

# Predykcja Opóźnień Lotów

## Uczenie Maszynowe - Projekt

K. Arkit, D. Chomiak, Ł. Guziczak, D. Sobótka

14 czerwca 2025

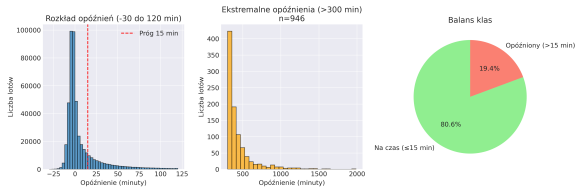
## Definicja problemu

## Cel projektu:

- Klasyfikacja binarna: opóźnienie  $>15$  minut
- Dataset: US Flight Delays 2015
- 5.8M rekordów  $\rightarrow$  500k do analizy
- 31 atrybutów o lotach

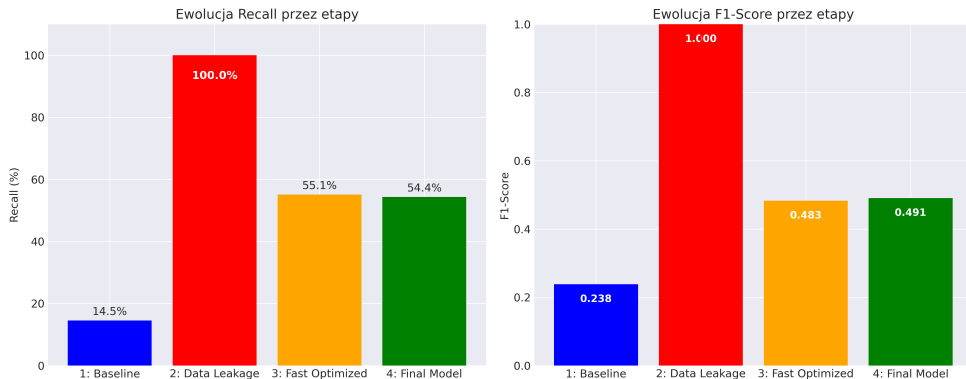
## Dlaczego ten problem?

- Duży, rzeczywisty dataset
- Problem klasyfikacji z niebalansowanymi klasami
- Możliwość testowania różnych technik ML



Rysunek 1. Rozkład opóźnień - 19.4% lotów opóźnionych

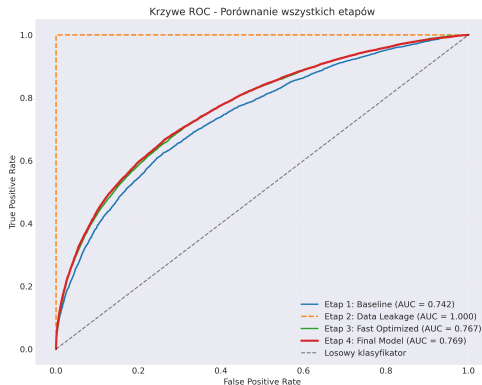
# Ewolucja modelu - 4 etapy



Rysunek 2. Porównanie metryk w kolejnych etapach rozwoju

- ❶ **Baseline:** Podstawowe cechy (12) → niski recall (14.5%)
- ❷ **Data Leakage:** Pokazanie pułapki wycieku danych

# Porównanie modeli - krzywe ROC

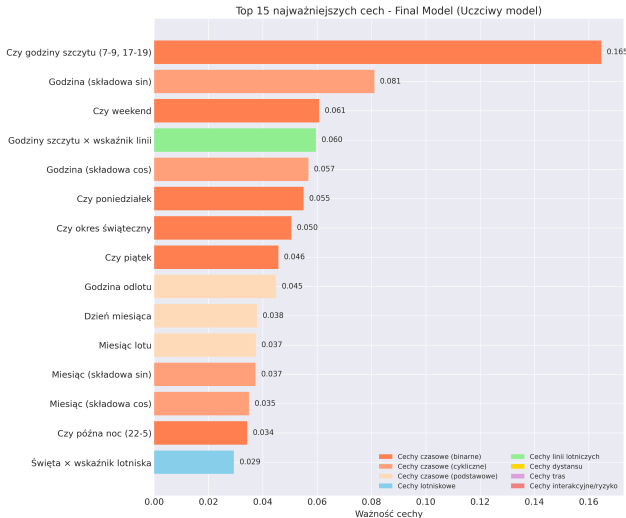


Rysunek 3. Krzywe ROC - ewolucja wydajności (AUC: 0.643  $\rightarrow$  0.769)

## Kluczowe obserwacje:

- Etap 2 (Data Leakage) - perfekcyjny AUC = ostrzeżenie!

# Analiza najważniejszych cech



## Dominują cechy czasowe:

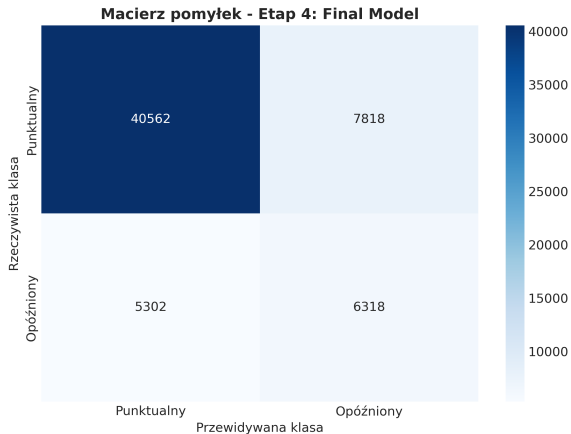
- IS\_RUSH\_HOUR (16.5%)
- HOUR\_SIN/COS (cykliczne kodowanie)
- IS\_WEEKEND, IS\_FRIDAY

## Inne ważne cechy:

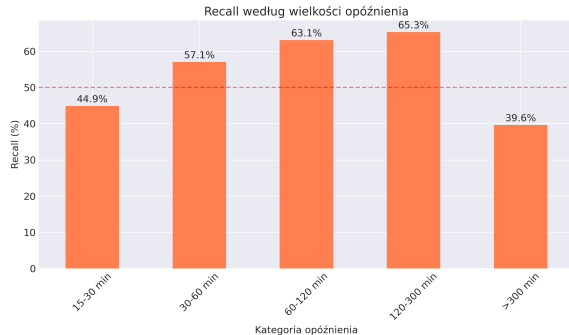
- Linia lotnicza (8.8%)
- Lotnisko wylotu (7.9%)
- Dystans lotu (7.6%)

*Wniosek: Czas odlotu ma największy*

# Wydajność finalna



Rysunek 5. Macierz pomyłek - Final Model



Rysunek 6. Model ma problem z ekstremalnymi opóźnieniami

# Podsumowanie

## Co udało się osiągnąć:

- Przetestowano 4 różne podejścia do modelowania
- Zidentyfikowano i naprawiono problem data leakage
- Feature engineering zwiększył liczbę cech z 12 do 28
- Najlepszy model: XGBoost z ROC-AUC = 0.769

## Napotkane wyzwania:

- Niezbalansowane klasy (19.4% opóźnień)
- Trade-off między recall a precision
- Słaba detekcja ekstremalnych opóźnień (>300 min)
- Brak danych o warunkach pogodowych

**Propozycje ulepszeń:** Model dwuetapowy, dane pogodowe, stacking ensemble, kalibracja prawdopodobieństw

# Dziękujemy za uwagę!

## Pytania?

Kod źródłowy: [GitHub/ML-FlightDelaysAndCancellations](#)