

ASEiED - projekty zaliczeniowe 2018

1. Stwórz własną implementację z wykorzystaniem Spark/Scala/Java/PySpark umożliwiającą obliczenie wartości średniej temperatury i wiatru na podstawie danych pochodzących ze stacji meteorologicznych w U.S.
 - a. Grupa A: <https://www.ncdc.noaa.gov/orders/qclcd/199607.tar.gz>
 - b. Grupa B: <https://www.ncdc.noaa.gov/orders/qclcd/199806.tar.gz>
 - c. Grupa C: <https://www.ncdc.noaa.gov/orders/qclcd/200003.tar.gz>
 - d. Grupa D: <https://www.ncdc.noaa.gov/orders/qclcd/200705.tar.gz>
2. Implementacja algorytmu w Spark/Scala/Java/PySpark polegająca na policzeniu osób o tym samym imieniu pracujących w tej samej firmie i tym samym mieście. Przedstawienie na wykresie statystyk dotyczących występowania określonych imion w danej lokalizacji.
 - a. Grupa A:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:grupaa_datamay-31-2017.json.zip
 - b. Grupa B:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:grupab_datamay-31-2017.json.zip
 - c. Grupa C:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:grupab_datamay-31-2017.json.zip
3. Zaimplementowanie algorytmów (selection-sort, quicksort) h z wykorzystaniem Spark/Scala/Java/PySpark. Proszę o optymalne wykorzystanie Contextu Sparka do realizacji tego zadania. Wyniki działania aplikacji prezentujemy w formie wygenerowanego pliku wynikowego. Dokonanie porównania czasu działania algorytmu.
 - a. Grupa A:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z3_grupaa_datamay-31-2017.json.zip
 - b. Grupa B:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z3_grupab_datamay-31-2017.json.zip
 - c. Grupa C:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z3_grupac_datamay-31-2017.json.zip
 - d. Grupa D:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z3_grupad_datamay-31-2017.json.zip

4. Implementacja naiwnego klasyfikatora Bayesowskiego z wykorzystaniem Spark/Scala/Java/PySpark. Proszę o optymalne wykorzystanie Contextu Sparka do realizacji tego zadania. Wyniki działania aplikacji prezentujemy w formie wykresu przedstawiającego zakwalifikowany punkt na tle zbiorów.
 - a. Grupa A:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z4_grupaa_datamay-31-2017.json.zip
 - b. Grupa B:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z4_grupab_datamay-31-2017.json.zip
 - c. Grupa C:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z4_grupac_datamay-31-2017.json.zip
 - d. Grupa D:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z4_grupad_datamay-31-2017.json.zip
5. Implementacja algorytmu k-means (https://en.wikipedia.org/wiki/K-means_clustering) z wykorzystaniem Spark/Scala/Java/PySpark. Aplikacja powinna przyjmować parametr pozwalający określić ilość grup. Wyniki działania aplikacji prezentujemy w formie wykresu (każda grupa zaznaczona innym kolorem).
 - a. Grupa A:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z5_grupaa_datamay-31-2017.json.zip
 - b. Grupa B:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z5_grupab_datamay-31-2017.json.zip
 - c. Grupa C:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z5_grupac_datamay-31-2017.json.zip
 - d. Grupa D:
http://ksdir.org/ac/lib/exe/fetch.php?media=ase:z5_grupad_datamay-31-2017.json.zip