

PROJEKT z przedmiotu Systemy BI
(realizacja w zespołach 2 osobowych)

1. Na podstawie wybranego zbioru danych z repozytorium „UCI Machine Learning” dokonać klasyfikacji binarnej.
<https://archive.ics.uci.edu/datasets>
2. Wybrany zbiór danych powinien mieć co najmniej 1000 obserwacji i co najmniej 10 atrybutów numerycznych.
3. Oprócz bazowych atrybutów w tabeli, proszę skonstruować kilka innych dodatkowych cech (min. 5) np. współczynniki cecha x_1/x_2 itp.
4. Zadanie polega na budowie modelu klasyfikacyjnego (binarnego) przy wykorzystaniu poznanych technik – drzewa klasyfikacyjne, k-najbliższych sąsiadów, sieci neuronowe czy inne (min. 2 różne techniki). Staramy się zbudować dobry model klasyfikacyjny biorąc pod uwagę następujące miary klasyfikacji: trafność, czułość, specyficzność, AUC, wszystko na zbiorze walidacyjnym.
5. Projekt należy przygotować w formie prezentacji/pdf – max 12 slajdów. W projekcie musi się znaleźć informacja o zbiorze danych, dodatkowo utworzonych atrybutach (min. 5), wyniki prezentujemy na zbiorze walidacyjnym (40% całego zbioru danych). Proszę zawrzeć w prezentacji różne spostrzeżenia związane z analizowanymi danymi i wykorzystywanymi technikami.
6. Proszę zwrócić uwagę na jakość prezentacji. Opracowanie powinno być czytelne, wraz z opisem danych użytych w projekcie, z wynikami klasyfikacji i wnioskami. Proszę także dołączyć kody/skrypty do przygotowania danych.
7. Czas realizacji: 10 grudnia 2025 (do końca dnia).
8. Projekt przesłać na tomasz_zabkowski@sggw.edu.pl, tytuł maila: Projekt SBI INF DZ 2025 (wystarczy, aby jedna osoba przesłała).
9. Informację o ocenie projektu każdy zespół otrzyma indywidualnie drogą mailową, do 20 grudnia 2025.