



# **Wstęp do analizy danych i uczenia maszynowego - projekt**

Koordynatorzy: Franciszek Saliński, Ludwik Madej



## Dla kogo?

- Każdego kto chce od podstaw nauczyć się jak przechodzić przez proces tworzenia modeli z klasycznego MLa, od poznawania i analizy danych do optymalizacji i wyjaśniania modelu



## O projekcie

- Project-based learning
- 4 etapy, rozłożone na około 2 miesiące
  - Zapoznanie się z danymi, ich analiza i wizualizacja
  - Przetwarzanie danych, tak, żeby nadawały się do modelowania
  - Stworzenie pierwszych modeli
  - Optymalizacja hiperparametrów, wybór najlepszego modelu, wnioski/ewaluacja
- Pracujemy w 2 lub 3 osobowych zespołach, ale...



## Jak będzie przebiegała praca w projekcie?

- Dzisiaj dobierzemy was w zespoły, w których będziecie pracować nad 1 etapem projektu przez najbliższe tygodnie
- Po każdym etapie spotykamy się wszyscy, przedstawiamy swoje wnioski, rozwiązania, łączymy wszystko w 1 rozwiązanie
- Następnie mieszamy zespoły, losowo i w taki sposób, żeby każdy pracował z dwoma nowymi osobami
- Do każdego etapu macie też dodatkowe zadanie, oprócz kodu i wykresów przygotować zdjęcie z wspólnego spotkania, gdzie przy piwie/kawie/herbacie pracujecie nad projektem:)



## Na jakich danych bedziemy pracować

- <https://www.kaggle.com/datasets/yasserh/loan-default-dataset>



# **Etap 1 - EDA**