

1. Celem projektu jest praktyczne sprawdzenie metod klasyfikacji.
2. Rzeczywiste dane pochodzące od jednej z firm telekomunikacyjnych.
3. Zmienne zostały zanonimizowane ze względu na ochronę danych osobowych.
4. Celem jest identyfikacja klientów którzy korzystają z oferty handlowej (zmienna `class=1`). Wartość `class=0` oznacza że klient nie skorzystał z oferty.
5. W pliku `train.txt` znajdują się dane treningowe (40000 klientów), które mogą być wykorzystane do budowy modelu.
6. Należy dokonać predykcji dla danych ze zbioru `testx.txt` (10000 klientów), każdemu klientowi przypisując **prawdopodobieństwo skorzystania z oferty**.
7. Wyniki należy zapisać do pliku, którego nazwa to: 3 pierwsze litery imienia+ 3 pierwsze litery nazwiska + ".txt". Przykładowo, student Jan Losowy zapisze wyniki do pliku `JANLOS.txt`. W pierwszej linijce powinien się znaleźć kod nazwiska (tu: "JANLOS") a w kolejnych prawdopodobieństwa odpowiadające kolejnym klientom ze zbioru `testx.txt`. Przykładowy plik: `JANLOS.txt`.
8. Dane potrzebne do wykonania projektu znajdują się a stronie: <https://home.ipipan.waw.pl/p.teisseyre/TEACHING/ZMUM/index.html>.
9. Należy przetestować co najmniej 4 metody klasyfikacji.
10. Projekty są wykonywane w zespołach 1 osobowych.
11. Ocena na podstawie:
  - jakości klasyfikacji mierzonej jako precyzja obliczona dla 10% najwyżej ocenionych przez model klientów (z najwyższymi prawdopodobieństwami), tzw. `prec @ 10%` (50 %),
  - prezentacji (5 minut) podsumowującej wyniki (25 %),
  - raportu (maksymalnie 3 strony a4) który zawiera: podsumowanie eksperymentów, uzasadnienie wyboru końcowej metody, opis przetwarzania danych (25 %).
12. Prezentacje odbędą się:
  - 7 maja na zajęciach projektowych (wtorek o 10.00),
13. Plik z wynikami, prezentacje, raport i kody programu należy wysłać na adres: `teisseyre(at)ipipan.waw.pl`. Ostateczny termin wysyłania:
  - 3 maja, godzina 23:59:59,Wysłanie po terminie będzie skutkowało naliczeniem karnych punktów.
14. Własne pomysły na użycie niestandardowych metod klasyfikacji oraz przetwarzania danych (metody selekcji zmiennych, przekształcania zmiennych, uzupełniania braków danych, itp.) będą mile widziane.
15. Projekt może być zrealizowany przy użyciu dowolnego narzędzia.