Autorzy:

Karolina Alisz (181227)  
Magdalena Pietrzak (182146)

Krystian Wachowski (190017)

Rekomender gier wideo

Projekt – Usługi Sieciowe

Spis treści

[Wstęp do problematyki 2](#_Toc169862239)

[Rozwinięcie problemu 3](#_Toc169862240)

[Złożoność wyboru i saturacja rynku 3](#_Toc169862241)

[Indywidualne preferencje i subiektywne oczekiwania 3](#_Toc169862242)

[Personalizacja rekomendacji i techniki uczenia maszynowego 3](#_Toc169862243)

[Wyzwania związane z jakością danych i integracją źródeł 3](#_Toc169862244)

[Ewaluacja skuteczności i zaufanie użytkowników 4](#_Toc169862245)

[Metodologia 5](#_Toc169862246)

[Wyniki 8](#_Toc169862247)

[Eksperymenty ALS 8](#_Toc169862248)

[Eksperymenty k-NN 8](#_Toc169862249)

[Dyskusja 8](#_Toc169862250)

# Wstęp do problematyki

Współczesna przemysł rozrywki elektronicznej, w szczególności gry wideo, rozwija się w zawrotnym tempie, oferując milionom użytkowników na całym świecie szeroki wybór tytułów do wyboru. Jednakże, z tak ogromną ilością gier dostępnych na rynku, znalezienie tych, które najlepiej odpowiadają indywidualnym preferencjom graczy, staje się wyzwaniem.

Rekomendera gier wideo można postrzegać jako narzędzie wspomagające decyzje użytkowników w procesie selekcji gier, które są najbardziej zgodne z ich gustami i oczekiwaniami. Takie systemy rekomendacji wykorzystują zaawansowane techniki analizy danych i uczenia maszynowego, aby sugerować gry na podstawie wcześniejszych preferencji użytkowników oraz charakterystyk samych tytułów.

Celem tego projektu jest eksploracja i implementacja rekomendera gier wideo opartego na danych. Projekt ten czerpie inspirację z istniejących rozwiązań oraz korzysta z publicznie dostępnych zbiorów danych dotyczących gier wideo. Analiza takich danych pozwoli na identyfikację wzorców i preferencji wśród użytkowników oraz umożliwi budowę modeli predykcyjnych, które mogą dokładnie przewidywać, jakie gry mogą się spodobać konkretnemu użytkownikowi.

W niniejszym raporcie przedstawimy proces tworzenia oraz ocenę skuteczności stworzonego rekomendera, zaczynając od zbierania danych i ich przetwarzania, przez wybór odpowiednich algorytmów i technik uczenia maszynowego, aż po testowanie i ocenę końcowych wyników. Ostatecznie, naszym celem jest stworzenie narzędzia, które nie tylko pomoże w lepszym odkrywaniu gier wideo, ale także poszerzy zrozumienie technik analizy danych w kontekście rekomendacji personalizowanych produktów.

# Rozwinięcie problemu

Rynek gier wideo, będący dynamicznie rozwijającym się sektorem przemysłu rozrywkowego, charakteryzuje się ogromną różnorodnością zarówno pod względem liczby dostępnych tytułów, jak i ich specyficznych cech, które determinują odbiór przez użytkowników. Zjawisko to, znane jako "problem nadmiaru wyboru", stawia przed graczami wyzwanie w doborze gier, które najlepiej odpowiadają ich indywidualnym preferencjom i oczekiwaniom rozrywkowym.

## Złożoność wyboru i saturacja rynku

Wzrost liczby gier dostępnych na rynku, obejmujący zarówno produkcje o wysokim budżecie, jak i niezależne tytuły, prowadzi do sytuacji, gdzie wybór odpowiedniej gry staje się kwestią niezwykle złożoną. Dla nowych graczy oraz dla doświadczonych entuzjastów, proces selekcji staje się wyzwaniem ze względu na nadmierną ilość opcji, co wymaga efektywnych narzędzi wspierających decyzje zakupowe i wybór gier.

## Indywidualne preferencje i subiektywne oczekiwania

Preferencje graczy są kształtowane przez różnorodne czynniki, takie jak gatunek gry (np. RPG, strzelanka, symulacja), mechanika rozgrywki (np. akcja, strategia, rozwój postaci), fabuła, estetyka graficzna oraz aspekty społeczne (np. współpraca czy rywalizacja). Złożoność tych preferencji sprawia, że jedna metoda oceny i rekomendacji gier nie jest wystarczająca, a konieczne stają się systemy personalizacji rekomendacji.

## Personalizacja rekomendacji i techniki uczenia maszynowego

Tradycyjne metody rekomendacyjne, takie jak filtrowanie kolaboratywne czy systemy oparte na treści, coraz częściej uzupełniane są przez zaawansowane techniki uczenia maszynowego. Takie podejścia analizują szeroki zakres danych behawioralnych i metadanych gier, aby dostosować rekomendacje do indywidualnych potrzeb użytkowników. Przykłady takich danych obejmują oceny użytkowników, historię gier zakupionych lub odtworzonych, czas spędzony na poszczególnych tytułach oraz interakcje społeczne w grach online.

## Wyzwania związane z jakością danych i integracją źródeł

Efektywne działanie systemów rekomendacyjnych wymaga nie tylko zaawansowanych algorytmów uczenia maszynowego, ale także dostępu do wysokiej jakości danych. Kluczowe jest integrowanie różnych źródeł danych, zapewniając jednocześnie ich dokładność i spójność. Wykorzystanie danych pochodzących z różnych platform sprzedaży gier, recenzji, społeczności graczy oraz danych behawioralnych staje się niezbędne do budowy wiarygodnych modeli rekomendacyjnych.

## Ewaluacja skuteczności i zaufanie użytkowników

Istotnym aspektem projektowania rekomenderów gier wideo jest ich ocena pod kątem skuteczności oraz akceptacji przez użytkowników. Skuteczność mierzona jest przez wskaźniki takie jak trafność (precision) i pokrycie (recall) rekomendacji, które powinny być adekwatne do różnych kontekstów i preferencji. Ponadto, kluczowe jest budowanie zaufania użytkowników poprzez transparentność działania systemu oraz odpowiednie zarządzanie ich oczekiwaniami w kontekście rekomendacji.

W kontekście realizacji projektu rekomendacji gier wideo opartego na danych, głównym celem jest nie tylko opracowanie skutecznego narzędzia wspierającego decyzje graczy, ale również wkład w rozwój technologii analizy danych i uczenia maszynowego w kontekście personalizowanych rekomendacji produktów w branży rozrywki elektronicznej. Poprzez integrację zaawansowanych technik analitycznych z głębszym zrozumieniem indywidualnych preferencji użytkowników, możliwe jest stworzenie bardziej satysfakcjonujących i zindywidualizowanych doświadczeń z grami wideo.

# Metodologia

Do realizacji projektu rekomendacji gier wideo użyto danych dostępnych na platformie github[[1]](#footnote-2). Są to dane pochodzące z serwisu Steam.

Użyto pliku "recommendations.csv" z następującymi danymi:

1. app\_id: Unikalny identyfikator aplikacji gry na platformie Steam.
2. helpful: Liczba "helpful" głosów użytkowników na temat tej rekomendacji.
3. funny: Liczba "funny" głosów użytkowników na temat tej rekomendacji.
4. date: Data, kiedy rekomendacja została wystawiona.
5. is\_recommended: Informacja logiczna (true/false), czy użytkownik zarekomendował tę grę.
6. hours: Liczba godzin, które użytkownik spędził grając w tę grę.
7. user\_id: Unikalny identyfikator użytkownika Steam, który wystawił rekomendację.
8. review\_id: Unikalny identyfikator recenzji użytkownika.

W ramach projektu rekomendacji gier wideo, dane z pliku "recommendations.csv" zostały łączone na podstawie identyfikatora aplikacji gry (app\_id) z innymi danymi, takimi jak cechy gry (np. tytuł, data wydania, oceny).

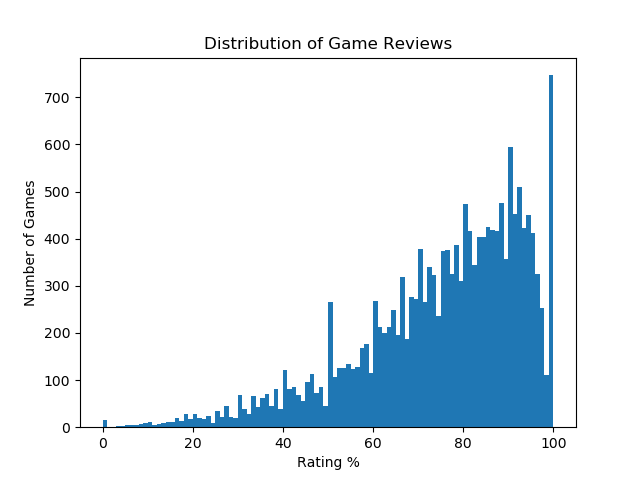
Do realizacji projektu użyto również pliku "games.csv" z następującymi danymi:

1. app\_id: Unikalny identyfikator aplikacji gry na platformie Steam.
2. title: Tytuł gry.
3. date\_release: Data wydania gry.
4. rating: Średnia ocena gry na platformie Steam.
5. positive\_ratio: Stosunek pozytywnych recenzji do wszystkich recenzji gry na Steamie.
6. user\_reviews: Liczba recenzji użytkowników gry na Steamie.
7. price\_final: Aktualna cena gry (w lokalnej walucie) na platformie Steam.
8. price\_original: Oryginalna cena gry przed zastosowaniem ewentualnej obniżki cenowej.
9. discount: Procentowa obniżka ceny w stosunku do oryginalnej ceny.
10. steam\_deck: wskazują, czy gra jest zgodna z urządzeniem Steam Deck.

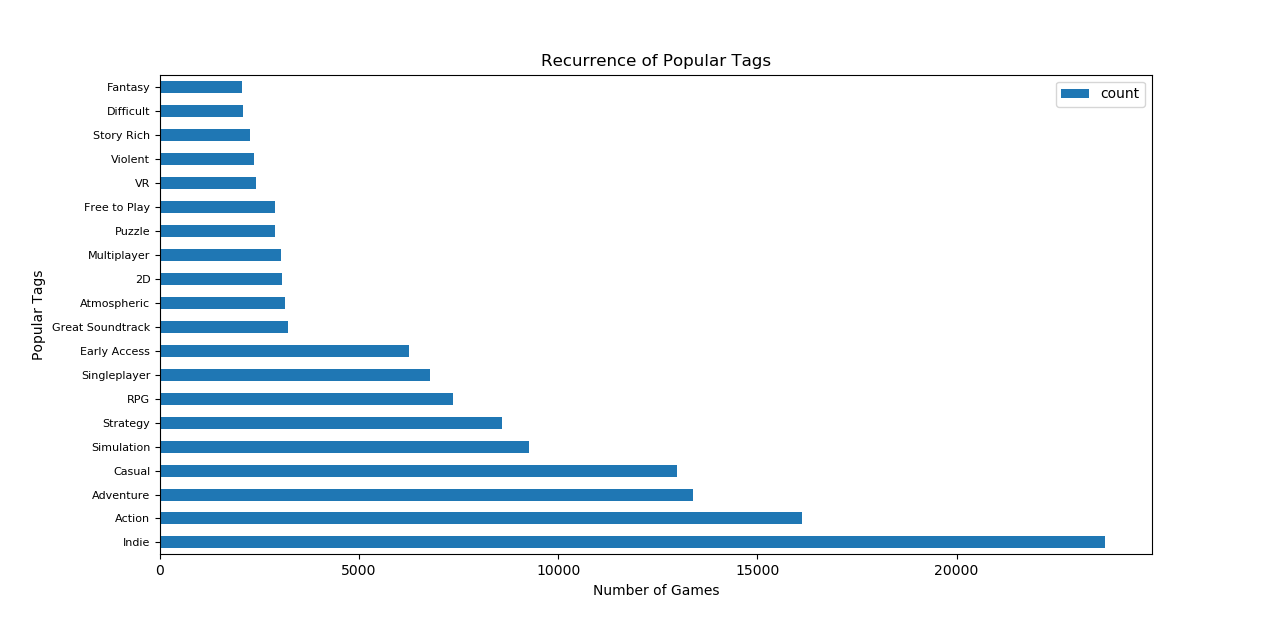
Plik "games.csv" zawierał informacje o grach, w tym tytuły oraz unikalne identyfikatory aplikacji (app\_id). W ramach projektu rekomendacji gier wideo, dane z tego pliku zostały łączone na podstawie identyfikatorów z innymi szczegółowymi informacjami dotyczącymi poszczególnych gier. Dzięki tej integracji danych możliwe było kompleksowe analizowanie i modelowanie cech oraz preferencji gier, co umożliwiło skuteczniejsze tworzenie spersonalizowanych rekomendacji dla użytkowników platformy Steam.

Do realizacji modelu rekomendacyjnego w projekcie wykorzystano model ALS (Alternating Least Squares). Model ALS jest popularną techniką w kontekście filtrowania kolaboratywnego, która jest efektywna w rekomendowaniu przedmiotów na podstawie macierzy ocen użytkowników i przedmiotów. W przypadku tego projektu, model ALS został zastosowany do analizy danych rekomendacyjnych użytkowników gier wideo na platformie Steam, co umożliwiło generowanie spersonalizowanych rekomendacji gier dla poszczególnych użytkowników na podstawie ich wcześniejszych preferencji i ocen. Model ALS działa poprzez iteracyjne optymalizowanie dwóch macierzy: jednej reprezentującej preferencje użytkowników i drugiej reprezentującej cechy przedmiotów (w tym przypadku gier), co prowadzi do lepszej predykcji ocen i preferencji użytkowników dla nowych przedmiotów.

Aby dokładniej zanalizować rozkład recenzji gier, stworzono wykres ilustrujący liczbę gier w zależności od procentowego udziału pozytywnych recenzji każdej z gier. Ten podejście pozwala nam lepiej zrozumieć, jak różne gry oceniane są przez użytkowników i jakie są tendencje w postrzeganiu tych gier na podstawie ich recenzji. Analiza ta jest kluczowa dla oceny popularności i opinii na temat poszczególnych tytułów w naszym zbiorze danych.



Poniższy wykres przedstawia wszystkie dostępne gatunki gier w zbiorze danych gier wraz z odpowiadającą im liczbą gier. Ten wykres jest przydatny, ponieważ umożliwia nam zrozumienie dystrybucji gatunków gier w zbiorze danych, co pozwala na analizę popularności poszczególnych typów gier oraz identyfikację dominujących trendów i preferencji wśród użytkowników. Dzięki tej analizie możemy lepiej dostosować nasze podejście do budowy rekomendacji gier oraz lepiej zrozumieć, jak różnorodne gatunki gier są reprezentowane w dostępnym zbiorze danych.



Najwięcej gier w zbiorze danych należy do gatunku "indie", natomiast najmniej reprezentowanym gatunkiem jest fantasy. Takie informacje są istotne przy analizie dystrybucji gatunków gier i pomagają zrozumieć ich różnorodność oraz popularność w zbiorze danych.

# Wyniki

Plik data manipulation, w którym wczytujemy dane z recommendations.csv i games.csv. Usuwamy niepotrzebne kolumny z obu dataframe. Łączymy dwa dataframe przez kolumnę app.id tak żeby stworzyć jeden dataframe z użytkownikiem, tytułem i godzinami gry.

Tworzymy tablicę przestawną, w której kolumnami są użytkownicy, indeksami są tytułu gier, wartościami są godziny. Następnie wypełniamy puste pola w tablicy zerami. Wszystko zapisujemy w formacie .pickle dla łatwego odtworzenia wyników.

Obraz zawierający zrzut ekranu, tekst

Opis wygenerowany automatycznie

Do stworzenia tabeli przestawnej wybieramy próbkę x użytkowników, gdzie x = 100 000. Posługujemy się komendami „sample”, „merge” i „pivot\_table” z biblioteki pandas.

## Eksperymenty ALS

Do eksperymenty als wczytujemy plik .pickle, tworzymy z niego macierz za pomocą komendy „csr\_matrix”. Dane zostają skompresowane, zera zostają usunięte. Tabela przestawna przed kompresją zajmuje około 7 GB dla 100 000 graczy, a po kompresji 1,17 MB. Matrix\_gen generuje nam macierz z parametrami pivot\_file i matrix\_file, gdzie pivot\_file jest nazwą tabeli przestawnej (format string), a matrix\_file jest nową nazwą stworzonej macierzy. Matrix\_load z parametrem matrix\_file wczytuje nam macierz z pliku matrix\_file. Wynikiem tej funkcji jest macierz csr. Model\_gen z parametrami matrix\_file i model\_file gdzie matrix\_file jest nazwą pliku z macierzą csr, a model\_file będzie nazwą stworzonego modelu. Generuje model ALS z parametrami factors=20, regularization=0.1, iterations=20. Model\_load z paramterem model\_file, który wczytuje model o nazwie podanej w parametrze model\_file. Wynikiem tej funkcji jest model ALS.

# Dyskusja

Metoda ALS, będąca jedną z technik filtrowania kolaboratywnego, okazała się skuteczna w identyfikowaniu wzorców preferencji użytkowników na podstawie macierzy ocen. Dzięki temu byliśmy w stanie rekomendować gry na podstawie wspólnych zainteresowań użytkowników. Jednakże, metoda ta ma pewne ograniczenia związane z koniecznością dostrojenia hiperparametrów oraz potencjalnymi problemami ze skalowalnością w przypadku bardzo dużych zbiorów danych.

Na podstawie przeprowadzonych analiz i porównań, nasze przyszłe działania będą skoncentrowane na kilku kluczowych obszarach:

* Optymalizacja Modeli: Planujemy dalszą optymalizację zarówno modelu ALS, jak i kNN, aby jeszcze bardziej poprawić dokładność rekomendacji. W przypadku ALS, szczególną uwagę poświęcimy dostrajaniu hiperparametrów, natomiast w przypadku kNN – efektywnemu zarządzaniu odległościami i wyborem k najbliższych sąsiadów.
* Integracja Hybrydowego Systemu: Biorąc pod uwagę zalety obu metod, rozważamy integrację hybrydowego systemu rekomendacyjnego, który łączy podejście kolaboratywne z filtrowaniem opartym na treści. Taki system mógłby wykorzystać mocne strony obu metod, prowadząc do jeszcze lepszych wyników.
* Rozszerzenie Zbiorów Danych: Planujemy wzbogacić nasze zbiory danych o dodatkowe informacje, takie jak dane demograficzne użytkowników, ich aktywność na platformie oraz szczegółowe opinie i komentarze. Dzięki temu będziemy w stanie dostarczać jeszcze bardziej spersonalizowane rekomendacje.
* Ewaluacja i Testy: Przeprowadzimy szeroko zakrojone testy i ewaluacje naszych modeli na różnych podzbiorach danych, aby ocenić ich skuteczność i niezawodność. Skupimy się na mierzeniu dokładności, precyzji i trafności rekomendacji w celu ciągłego doskonalenia naszych algorytmów.
* Interaktywność Systemu: Pracujemy nad wdrożeniem mechanizmów umożliwiających użytkownikom interakcję z systemem rekomendacji, takich jak oceny, opinie zwrotne i możliwość dostosowywania preferencji. Pozwoli to na dynamiczne dostosowywanie rekomendacji do zmieniających się gustów użytkowników.

Podsumowując, równoległe zastosowanie metody ALS i kNN oraz dalsza praca nad ich optymalizacją i integracją w hybrydowym systemie rekomendacyjnym stanowić będą kluczowe elementy naszych przyszłych działań. Dzięki temu będziemy mogli dostarczać użytkownikom bardziej trafne i spersonalizowane rekomendacje gier wideo, co przyczyni się do zwiększenia ich satysfakcji i zaangażowania.

1. <https://github.com/AudreyGermain/Game-Recommendation-System/tree/master> [↑](#footnote-ref-2)