



Adam Piróg | Mateusz Stolarski | dr hab. Piotr Bródka

Katedra Sztucznej Inteligencji

Znajdź swojego influencera

Czyli jak szybko zidentyfikować kluczowy węzeł w sieci

Problem Badawczy

Identyfikacja wpływowych węzłów jest jednym z kluczowych zadań nauki o sieciach, mających jednocześnie praktyczne zastosowanie w wielu problemach biznesowych.



Przykładowo: chcąc przeprowadzić efektywną kampanię reklamową i skutecznie rozpowszechnić informację o naszym produkcie, trzeba zdecydować kogo zatrudnić do współpracy reklamowej, aby uzyskać największe zasięgi.

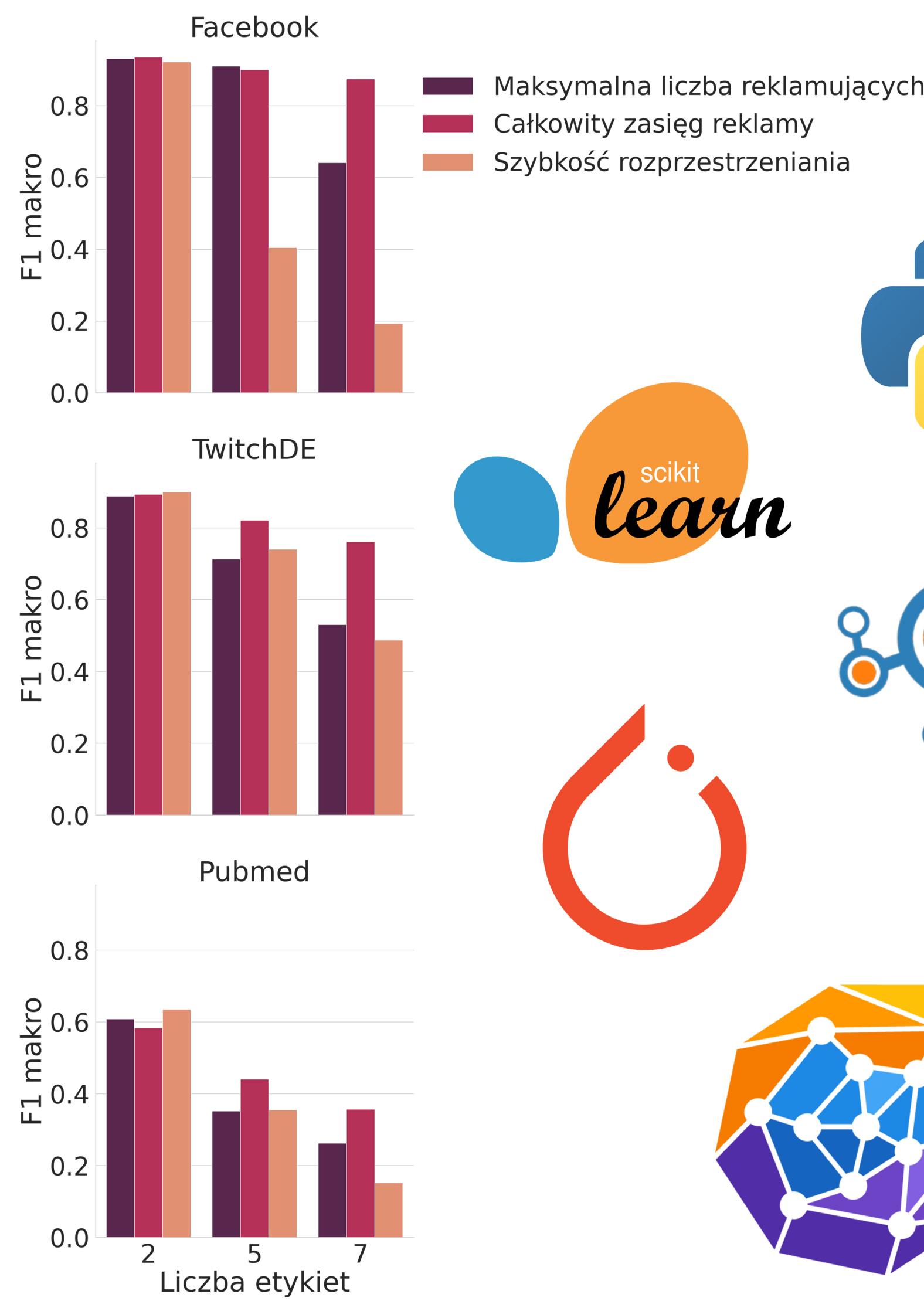
Rozwiązań aktualnie stosowane w tym obszarze opierają się na wielokrotnych symulacjach, co gwarantuje wysoką dokładność, ale jest bardzo kosztowne obliczeniowo.

W ramach naszej pracy wykorzystujemy techniki uczenia maszynowego, do stworzenia systemu oceniającego daną osobę w sieci pod trzema kluczowymi względami:

1. Całkowitego zasięgu, jaki jest w stanie wygenerować (**całkowity zasięg reklamy**)
2. Liczby innych użytkowników, których zachęci do aktywnego reklamowania produktu (**maksymalna liczba reklamujących**)
3. Szybkości, z jaką rozprzestrzenią informacje o produkcie (**szybkość rozprzestrzeniania**)

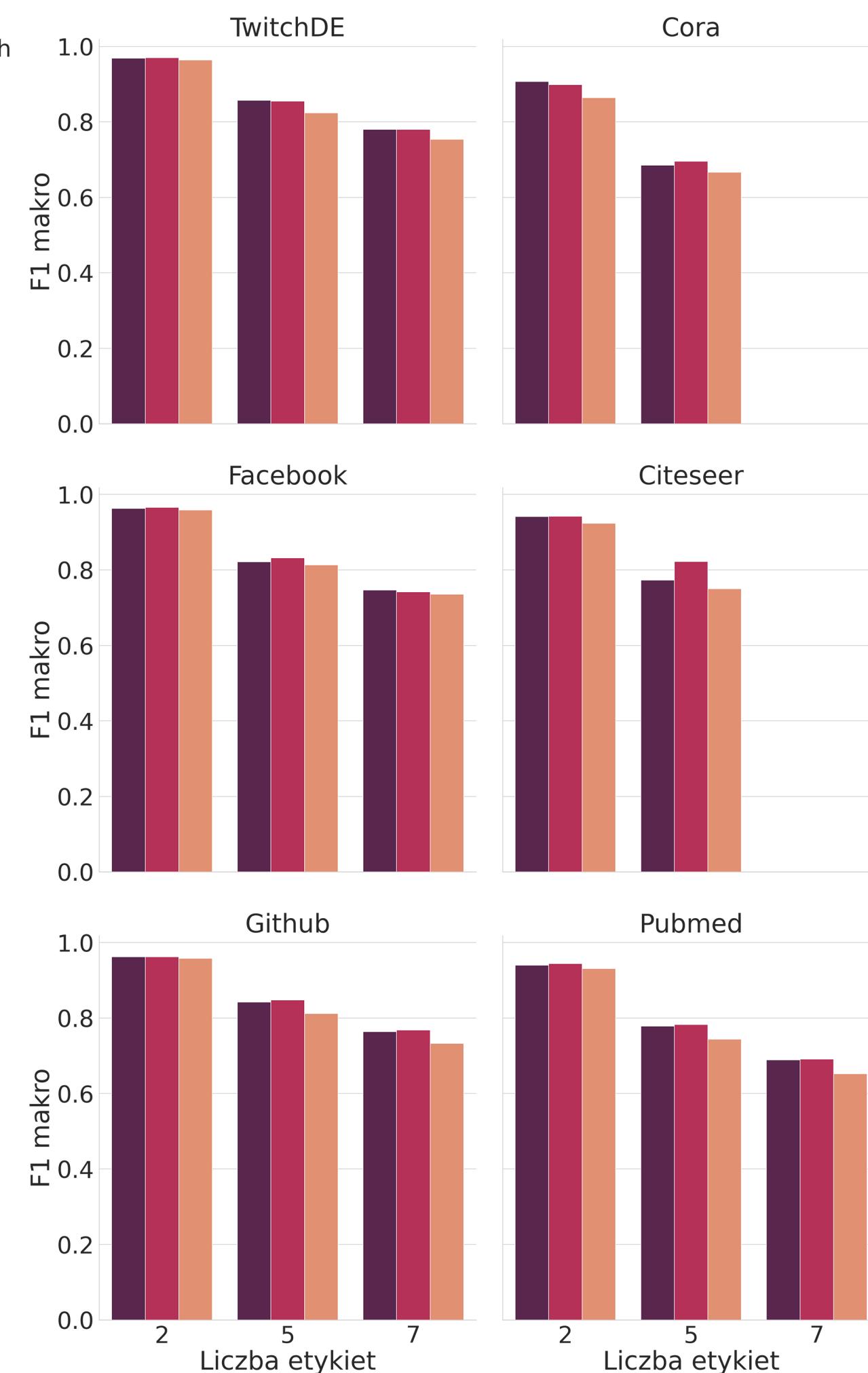
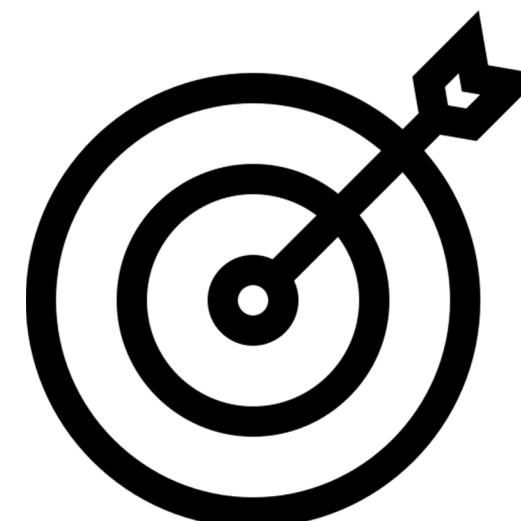
Generalizacja

Model uczyony na sieci Github



Rezultaty

- Maksymalna liczba reklamujących
- Całkowity zasięg reklamy
- Szybkość rozprzestrzeniania



Czas oceny sieci

1.6s
model AI

28.2h
symulacja

Metodologia

Zbiór danych

Sieci społecznościowe z serwisów:

- Twitch (DE) - 9 tys. wierzchołków
- Facebook - 22 tys. wierzchołków
- Github - 37 tys. wierzchołków

Sieci cytowań:

- Citeseer - 3 tys. wierzchołków
- Pubmed - 20 tys. wierzchołków

Modele były testowane zarówno na podzbiorze wierzchołków grafu treningowego, jak i na całkowicie nowych grafach.

Model symulacji

Model niezależnych kaskad - w danej iteracji aktywny wierzchołek ma dokładnie jedną szansę by aktywować każdego ze swoich sąsiadów z danym prawdopodobieństwem *beta*.

Podsumowanie

- Udało się stworzyć model AI, potrafiący ocenić potencjał influencera z dokładnością ~80%
- Dzięki technikom uczenia maszynowego, udało się przyspieszyć proces oceny ponad 60 000 razy
- Projekt jest w trakcie dalszego rozwoju:
 - Wykorzystanie cech indywidualnych wierzchołka
 - Stworzenie zaawansowanych architektur opartych o grafowe sieci neuronowe
 - Zbadanie umiejętności generalizacji dla silnie zróżnicowanych sieci
 - Analiza wpływu poszczególnych miar centralności na uzyskiwane wyniki



Dane wejściowe modelu

Miary centralności każdego wierzchołka:

- closeness
- pagerank
- betweenness
- eigenvector
- clustering coefficient
- core number
- harmonic

Etykiety

1. Symulacja z użyciem modelu niezależnych kaskad*
2. Dyskretyzacja wyników symulacji przy użyciu technik uczenia nienadzorowanego

Modele sztucznej inteligencji

- Support Vector Machine
- Random Forest
- Graph Neural Networks



Fundusze Europejskie
Polska Cyfrowa



Rzeczpospolita
Polska



Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



*Obliczenia wykonano przy użyciu zasobów udostępnionych przez Wrocławskie Centrum Sieciowo-Superkomputerowe (<http://wcss.pl/>)

Projekt współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego.

Program Operacyjny Polska Cyfrowa na lata 2014-2020.

Oś priorytetowa nr 3 „Cyfrowe kompetencje społeczeństwa”, działanie nr 3.2 „Innowacyjne rozwiązania na rzecz aktywizacji cyfrowej”.

Tytuł projektu: „Akademia Innowacyjnych Zastosowań Technologii Cyfrowych (AI Tech)”.