addepto

Zadanie dla kandydata na stanowisko Data Science

Wstęp

Przygotowane zadanie ma na celu sprawdzenie umiejętności kandydatów aplikujących do na stanowisko Data Science w dziedzinie budowania modeli Machine Learning. Zadanie to ma przede wszystkim pokazać pewną świadomość działań podejmowanych przez kandydata w procesie analizy danych i budowania modelu statystycznego.

Opis zadania

W pliku ***„house.csv”*** znajdują się dane dotyczące cen nieruchomości o danych cechach. Należy wykonać analizę danych prowadzącą do skonstruowania najlepszego modelu, który na podstawie cech danej nieruchomości dokona predykcji jej ceny, zarówno ciągłej (price) jak i binarnej (price_bin). Preferowanym językiem do rozwiązania tego zadania jest Python. Można korzystać z dowolnych bibliotek do analizy danych statystycznych i ML. Rozwiązanie problemu powinno być przedstawione w postaci analizy dojścia do najlepszego modelu, z opisami poszczególnych kroków.

Dane

W pliku „house.csv” znajdują się nieruchomości z ich atrybutami. Analizę modelu należy wykonać korzystając z następujących danych:

• 'id':str - ID wpisu w bazie,
• 'date':str - data wpisu,
• 'price':float - cena
• 'price_bin':int - binarna zmienna, gdzie ‘1‘ występuje wtedy, kiedy cena jest wyższa od 1mln
• 'bedrooms':float - ilość sypialni,
• 'bathrooms':float - ilość łazienek,
• 'sqft_living':float - powierzchnia użytkowa,
• 'sqft_lot':int - powierzchnia działki,
• 'floors':float - ilość pięter,
• 'waterfront':int - położenie na nabrzeżu {0,1},
• 'view':int - widok (0:4),
• 'condition':int - stan nieruchomości (1:5),
• 'grade':int - nachylenie działki (1:13),
• 'sqft_above':int - powierzchnia poddasza,
• 'sqft_basement':int – powierzchnia piwnicy,
• 'yr_built':int - rok budowy

Sposób prezentacji wyników

Wyniki powinny zostać przesłane w postaci wykonywalnego kodu Python (lub R) z

komentarzami wykonywanych czynności (np. w jupiter notebook, markdown itp.). Dodatkowo należy dodać krótki opis zrozumienia i analizy danego problemu. Także warto wzbogacić raport o przemyślenia na temat potencjalnego rozwoju danego rozwiązania lub zupełnie innych podejść (wystarczy na koniec umieścić w komentarzach).

Wyniki należy wysłać na adres: contact@addepto.com ze stopką **[ZADANIE DATA SCIENCE]**. Z kandydatami, którzy prześlą najlepsze rozwiązania będziemy się kontaktować w celu przeprowadzania dyskusji nad opracowaną analizą.

W razie jakichkolwiek pytań lub problemów z danymi prosimy o kontakt mailowy na powyżej podany adres.