

PRACA INŻYNIERSKA

System rekomendacji produktów

wykorzystujący filtrację opartą na treści, logikę
rozmytą i modele probabilistyczne

Piotr Smoła

Nr albumu: 125162 • Kierunek: Informatyka

Promotor: dr inż. Piotr Grochowalski

Rzeszów, 2026



Problem i motywacja

29%

wzrost przychodów Amazon
po wdrożeniu rekomendacji

Przeładowanie informacyjne

Tysiące produktów → klient porzuca zakupy

Czarne skrzynki

Amazon Personalize, Google AI - brak kontroli nad algorytmami

Brak interpretowalności

Gotowe rozwiązania nie wyjaśniają "dlaczego" polecono produkt



Rozwiązanie

Modułowy system łączący 3 komplementarne metody rekomendacji - implementacja od podstaw bez gotowych bibliotek ML (TensorFlow, Surprise), zapewniająca pełną kontrolę, interpretowalność i dostosowanie do specyfiki e-commerce.

Praca zespołowa z Dawidem Olko - Filtracja współpracy, analiza sentymentu, reguły asocjacyjne

Cel i zakres pracy

2



Cel główny:

Opracowanie i wdrożenie zaawansowanego mechanizmu rekomendacyjnego w ramach funkcjonalnej platformy e-commerce

Filtracja oparta na treści

Ważone wektory cech +
podobieństwo kosinusowe

Logika rozmyta

System TSK
6 reguł IF-THEN

Modele probabilistyczne

Łańcuch Markowa + Naive Bayes

Dodatkowe cele realizacyjne:



Architektura trójwarstwowa: Django + React + PostgreSQL



Optymalizacja wydajności (cache, bulk operations, threshold pruning)



Ewaluacja na zbiorze: 500 produktów, 20 użytkowników, 200 zamówień

Filtracja oparta na treści

$$\cos(\theta) = (A \cdot B) / (||A|| \times ||B||) \rightarrow \text{wynik} \in [0, 1]$$

Wagi cech produktu:

Kategoria 40%

Tagi 30%

Cena 20%

Słowa
kluczowe
10%

Proces:

1. Ekstrakcja ważonych cech \rightarrow wektor rzadki
2. Obliczenie cosinus similarity dla par produktów
3. Zapis par z similarity > 0.2 (próg 20%)
4. Bulk insert: partie po 1000 rekordów

Wyniki optymalizacji:

~97%

redukcja tabeli
(threshold pruning)

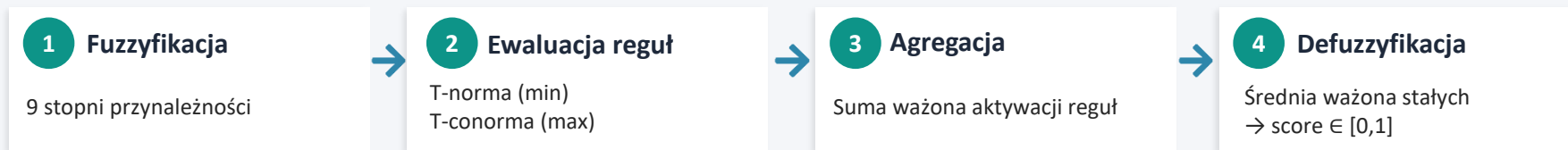
~80x

szybszy zapis
(bulk insert)

Rozwiązuje problem zimnego startu dla nowych produktów - wystarczą cechy (kategoria, tagi, cena)

Logika rozmyta - Sugeno (TSK)

4



Zmienne wejściowe (3 × 3 zbiory rozmyte):

Cena

cheap · medium · expensive

Progi: 100, 500, 1200, 2000 PLN

Jakość

low · medium · high

Rating: 2.5 / 3.5 / 4.5

Popularność

low · medium · high

Zamówienia: 2 / 10 / 30

6 reguł IF-THEN - wybrane przykłady:

Reguła	Warunek IF	Waga
R1: High Quality Bargain	quality HIGH \wedge price CHEAP/MEDIUM	0.9
R2: Popular in Category	category_match HIGH \wedge popularity HIGH	0.7
R4: Category Quality	category_match HIGH \wedge quality MED/HIGH	0.85
R5: Premium Match	user.sensitivity < 0.4 \wedge price EXP \wedge quality HIGH	0.8

+ 2 dodatkowe reguły: Price Sensitive (0.6), Quality-Price Balance (0.75)

Klasyfikacja: Sugeno zerowego rzędu (TSK-0)

Poprzedniki rozmyte (jak Mamdani), ale następniki to stałe liczbowe → średnia ważona zamiast centroidu.
Sugeno-0 jest szczególnym przypadkiem Mamdaniego (Ross, 2010: "Fuzzy Logic with Engineering Applications")

Modele probabilistyczne

Łańcuch Markowa (1. rzędu)

Predykcja sekwencji zakupowych

$$P(s_j | s_i) = \text{count}(s_i \rightarrow s_j) / \sum_k \text{count}(s_i \rightarrow s_k)$$

- Stany: 48 kategorii produktów
- Trening: pary sąsiadujących stanów z zamówień
- Predykcja: sortowanie wg $P(\text{przejścia})$, top-k
- Rozkład stacjonarny: $\pi = \pi P$
- Wykrywanie cykli (przerywanie >2 powtórzeń)

Naiwny klasyfikator Bayesa z wygładzaniem Laplace'a

Predykcja zakupu + ryzyka odejścia

$$P(C|X) = P(C) \times \prod P(x_i | C) / P(X)$$

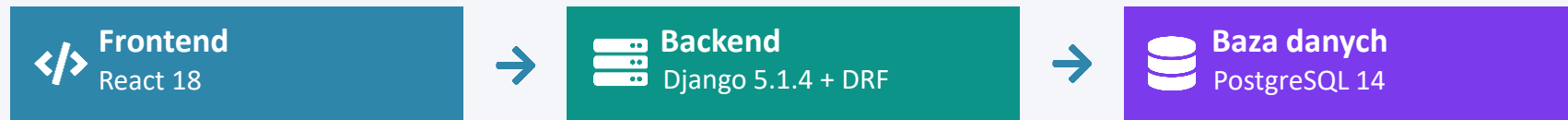
- 5 cech: zamówienia, avg_order, dni_od_ostatniego, kategoria, częstotliwość
- Wygładzanie Laplace'a: $(\text{count}+1)/(\text{total}+|V|)$
- Obliczenia w log-przestrzeni (zapobieganie niedomiaru zmiennoprzecinkowego)
- Predykcja: will_purchase / will_not_purchase
- Ryzyko odejścia: will_churn / will_not_churn

Integracja:

$\text{score} = 0.6 \times P_{\text{category}}(\text{Markov}) + 0.4 \times P_{\text{purchase}}(\text{NB}) \rightarrow$ Markov lepiej przewiduje kategorię, NB moduluje prawdopodobieństwem zakupu

Architektura systemu

6



Docker Compose — 3 kontenery | docker-compose up | Zero konfliktów zależności | Izolacja serwisów

Mechanizmy optymalizacyjne:

Cache 3-poziomowy

5min / 30min / 2h

Bulk operations

~80× szybszy zapis

Threshold pruning

Redukcja ~97%

Prefetching

Eliminacja N+1

Komplementarność metod

7

Kryterium	Content-Based Filtering	Logika rozmyta	Modele probabilistyczne
Źródło danych	Cechy produktów (kategoria, tagi, cena)	Cechy + profil użytkownika	Sekwencje zakupowe + cechy behawioralne
Problem zimnego startu	Rozwiązany	Profil domyślny	Wymaga historii
Interpretowalność	Wysoka (wspólne cechy)	100% (reguły IF-THEN)	Średnia (prawdopodobieństwa)
Personalizacja	Niska (cechy produktu)	Wysoka (profil cenowy)	Najwyższa (analiza behawioralna)
Najlepsze zastosowanie	Nowe produkty bez historii	Wrażliwość cenowa + interpretowalność	Predykcja sekwencji + analiza odejścia

Kluczowy wniosek

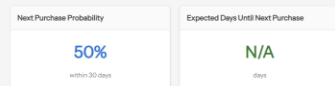
Każda metoda adresuje inny wymiar problemu rekomendacji. CBF dla nowych produktów, Fuzzy Logic dla personalizacji z interpretowalnością, modele probabilistyczne dla głębokiej analizy behawioralnej. Optymalne rezultaty osiąga się przez ich synergiczne zastosowanie.

Prezentacja aplikacji

8

Next Purchase Predictions

Our Markov chain models analyze your shopping sequences to predict what you're likely to buy next.



Likely Next Products

IM Inno Cruiser 2 BMP \$179.99 Prediction Score: 100%	IM Inno Cruiser Dual BMP \$229.99 Prediction Score: 100%	A4 A4Tech HD PK-810P USB Black \$39.99 Prediction Score: 100%	CR Creative Live! Cam Syno 1080p V2 \$49.99 Prediction Score: 100%
BA Baseus USB-C - USB-C (PD 100W, 2m) \$10.99 Prediction Score: 100%	SI Silver Monkey Adapter USB-C - minijack 3.5 W \$9.99 Prediction Score: 100%		

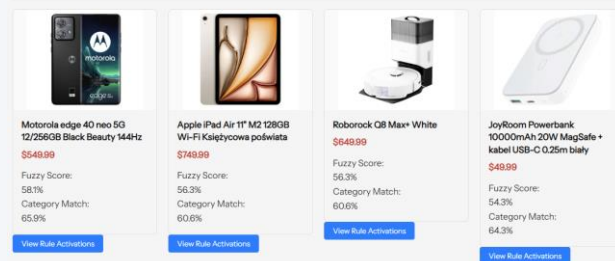
Rekomendacje w panelu klienta - Markov

Recommended For You (Content-Based)

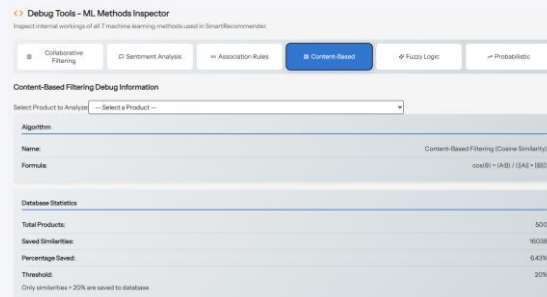


Filtracja oparta na treści - strona główna

Recommended Products



Rekomendacje w panelu klienta - Logika rozmyta



Panel debugowania - administrator

Wyniki i ewaluacja

9

500

produktów

20

użytkowników

200

zamówień

48

kategorii
(stany Markova)

~4000

par podobieństw
(CBF)

Kluczowe osiągnięcia:

Implementacja od podstaw

Bez gotowych bibliotek rekomendacyjnych — pełne zrozumienie algorytmów

3 komplementarne metody

CBF + Fuzzy Logic + Probabilistyczne — pokrywają wzajemne słabości

Pełna interpretowalność Fuzzy Logic

Każda rekomendacja wyjaśniona regułami IF-THEN (GDPR)

Wydajne optymalizacje

Cache 3-poziomowy, bulk ops, threshold pruning — real-time odpowiedzi

Kompletna platforma e-commerce

Katalog, koszyk, zamówienia, opinie, panel admin + debug algorytmów

Podsumowanie i rozwój

10

Zidentyfikowane ograniczenia

Cold start:

Markov + NB wymagają historii zakupów (CBF kompensuje)

Filter bubble CBF:

Rekomenduje tylko podobne - brak odkrywania nowych kategorii

Skalowalność >10k:

$O(n^2)$ CBF wymaga ANN (LSH/HNSW) dla dużych katalogów

Brak sezonowości:

Nie uwzględnia kontekstu czasowego

Kierunki dalszego rozwoju

Deep learning:

Autoencodery, LSTM/GRU dla lepszych sekwencji

Meta-learner:

Automatyczny dobór najlepszej metody per kontekst

Real-time rekomendacje:

Celery + Redis, opóźnienie 30-60s

Markov wyższego rzędu:

Uwzględnienie 2-3 ostatnich zakupów

Wniosek końcowy

Implementacja od podstaw jest możliwa i celowa w kontekście edukacyjnym. Zrealizowany system stanowi kompletne rozwiązanie e-commerce z trzema komplementarnymi metodami rekomendacyjnymi, gotowe do wdrożenia w środowisku produkcyjnym.

Dziękuję za uwagę

Piotr Smoła • 125162

System rekomendacji produktów wykorzystujący filtrację opartą na treści,
logikę rozmytą i modele probabilistyczne

Promotor: dr inż. Piotr Grochowalski • Uniwersytet Rzeszowski, 2026

