WOJSKOWA AKADEMIA TECHNICZNA IM. JAROSŁAWA DĄBROWSKIEGO W WARSZAWIE

***Projekt z przedmiotu „Metody uczenia maszynowego II”***

**Autor:** *Jakub Kapusta*

**Grupa:** *I9B2S4*

**Prowadzący*:*** *dr Jarosław Olejniczak*

**Wybrany język:** R

Spis treści:

[1. Opis wybranego zbioru danych 3](#_Toc61343139)

[2. Wnioski końcowe 3](#_Toc61343140)

# Opis wybranego zbioru danych

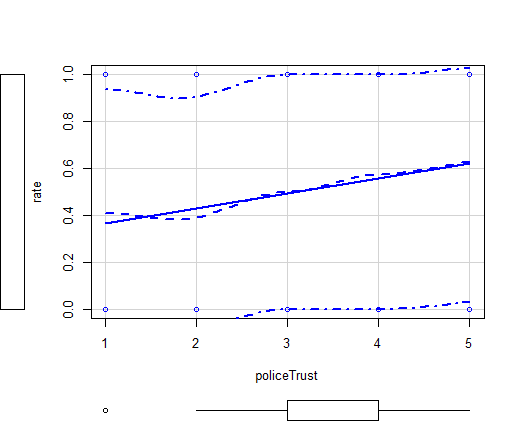
Jako zbiór danych wybrałem **Somerville Happiness Survey Data Set** pobrany ze strony <https://archive.ics.uci.edu/ml/datasets/Somerville+Happiness+Survey>. Zbiór ten przedstawia wyniki ankiety przeprowadzonej w 2015 roku wśród losowych mieszkańców Somerville. Pytania dotyczyły oceny mieszkańców ich szczęścia i satysfakcji ze służb miejskich.

Opis atrybutów podsumowujących wyniki ankiety przedstawia Tabela 1.

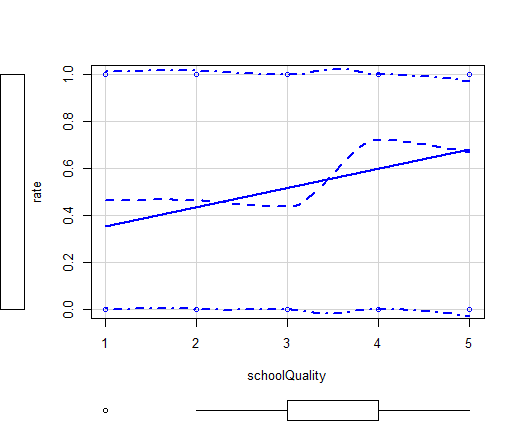
Tabela 1 Opis atrybutów opisujących zbiór danych

|  |  |  |  |
| --- | --- | --- | --- |
| **Oznaczenie atrybutu w pliku** | **Oznaczenie atrybutu w tworzonych modelach** | **Znaczenie atrybutu** | **Możliwe wartości** |
| D | rate | Ocena mieszkańca (czy jest szczęśliwy, czy nie) | 0 (nieszczęśliwy)  1(szczęśliwy) |
| X1 | cityServiceInfoAvailability | Dostępność informacji o służbach miejskich | 1  2  3  4  5 |
| X2 | housingCost | Koszty zakwaterowania |
| X3 | schoolQuality | Jakość szkół publicznych |
| X4 | policeTrust | Zaufanie do lokalnej policji |
| X5 | infrastructureMaintance | Utrzymanie ulic i chodników |
| X6 | eventsAvailability | Dostępność wydarzeń społecznościowych |

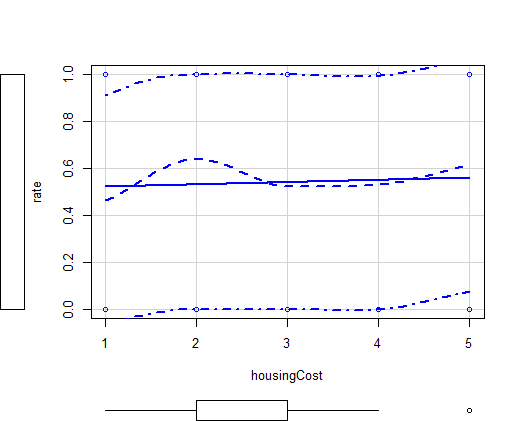
# Analiza wstępna (wizualna) zbioru



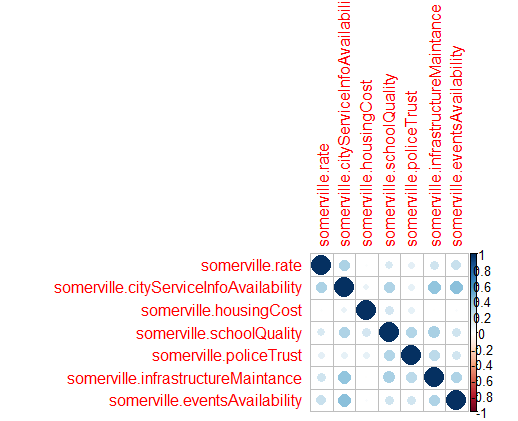
Rysunek 1 Wykres zależności rate od policeTrust. Z wykresu wynika, że wraz ze wzrostem policeTrust wzrasta również rate.



Rysunek 2 Wykres zależności rate od schoolQuality. Z wykresu wynika, że wraz ze wzrostem schoolQuality wzrasta również rate.

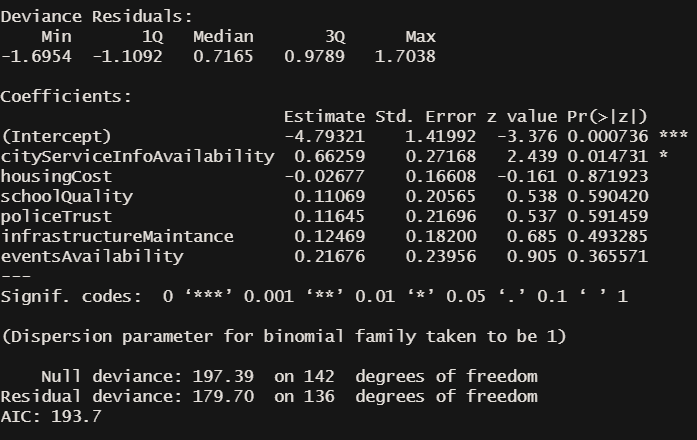


Rysunek 3 Wykres zależności rate od housingCost. Z wykresu wynika, że housingCost nie wpływał na rate.

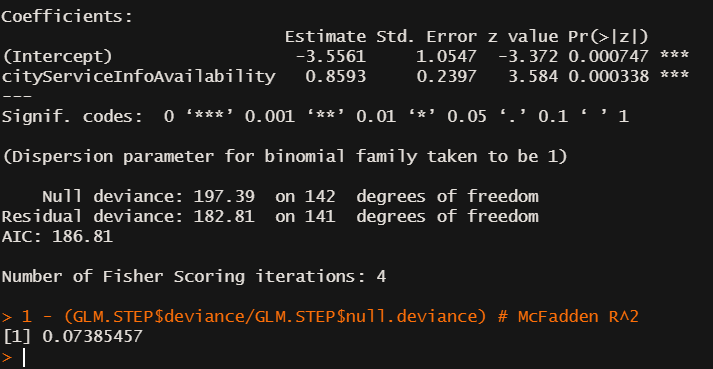


Rysunek 4 Korelacja zmiennych. Z rysunku wynika, że zmienne nie są ze sobą silnie skorelowane. Najsilniej skorelowane ze zmienną objaśnianą jest cityServiceInfoAvailibility.

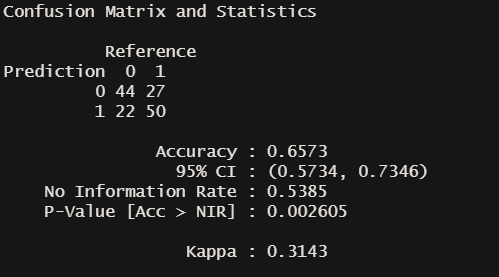
# Liniowy model regresji



Rysunek 5 Podsumowanie liniowego modelu regresji.



Rysunek 6 Linowy model regresji po redukcji zmiennych nieistotnych.



Rysunek 7 Macierz pomyłek dla modelu regresji liniowej (po redukcji zmiennych nieistotnych).

# Sieci neuronowe

# Podsumowanie utworzonych modeli

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Pakiet** | **Współczynnik Accuracy** | **Uwagi** |
| Regresji liniowej | - | 0.6573 | Model po redukcji zmiennych nieistotnych przy użyciu algorytmu SVS |
| Sieci neuronowe | Neuralnet |  |  |