

## Zadanie zaliczeniowe

W repozytorium plików znajdziesz zbiór danych o nazwie *weather.csv*. Zawiera on 366 codziennych obserwacji pogodowych, zebranych w ciągu jednego roku ze stacji meteorologicznej w Canberze (Australia). Oprócz informacji identyfikujących (data, lokalizacja) zawiera on takie atrybuty, jak:

- minimalna i maksymalna temperatura w ciągu dnia,
- ilość opadów deszczu i tzw. współczynnik parowania,
- liczbę godzin słonecznych w ciągu dnia,
- kierunek i szybkość wiatru oraz jego porywów w różnych porach dnia,
- wilgotność, ciśnienie, stopień zachmurzenia i temperaturę w różnych porach dnia,
- inne zmienne ("RainToday", "RISK\_MM"), nieistotne w tym zagadnieniu,
- **zmienną celu**, mówiąca o tym, czy następnego dnia po obserwacji padał deszcz.

Twoim zadaniem będzie zbudowanie modeli predykcyjnych, opartych na powyższych danych, mających przewidzieć, czy jutro będzie padać, znając warunki atmosferyczne dnia dzisiejszego.

Swoją pracę zacznij od przeanalizowania i zrozumienia zbioru danych – identyfikacji atrybutów, w jakich jednostkach są wyrażone („Use the Google, Luke!”), czy są braki w danych lub obserwacje błędne/odstające, jakie są korelacje zmiennych itp. Jeżeli zajdzie taka potrzeba, dokonaj odpowiedniej transformacji/selekcji danych, wyboru/określenia roli atrybutów itp.

Następnie stwórz kilka klasyfikatorów, bazujących na poznanych w trakcie zajęć metodach, wytrenuj je i dokonaj ich oceny oraz porównania skuteczności (również stosując znane Ci metody). Swoje wyniki przedstaw bądź w postaci samodzielnego raportu w formie pliku PDF (ale zawierającego kod, użyty do jego sporządzenia), bądź – lepiej – w formie notatnika Jupyter Notebook / Google Colab i udostępnienie go (np. na GitHub-ie lub Google Drive).