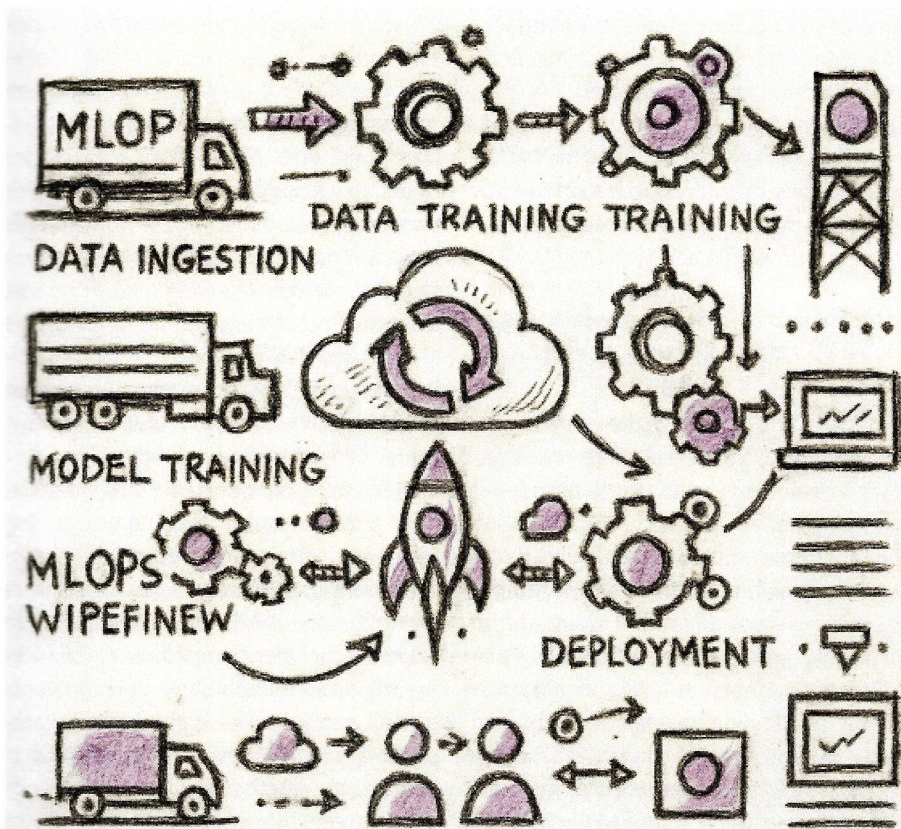


System rekomendacji w młynie (cz. 2)

Rekomendacje w chmurze AWS – Amazon Personalize

Amazon Personalize to zarządzana usługa, która na podstawie danych o interakcjach użytkowników z produktami generuje spersonalizowane rekomendacje. Integruje się z innymi usługami AWS – np. S3 do przechowywania danych, AWS Glue/Athena do przygotowania danych, czy DynamoDB i Lambda/ECS do udostępniania wyników w aplikacji.



Schemat referencyjny personalizacji na AWS (Retail Personalization).

Amazon Personalize automatyzuje trenowanie modeli ML na danych dostarczonych przez klienta i udostępnia wyniki poprzez proste API. Jej zaletą jest to, że nie musimy sami implementować algorytmów – wystarczy odpowiednio załadować dane (historię interakcji, katalog produktów, ewentualnie profile użytkowników), a usługa sama zajmie się uczeniem modelu i udostępnieniem rekomendacji. Amazon Personalize wywodzi się z doświadczeń tworzenia systemu rekomendacji w sklepie Amazon.com i udostępnia m.in. następujące przykłady praktyczne: *Recommended for you* (personalizacja ogólna), *Customers*

who bought X also bought Y (rekomendacja produktów komplementarnych), *More like this* (podobne do oglądane-go) oraz listy trendów czy segmentację użytkowników.

Jak działa Amazon Personalize?

Amazon Personalize wykorzystuje głównie dane o interakcjach użytkowników z produktami – np. przeglądanie, dodawanie do koszyka, zakup, ocenianie itp. (to tzw. *item interaction data*). Te dane mogą pochodzić zarówno z historycznych logów (wsadowo), jak i być wysyłane w trybie real-time (np. z aplikacji poprzez API). Dodatkowo można

dostarczyć metadane produktów (katalog: np. kategorie, ceny) i metadane użytkowników (np. wiek, płeć, segment lojalnościowy), które niektóre algorytmy Personalize także potrafią wykorzystać w modelowaniu. W odróżnieniu od podejścia Azure, tutaj nie musimy wybierać algorytmu ML bezpośrednio – zamiast tego wybieramy tzw. *recipe* lub *use case*, czyli predefiniowany schemat modelowania:

- *Recipe User-Personalization* (w domyślnym use-case Retail) trenuje model uwzględniający sekwencje interakcji użytkowników (bazujący na sieciach neuronowych RNN),
- *Recipe Similar-Items* tworzy model podobieństwa produktów na podstawie współwystępowania,
- *Recipe Personalized* – ranking do rerankingu listy itemów dla konkretnego usera itp.

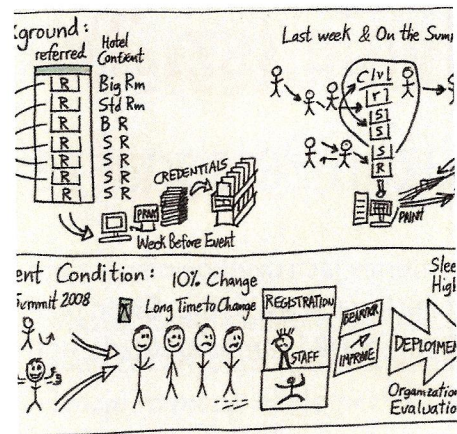
Możliwe jest również stworzenie tzw. custom solution z własną konfiguracją, ale zwykle korzysta się z gotowych receptur dopasowanych do scenariusza biznesowego.

Przykład użycia w sklepie internetowym z mąką

Krok pierwszy: tworzenie danych

Aby użyć Amazon Personalize, najpierw definiujemy schematy danych i tworzymy tzw. Dataset Group w serwisie. Potrzebne będą trzy zbiory danych.

- Historia interakcji użytkowników z produktami (*interactions dataset*). W najprostszym ujęciu jest to tabela (CSV/JSON) z kolumnami: *user_id*, *item_id*, *timestamp* oraz opcjonalnie *event_type* i *event_value*. Dla sklepu



z mąką wydarzeniami mogą być np. *Purchase* (zakup), *View* (obejrzenie strony produktu), *AddToCart* itp. Każdy wiersz to jedna akcja użytkownika na konkretnym produkcie w określonym czasie. Wymagane jest posiadanie wystarczającej ilości takich danych (AWS zaleca >1000 użytkowników i >50k interakcji dla sensownych rezultatów). Nasza baza transakcji sprzedaży dostarczy część tych interakcji (zakupy), ale warto by też uwzględnić inne zachowania (jeśli są dostępne) – choć w hurtowni spożywczej głównym sygnałem będą zakupy. Interakcje te importujemy do Personalize – najczęściej poprzez załadowanie pliku do pojemnika na dane S3, a następnie wskazanie go usłudze importującej dane do datasetu.

- Katalog produktów (*items dataset*). To opcjonalny zbiór zawierający informacje o produktach. Dla naszego sklepu z mąkami moglibyśmy tu umieścić atrybuty jak typ zboża, kraj pochodzenia, czy produkt jest bezglutenowy, itp. W Amazon Personalize schemat itemów wymaga unikalnego *ITEM_ID* i ewentualnie pól metadanych (stringi, liczby lub kategorie). Te informacje mogą pomóc modelowi lepiej zrozumieć relacje między produktami (np. rekomendować produkty z tej samej kategorii) i lepiej radzić sobie z tzw. cold start (wprowadzeniem nowego produktu bez wielu interakcji).

- Zbiór z danymi statycznymi użytkowników (*users dataset*): również opcjonalny. Tutaj moglibyśmy umieścić np. wiek, płeć klienta, rodzaj prowadzonej działalności czy kraj zamieszkania. Amazon Personalize pozwala definiować kilka takich atrybutów użytkownika, które mogą być uwzględnione

przez algorytm (np. segmentacja). Nie wszystkie recipes wykorzystują te dane, ale np. do tworzenia segmentów użytkowników mogą się one przydać. W naszym przypadku te dane statyczne są w bazie klientów, więc możemy je zaimportować. Warto jednak zaznaczyć, że zaawansowane, niestandardowe cechy typu „skłonność do klikania banerów” nie mają od razu oczywistego miejsca w tym schemacie – moglibyśmy je zakodować np. jako liczbowy atrybut użytkownika (o określonej skali) i zaimportować do Users dataset. Amazon Personalize prawdopodobnie potraktuje je jak każdą inną cechę kategorii lub liczbową i potencjalnie użyje podczas trenowania modelu.

Uwaga: główny nacisk w recipes AWS jest położony na interakcji użytkowników z produktami. Oczywiście możemy je odczytać z logów aplikacyjnych.

Krok drugi: wybór odpowiedniego AWS recipe i uruchomienie treningu

Kiedy dane są przygotowane i załadowane do S3 Personalize, tworzymy tzw. Solution i wybieramy algorytm (recipe).

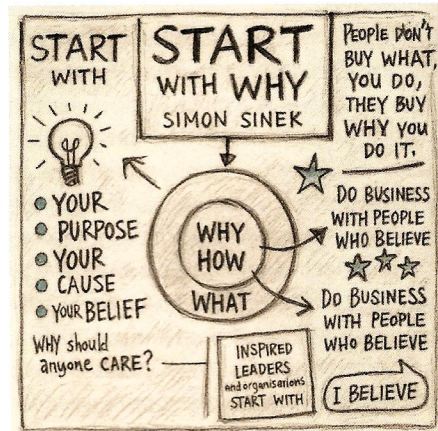
Dla spersonalizowanych rekomendacji ogólnych wybierzemy np. recipe *aws-user-personalization*. Następnie inicjujemy trenowanie modelu – Personalize automatycznie skalibruje model na podstawie dostarczonych danych. Można podać hiperparametry, ale nie jest to wymagane, można działać na domyślnych ustawieniach. Trening trwa od kilkunastu minut do kilku godzin. Po ukończeniu otrzymujemy Solution Version – wytrenowany model, który możemy wykorzystać na dwa sposoby.

- W czasie rzeczywistym używając Kampania (campaign). Uruchamiając kampanię, rezerwujemy zasoby w AWS, które hostują nasz model i umożliwiają wykonywanie zapytań w czasie rzeczywistym. Następnie możemy wywoływać API Personalize *GetRecommendations* z podaniem np. *userId* (dla listy rekomendacji „dla Ciebie”) lub *itemId* (dla rekomendacji podobnych do danego itemu, jeśli użyliśmy recipe *SIMS*). W odpowiedzi dostaniemy listę rekomendowanych *item_id* wraz ze wskaźnikami

prawdopodobieństwa. W naszym scenariuszu nie chcemy generować rekomendacji *na żywo* przy każdym wejściu użytkownika, bo przyjęliśmy aktualizację raz na tydzień, dlatego kampania nie jest konieczna – zamiast tego skorzystamy z trybu batch.

- W plikach wsadowych Batch inference job. Amazon Personalize umożliwia uruchomienie zadania, które wygeneruje rekomendacje dla wszystkich lub wybranej grupy użytkowników naraz i zapisze wyniki do pliku na S3. Jest to idealne do zastosowań np. personalizowanych mailingów lub okresowego odświeżania rekomendacji w bazie. Możemy zlecić takie zadanie raz na tydzień po zaktualizowaniu modelu. Usługa wygeneruje np. plik CSV zawierający dla każdego użytkownika listę *N* polecanych produktów. Taki plik możemy następnie zaimportować do naszej bazy danych sklepu i serwować z niej rekomendacje. Dzięki temu zachowujemy tryb offline – przez tydzień wyświetlamy użytkownikowi stały zestaw poleceń, a co tydzień aktualizujemy na podstawie najnowszych danych transakcyjnych.

AWS ma gotowe rozwiązania ułatwiające automatyzację tego procesu. Istnieje np. rozwiązanie *Maintaining Personalized Experiences with ML*, które wykorzystuje AWS Step Functions do orkiestracji całego pipeline: regularnego importu nowych danych, trenowania nowej wersji modelu, aktualizacji kampanii oraz uruchamiania zadań batch. Można to dostosować do harmonogramu tygodniowego. Jeśli zaś chodzi o integrację wyników z aplikacją – AWS często pokazuje architekturę, gdzie API AppSync (GraphQL) lub Lambda pobiera rekomendacje z Personalize i łączy



e- zdanymi z bazy, np. aby zwrócić pełne informacje o produk-
cie. W naszym tygodniowym batchowym podejściu wystar-
czy jednak zaimportować gotowe wyniki do bazy sklepu lub
statycznie wyświetlać je z pliku.

Podsumowanie i różnice w stosunku do Azure

Amazon Personalize to rozwiązanie black-box – bardzo
upraszcza wdrożenie, ale ogranicza nam wgląd w model.
Nie musimy pisać kodu ML, ale też np. nie wiemy dokładnie
czy i jak silnie wykorzystane zostały nasze niestandardo-
we atrybuty użytkowników sklepu. Mamy ograniczony ze-
staw algorytmów do wyboru. W kontekście cech behawio-
ralnych klientów: w Personalize możemy je zaimportować
jako dodatkowe dane, lecz nie mamy gwarancji, że wybrany
model ich użyje – prawdopodobnie użyje, ale w jakiś nie-
jawny sposób.

Azure daje pełną kontrolę – tu oddajemy stery Amazonowi.
Z drugiej strony, Personalize łatwo integruje się z infrastruktu-
rą AWS (S3, Lambda, DynamoDB, Pinpoint do maili itp.) i jest
zaprojektowany pod skalę i aktualizację w miarę napływu no-
wych danych. Np. można wysyłać zdarzenia streamingowo
i model AWS będzie je brał pod uwagę niemal od razu dzie-
ki mechanizmom incrementalu. W naszym projekcie sklepu
z mąką nie potrzebujemy streamingu, ale gdyby jednak ruch
wzrósł i wymagano bardziej dynamicznej personalizacji, Per-
sonalize oferuje tę możliwość konfiguracji.

Personalize nalicza opłaty za czas trenowania modelu
i za zapytania – przy 20k klientów i cotygodniowej aktuali-
zacji powinno to być dość przewidywalne. Można też przez
pierwsze dwa miesiące korzystać z darmowego prognozy.

Przykład efektu: po wdrożeniu Personalize moglibyśmy
na stronie głównej sklepu wyświetlać sekcję „Polecane dla
Ciebie” z listą np. 5 produktów, czyli różnych rodzajów mąk,
które Personalize uzna za najbardziej interesujące dla dane-
go zalogowanego klienta na podstawie jego historii zaku-
pów i podobieństwa do innych klientów. Na stronach kon-
kretnych produktów moglibyśmy dodać sekcję „Klienci, któ-
rzy kupili ten produkt, kupili również:” wykorzystując osobno
model typu SIMS (Similar Items). Całość aktualizowalibyśmy
co tydzień – np. w poniedziałek. Personalize przetrenuje mo-
del z nowymi danymi z całego poprzedniego tygodnia, a od
wtorku klienci widzą odświeżone rekomendacje. Takie podej-
ście spełnia wymaganie, by system nie działał ściśle w czasie
rzeczywistym, a jednocześnie korzystał z zaawansowanej in-
frastruktury ML AWS.

Wojciech Moszczyński



Wojciech Moszczyński – absolwent Katedry
Ekonometrii i Statystyki Uniwersytetu Miko-
łaja Kopernika w Toruniu, specjalista z zakre-
su ekonometrii, finansów, data science oraz
rachunkowości zarządczej. Specjalizuje się
w optymalizacji procesów produkcyjnych i lo-
gistycznych. Prowadzi badania w obszarze
rozwoju i zastosowania sztucznej inteligencji.
Od lat zaangażowany w popularyzację machine learning oraz
data science w środowiskach biznesowych.

Możesz nas zaprenumerować:

e-mailem: prenumerata@sigma-not.pl
telefonicznie: 22/840 30 86, 22/840 35 89
przez internet: www.sigma-not.pl www.pzmlyn.pl
listownie: Zakład Poligrafii i Kolportażu
Wydawnictwa SIGMA-NOT Sp. z o.o.
ul. Ks. J. Popiełuszki 21, 01-595 Warszawa
wpłata na konto: Wydawnictwo SIGMA-NOT Sp. z o.o.
PKO BP 24 1020 1026 0000 1002 0250 0577
z dopiskiem:
prenumerata „Przeglądu Zbożowo-Młynarskiego”

Ceny brutto „Przeglądu Zbożowo-Młynarskiego” w 2025 r.

- 1 egz. 50 zł
- roczna wersja papierowa 288 zł + koszt wysyłki 21 zł
- roczna cyfrowa 240 zł
- roczna PAKIET PLUS wersje papierowa i cyfrowa
+ dostęp do archiwalnych treści 384 zł

UWAGA! Okres prenumeraty nie ulega skróceniu.

WAŻNE. Dostęp do pojedynczych artykułów i całych archiwal-
nych numerów „Przeglądu Zbożowo-Młynarskiego” na Portalu
Informacji Technicznej www.sigma-not.pl Każdy prenumeratorki
wersji cyfrowej i wersji PLUS dostaje indywidualny link aktyw-
acyjny.

W przypadku zmiany stawki VAT na czasopismo i – w kon-
sekwencji – zmiany ceny brutto prenumeraty, prenumerato-
rzy są zobowiązani do dopłaty różnicy.

Prenumerata zagraniczna. Dla prenumeratorów zagranicz-
nych obowiązuje cena według kursu waluty NBP z dnia bez-
pośrednio poprzedzającego datę wystawienia faktury plus
koszty wysyłki.

Informacje dla Autorów. Redakcja przyjmuje do publikacji
tylko prace oryginalne, niepublikowane wcześniej w innych
czasopismach. Autor przysyłając do redakcji niezamówio-
ną przez nią publikację jednocześnie udziela Wydawnictwu
SIGMA-NOT Sp. z o.o. nieodpłatnej i niewyłącznej licencji do
utworu. W przypadku publikacji zamawianych przez redak-
cję, Autor otrzymuje do podpisania umowę z Wydawnictwem
SIGMA-NOT Sp. z o.o. o przeniesieniu praw autorskich na wy-
łącznie Wydawcy lub umowę licencyjną – do wyboru Autora.
Z chwilą przyjęcia artykułu przez redakcję następuje przenie-
sienie praw autorskich na Wydawcę, który ma odtąd prawo do
korzystania z utworu, rozporządzania nim i zwielokrotnienia
dowolną techniką, w tym elektroniczną oraz rozpowszechnia-
nia dowolnymi kanałami dystrybucyjnymi. Redakcja nie zwraca
materiałów niezamówionych oraz zastrzega sobie prawo reda-
gowania, skracania tekstów i dokonywania streszczeń. Za mery-
toryczną treść artykułów odpowiadają Autorzy.