

1. Celem projektu jest praktyczne sprawdzenie metod klasyfikacji.
2. Dane dotyczą recenzji leków.
3. Do każdej recenzji mamy przypisaną ocenę (zmienne rate i rate1) dotyczącą leku. Zmienna rate przyjmuje wartości 1-10 (10 maksymalna ocena). Zmienna rate1 przyjmuje wartości: low, medium, high. Przykładowy rekord danych:

name	condition	opinion	rate	rate1
Ethosuximide	Seizures	This medicine is very good at controlling me so I can work part time Got 3 weeks supply left then guess they switch me and I go through seizures again	9	high

4. Celem jest zbudowanie modelu na podstawie danych treningowych który będzie przewidywał ocenę (zmienne rate i rate1) na podstawie zmiennych name, condition, opinion.
5. W pliku **Train.csv** znajdują się dane treningowe (150K recenzji), które mogą być wykorzystane do budowy modelu (dla tych danych są podane wartości zmiennych rate i rate1).
6. Należy dokonać predykcji dla danych ze zbioru **TestX.csv** (10K recenzji), każdej recenzji przypisując opinie (wartość low, medium lub high).
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 kolumnie należy zapisać wyniki dla rate a w drugiej dla zmiennej rate1. W pierwszym wierszu znajdują się nagłówki: rate i rate1. W kolejnych wierszach znajdują się przewidywane oceny dla poszczególnych recenzji ze zbioru **testX.csv**. Przykładowy plik: **JANLOS.txt**.
8. Dane potrzebne do wykonania projektu znajdują się na stronie: <https://home.ipipan.waw.pl/p.teisseyre/TEACHING/ZMUM/index.html>.
9. Projekty są wykonywane w zespołach (maksymalnie) 2 osobowych.
10. Ocena na podstawie:
 - dokładności predykcji i klasyfikacji dla zmiennych rate i rate1 (50 %),
 - prezentacji (10 minut) podsumowującej wyniki (25 %),
 - raportu (maksymalnie 3 strony a4) który zawiera: podsumowanie eksperymentów, uzasadnienie wyboru końcowej metody, opis przetwarzania danych (25 %).
11. Prezentacje odbędą się:
 - 26 maja na zajęciach projektowych (wtorek o 10.00),
12. Plik z wynikami, prezentacje, raport i kody programu należy wysłać na adres: [teissey-rep\(at\)ipipan.waw.pl](mailto:teissey-rep(at)ipipan.waw.pl). Ostateczny termin wysyłania:
 - 22 maja, godzina 23:59:59,

Wysłanie po terminie będzie skutkować naliczeniem karnych punktów.

13. Własne pomysły na użycie niestandardowych metod klasyfikacji oraz przetwarzania danych (definicje zmiennych, metody selekcji zmiennych, przekształcania zmiennych itp.) będą mile widziane.
14. Projekt może być zrealizowany przy użyciu dowolnego narzędzia.