scripts/

│

├── \_\_init\_\_.py <- żeby traktować to jak paczkę Pythona

│

├── clean\_missing.py <- obsługa braków

├── encode\_features.py <- zamiana kategorii na liczby

├── scale\_normalize.py <- standaryzacja i normalizacja

├── feature\_engineering.py <- tworzenie nowych cech

├── outlier\_detection.py <- wykrywanie i obsługa wartości odstających

├── data\_split.py <- train/test/validation split

├── data\_loader.py <- wczytywanie danych z csv, json, parquet, sql

├── imbalance\_handler.py <- balansowanie klas (oversampling, undersampling, SMOTE)

├── text\_preprocessing.py <- tokenizacja, czyszczenie tekstu, stemming/lemmatyzacja

├── image\_preprocessing.py <- augmentacja obrazów, resize, normalizacja

├── pipeline\_utils.py <- łączenie kroków w pipeline (np. jak w scikit-learn)

└── save\_load.py <- zapisywanie i wczytywanie przetworzonych danych

 **clean\_missing.py** – strategie uzupełniania braków (mean, median, mode, constant, interpolation).

 **encode\_features.py** – one-hot, label encoding, target encoding.

 **scale\_normalize.py** – StandardScaler, MinMax, log transform, robust scaling.

 **feature\_engineering.py** – generowanie nowych cech (np. interakcje, binning, extractions typu z daty → dzień tygodnia).

 **outlier\_detection.py** – IQR, z-score, DBSCAN do wycinania outlierów.

 **data\_split.py** – proste train\_test\_split, ale też k-fold cross validation.

 **data\_loader.py** – żeby nie powtarzać 100x pd.read\_csv; robisz jedną funkcję load\_data("titanic.csv") i masz gotowe.

 **imbalance\_handler.py** – oversampling (dupienie przykładów), undersampling, SMOTE.

 **text\_preprocessing.py** – lowercasing, usuwanie stopwords, tokenizacja, stemming/lemmatyzacja (NLTK/spacy).

 **image\_preprocessing.py** – dla deep learningu: obrót, przycięcie, normalizacja, konwersja kolorów.

 **pipeline\_utils.py** – spina wszystkie kroki w sekwencję (np. pipeline = [clean\_missing, encode, scale]).

 **save\_load.py** – zapisujesz gotowe dane do pliku .pkl albo .parquet i odpalasz w innym projekcie.