

# DDDD & DD

Dynamicznie Dedukowane Drzewa Decyzyjne & Detekcja Dyferencji

Wspomaganie Podejmowania Decyzji w Sprzedaży Biżuterii  
Narzędzie Wizualizacji Preferencji Klientów

Patryk Lange Uniwersytet SWPS

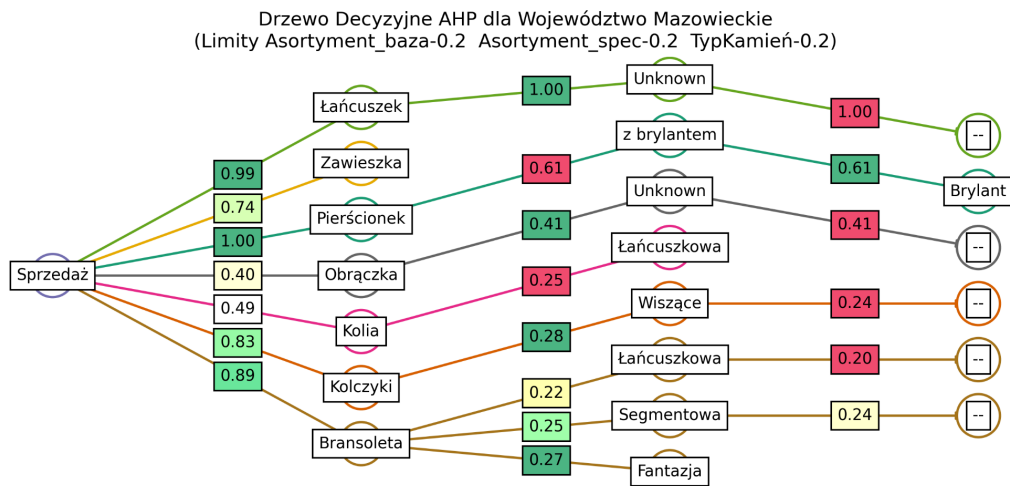
W branży jubilerskiej, gdzie decyzje o zarządzaniu asortymentem i pozycjonowaniu często pojedynczych sztuk produkt bywa działaniem instynktownym, zastosowanie zaawansowanych metod analizy danych, takich jak Proces Hierarchii Analitycznej (AHP), nabiera szczególnego znaczenia. Wraz z postępem narzędzi analitycznych w biznesie, szczególnie w erze Sztucznej Inteligencji, odchodzimy od interpretowania średnich i median na rzecz złożonych metod statystycznych czy nie wyjaśnialnego uczenia maszynowego. W oparciu o metodę AHP, prezentowane narzędzie wizualizuje hierarchię preferencji klientów oraz umożliwia porównanie tych preferencji pozwalając użytkownikowi na swobodny dobór poziomów drzewa, zakresów tolerancji czy sposobów reprezentacji danych.

### Metodologia

Model AHP został zaimplementowany przy użyciu języka skryptowego Python, korzystając z bibliotek takich jak pandas, numpy i matplotlib. Model opiera się na trójstopniowej hierarchii preferencji, którą użytkownik może ustawić na jakąkolwiek kombinację dostępnych kolumn. Na potrzeby omówienia, za przykład weźmy wybór następujących trzech kolumn:

- Poziom 1 'Typ Biżuterii'                      np. Kolczyki, Zawieszka
- Poziom 2 'Podtyp'                              np. Wiszące, Kółka, Zodiak, Medalik
- Poziom 3 'Kamienie'                           np. Cyrkonia, Brylant, Gładkie(brak)

Użytkownik wybiera również Kolumnę agregującą oraz typ agregacji - w naszym przypadku będzie to 'Cena Brutto' oraz 'Suma'. Wewnątrz każdego poziomu dane są agregowane dla każdej grupy wewnątrz każdego poziomu oraz skalowane wybranym algorytmem reprezentując wagi decyzyjne klientów, które są następnie wizualizowane w postaci drzewa AHP.



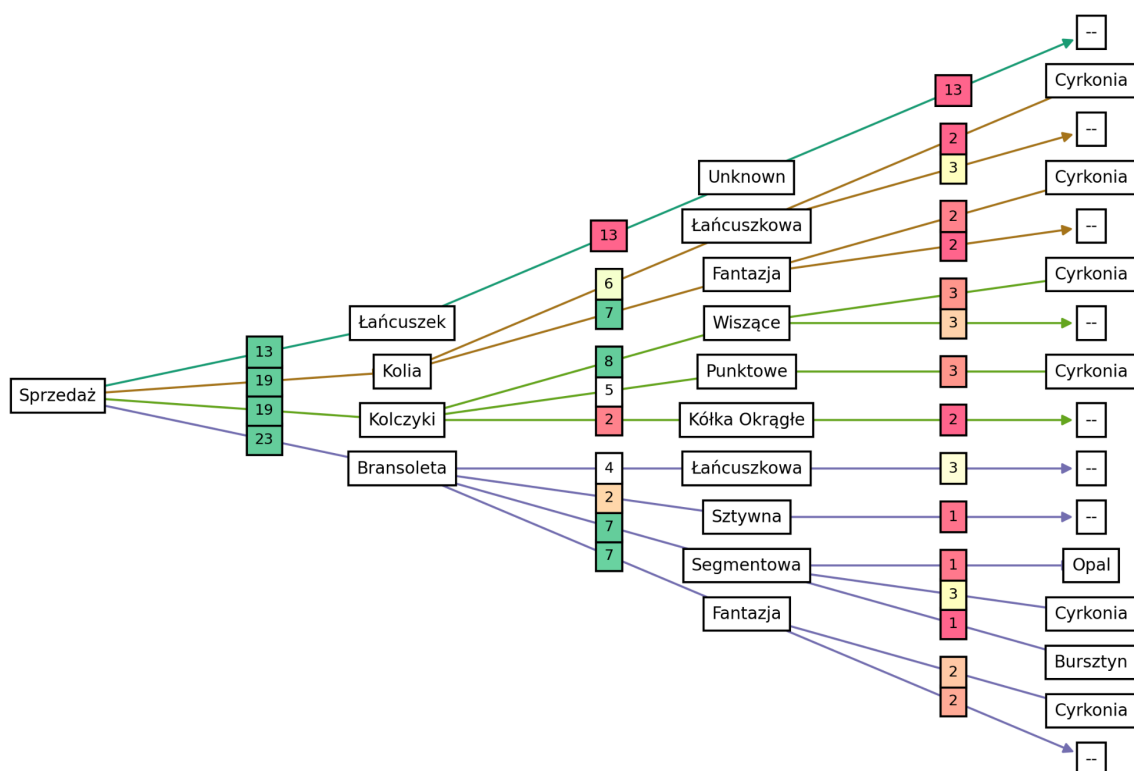
W zależności od potrzeb obliczone już wagi można reprezentować w różnych skalach lub nawet surowych wartościach. Ważnym jest również zauważyć, że każda kolejna grupa jest mniejsza - obliczając wagi dla kolejnych poziomów zawężamy ilość dostępnych danych. Na powyższym przykładzie widzimy zagregowane sumy brutto przeskalowane algorytmem `MinMaxScaler()`. Powyższe wagi jednak mogą mylić, gdyż wkład 'łańcuszka' oraz 'Pierścionka' jest na podobnym poziomie - oba otrzymują wagę bliską 1 (Maksymalnej). Problem jest również widoczny na dalszych poziomach, w których wagi odnoszą się do wartości Maksymalnej wewnątrz swojego rodzica, w efekcie czego użytkownik, aby dogłębnie zinterpretować wizualizacje musi przeliczać relatywny udział podkategorii w głowie.

Całe szczęście program posiada również przyjaźniejsze dla oka ustawienia. Zastosowane rozwiązanie skaluje wagi w taki sposób, aby dawały łączną sumę `VOTE_AMOUNT`. Poniżej skonstruujemy drzewo które reprezentuje zagregowaną kolumnę za pomocą przechodnich na dzieci 'głosów'.

```
Python
SET_SCALER = RobustScaler()
VOTE_SCALER_INSTEAD = True
SHARE_VOTES_WITH_PARENT = False
VOTE_AMOUNT = 100
GOAL_NAME, COL_GOAL = 'Sprzedaż', 'Cena brutto'
COL_LV_1, COL_LV_2, COL_LV_3 = 'Asortyment_baza', 'Asortyment_spec', 'TypKamień'
GOAL_FUNC = 'sum'
AHP_LIM1, AHP_LIM2, AHP_LIM3 = 10, 2.5, 0.5
CONDITION_COLUMN, DATASET_1_CONDITION, DATASET_2_CONDITION = "Województwo", "Mazowieckie", "Śląskie"
```

Liczba Głosujących jest dedykowanym rozwiązaniem które skaluje zagregowaną wartość od 0 do 100 która jest przechodnia z wyższymi poziomami. W efekcie otrzymujemy ścieżki decyzyjne dla naszej setki klientów w bardziej instynktownej wizualizacji.

DrzewoDecyzyjne AHP dla Województwo Mazowieckie  
(Limity Asortyment\_baza-10 Asortyment\_spec-2 TypKamień-1)  
100 Głosujących



Teraz historia jest jasna. Na sto klientów w województwie Mazowieckim 19 z nich sięgnie po Kolczyki, dalej 8 z 19 weźmie Wiszące a wśród wiszących 3 z nich będzie z cyrkonią a 3 bez. Zauważamy natychmiast że sumy widocznych kategorii nie składają się na całość wartości rodzica. Do użytkownika należy ustawienie progów dla każdego poziomu od których będziemy definiować 'ważne' decyzje. Limity te są widoczne w tytule i służą jako pomoc w odfiltrowywaniu niepotrzebnych danych. Ważnym jest ich świadomy dobór w zależności od reprezentowanych wartości - dla wag przeskalowanych MinMaxScaler() należałoby podać wartość w przedziale od 0

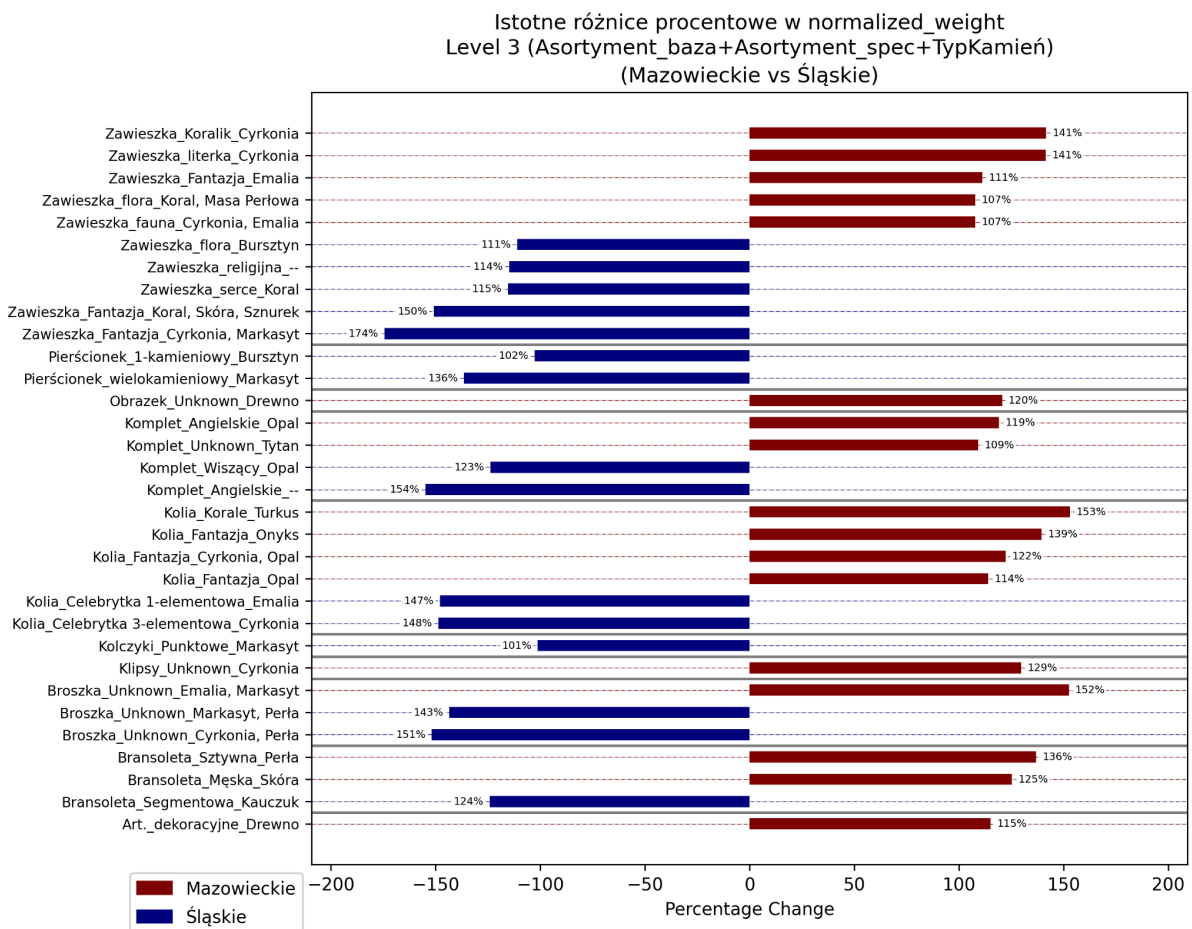
Figure 1 displays three decision trees for AHP analysis, each representing a different region and voter profile:

- Left Tree (Mazowieckie):** DrzewoDecyzyjne AHP dla Województwa Mazowieckiego (Limity Asortyment\_baza=10, Asortyment\_spec=10, TypKamień=5). The tree shows a hierarchy starting from a root node, branching into nodes like 'Lafocuzan', 'Kula', 'Kokcyli', and 'Błociszta', with numerical values (e.g., 15, 10, 20) and colored edges (green, orange, purple).
- Middle Tree (Śląskie):** DrzewoDecyzyjne AHP dla Województwa Śląskiego (Limity Asortyment\_baza=0.1, Asortyment\_spec=0.1, TypKamień=0.1). The tree shows a hierarchy starting from a root node, branching into nodes like 'Lafocuzan', 'Kula', 'Kokcyli', and 'Błociszta', with numerical values (e.g., 15, 10, 20) and colored edges (green, orange, purple).
- Right Tree (Mazowskie):** DrzewoDecyzyjne AHP dla Województwa Mazowskiego (Limity Asortyment\_baza=10, Asortyment\_spec=0, TypKamień=0). The tree shows a hierarchy starting from a root node, branching into nodes like 'Lafocuzan', 'Kula', 'Kokcyli', and 'Błociszta', with numerical values (e.g., 15, 10, 20) and colored edges (green, orange, purple).

Kluczowym dla całego narzędzia, poza samą wizualizacją ścieżek decyzyjnych jest automatyczny zapis, odczyt i podział danych, który pozwala na przekazywanie obliczonych wag do porównania. Wśród wspomnianych ustawień użytkownika pojawia się wyraźne zapytanie o podział w kolumnie Województwo na Mazowieckie i Śląskie. Dla każdego z warunków, dane wejściowe są czytane, filtrowane, agregowane i ważone oddzielnie, co pozwala nam porównywać ze sobą nierówne lub wybrakowane zbiory, gdyż wagi są relatywne tylko wewnątrz zbioru (możemy wpuścić 100 klientów na 5 par kolczyków). Ciężko jednak sobie zwizualizować nałożenie na siebie dwóch drzew w celu

identyfikacji różnic między nimi. Z tego właśnie powodu surowe wagi obliczone dla drzew są przekazywane dalej i porównywane statystycznie.

Wykres słupkowy, który jest rezultatem drugiej części kodu, ukazuje procentową zmianę w wagach decyzyjnych dla każdego z poziomów. Jest to przedstawione za poziomych słupków, gdzie każdy kolor reprezentuje jedno z dwóch kryteriów, tu - województwo. Rozciąganie słupka na prawo lub lewo wskazuje na odpowiednio wyższą różnicę w wagach pomiędzy oboma zbiorami. Ten wykres jest generowany na podstawie obliczonych dla AHP wag, jednak nie jest ograniczony limitami które ustawialiśmy do wizualizacji drzewa. Wykres słupkowy przyjmuje oddzielne wartości limitów od których zacznie wizualizację. Powyżej jest przedstawione porównanie wag drugiego poziomu, które wygenerowaliśmy wraz z wcześniejszymi dwoma drzewami. Dla uproszczenia, wyświetlam odchylenia powyżej 80%



Nie wchodząc zbytnio w szczegóły, zauważamy na przykład to, że Mazowszanie preferują prostsze zawieszki z Cyrkoniami oraz unikają pereł a Ślązacy preferują zawieszki wymyślne i lżejsze bransoletki celebrytki

Wykorzystanie dynamicznych drzew decyzyjnych pozwala na głęboką i instynktowną analizę preferencji klientów, identyfikując kluczowe różnice w decyzjach zakupowych i nie tylko. Nie trudno wyobrazić sobie porównanie pomiędzy tym co do sklepu się kupuje a tym, co kupują w nim klienci. Można też porównać jeden sklep wobec wszystkich innych i zyskać w prosty sposób informacje które wspomogą przemyślaną decyzję. Narzędzie to może być użyte do optymalizacji zapasów, planowania promocji, ustalania strategii cenowych, a nawet projektowania układu sklepów. Dzięki dynamicznemu charakterowi modelu AHP, narzędzie to ma ogromny potencjał w najróżniejszych zastosowaniach biznesowych.

#### Przypadki Użycia i Korzyści

Zastosowanie narzędzia AHP w sprzedaży biżuterii nie ogranicza się jedynie do analizy danych sprzedażowych. Może ono również służyć do przewidywania trendów, optymalizacji asortymentu w sklepie w zależności od sezonu czy okazji, a także do personalizacji oferty dla klientów. Korzyści z jego zastosowania obejmują:

- Lepsze zrozumienie potrzeb i preferencji klientów.
- Zwiększenie efektywności zarządzania zapasami.
- Wzmocnienie strategii marketingowych poprzez ukierunkowanie działań na najbardziej dochodowe kategorie produktów.
- Optymalizacja cen i promocji w oparciu o dokładną analizę danych.

Rozwój narzędzia AHP do analizy danych w branży jubilerskiej stanowi przełom w sposobie podejmowania decyzji. Jego wdrożenie pozwala na systematyczne podejście do danych sprzedażowych, przekształcając je w użyteczną wiedzę, która wspiera rozwój biznesu. Rekomenduje się dalszy rozwój i dostosowanie narzędzia do specyficznych potrzeb branży, co może przyczynić się do jeszcze lepszego zrozumienia rynku i skuteczniejszego reagowania na zmieniające się warunki.

Narzędzie to zostało opracowane przez zespół ekspertów IT specjalizujących się w branży jubilerskiej, z pasją do tworzenia rozwiązań, które ułatwiają podejmowanie informowanych decyzji biznesowych. Projekt został zrealizowany przy współpracy z lokalnymi przedsiębiorstwami jubilerskimi, które dostarczyły dane i wsparcie w testowaniu funkcjonalności narzędzia.

#### Informacje Kontaktowe

[Zespół IT - nazwa firmy]

[Adres e-mail]

[Numer telefonu]

[Adres strony internetowej]

#### Deklaracja

Prezentowane wyniki i analizy oparte są na rzeczywistych danych sprzedażowych, zanonimizowanych w celu zachowania poufności informacji biznesowych. Projekt został zrealizowany z zachowaniem najwyższych standardów etycznych i profesjonalnych.