

MMDS Challenge

Raport I

Imię Nazwisko Indeks

Imię Nazwisko Indeks

Imię Nazwisko Indeks

22 października 2017

1 Opis zadania

Celem pracy nad raportem jest zapoznanie się z danymi. Raport powinien stanowić podsumowanie działań podjętych w celu zrozumienia zjawiska jakiego dotyczą dane. Powinien zawierać otrzymane rezultaty, wraz z ich analizą, komentarzem i odpowiednimi wnioskami.

2 Oddawanie raportów

Raporty należy przygotować w formacie PDF. Do wykonania raportów należy użyć dostarczonego szablonu w \LaTeX .

Termin oddania raportów to 22 listopada, godzina 9 rano, (dla grupy środowej) i 24 listopada, godzina 9 rano, (dla grup piątkowych). Raporty należy przesłać mailem na adres kjasinska@cs.put.poznan.pl, w tytule maila powinny się znajdować "[MMDS] Raport 1". W raporcie należy umieścić imiona, nazwiska oraz numery indeksu wszystkich autorów raportu.

Przypominam, że w przypadku niedostarczenia raportu na czas otrzymuje się za niego 0 punktów, z możliwością przyniesienia raportu w późniejszym terminie (do 17 stycznia 2018). Ostateczna punktacja za raport jest wtedy średnią z punktów z obu terminów.

Na zajęciach 22/24 listopada studenci będą przygotowywać na podstawie raportów wspólne prezentacje na wykład. Na zajęcia każda grupa powinna mieć ze sobą wyniki własnej pracy nad raportem, aby sprawnie opracować wspólną prezentację grupy laboratoryjnej.

3 Przykładowe zagadnienia

Przykładowe zagadnienia jakie można zgłębić podczas eksploracji danych podane są poniżej. Punkty oznaczone * są obowiązkowe dla wszystkich, po-

zostałe stanowią przykłady i mogą być zastąpione innymi wybranymi przez studenta.

- Histogramy (łącznie z wykresem) liczby wyświetleń, odpowiedzi oraz sprzedaży (*),
- Czy istnieją pewne zależności pomiędzy liczbą wyświetleń, odpowiedzi oraz sprzedażą ogłoszenia? (najprościej: korelacje + wykres) (*)
- Czy istnieje cyklicznie powtarzalny tygodniowy wzorec liczby zapytań w systemie? (*) Czy jest on różny dla pewnych kategorii lub zapytań (proszę pominąć rzadkie kategorie i zapytania)?
- Czy pewne zapytania są bardziej popularne w różnych sezonach? (*)
- Czy można zaobserwować różnice pomiędzy danymi treningowymi a testowymi (np. liczba ogłoszeń w różnych kategoriach, częstość określonych zapytań)? (*)
- Czy ceny w ogłoszeniach z wybranych kategorii zmieniają się zależnie od sezonu (np. ceny rowerów czy liczba ogłoszeń dotyczących sprzedaży rowerów w zależności od miesiąca),
- Czy można zauważyć pojawienie się nowych produktów analizując te dane? Czy pojawiają się w nowe częste zapytania, które nie miały miejsca wcześniej?

Raport powinien zawierać przynajmniej jedną dodatkową, zaproponowaną przez projektową grupę studentów, analizę lub przynajmniej jedną z zaproponowanych (bez *). Na zajęciach po oddaniu raportów grupy laboratoryjne będą wspólnie przygotowywać prezentacje na wykład (jedna prezentacja od każdej grupy). Grupy zostaną wyróżnione dodatkowymi punktami za najciekawsze prezentacje.

4 Uwagi

Do wykonania zadania należy użyć **wszystkich danych** udostępnionych w ramach konkursu, tzn. danych ze wszystkich dostępnych miesięcy.

Należy podać nie tylko otrzymane wyniki, ale również zwrócić uwagę na **czas przetwarzania**. Jakie kroki zostały podjęte aby zapewnić możliwie krótki czas?

Należy zwrócić szczególną uwagę na **istotność** otrzymanych wyników:

- czy nie wyciągamy wniosków z szumu i losowych zjawisk (patrz pierwsze zajęcia z MMDS),

- czy zaobserwowane zjawiska są powtarzalne (przykładowo w przypadku wahań cyklicznych, np tygodniowych wzorców liczby zapytań w systemie),
- czy dane w różnych okresów są istotnie różne.

Raport powinien zawierać krótki opis podjętych działań i przeprowadzonych badań. Nie należy opisywać użytych narzędzi czy użytych metod. Należy przytoczyć ich nazwę, najlepiej także anglojęzyczną aby uniknąć nieśpójności wynikających w różnych tłumaczeń.

5 Punktacja

Za raport można uzyskać do 50 punktów. Oceniane będą:

- staranność wykonania,
- poprawność wyników,
- czy zostały przeprowadzone dodatkowe analizy, nie zaproponowane na zajęciach,
- jakość analizy wyników i wniosków.

Całość raportu nie może przekraczać 10 stron.