



# Projekt Końcowy

## Bootcamp Data Science

Projekt może dotyczyć dowolnego zagadnienia z obszaru data science. Temat projektu jest proponowany przez każdego uczestnika indywidualnie, ale wymaga akceptacji opiekuna bootcampu. Uzyskanie pozytywnej opinii z projektu jest warunkiem koniecznym do uzyskania certyfikatu ukończenia bootcampu Data Science.

1. Projekt realizowany jest samodzielnie.
2. Forma przekazania projektu do oceny: **projekt musi być umieszczony w repozytorium na githubie**. Repozytorium może być publiczne lub prywatne - wówczas dodajemy do repozytorium użytkownika "martyspiew". Projekt musi zawierać plik w formacie .ipynb (notebook Jupyter'owy), w którym znajdować się będzie cała realizacja projektu lub główna jego część przedstawiająca kluczowe elementy (dopuszczalne są pliki pomocnicze pomagające zachować przejrzystość pliku głównego). Dane mogą, ale nie muszą znajdować się w repozytorium. Inne formy realizacji i przekazania projektu nie są dopuszczalne - w szczególności **przesłanie plików z rozwiązaniem nie jest traktowane jako przekazanie rozwiązania**. Link do repozytorium z projektem należy przesłać opiekunowi bootcampu w wiadomości prywatnej na slacku **wraz z tytułem projekt w języku polskim**.

3. Termin oddania projektów: **4 tygodnie od ostatniego dnia zajęć bootcampu.**
4. Aby projekt został zaliczony, musi on:
  - a. być poprawny pod względem merytorycznym (oznacza to np: poprawną metodologicznie ocenę modeli czy właściwe przygotowanie danych, bez wyraźnych błędów w sztuce),
  - b. zawierać wstęp z opisem **problemu** oraz **danych**
  - c. mieć czytelną, przejrzystą formę umożliwiającą osobie sprawdzającej zrozumienie jego przebiegu. W szczególności **niezbędne są komentarze i objaśnienia**, które wyjaśniają poszczególne kroki. **Projekty nieestetyczne/trudne w odbiorze/bez objaśnień słownych będą zwracane do poprawki.**
  - d. być zrealizowany języku angielskim
5. Projekt może być zrealizowany na danych, które pojawiły się na zajęciach, ale wówczas jego zawartość musi wykraczać istotnie poza elementy zrealizowane podczas zajęć.
6. Projekt, który nie zostanie pozytywnie oceniony, zostanie zwrócony do uczestnika wraz z komentarzem, które elementy wymagają korekty. Po wprowadzeniu poprawek projekt należy przesłać ponownie.

## Uwagi do zawartości

Jeżeli projekt dotyczy problemu predykcji, to koniecznym elementem jest analiza porównawcza różnych rozwiązań (porównanie różnych modeli, optymalizacja ich parametrów, ew. porównanie różnych reprezentacji danych, metod redukcji wymiaru, itp).

Zalecanym jest zrobienie projektu dotyczącego lub wykorzystującego uczenie maszynowe (ogólnie: modelowanie). Natomiast nie jest to absolutnie konieczne.

Projekt może mieć dowolną formułę - rozwiązanie problemu klasyfikacji, pogrupowanie danych, eksploracja i wizualizacja danych (nietrywialna), aplikacja interaktywna, system rekomendacyjny, itp.

### Uwagi ogólne

- Realizacja projektu jest częścią nauki w ramach bootcampu. Dlatego warto to potraktować w ten właśnie sposób, a nie jako przykry obowiązek do odhaczenia.
- Wykonanie projektu zajmuje trzy razy więcej czasu niż się wydaje. Dlatego nie należy odkładać rozpoczęcia prac zbyt długo.
- Liczy się jakość a nie ilość.
- Projekt ten stanowi podstawę Waszego portfolio i może być przedmiotem rozmowy podczas rozmów rekrutacyjnych. Dlatego warto się do niego przyłożyć.

Polecane źródła danych i problemów do rozwiązania:

1. <https://www.kaggle.com/competitions>
2. <https://archive.ics.uci.edu/ml/datasets.php>

Przykładowe projekty:

- Klasyfikacja wydźwięku tweetów na temat linii lotniczych.
- Detekcja nazw marek w tytułach ofert sprzedażowych zamieszczanych na portalach aukcyjnych.
- Przewidywanie zmian indeksu giełdowego WIG20 na podstawie doniesień prasowych.
- Rozpoznawanie logotypów marek samochodów - klasyfikacja obrazów przy użyciu konwolucyjnych sieci neuronowych.
- Detekcja kłamliwych nagłówków - klasyfikacja artykułów polegająca na określaniu zgodności tytułu z treścią.
- Predykcja cen kryptowaluty Bitcoin na podstawie danych historycznych.
- Modelowanie ceny piw.
- Modelowanie cen mieszkań.
- Przewidywanie churnu.
- Przewidywanie wypłacalności kredytobiorcy.
- Rozpoznawanie rodzaju obiektów na powierzchni na podstawie danych LIDAR.
- Profilowanie autora ze względu na płeć na podstawie tweetów.