

Analiza adaptacji studentów do nauki zdalnej z pomocą sieci bayesowskiej

Krawiec Piotr
Inżynieria i analiza danych, 3 Rok

29/12/2021

Wstęp

W roku 2020 większość placówek edukacyjnych rozpoczęła naukę w formie zdalnej. Spowodowało to wiele trudności zarówno po stronie uczniów jak i prowadzących. Celem tej pracy jest analiza czynników, które miały wpływ na dopasowanie się uczniów do nowej sytuacji z pomocą sieci bayesowskiej.

Agenda

- Dane
- Tworzenie struktury sieci
- Analizy

Dane

Dane to zbiór zawierający wyłącznie dane kategoryczne (factors w R). I składa się z kolumn:

- Gender (Girl/Boy) - płeć ucznia
- Age (1 to 5/6-10/11-15/16-20/21-25/26-30/30+) - przedział wiekowy
- Education.Level (School/College/University) - poziom edukacji
- Institution.Type (Non Government/Government) - typ szkoły
- IT.Student (Yes/No) - czy to student IT
- Location (Yes/No) - czy uczy się i mieszka w tym samym mieście
- Load.shedding (Low/High) - niestabilność sieci elektrycznej, częstotliwość zaników prądu

Dane - ciąg dalszy

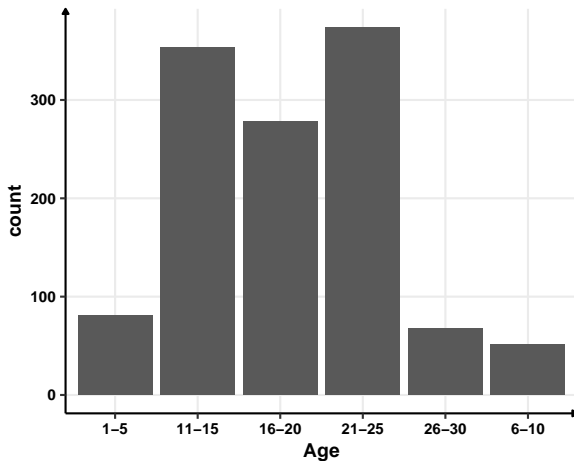
- Financial.Condition (Poor/Mid/Rich) - kondycja finansowa ucznia
- Internet.Type(2G/3G/4G) - rodzaj połączenia internetowego wykorzystywanego do nauki
- Device (Tab/Mobile/Computer) - urządzenie wykorzystywane podczas zajęć
- Network.Type (Mobile Data/Wifi) - rodzaj połączenia z internetem
- Class.Duration (0/1-3/3-6 hours) - ilość godzin lekcyjnych dziennie
- Self.Lms (Yes/No) - czy szkoła ma własny e-learning
- Adaptivity.Level (Low/Moderate/High) - poziom adaptacji ucznia

Załadowanie danych w R

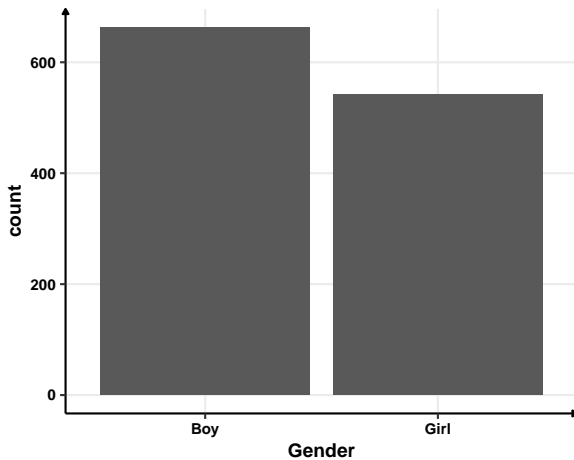
```
library(tidyverse)
library(tibble)
library(bnlearn)
library(lattice)
library(Rgraphviz)

df <- read.csv("datasets/dataset.csv")
col_names <- names(df)
df[] <- lapply(df[col_names], as.factor)
```

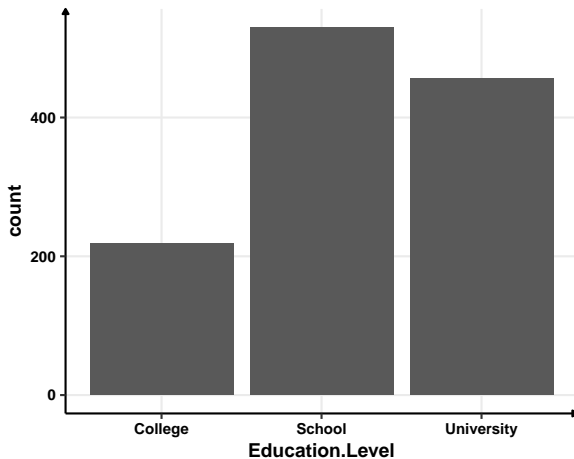
Wizualizacje



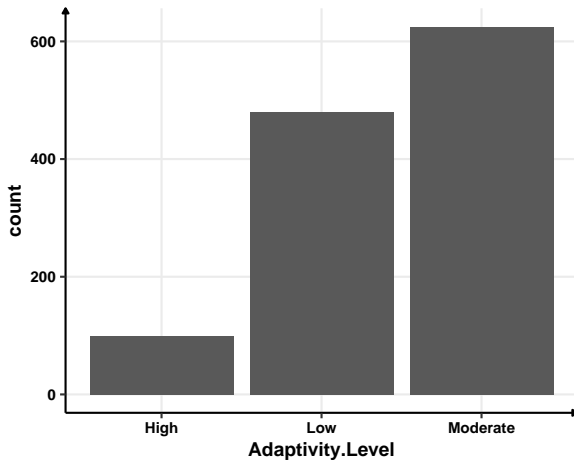
Wizualizacje



Wizualizacje



Wizualizacje

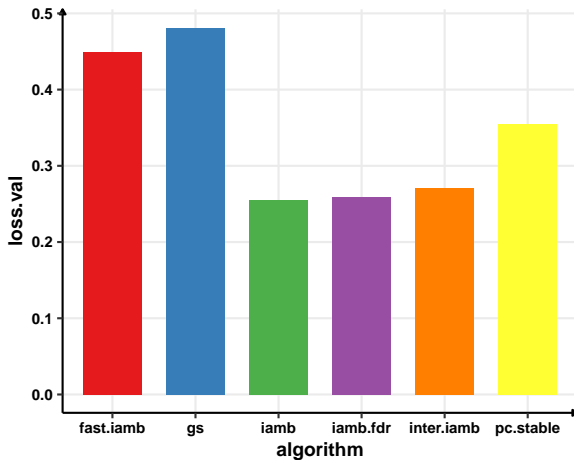


Tworzenie struktury sieci

Aby znaleźć najlepszą strukturę sieci wykorzystałem fakt, iż sieć może zostać wykorzystana nie tylko do wnioskowania, ale też predykcji. Wybiorę sieć, która dokonuje najlepszej predykcji.

Przy czym skorzystam też z czarnej listy i nie pozwolę na utworzenie krawędzi między wierzchołkami: Gender, Age, Class.Duration. Gdyż uważam, że zmienne te powinny być niezależne.

Porównanie algorytmów na zbiorze danych



Otoczka Markova

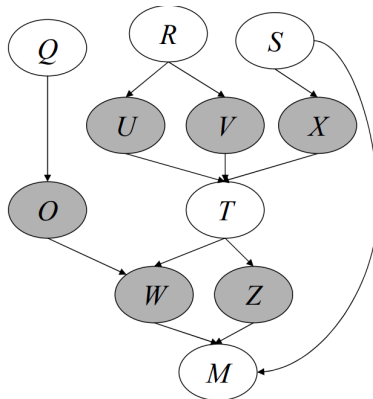
Dla zmiennej losowej T **otoczką markowa** (Markov Blanket) nazywamy zbiór zmiennych losowych, które wnoszą informację o zmiennej T . Usuwając jakąkolwiek zmienną z otoczki Markova tracimy informację o zmiennej T .

Mając zmienną losową T oraz zbiór $S = X_1, \dots, X_n$, otoczką Markova jest dowolny podzbiór S_1 , który spełnia warunek:

$$T \perp\!\!\!\perp S \setminus S_1 \mid S_1$$

W sieciach Bayesowskich, otoczka markowa wierzchołka T zawiera jego rodziców, dzieci oraz wszystkich rodziców jego dzieci.

Otoczka Markowa



Rysunek 1: Otoczka Markowa wierzchołka T , $MB(T)$

Algorytm IAMB (Incremental Association Markov Blanket)

Faza pierwsza. Szacujemy, które wierzchołki mogą należeć do **otoczki markowa** wierzchołka T tj. $MB(T)$, umieszczając je będziemy w CMB . Wierzchołek X umieścimy w CMD jeżeli:

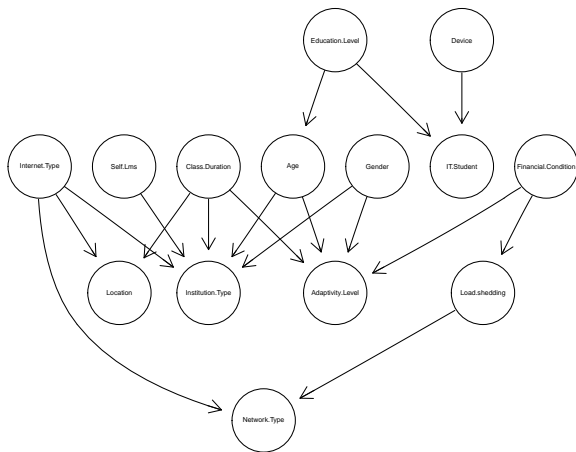
- maksymalizuje (zwiększa) on funkcję $f(X, T|CMB)$ będącą **Informacją wzajemną** zmiennych X i T pod warunkiem CMB
- oraz $X = S - T - CMB$ nie jest niezależna warunkowo od T pod warunkiem CMB .

W fazie drugiej usuwamy z CMD wszystkie wierzchołki X , dla których zachodzi:

$$I(X; T|CMB - \{X\})$$

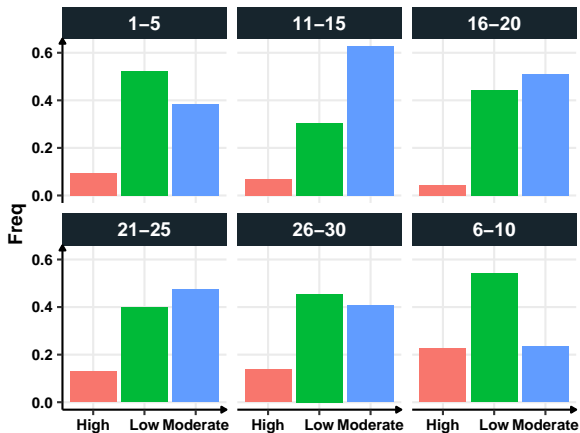
$I(X, T|Z)$ oznacza warunkową niezależność X od T pod warunkiem Z .

Struktura utworzonej sieci



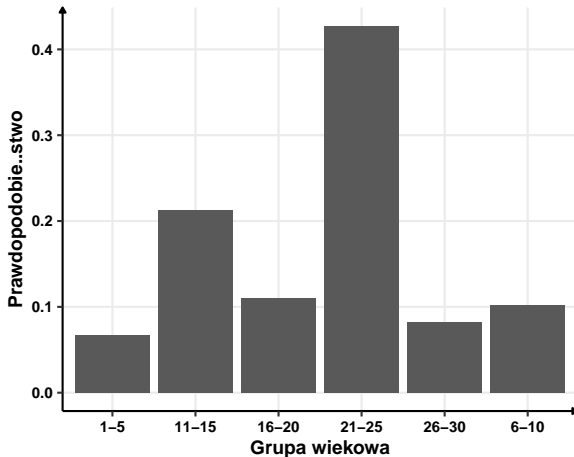
Wnioskowanie

W którym przedziale wiekowym poziom adaptacji jest największy?



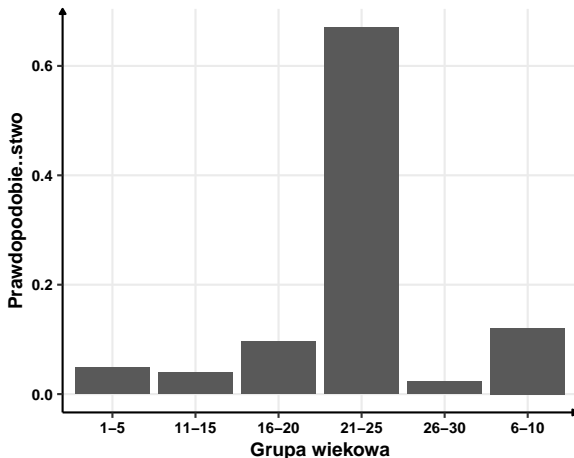
Wnioskowanie - wiek

Wiedząc, że poziom adaptacji jest “High”, do jakiej grupy wiekowej należała dana osoba?



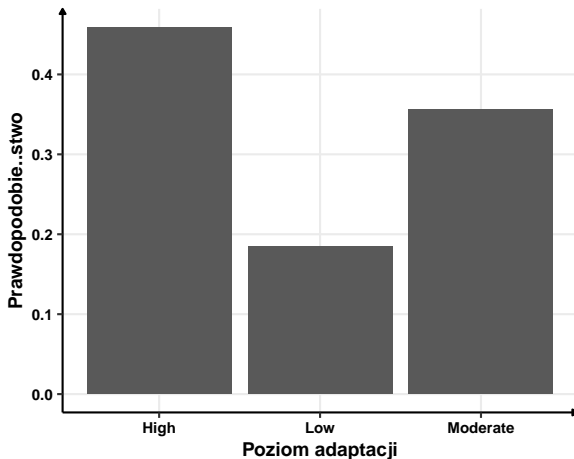
Wnioskowanie - wiek i kondycja finansowa

Wiedząc, że poziom adaptacji jest “High” oraz, że jej kondycja finansowa jest “Poor” do jakiej grupy wiekowej należała dana osoba?



Wnioskowanie - kondycja finansowa

Wiemy, że kondycja finansowa studenta/ucznia to “Rich”. Jaki jest rozkład prawdopodobieństwa poziomu adaptacji?



Wnioskowanie - kondycja finansowa

Wiemy, że kondycja finansowa studenta/ucznia to “Rich”. Jaki jest rozkład prawdopodobieństwa poziomu adaptacji pod warunkiem każdego z przedziałów wiekowych?

