**A picture containing text, clipart

Description automatically generatedA close-up of a logo

Description automatically generated with low confidenceУниверзитет у Нишу**

**Електронски факултет**

**Семинарски рад**

***Прикупљање и предобрада података за Машинско учење***

**Трансформација података**

**Петар Петковић**

**Ниш, 2023**.

Contents

[Увод 3](#_Toc128849272)

[Технике обраде података 4](#_Toc128849273)

[Идентификација проблема и дефинисање приступа 5](#_Toc128849274)

[Анализа података 6](#_Toc128849275)

[Закључак 7](#_Toc128849276)

# Увод

По дефиницији, трансформација података је процес претварања, чишћења и структурирања података у употребљив формат који се може анализирати и који је основа за креирање доброг модела [1].

У данашњем времену, дневно се генерише и до 118 зетабајта података [2]. Подаци се генеришу са разних сензора и на различите начине и начин записивања података у тренутку генерисања најчешће је у облику који није употребљив. Због тога, трансформација података постаје један од најважнијих корака у припреми података за анализу. Трансформација података обухвата низ техника које се примењују на сирове податке како би се побољшао њихов квалитет, време, прецизност и поузданост анализе. Трансформација података обично подразумева неколико техника, као што су:

* Чишћење података
* Нормализовање података
* Скалирање података
* Дискретизација података
* Агрегација података
* Промена расподеле
* Енкодирање категоричких атрибута
* Уклањање екстремних вредности (eng. outliers)

Циљ овог семинарског рада јесте дефинисати све могуће проблеме и аномалије који се могу наћи у почетном скупу података а затим детаљно дефинисање свих техника трансформација, када се примењују, у којим случајевима, који је резултат трансформације и због чега је тај облик погоднији за анализу. Након тога, проучити све примере примене одређене трансформације на одређену аномалију, добре и лоше стране и то све у циљу каснијег препознавања при реалним проблемима.

# Технике обраде података

Нормализација

Скалирање

Промена расподеле

Енкодирање категоричких атрибута

Дискретизација

Аутлајери

# Идентификација проблема и дефинисање приступа

# Анализа података

# Закључак

# References

|  |  |
| --- | --- |
| [1] | "Geeks for geeks," [Online]. Available: https://www.geeksforgeeks.org/data-transformation-in-data-mining/. |
| [2] | "FinancesOnline," [Online]. Available: https://financesonline.com/how-much-data-is-created-every-day/#:~:text=Every%20day%20Big%20Data%20statistics,(2021%2C%20February%209).. |