

**PROJEKT** z przedmiotu Wprowadzenie do Systemów 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/ml/datasets.php>
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).
4. Zadanie polega na budowie modelu klasyfikacyjnego przy wykorzystaniu poznanych technik – drzewa klasyfikacyjne, k-najbliższych sąsiadów, sieci neuronowe czy inne. Staramy się zbudować dobry model klasyfikacyjny biorąc pod uwagę następujące miary klasyfikacji: trafność, czułość, specyficzność, przyrost (lift) na zbiorze walidacyjnym.
5. Można pracować w dowolnym narzędziu (SAS EM, Phyton, R-project, różne darmowe pakiety do drzew decyzyjnych /klasyfikacyjnych/sieci neuronowych).
6. Projekt należy przygotować w formie prezentacji – max 12 slajdów, nie więcej niż 3MB. W ppt 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.
7. 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.
8. Czas realizacji: 4 grudnia 2022 (do końca dnia).
9. Projekt przesać na [tomasz\\_zabkowski@sggw.edu.pl](mailto:tomasz_zabkowski@sggw.edu.pl), tytuł maila: Projekt INF DZ 2022.