Adrian Dybowski 77327

<https://github.com/adrian98dybowski/MasterThesisSGH>

Szanowny Panie Doktorze,

Poniższy dokument napisałem, aby przybliżyć Panu moją dotychczasową pracę nad pracą magisterską (temat, koncept, wykorzystane dane, techniki i języki programowania).

Celem mojej pracy magisterskiej jest stworzenie aplikacji, pozwalającej użytkownikowi wybrać filmu do obejrzenia na podstawie ostatniego filmu, który obejrzał (użytkownik wybiera z listy film, który ostatnio oglądał i po chwili otrzymuje listę 10 filmów polecanych mu do obejrzenia).

Dane (filmy i oceny użytkowników) pochodzą z kaggle (<https://www.kaggle.com/shubhammehta21/movie-lens-small-latest-dataset>) – zbiór liczy około 10 000 filmów i ponad 100 000 ocen różnych użytkowników. W dalszym etapie zostały usunięte filmy z małą ilością ocen i użytkownicy, którzy ocenili mało filmów.

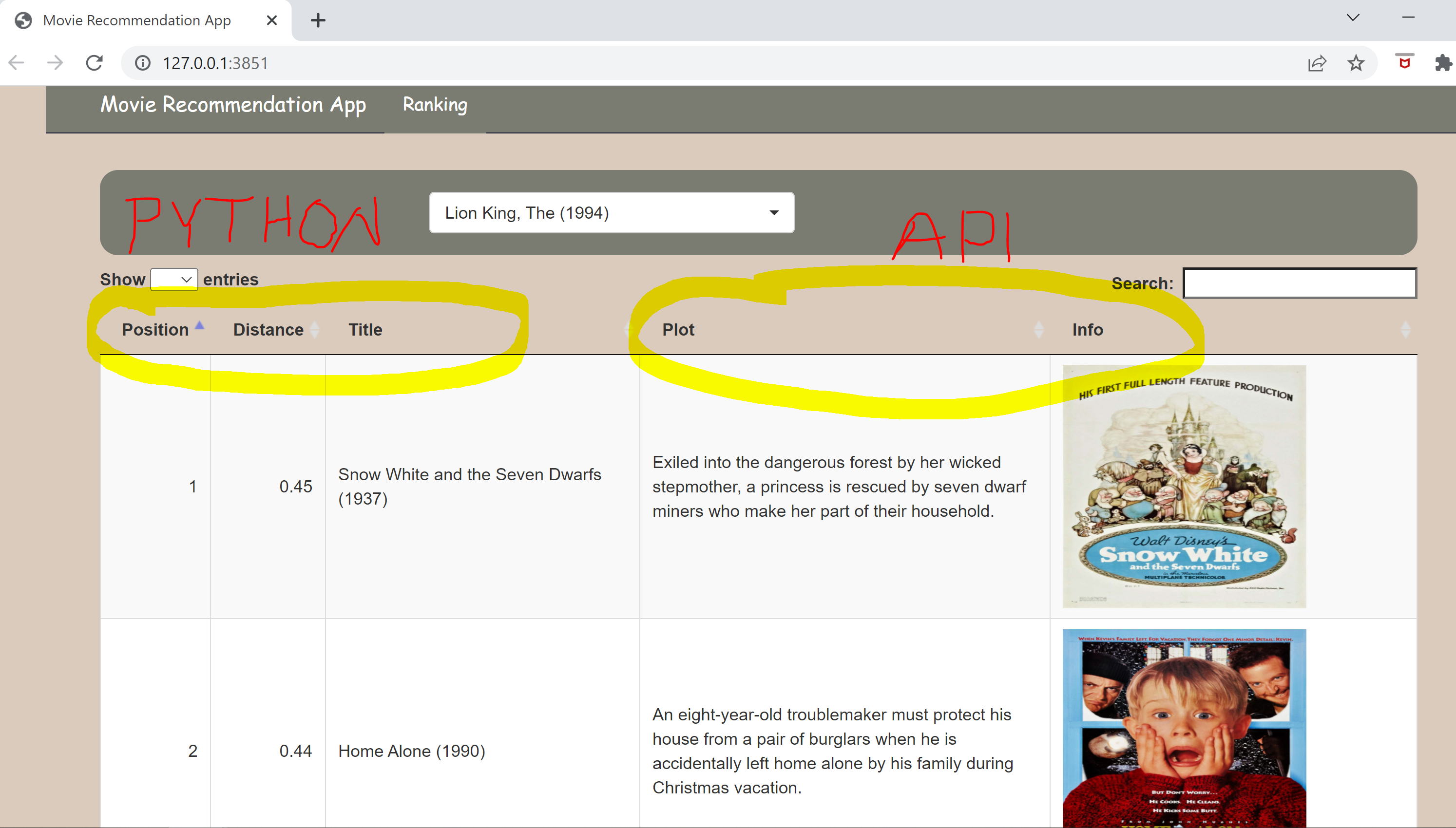
Uogólniając, system rekomendacji bazuje na metodzie najbliższych sąsiadów. W skrócie, analizuje on użytkowników na zasadzie podobnych preferencji. Przykładowo, jeśli użytkownik X obejrzał filmy A, B i C, to użytkownikowi Y, który także obejrzał filmy A i B, system zarekomenduje obejrzenie filmu C (w dużym uproszczeniu).

System rekomendacji został napisany w języku python (wykorzystując odpowiednie biblioteki).

Następnie przeniosłem pracę do R, a dokładniej do R shiny, aby zwizualizować otrzymane wyniki. Za pomocą odpowiednich bibliotek wczytałem plik .py do R, a potem wykonałem odpowiednie działania do stworzenia aplikacji (pliki ui, tabs, global, server).

Na koniec wykorzystałem jedno z darmowych API internetowych ze strony rapidapi o filmach, aby pobrać z niego krótki opis i plakat do każdego z 10 filmów wyświetlanych w aplikacji (<https://rapidapi.com/hmerritt/api/imdb-internet-movie-database-unofficial/> ).

Finalnie aplikacja wygląda tak, jak na poniższym zdjęciu. Pierwsze kolumny, to wynik analizy z pliku pythoowego, a pozostałe dwie, to dołączenie informacji z API.



Taki jest mój dotychczasowy postęp nad pracą magisterską. Zacząłem od części technicznej, czyli przygotowania aplikacji. Dlaczego taki temat? Dość długo myślałem nad tematem pracy. Próbowałem wymyślić coś z analizą danych w czasie rzeczywistym, którą Pan mi zaproponował podczas ostatniej rozmowy na Teamsach, oraz z wizualizacją danych w R shiny, jednak ne potrafiłem nic dopasować. W międzyczasie zainteresował mnie ten temat rekomendacji filmów do obejrzenia (po przeczytaniu jednego z artykułów) co udało mi się połączyć z zwizualizowaniem tych wyników w prostej aplikacji.

Podsumowując, wykorzystałem dane z Kaggle i dane z API internetowego oraz języki Python i R (szczególnie R Shiny) do pracy nad danymi. Dla lepszego wyglądu aplikacji (kolory, czcionka, detale) posłużył CSS. Python był mi potrzebny do przygotowania systemu rekomendacji filmów. Nie potrafiłem tego zrobić bezpośrednio w R, do przygotowania w Pythonie było więcej przydatych inofrmacji w internecie, jak to zrobić.

Tak więc, taka jest moja dotychczasowa praca. Będę wdzięczny, jeśli w wolnej chwili przeanalizuje Pan mój postęp i poda wszelkie uwagi i wątpliwości, rzeczy do poprawy i ewentualnie jakie dalsze kroki podjąć. Wszystkie kody i dane znajdują się na moim githubie w repozytorium dotyczącym pracy magisterskiej:

<https://github.com/adrian98dybowski/MasterThesisSGH>

W razie pytań, jestem do dyspozycji na Teamsach lub mailowo. Jeśli będzie taka potrzeba, jestem skłonny umówić się na spotkanie na Teamsach (nie mogę stawić się osobiście, ponieważ przebywam na wymianie zagranicznej).

Pozdrawiam,

Adrian Dybowski