

Projekt IUM – 2023L

v.1.0

Kontekst

W ramach projektu wcielamy się w rolę analityka pracującego dla portalu „Pozytywka” – serwisu muzycznego, który swoim użytkownikom pozwala na odtwarzanie ulubionych utworów online. Praca na tym stanowisku nie jest łatwa – zadanie dostajemy w formie enigmatycznego opisu i to do nas należy doprecyzowanie szczegółów tak, aby dało się je zrealizować. To oczywiście wymaga zrozumienia problemu, przeanalizowania danych, czasami negocjacji z szefostwem. Same modele musimy skonstruować tak, aby gotowe były do wdrożenia produkcyjnego – pamiętając, że w przyszłości będą pojawiać się kolejne ich wersje, z którymi będziemy eksperymentować.

Jak każda szanująca się firma internetowa, Pozytywka zbiera dane dotyczące swojej działalności – są to (analitycy mogą wnioskować o dostęp do tych informacji na potrzeby realizacji zadania):

- lista dostępnych artystów i utworów muzycznych,
- baza użytkowników,
- historia sesji użytkowników,
- techniczne informacje dot. poziomu cache dla poszczególnych utworów.

Zadania

Prowadzący - Paweł Zawistowski

pawel.zawistowski@pw.edu.pl konsultacje <https://calendly.com/pawel-zawistowski>

1. “Większość serwisów udostępniających muzykę czy filmiki, poleca coś swoim użytkownikom. Przyszedł czas, abyśmy zaczęli robić tak samo”
2. “Nasi użytkownicy chcą słyszeć muzykę możliwie najszybciej po uruchomieniu odtwarzania utworu. Nie możemy jednak całej muzyki utrzymywać w naszym najszybszym cache’u – jest to zbyt drogie. Pracujemy nad rozwiązaniem, które każdego wieczora będzie rozmieszczało muzykę (na następne 24h) pomiędzy różnymi poziomami cache – tak, aby zminimalizować koszty. Pytanie: skąd mamy wiedzieć, gdzie dany utwór umieścić?”
3. “Może bylibyśmy w stanie wygenerować playlistę, która spodoba się kilku wybranym osobom jednocześnie? Coraz więcej osób używa Pozytywki podczas różnego rodzaju imprez i taka funkcjonalność byłaby hitem!”
4. “Mamy pomysł na świetne narzędzie dla naszych użytkowników – zaczniemy generować listę top utworów, ale z wyprzedzeniem tygodniowym! Będziemy szybsi od konkurencji i zwiększymy zaangażowanie.”
5. “Fajnie byłoby rozszerzyć nasz serwis o generowanie popularnych playlist – zestawów pasujących do siebie utworów, których słuchaniem zainteresowane będzie wiele osób”

Prowadzący - Łukasz Neumann
lukasz.neumann@pw.edu.pl

6. “Nasz serwis zarabia wtedy, gdy ludzie słuchają muzyki. Jeśli nie wiemy ile czasu będą słuchali jej w przyszłości, to trudno nam rozliczać się z artystami i negocjować z nimi stawki”
7. “Gdybyśmy tylko wiedzieli, kiedy użytkownik będzie chciał przesłuchać bieżący utwór w całości, a kiedy go przewinie – moglibyśmy lepiej zorganizować nasz cache”
8. “Nasza baza utworów jest całkiem bogata – opisywane są one wieloma ciekawymi parametrami. Dlaczego jednak do tej pory nikt nie oznaczył czy są w skali dur czy mol? Musimy zmienić tę sytuację!”
9. “Nie wszyscy nowo dodawani wykonawcy do naszej bazy mają przypisany gatunek muzyczny – musimy jakoś temu zaradzić!”
10. “Jakiś czas temu wprowadziliśmy konta premium, które uwalniają użytkowników od słuchania reklam. Nie są one jednak jeszcze zbyt popularne – czy możemy się dowiedzieć, które osoby są bardziej skłonne do zakupu takiego konta?”

Zasady

- 1) Projekt **realizujemy w parach** – każda para otrzymuje do realizacji jedno z zadań z listy powyżej.
- 2) W trakcie trwania projektu prowadzący pełni rolę klienta, dla którego realizujecie Państwo zadanie biznesowe.
- 3) **Zgłoszenia do zadań realizujemy poprzez serwer do zgłoszeń:**
 - a) Adres: http://eve.ii.pw.edu.pl:9009/select_course/
 - b) Przedmiot: IUM
 - c) Dwa dni po zgłoszeniu zespołu proszę upewnić się, że projekt został Państwu przydzielony (na wypadek, gdyby było więcej niż jeden zespół do konkretnego tematu).
- 4) W ramach projektu trzeba dostarczyć:
 - a) etap 1 – możliwe, że będzie konieczne wykonanie więcej niż jednej iteracji tego etapu – pierwsza powinna zakończyć się do **2023.04.07 (0-20 pkt)**:
 - i) definicję problemu biznesowego, zdefiniowanie zadania/zadań modelowania i wszystkich założeń, zaproponowania kryteriów sukcesu),
 - ii) analizę danych z perspektywy realizacji tych zadań (trzeba ocenić, czy dostarczone dane są wystarczające – może czegoś brakuje, może coś trzeba poprawić, domagać się innych danych, ...),
 - b) etap 2 – do **2023.06.02** (z zastrzeżeniem, że aby możliwa była dodatkowa iteracja poprawiająca finalną punktację – rozwiązanie trzeba wysłać wcześniej do **2023.05.19**):

- i) dwa modele: model bazowy (najprostszy możliwy dla danego zadania) i bardziej zaawansowany model docelowy, oraz raport z pokazujący proces budowy modelu i porównujący wyniki **(0-15 pkt)**,
 - ii) implementację aplikacji (w formie mikroserwisu), która pozwala na **(0-15 pkt)**:
 - (1) serwowanie predykcji przy pomocy danego modelu,
 - (2) realizację eksperymentu A/B – w ramach którego porównywane będą oba modele i zbierane dane niezbędne do późniejszej oceny ich jakości,
 - iii) materiały pokazujące, że implementacja działa.
- 5) “Oddanie” etapu składa się z dwóch kroków:
- a) wysyłamy rozwiązanie mailem do Prowadzącego:
 - i) raporty/dokumentację w formie PDF lub **notatników [jupyter](#)** (do których **zachęcamy**),
 - ii) implementację – w pliku **zip**,
 - iii) proszę nie wysyłać formatu **rar** – serwer PW często oznacza takie wiadomości jako spam.
- 6) Zachęcam do korzystania **konsultacji** w przypadku jakichkolwiek wątpliwości – najpóźniej do **2023.05.26r** – ostatni tydzień przed terminem traktujemy jako „finalny sprint”, podczas którego wszystkie ustalenia są już zamrożone.
- 7) Niedotrzymanie terminów (czyli niewysłanie rozwiązania do Prowadzącego –konsultacje mogą być później) skutkuje karą za spóźnienia: -2 pkt za każdą rozpoczętą dobę.