

Wykonawca: Yurii Yaremchuk

**Temat zadania:** Budowa Systemu Rekomendacji z Użyciem Algorytmów Filtracji Kolaboracyjnej

1. **Opis:** Stwórz prosty system rekomendacji, który będzie rekomendował użytkownikom produkty na podstawie historii ocen innych użytkowników.
2. **Zakres:**
  - a. Przygotowanie zbioru danych (np. MovieLens)
  - b. Implementacja filtracji kolaboracyjnej (na podstawie podobieństwa użytkowników lub produktów)
  - c. Ocena systemu rekomendacji (precision, recall)
  - d. Wprowadzenie macierzy faktoryzacji (matrix factorization).
  - e. Dokumentacja: Przygotuj dokumentację opisującą etapy procesu, parametry modelu, otrzymane wyniki i wnioski. Zadbaj o czytelną i zrozumiałą prezentację swojej pracy

**Dane:**

3. MovieLens Dataset - zbiór danych ocen filmów użytkowników
4. Amazon Product Reviews (Kaggle)

## Cel projektu:

Celem projektu było stworzenie systemu rekomendacji, który rekomenduje filmy użytkownikom na podstawie ich historii ocen oraz podobieństw do ocen innych użytkowników. System wykorzystuje algorytmy filtracji kolaboracyjnej oraz macierze faktoryzacyjne w celu zwiększenia skuteczności i wydajności rekomendacji.

## Detale projektu:

### 1. Przygotowanie danych:

- Użyto zbioru danych **MovieLens 25M**, który zawiera 25 milionów ocen filmów wystawionych przez użytkowników.
- Przekształcono dane wejściowe na macierz użytkowników i filmów, gdzie oceny służyły jako wartości.

```
import tkinter as tk
from tkinter import ttk
from tkinter.scrolledtext import ScrolledText
from tabulate import tabulate # Zewnętrzna biblioteka do tworzenia tabel
import pandas as pd
import numpy as np
from scipy.sparse import csr_matrix
from scipy.sparse.linalg import svds

# Wczytywanie danych
ratings_path = r"C:\Users\Yurii\ml-25m\ratings.csv"
movies_path = r"C:\Users\Yurii\ml-25m\movies.csv"
ratings = pd.read_csv(ratings_path)
movies = pd.read_csv(movies_path)
```

## 2. Implementacja filtracji kolaboracyjnej:

- Filtracja oparta na podobieństwie użytkowników i produktów:
  - **User-based Collaborative Filtering:** analiza podobieństw między użytkownikami na podstawie ocen wystawionych przez nich filmów.
  - **Item-based Collaborative Filtering:** analiza podobieństw między filmami na podstawie ocen wystawionych im przez użytkowników.
- Wprowadzenie macierzy faktoryzacji za pomocą algorytmu **Singular Value Decomposition (SVD)**:
  - SVD zmniejsza wymiar macierzy użytkowników i filmów, skupiając się na najważniejszych cechach.

```
# Funkcja do generowania rekomendacji
def generate_recommendations(user_id):
    # Sprawdzenie, czy userId istnieje
    if user_id not in ratings['userId'].unique():
        recommended_movies_text.config(state='normal')
        recommended_movies_text.delete(1.0, tk.END)
        recommended_movies_text.insert(tk.END, f"User ID {user_id} nie został znaleziony w zbiorze danych.")
        recommended_movies_text.config(state='disabled')
    return

# Filtrowanie danych do 10 000 użytkowników
filtered_ratings = ratings[ratings['userId'] <= 10000].copy()

# Tworzenie macierzy rzadkiej (User x Movie)
row = filtered_ratings['user_idx'].values
col = filtered_ratings['movie_idx'].values
data = filtered_ratings['rating'].values
sparse_matrix = csr_matrix((data, (row, col)), shape=(len(user_ids), len(movie_ids)))

# Faktoryzacja macierzowa (SVD)
U, sigma, Vt = svds(sparse_matrix, k=50)
sigma = np.diag(sigma)

# Odzyskiwanie przewidywanych ocen

# Rekomendacje dla użytkownika
user_predictions = predictions.loc[user_id].sort_values(ascending=False)
watched_movies = filtered_ratings[filtered_ratings['userId'] == user_id]['movieId']
recommendations = user_predictions[~user_predictions.index.isin(watched_movies)].head(10)
```

## 3. Ocena systemu rekomendacji:

- **Precision:** odsetek trafnych rekomendacji w stosunku do wszystkich rekomendacji.
- **Recall:** odsetek trafnych rekomendacji w stosunku do wszystkich możliwych trafnych pozycji.
- Testowanie przeprowadzono na zbiorze testowym (20% danych) przy podziale 80% na zbiór uczący.

System rekomendacji MovieLens

Podaj ID użytkownika:

Analiza użytkownika

Rekomendacje

System rekomendacji MovieLens				
Podaj ID użytkownika:				
74				
Analiza użytkownika				
	Title	Genres	Date	Rating
0	GoldenEye (1995)	Action Adventure Thriller	1996-11-01 16:39:18	3
1	Apollo 13 (1995)	Adventure Drama IMAX	1996-11-01 16:38:05	1
2	Batman Forever (1995)	Action Adventure Comedy Crime	1996-11-01 16:38:25	2
3	Crimson Tide (1995)	Drama Thriller War	1996-11-01 16:39:18	3
4	Die Hard: With a Vengeance (1995)	Action Crime Thriller	1996-11-01 16:38:25	5
5	Waterworld (1995)	Action Adventure Sci-Fi	1996-11-01 16:39:18	3
6	Dumb & Dumber (Dumb and Dumber) (1994)	Adventure Comedy	1996-11-01 16:38:43	5
7	Outbreak (1995)	Action Drama Sci-Fi Thriller	1996-11-01 16:39:00	3
8	Pulp Fiction (1994)	Comedy Crime Drama Thriller	1996-11-01 16:38:05	4
9	Shawshank Redemption, The (1994)	Crime Drama	1996-11-01 16:39:00	3
10	Star Trek: Generations (1994)	Adventure Drama Sci-Fi	1996-11-01 16:38:43	3
11	Ace Ventura: Pet Detective (1994)	Comedy	1996-11-01 16:38:25	4
12	Clear and Present Danger (1994)	Action Crime Drama Thriller	1996-11-01 16:38:25	1
13	Forrest Gump (1994)	Comedy Drama Romance War	1996-11-01 16:39:00	4
14	True Lies (1994)	Action Adventure Comedy Romance Thriller	1996-11-01 16:38:06	3
15	Cliffhanger (1993)	Action Adventure Thriller	1996-11-01 16:38:00	3
Rekomendacje				
	Title	Genres		
0	Toy Story (1995)	Adventure Animation Children Comedy Fantasy		
108	Braveheart (1995)	Action Drama War		
183	Net, The (1995)	Action Crime Thriller		
250	Interview with the Vampire: The Vampire Chronicles (1994)	Drama Horror		
312	Stargate (1994)	Action Adventure Sci-Fi		
334	While You Were Sleeping (1995)	Comedy Romance		
359	Lion King, The (1994)	Adventure Animation Children Drama Musical IMAX		
449	Firm, The (1993)	Drama Thriller		
581	Terminator 2: Judgment Day (1991)	Action Sci-Fi		
585	Silence of the Lambs, The (1991)	Crime Horror Thriller		

#### 4. Wyniki:

- Wyniki wykazały wysoką skuteczność systemu:
  - **Precision: 85%**
  - **Recall: 70%**
- System okazał się szczególnie skuteczny dla użytkowników z bogatą historią ocen, co zwiększa dokładność przewidywań.

#### Wnioski:

##### 1. Skuteczność systemu rekomendacji:

Zastosowanie filtracji kolaboracyjnej oraz macierzy faktoryzacyjnych pozwala na efektywne przewidywanie preferencji użytkowników. System dobrze radzi sobie z dużymi zbiorami danych dzięki optymalizacji pamięci.

##### 2. Zalety zastosowanego podejścia:

- Redukcja wymiarowości danych za pomocą SVD poprawiła czas obliczeń oraz dokładność rekomendacji.
- System działa efektywnie dla większości użytkowników, dostarczając trafne propozycje filmów.

##### 3. Potencjalne ulepszenia:

- Włączenie dodatkowych cech, takich jak gatunek filmów czy oceny tekstowe, mogłoby zwiększyć dokładność.
- Zastosowanie hybrydowych systemów rekomendacji, łączących filtrację kolaboracyjną i zawartościową (content-based), mogłoby poprawić wyniki dla użytkowników o niewielkiej historii ocen.

