

# WTUM: Przewidywanie "śmieszności" żartu

Maciej Łukasik, Kajetan Larwa, Anton Yushkevich, Maksim Makaranka

March 2022

## 1 Wstęp

Wiele portali internetowych polega na opiniach i ocenach klientów. Analiza opinii jest szczególnie ważna w branży rozrywkowej i e-commerce, gdzie te oceny mają wpływ na zaangażowanie klientów. Na przykład, Netflix opiera się na takich danych ocenach, aby zasilać swój silnik rekomendacji, w celu zapewnienia jak najlepszych rekomendacji filmów i seriali, które są spersonalizowane i najbardziej odpowiednie dla użytkownika. Ten problem praktyczny zmusza uczestników do przewidzenia ocen dowcipów wystawionych przez użytkowników, mając do dyspozycji bazę dowcipów, oceny tego użytkownika pod innymi żartami i oceny pozostałych użytkowników. Analizowany zestaw danych pochodzi ze słynnego zestawu danych internetowego systemu Joke Recommender.

## 2 Technologie i założenia

Postaramy się zbudować model podobny do silnika rekomendacyjnego, który będzie w stanie przewidywać ocenę danego dowcipu na podstawie ocen innych użytkowników oraz na podstawie ocen tego użytkownika pod innymi dowcipami. Obiecującą ścieżką wydaje się próba wyznaczenia pewnej miary podobieństwa między dowcipami i dawanie przewidywanej oceny tym wyższej im wyższe były poprzednie oceny użytkownika pod podobnymi żartami. W celu obliczenia takowej metryki użyte zostałyby techniki NLP. W miarę rozwoju projektu możliwe, że ścieżka ta ulegnie znaczącej modyfikacji lub podejdziemy do problemu od innej strony. Jako język programowania zdecydowaliśmy się wbrać Python 3 wraz z popularnymi bibliotekami używanymi do analizy danych (numpy, pandas). Inną techniką jest użycie algorytmu popularnościowego i jego późniejsze złączenie z poprzednio wspomnianym modelem.

### 3 Podział pracy

- Zaprojektowanie miary podobieństwa żartów - Maciej Łukasik, Kajetan Larwa
- Zbudowanie modelu filtrowania zespołowego, który ma za zadanie wyznaczyć ocenę dowcipu na podstawie ocen danego użytkownika pod podobnymi dowcipami, przy czym podobieństwo dowcipów powinno być obliczane korzystając z miary podobieństwa żartów opracowanej w punkcie wyżej - Maciej Łukasik, Kajetan Larwa
- Zbudowanie modelu popularnościowego, który ma za zadanie wyznaczyć ocenę dowcipu na podstawie ocen wszystkich użytkowników, którzy ocenili ten żart - Anton Yushkevich, Maksim Makaranka
- Zbudowanie modelu, który będzie połączeniem powyższych modeli, czyli ma za zadanie przewidzieć ocenę dowcipu, na podstawie ocen użytkownika, którego ocenę chcemy przewidzieć, pod podobnymi dowcipami oraz ocen tego dowcipu wystawionych przez innych użytkowników - Anton Yushkevich, Maksim Makaranka