

Wprowadzenie

11 października 2016

1 Informacje organizacyjne

- Kontakt:
 - kjasinska . . .
 - prefix tytułu emaila: [MMDS]
 - dyżur: wtorki, 15:40 - 16:40, 2.7.5 BT
- Strona przedmiotu <http://www.cs.put.poznan.pl/kdembczynski/lectures/mmds-challenge>
- Materiały na zajęcia projektowe <http://www.cs.put.poznan.pl/kjasinska/lectures/mmds-challenge/>

Forma zajęć

- Podpowiedzi dotyczące kolejnych etapów pracy podczas konkursu,
- Dyskusje:
 - możliwe rozwiązania zadań konkursowych,
 - użyte algorytmy (i osiągnięte rezultaty),
 - napotkane problemy (i ich rozwiązania?),
 - otwarte problemy i pojawiające się pytania.

Plan przedmiotu

- Przetwarzanie danych (Raport 1)
- Modele predykcyjne (Raport 2)

2 Background

- Doświadczenie z uczeniem maszynowym czy przetwarzaniem danych,
- Znane oraz preferowane języki programowania,
- Narzędzia do przetwarzania danych czy uczenia maszynowego.

3 Podstawowe informacje o zbiorze danych ★

Treść

Pobierz dane udostępnione pod adresem <http://www.cs.put.poznan.pl/kjasinska/lectures/mmds-challenge/data/lab-01/>: cechy przykładów X.csv oraz ich etykiety y.csv. Z użyciem wybranych przez siebie narzędzi oblicz i przedstaw poniższe informacje o zbiorze danych:

1. liczba cech i liczba przykładów,
2. liczba niezerowych wartości cech,
3. dominująca klasa oraz jej prawdopodobieństwo,
4. dla każdej cechy sprawdź średnie i wariancje, liczbę unikalnych wartości oraz przedstaw histogram przyjmowanych wartości. Wskaż cechy z jedną dominującą wartością, binarne i ciągłe.

Zwróć uwagę na czas przetwarzania. Na następnych zajęciach, 18 października, przedstawiamy wyniki i wnioski z zadania: napotkane problemy i ich rozwiązania.