Zadanie rekrutacyjne – eksploracja danych i model ML.

# Treść zadania.

Celem zadania jest sprawdzenie umiejętności przeprowadzania eksploracji danych oraz umiejętności budowy modelu machine learningowego. Problem można rozwiązać przy pomocy dowolnych narzędzi (np. R, Python, Julia).

Eksplorację i trenowanie modelu należy wykonać na załączonym zbiorze *train.csv.* Nazwy zmiennych są zakodowane, predykcję należy wykonać dla kolumny o nazwie „Zmienna\_20”.

Zadanie składa się z 3 części:

1. **Eksploracja. (25% oceny)**

* Wskazanie które zmienne są najsilniej powiązane ze zmienną 20 i wizualizacja kluczowych zależności między zmiennymi objaśniającymi, a zmienną objaśnianą.
* (opcjonalne) feature engineering (np. redukcja wymiarów).

1. **Model predykcyjny. (35% oceny)**

Przygotowanie danych pod model predykcyjny. Budowa modelu regresji przy pomocy dowolnego, znanego algorytmu ML. Dodatkowo należy wykonać predykcję zmiennej 20 na zbiorze testowym *test.csv.*

1. **Ocena modelu. (20% oceny)**

* Ocena modelu przy użyciu miary RMSE (błąd średniokwadratowy), MAE (średni błąd absolutny) i R^2.
  + (opcjonalne w przypadku modelu innego niż regresja liniowa) Interpretacja modelu – które zmienne mają największy wpływ na wynik predykcji stworzonego modelu. Prezentacja zależności między zmiennymi.

Dodatkowo model zostanie przetestowany na zbiorze *test.csv* przy użyciu miary MAE. **(20% oceny)**

# Forma rozwiązania.

Rozwiązanie powinno zawierać:

* + Kod wykorzystany do rozwiązania zadania.
  + Raport z przeprowadzonej eksploracji oraz budowy modelu predykcyjnego.
  + Plik .csv z wynikami predykcji dla zbioru *test.csv.* Tabela powinna zawierać 2 kolumny: *ID* oraz *Zmienna\_20.*

(Jeśli rozwiązanie będzie w formie jupyter notebook lub rmarkdown, to nie jest potrzebny osobny plik z raportem).