

Eksploracja masywnych danych  
**Projekt – Sprawozdanie II**  
Modele predykcyjne

Imię Nazwisko    Imię Nazwisko

10 stycznia 2017

## 1 Wprowadzenie

Celem raportu jest uzyskanie w metodyczny sposób rozwiązania obu problemów omawianych podczas zajęć..

## 2 Let's Roq

Należy wykonać zadania opisane w sekcji 4 z użyciem danych udostępnionych w konkursie Let's Roq pod adresem <https://inclass.kaggle.com/c/letsroq>.

## 3 Data Ninja

Należy wykonać zadania opisane w sekcji 4 z użyciem danych udostępnionych w konkursie Data Ninja pod adresem [dataninja.olx.pl](http://dataninja.olx.pl).

## 4 Wymagania

Treść raportu dotycząca każdego z problemów ma zawierać następujące sekcje:

### 4.1 Sformułowanie problemu

Krótki i konkretny opis **zadania** konkursowego. Proszę również dokładnie zdefiniować miary, według której oceniane są rozwiązania w konkursie.

(maks. 1/2 strony)

## 4.2 Rozwiązanie bazowe

Proszę opisać co jest rozwiązaniem bazowym dla danego problemu. Jako rozwiązanie bazowe rozumiemy tutaj model predykcyjny mający do dyspozycji jedynie informację o klasach w zbiorze treningowym, bez informacji o cechach, podejmujący decyzję zgodnie z wiedzą na temat dostępnych danych treningowych oraz zgodnie z miarą oceny użytą w konkursie. Proszę uzasadnić wybór rozwiązania bazowego.

(maks. 1 strona)

## 4.3 Opis metody predykcyjnej

Szczegółowy metod wykorzystywanych do rozwiązania konkursu. Należy użyć przynajmniej jednego algorytmu uczenia modelu predykcyjnego. Dla każdego użytego algorytmu należy, poprawnie metodyczne i sensownie, przeprowadzić dobór parametrów.

Należy zwrócić szczególną uwagę na jakość metodyki przeprowadzenia doboru parametrów i walidacji wyników.

Wykonując predykcje należy uwzględnić miarę jakości użytą w konkursie. Proszę opisać i wyjaśnić sposób uwzględnienia miary jakości (lub funkcji straty).

Wyniki zaproponowanej metody należy odnieść do wyników osiąganych przez rozwiązanie bazowe.

(maks. 4 strony)

## 4.4 Wyniki eksperymentalne

W tym punkcie należy umieścić i opisać wyniki uzyskane w konkursie. Proszę również umieścić wyniki uzyskiwane podczas pracy nad rozwiązaniem na danych treningowych i przy doborze parametrów algorytmów uczących (np. przy wykorzystaniu walidacji krzyżowej lub zbioru walidującego). Należy zamieścić krótkie komentarze oraz tabele z wynikami (Tabela 1 jest przykładem jak stworzyć tabelę w  $\text{\LaTeX}$ ’u).

Wyniki mają dotyczyć porównania paru wariantów jednej metody lub porównania wariantów różnych metod:

- należy podać wyniki uzyskane podczas walidacji na dostępnych danych oraz wyniki otrzymane na zbiorze treningowym,
- należy przedstawić wyniki metod na mierze wykorzystanej w konkursie,
- należy odnieść uzyskane wyniki do rozwiązania bazowego,
- należy również podać czas uczenia się metody i czas obliczania predykcji.

To są kolumny	zawierające np.	wartość RMSE itp.
Opis i liczby:	99.99%	1.00
Opis i liczby:	99.99%	0.92
Opis i liczby:	99.99%	0.88

Tablica 1: Wyniki eksperymentalne

Do utworzenia modelu predykcyjnego można użyć wszystkich danych dostępnych w konkursie. Wartości miary oceny w konkursie mają dotyczyć zbioru testowego, na podstawie wartości podanej przez sprawdzarki udostępnionej w konkursie.

(maks. 3 strony)

#### 4.5 Podsumowanie

Na końcu jest zawsze miejsce na krótkie podsumowanie (maks. 1/2 strony).

**Całość raportu nie może przekraczać 16 stron.**