SWT

1. Introduction (SELINA)

ο Introduction of the topic: what is the problem?

ο Motivation: why the chosen topic is interesting and relevant?

ο Goals: what is it the paper strives to achieve? What is the contribution?

ο Outline: short presentation of the approach and structure of the paper KDD, erstes Protokoll, ausführlicher besprechen wie wir vorgehen/gedanken gemacht haben 1.Iteration, 2. Iteration

1. Preprocessing (Explorative Datenauswertung/ Analyse (nicht für beide Iterationen), Statistik, Ziel variable verteilung etc. FELIX/ALEX
2. 1. Iteration
   1. Transformation (N/A valuesFELIX und OutliersALEX, Data reduction(PCA ALEX, FEATURE SELECTIONSELINA, Encoding (scaling)MAX,
   2. Data mining (Anwendung der Algos, auswahl der algorithemn, Robust vs. Unrobost (2. Trainingssets) VERSTÄNDNISS AUFZEIGEN, HABEN GANZ VIEL GELERNT GASNZ TOLL FREDDY
   3. Evaluation (Maschinelle Sprachverarbeitung, argumentation, wo hackt es STARK ARGUMENTIEREN; SUPER GEIL) FREDDY/ MAX
3. 2. Iteration
   1. Transformation (N/A values und Outliers, Data reduction, Encoding (scaling),
   2. Data mining (ENSAMBLE, Anwendung der Algos, auswahl der algorithemn, Robust vs. Unrobost (2. Trainingssets) VERSTÄNDNISS AUFZEIGEN, HABEN GANZ VIEL GELERNT GASNZ TOLL GELESEN BESONDER VON LESSY
   3. Evaluation (Was hat sich verändert,)

6. Conclusion: 2.Seiten

Results

Discussion

ο Short summing up of the paper − What has been done? − What has been found?

ο Discussion of key implications (take-aways) ο Limitations/ ideas of future research

Outlook

# Introduction

* What is the problem?
* Motivation: why the chosen topic is interesting and relevant?
* Goals: what is it the paper strives to achieve? What is the contribution?
* Outline:
  + short presentation of the approach and structure of the paper / welche Gedanken gemacht / wie genau vorgehen
  + KDD (Source: The KDD Process for Extracting Useful Knowledge from Volumes of Data; U s a m a F a y y a d , G r e g o r y P i a t e t s k y - S h a p i r o , a n d P a d h r a i c S m y t h)
    - Framework for getting from data to knowledge
    - Keep in mind that it is not a binding sequence, but includes numerous iterations (for optimizations etc .)
    - Selection: select out of data the target data which should be discovered/analyzed
    - Preprocessing: exploratory data analysis and data cleaning (handle noise, outliers, missing values, transform 99 values)
    - Transformation: transform preprocessed data into transformed data (Data reduction: PCA, feature selection), encoding and projection (scaling…)
    - Data mining: choose DM function (clustering), algorithms (which to use for pattern search in data, models and parameters)
    - Interpretation: interpret results from data mining and possibly connections to previous steps: what can be improved?
  + 1.Iteration,
    - Focus on rather simple and fast methods
    - Decide with methods work well and which should be replaced by more complex ones
  + 2. Iteration
    - Try more complex methods
    - E.g. multidimensional outlier detection

# Preprocessing: Exploratory data analysis

# First iteration

## Preprocessing: Data cleaning

## Transformation

## Data mining

## Evaluation

# Second iteration

## Preprocessing: Data cleaning

## Transformation

## Data mining

## Evaluation

# Conclusion

## Results

## Discussion

## Outlook