Zadanie 5

Clusters

Bc. Peter Tokovics

Obsah

[2 Social Networking Ads 3](#_Toc130326416)

[2.1 Logická regresia 3](#_Toc130326417)

[2.1.1 Training set 3](#_Toc130326418)

[2.1.2 Testing set 4](#_Toc130326419)

[2.2 Neurónová sieť 4](#_Toc130326420)

[3 Churn Modeling 5](#_Toc130326421)

[3.1 Logická regresia 5](#_Toc130326422)

[3.1.1 Training set 5](#_Toc130326423)

[3.1.2 Testing set 5](#_Toc130326424)

[3.2 Neurónová sieť 6](#_Toc130326425)

# Kód

Obrázok, na ktorom je text

Automaticky generovaný popis

Na začiatku načítame dataset, následne dataset osekáme podľa hodnoty z cyklu. Hodnota z cyklu hovorí o tom koľko percent z datasetu chceme použiť. Potom vypočítame WCSS a pomocou neho vypočítame optimálny počet klastrov. Vykreslíme si WSCC kde budeme schopný vyčítať optimálny počet klastrov. Potom pomocou optimálneho počtu klastrov vytvoríme závislí premennú ktorú použijeme pri vizualizácií.

# Vizualizácie

## 10 %

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

## 20 %

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

## 35 %

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

## 50 %

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

## 70 %

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

## 100 %

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

Obrázok, na ktorom je tabuľka

Automaticky generovaný popis

# Záver

Podľa výsledných vizualizácií môžeme vidieť že ako sa nám zväčšoval počet záznamov z datasetu, ktoré sme používali za zväčšovala aj presnosť jednotlivých klastrov. Samozrejme to bolo ovplyvnené aj tým že sme vyberali náhodné záznamy bola by tam šance že vyberieme zrovna tie, ktoré nám vytvoria presne centroidy.

Ideálny počet klastrov sa vo všetkých prípadoch približoval k 5-tim. To nemusí platiť pri inom datasete.