

PROJEKT

Całość należy wykonać korzystając z Pythona (algorytmy w Scikit learn) i przedstawić raport z wynikami w Jupyter Notebook wraz z opisami i wykorzystaniem możliwości Markdown.

Projekt w formie Jupyter Notebook należy zamieścić w kanale prywatnym zespołu w MS Teams.

W pliku dane_projekt2022.xlsx zawarto następujące dane:

- asortyment: zadowolenie z asortymentu w markecie X (z ankiety) 0-100 pkt.
- obsługa: zadowolenie z obsługi w markecie X (z ankiety) 0 -100 pkt.
- dni: liczba dni w tygodniu, kiedy klient robi zakupy
- dochod: przeciętny dochód
- inne: liczba innych marketów, w których robi zakupy
- internet: czy robi zakupy przez Internet (0 - nie, 1 - tak)
- kredytowa: czy posiada kartę kredytową (0 - nie, 1 - tak)
- wiek: wiek w latach
- płeć: 0 - mężczyzna, 1 - kobieta
- pożyczka – czy ma kredyt/pożyczkę/chwilówkę (0 - nie, 1 - tak)
- odległość: odległość miejsca zamieszkania (km)
- zakupy: czy w przyszłości zamierza robić zakupy w markecie X (0 - nie, 1 - tak)

Analizujesz dane bez obserwacji.... Każdy zespół otrzyma informację, które obserwacje ma usunąć.

1. Wczytaj **wszystkie dane z pliku dane_projekt2023**, a następnie usuń obserwacje wskazane dla zespołu. Sprawdź, czy są braki danych. Jeśli tak, to usuń obserwacje z brakami.
2. Dokonaj wstępnej analizy danych – zastosuj stosowne statystyki opisowe oraz wykorzystaj metody wizualizacji danych (wizualizacje zrób tylko dla kilku wybranych zmiennych). Skomentuj wyniki.
3. Rozwiąż problem klasyfikacyjny trzema wybranymi przez siebie algorytmami. Zmienna zakupy jest zmienną zależną. Pozostałe to potencjalne zmienne niezależne. Celem jest predykcja, czy osoba o danych charakterystykach zamierza w przyszłości robić zakupy w markecie X czy nie.
4. Rozwiąż problem regresyjny trzema wybranymi przez siebie algorytmami. Zmienna zadowolenie z obsługi w markecie jest zmienną zależną. Pozostałe to potencjalne zmienne niezależne. Celem jest predykcja, czy osoba o danych charakterystykach jest zadowolona z obsługi czy nie.
5. Podaj wnioski.

Istotne kwestie:

1. Dla każdego algorytmu omów sposób postępowania, oceń jakość klasyfikacji/regresji. Przynajmniej w jednym algorytmie klasyfikacyjnym i regresyjnym przeprowadź strojenie hiperparametrów i omów tę procedurę. Oceń znaczenie cech.
2. Porównaj wyniki otrzymane różnymi metodami i podaj swoje rekomendacje (czy da się przewidywać zmienną zakupy oraz zmienną zadowolenie z obsługi w markecie, jeśli tak, to jaki algorytm proponujesz i dlaczego).
3. W opisach wykorzystaj możliwości Markdown.
4. Wybór zmiennych niezależnych zależy od zespołu – przesłanek merytorycznych, uzasadnionych kwestii formalnych, podejścia statystycznego...

Kwestie techniczne:

1. Projekt można wykonać w zespołach dwuosobowych – takich, jak na zajęciach.
2. Projekt należy przesłać najpóźniej do 10 grudnia. Każdy dzień zwłoki skutkuje odjęciem punktów.