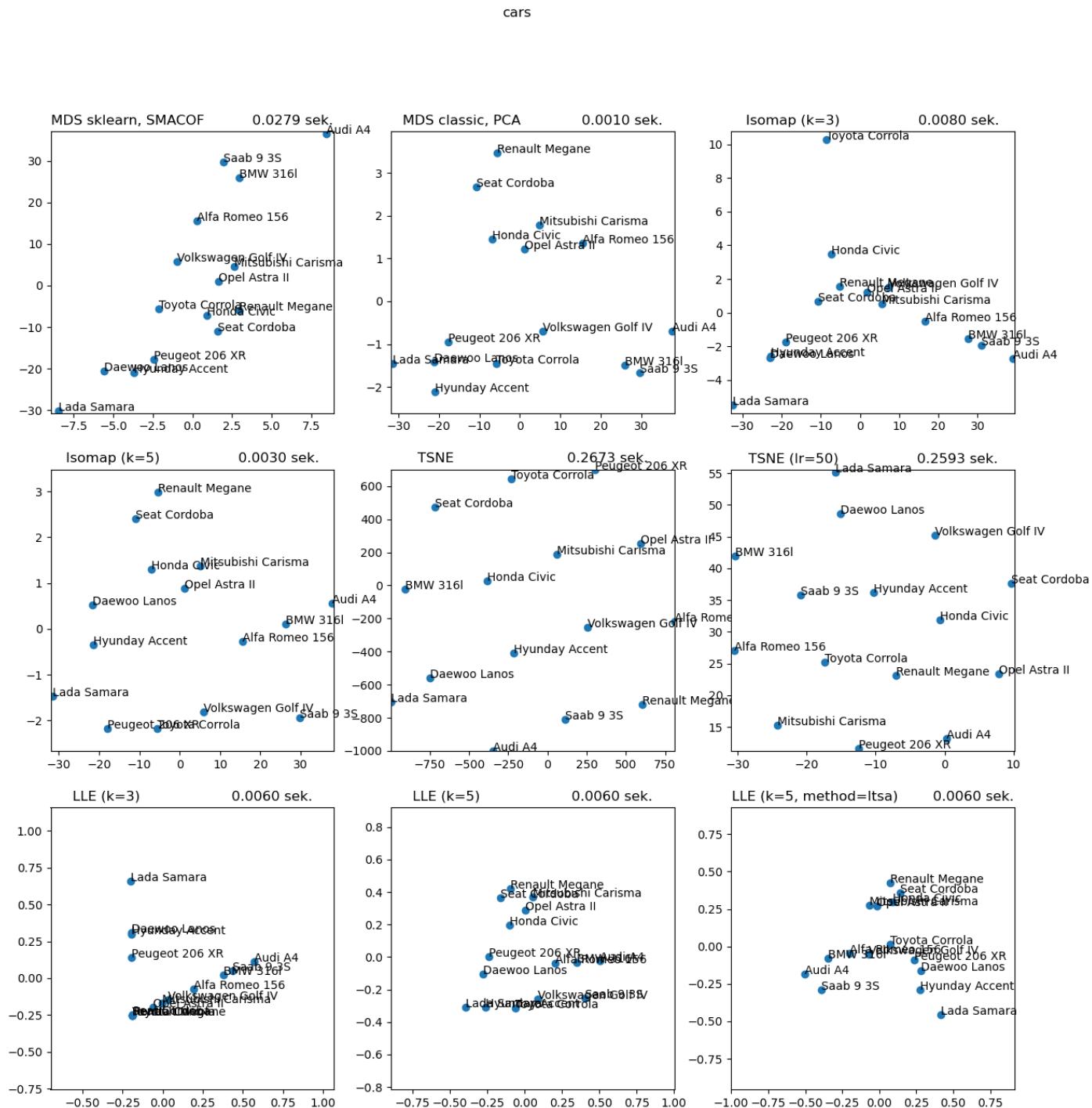


Sprawozdanie z laboratorium nr 5

Prowadzący: Iwo Błądek

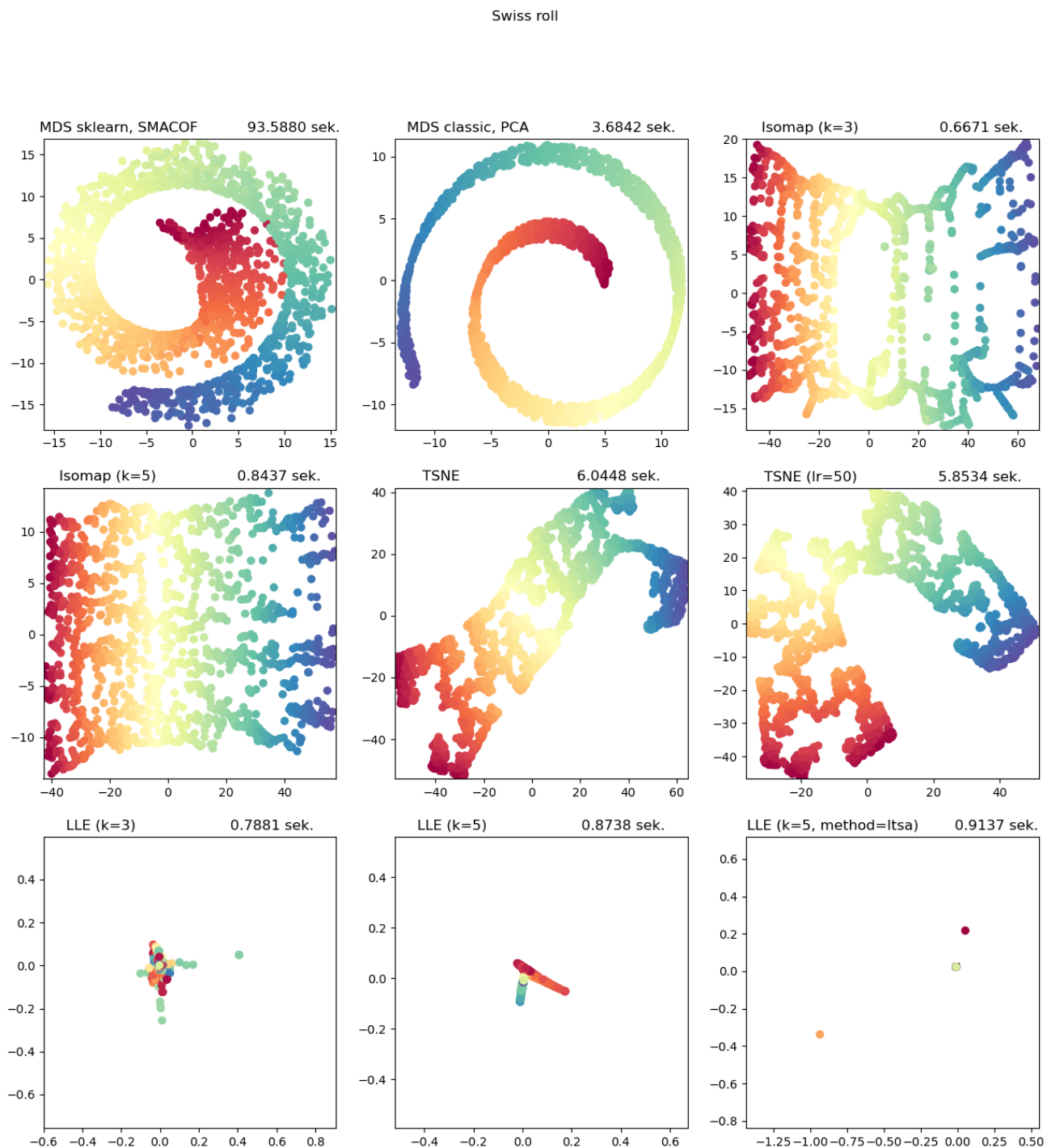
Imię i nazwisko: Jakub Tomczak, nr indeksu: 127083

1. Wykresy dla zbioru 'cars.csv'.



Rysunek 1: Wykresy dla zbioru 'cars.csv'.

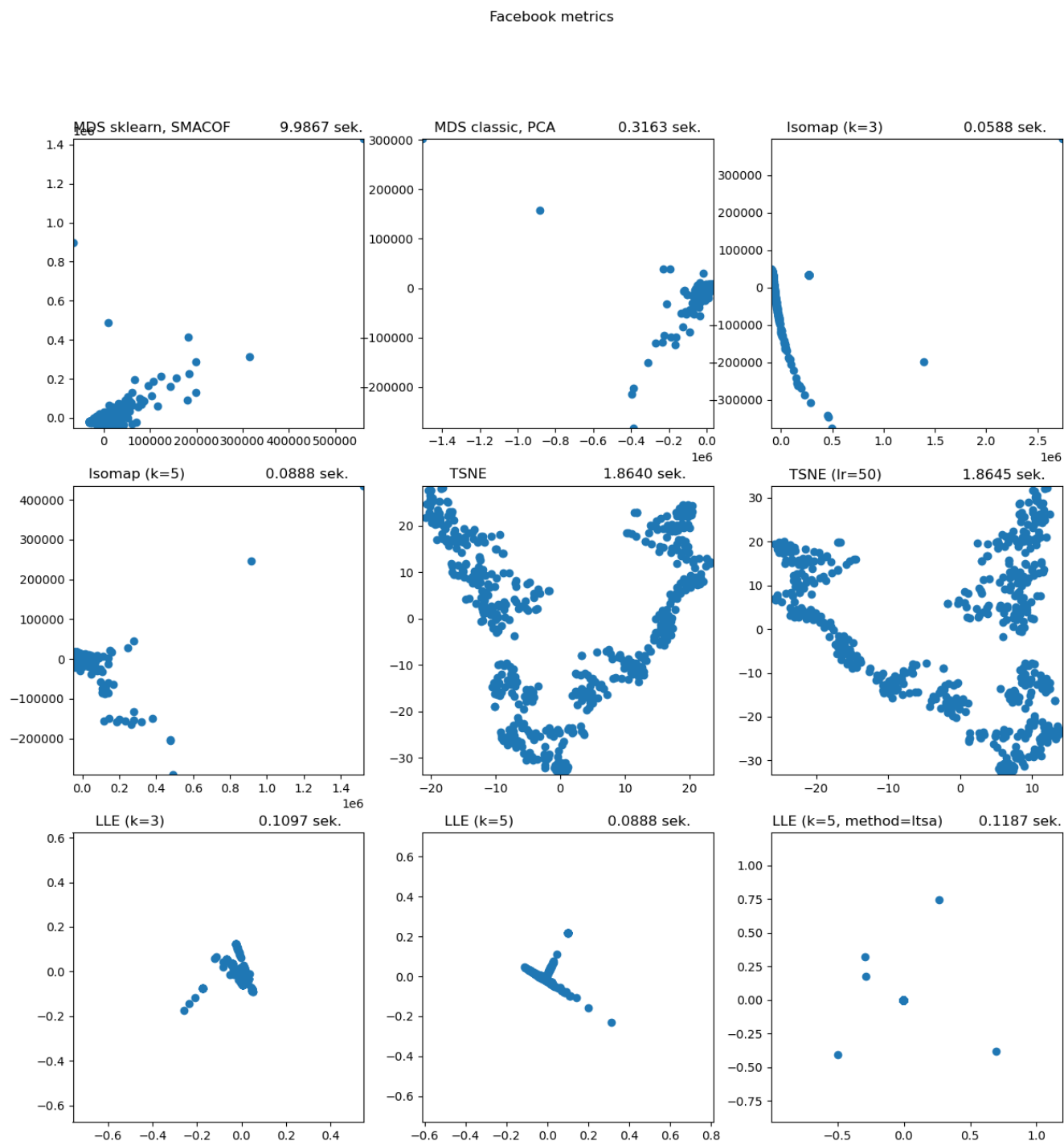
2. Wykresy dla Swiss rolls.



Rysunek 2: Wykresy dla Swiss roll.

Zbiór danych zawierał 1600 próbek dla krzywej S.

3. Wykresy dla zbioru Facebook metrics [2] [1].



Rysunek 3: Wykresy dla zbioru 'Facebook metrics'.

Zbiór ten został pobrany ze strony UCI [2]. Dane w nim zawarte są różnymi metrykami z platformy Facebook zbieranymi dla stron kosmentycznych. Dane służyły do predykcji metryk zasięgu postów upublicznianych na stronach firmowych.

4. Podsumowanie.

Po zastosowaniu 5 różnych podejść, które dały 9 różnych konfiguracji nie widać jednego najlepszego i ogólnego sposobu na redukcję wymiarowości. Oczywiście do pełnej oceny redukcji należałoby dodać jakieś pomiary jakości, jednakże po wykresach danych przetransformowanych widać duże dysproporcje pomiędzy zbiorami danych. Dla przykładu, MDS z użyciem PCA dosyć dobrze odseparowuje dane dla zbioru 'cars' ale dla zbioru 'Facebook metrics' już tworzy jedną, dosyć mocno zbitą grupę. Z drugiej strony, algorytm Isomap wymaga dostrojenia parametru k - związanego z liczbą sąsiadów, dla zbioru cars widać, że jest duża różnica między $k=3$, a $k=5$.

Dodatkowo, występują duże różnice między czasem uruchomienia algorytmów. Najwolniej zawsze wypada MDS korzystający ze SMACOF. Drugi jest algorytm TSNE, a reszta ma dosyć podobne czasy uruchomienia. Jeżeli chodzi o PCA, które jest moją implementacją może być pewna różnica jeżeli użyjemy tego z biblioteki sklearn.

Literatura

- [1] Sérgio Moro, Paulo Rita, and Bernardo Vala. Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. *Journal of Business Research*, 69(9):3341 – 3351, 2016.
- [2] P. Rita S. Moro and B. Vala. Facebook metrics dataset. <http://archive.ics.uci.edu/ml/datasets/Facebook+metrics>.