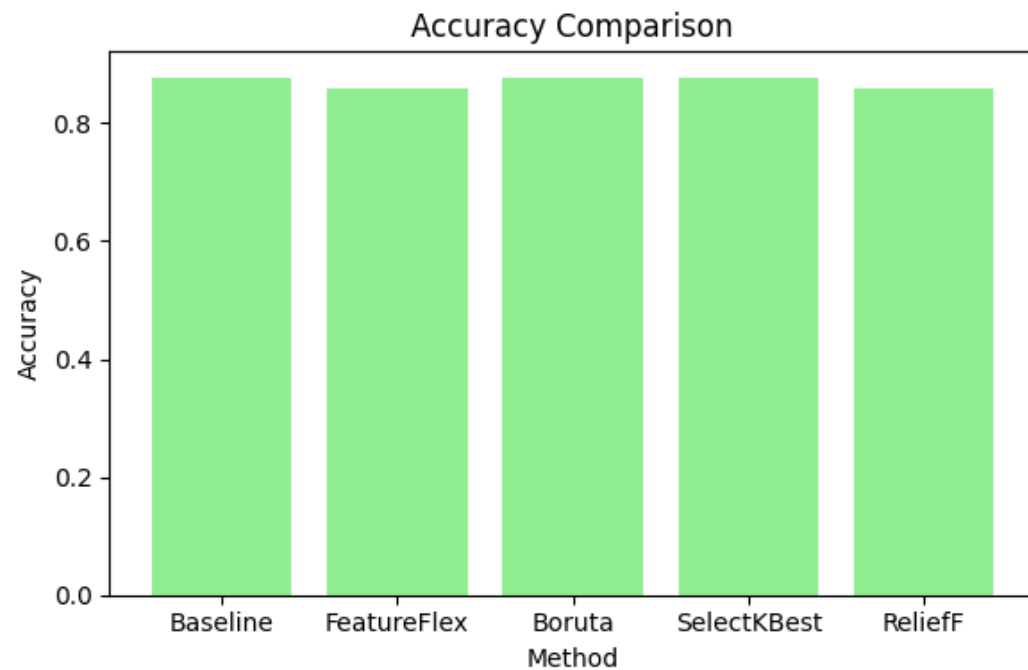
# EWALUACJA WYNIKÓW

Pakiet dostarcza pełny zestaw metryk takich jak **AUC**, **dokładność**, **precyzja**, **czułość** i **F1-score**, a także wizualizacje w postaci macierzy konfuzji i krzywych ROC oraz Precision-Recall.

Raporty mogą być wyświetlane w konsoli lub zapisywane jako pliki HTML z dodatkowymi wykresami (macierz błędów, krzywa ROC, krzywa Precision-Recall).

WYNIKI

# PORÓWNANIE





KONIEC