# Juliusz Wasieleski Informatyka, III rok, grupa 6 marzec 2024

Podstawy uczenia macierzowego – laboratorium nr 2 – sprawozdanie

# Opis ćwiczenia

Naszym zadaniem było wykonanie analizy zbioru Bone marrow mononuclear cells with AML w programie Orange. Trzeba było:

1. Zaznajomić się ze zbiorem
2. Odpowiedzieć na pytanie dlaczego taki zbiór nie kwalifikuje się do użycia modeli ML
3. Zwizualizować zbiór i podać najlepszy wynik jaki uzyskał model MLP
4. Wyznaczyć funkcję złożoności modelu w zależności od zmniejszającej się liczby cech
5. Zbadać zależność jakości klasyfikacji od
   1. Zmniejszającej się liczby przykładów
   2. Zwiększającej się liczby cech

# Środowisko, biblioteki, założenia oraz użyte narzędzia

Ćwiczenie wykonałem w programie Orange Data Mining oraz przy użyciu arkuszy kalkulacyjnych Excel*.*

Wszystkie obliczenia prowadziłem na komputerze Lenovo Y50-70 z systemem Windows 10 Pro w wersji 10.0.19045, procesor Intel Core i7-4720HQ 2.60GHz, 2601 MHz, rdzenie: 4, procesory logiczne: 8.

# Rozwiązanie

## Analiza zbioru

Zbiór zbioru *Bone marrow mononuclear cells with AML* opisuje ekspresję genów w komórkach jednojądrzastych szpiku kostnego od pacjenta z ostrą białaczką szpikową (AML) i dwóch zdrowych dawców użytych jako kontrola. Dane obejmują próbkę 1000 komórek i 1000 genów o największej dyspersji. Są to przykładowe dane dostarczane z przeglądarką Loupe Cell Browser i obejmują komórki z trzech oddzielnych eksperymentów z zestawami danych opublikowanymi na stronie zestawów danych pojedynczych komórek 10x Genomics: AML027 BMMCs przed przeszczepem, zamrożone BMMCs (Healthy Control 1) i zamrożone BMMCs (Healthy Control 2).

Rozmiar: 1000 wierszy 1004 kolumny

Cechy: 1000 cech numerycznych

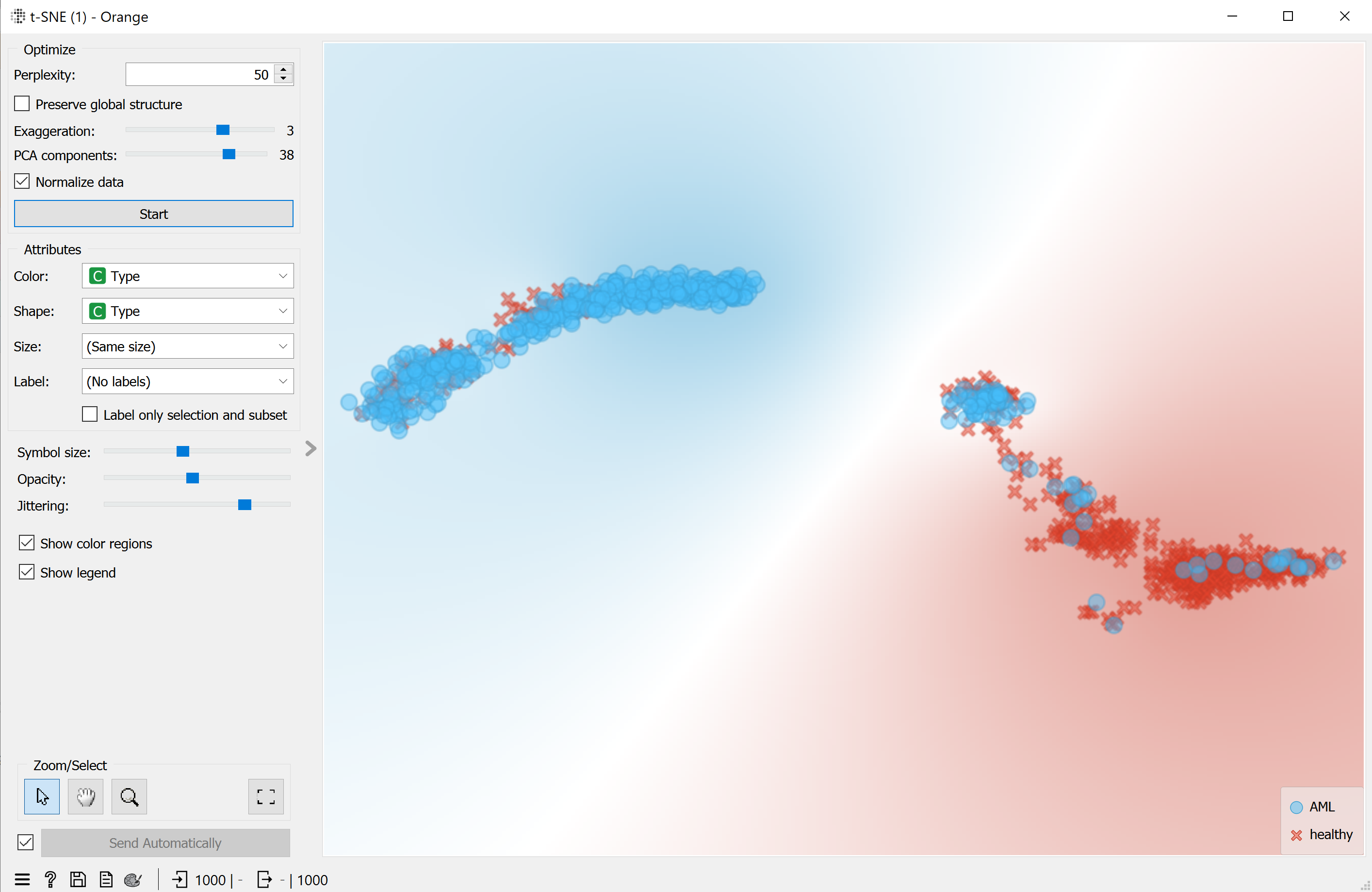
Cel: Kategoryzacja na dwie klasy

Jako, że dwie klasy danych to pacjent zdrowy i pacjent z białaczką, trenując model na podstawie tego zbioru możemy chcieć aby odpowiadał on na pytanie czy na podstawie występowania danego genu pacjent jest chory czy nie.

## 3.2 Dlaczego ten zbiór nie kwalifikuje się do użycia ML

## Wizualizacja zbioru i najlepszy model

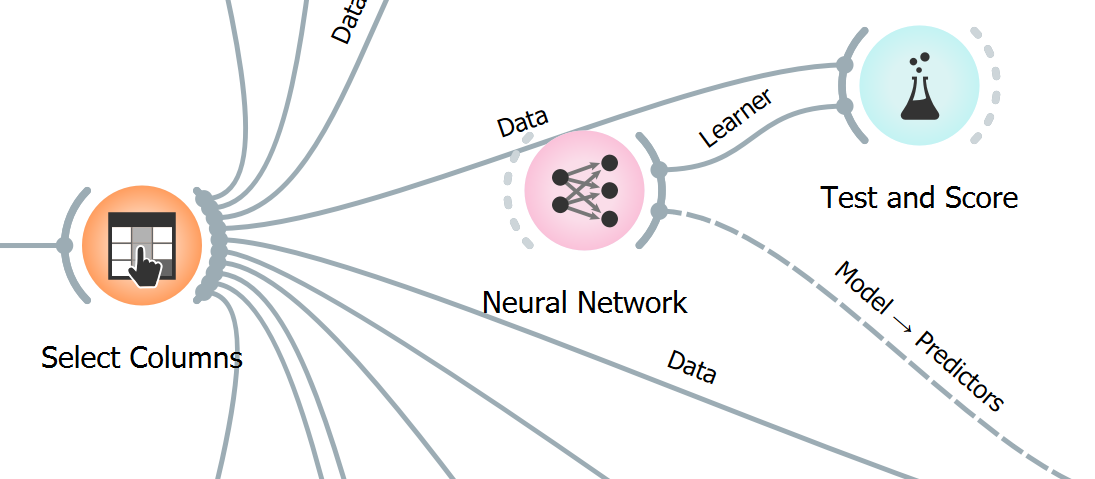
### 3.3.1 Rozseparowanie klas

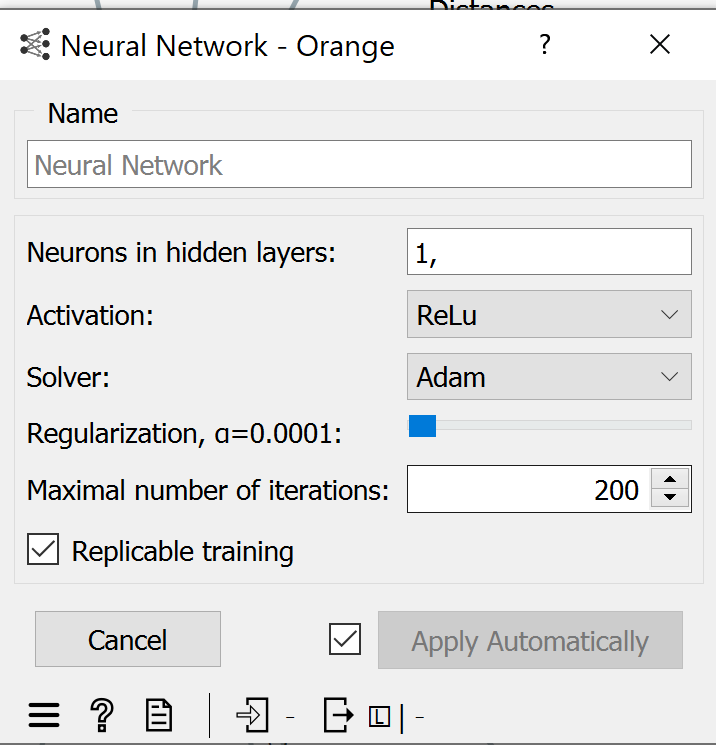
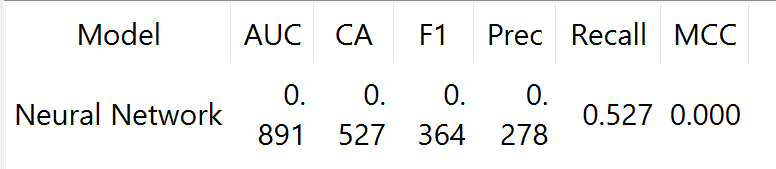


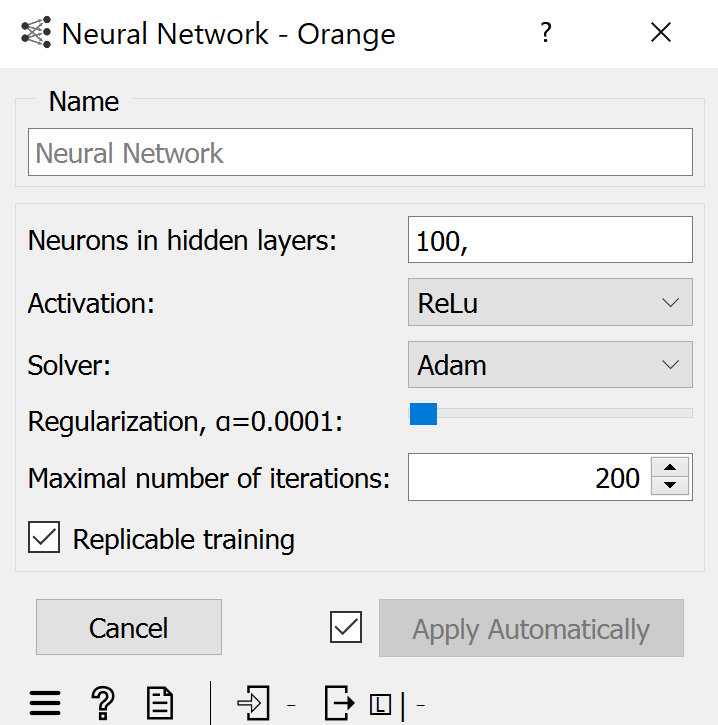
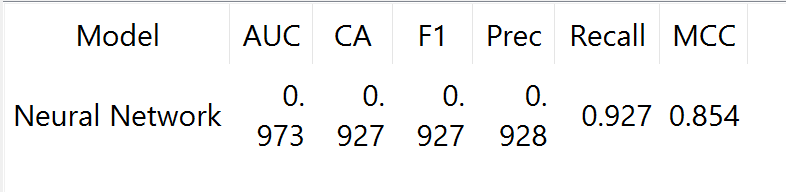
Rys.1 Wykres rozseparowania klas wśród danych

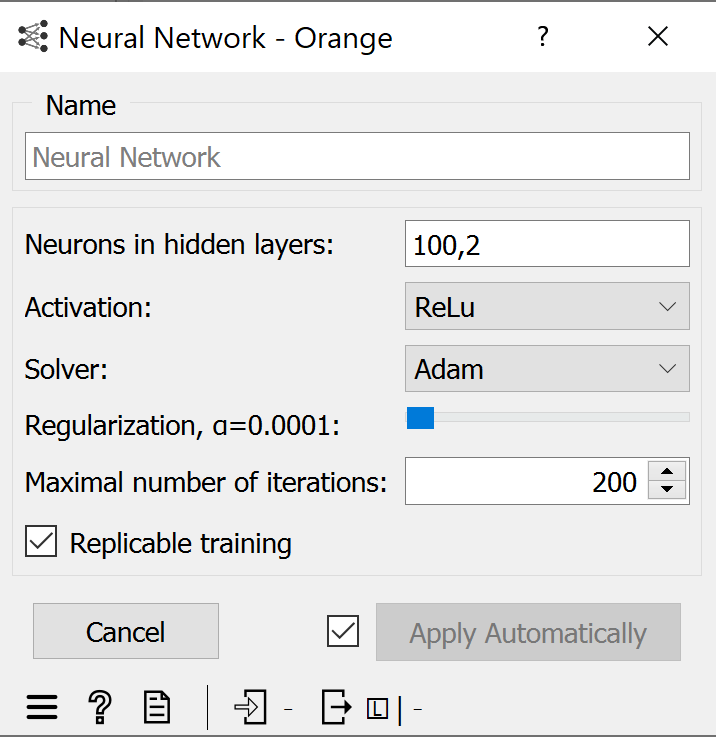
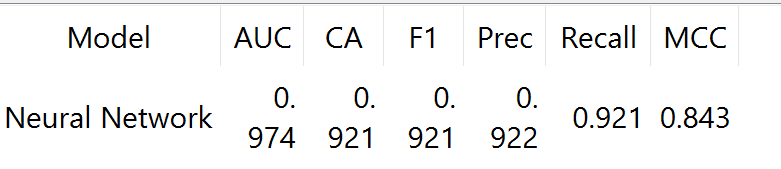
Nie jest tak źle z tym rozseparowaniem i dalszy opiiis.

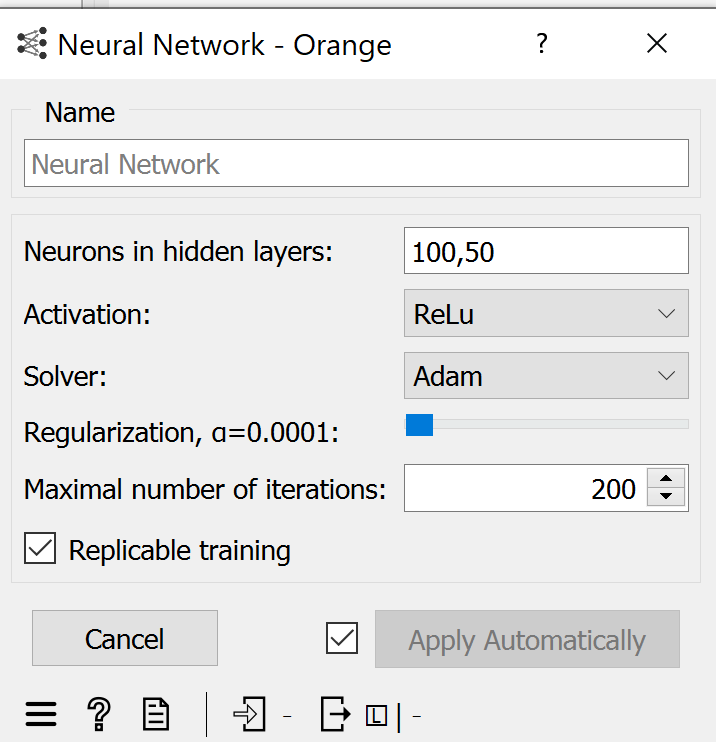
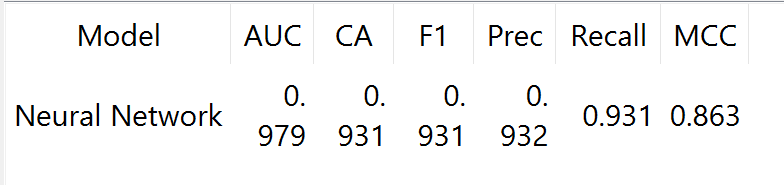
### 3.3.2Najlepszy model

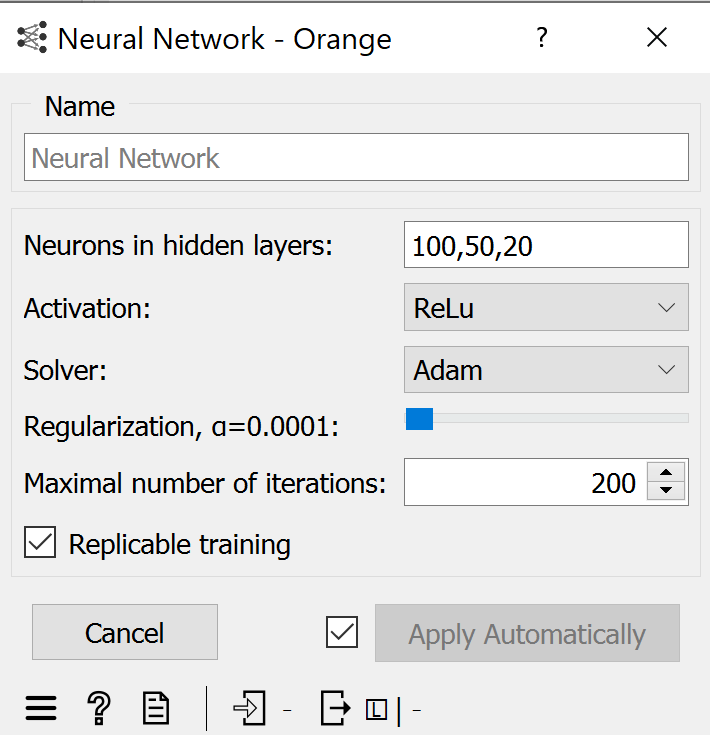
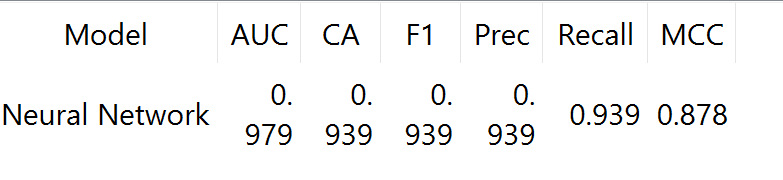


## Zależność złożoności modelu od liczby cech

## Analiza jakości klasyfikacji

### 3.5.1

### 3.5.2

# Wnioski