

Jakub Dębiński 251501
Łukasz Murza 251592

Wtorek 10:30
10.06.2025

Tytuł: Sztuczna inteligencja i systemy ekspertowe
Podtytuł: Zadanie 3: Drzewa decyzyjne

Cel

Przewidywanie cen mieszkań na podstawie cech ogłoszeń — zadanie regresji na danych o apartamentach.

Wyniki

Narzędzie, z którego korzystaliśmy do obliczeń oraz do generowania wyników to język interpretowany Python.

Dane

Plik zawiera dane 1 000 mieszkań z takimi cechami jak powierzchnia (m^2), liczba pokoi, piętro, rok budowy, odległość od centrum (km), kategoria lokalizacji, dostępność windy oraz odpowiadającą im cenę w euro.

Metoda

Drzewa decyzyjne klasyfikacji to algorytm, który iteracyjnie dzieli przestrzeń cech, wybierając w każdym kroku podział maksymalizujący czystość podzbiorów (poprzez zysk informacyjny lub spadek Giniego). Dzięki prostej strukturze drzewiastej efektywnie algorytmy te uczą się reguł decyzyjnych i są łatwe do interpretacji, ale wymagają kontroli złożoności (przycinania, walidacji), aby dobrze klasyfikować nowe dane.

Wyniki

1. Metryki jakości modelu

- **Mean Squared Error (MSE):** 1 866 516 040.59
Pierwiastek z MSE to ok. **43 200 €**, co oznacza, że średnio o tyle „chybia” przewidywania modelu względem rzeczywistych cen.
- **Współczynnik determinacji (R^2):** 0.97
Model wyjaśnia **97 %** zmienności cen mieszkań w zbiorze testowym. To bardzo wysoki wynik, świadczący o dobrym dopasowaniu i niewielkim overfittingu. Udało się to osiągnąć po ograniczeniu głębokości drzewa.

2. Ważności cech

Cecha	Waga	Udział wyjaśnianej wariancji
area_sqm	0.7715	77.2 %
location_outskirts	0.1518	15.2 %
location_suburbs	0.0765	7.7 %
floor	0.0001	0.01 %
year_built	0.00007	0.007 %
rooms	0.0000	≈ 0 %
has_elevator	0.00000	≈ 0 %
distance_to_center_km	0.00000	≈ 0 %

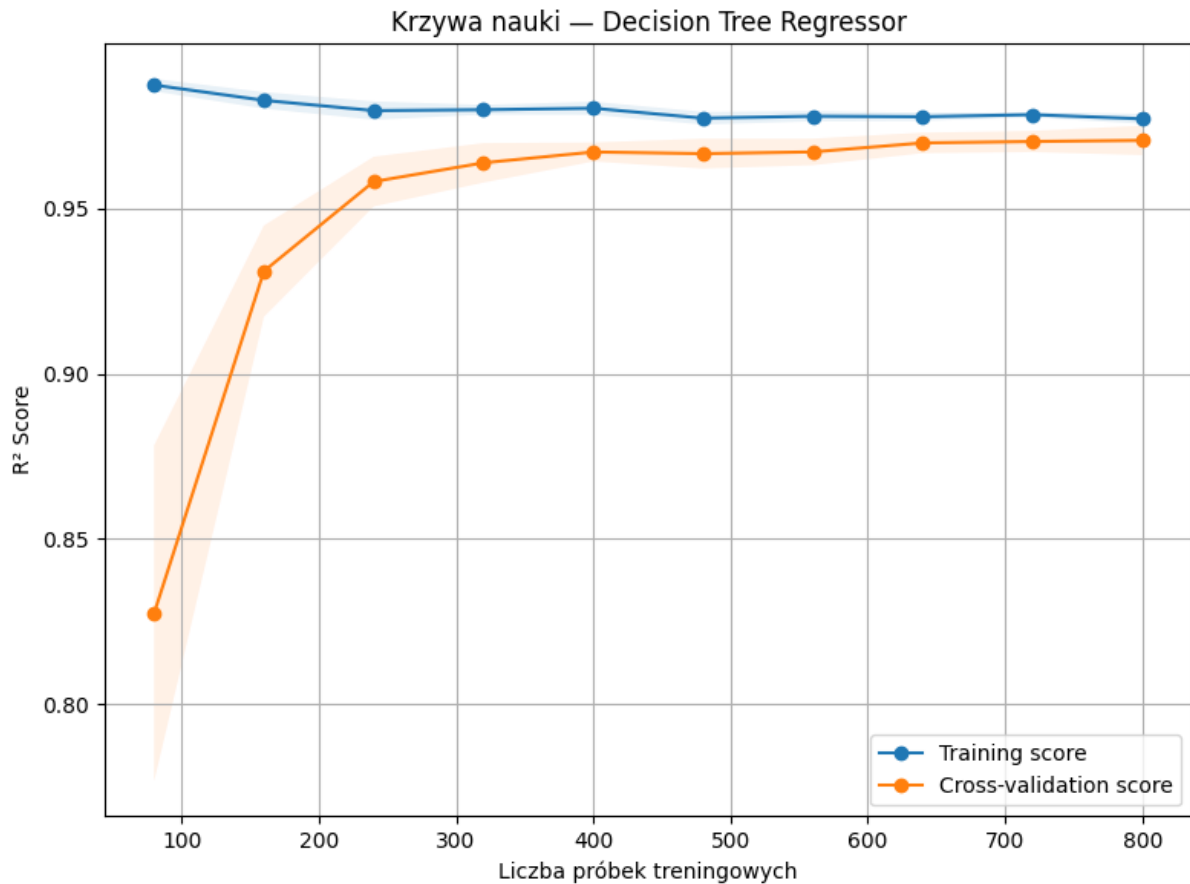
- powierzchnia (area_sqm) odpowiada za 77 % decyzji modelu – to najważniejszy czynnik cenotwórczy.
- Kategorie lokalizacji („outskirts” i „suburbs”) razem wyjaśniają blisko **23 %** wariancji.
- Pozostałe cechy mają praktycznie zerowy wpływ w uproszczonym drzewie.

3. Przykładowe przewidywania

Indeks	Cena rzeczywista (€)	Cena przewidywana (€)	Różnica (€)	Błąd relatywny
521	412 830.24	469 298.10	+56 467.86	+13.7 %
737	713 350.00	773 030.61	+59 680.61	+8.4 %
740	607 791.90	616 003.23	+8 211.33	+1.4 %
660	382 278.65	384 531.65	+2 253.00	+0.6 %
411	1 191 652.17	1 167 159.77	-24 492.40	-2.1 %

- Większość przewidywań mieści się w granicach ± 5 % rzeczywistej wartości, co w kontekście cen nieruchomości jest dość dobrym wynikiem.
- Największe odchylenie (≈ 13.7 %) wystąpiło dla próbki 521, ale pozostałe błędy są znacznie mniejsze.

Wykres krzywej nauki



Wnioski

- **Model działa bardzo dobrze:** $R^2 = 0.97$ i niski MSE po ograniczeniu głębokości do 5 wskazują na dobry kompromis pomiędzy błędem bias a wariancją.
- **Główne czynniki:** powierzchnia mieszkania i lokalizacja.