Mining the Success for Movies

Student Project Data Mining HWS17
Team 6

Presented by

Steffen Jung
Adrian Kochsiek
Martin Koller
Marvin Messenzehl
Daniel Szymkowiak

October - December 2017

Contents

List of Algorithms

List of Figures
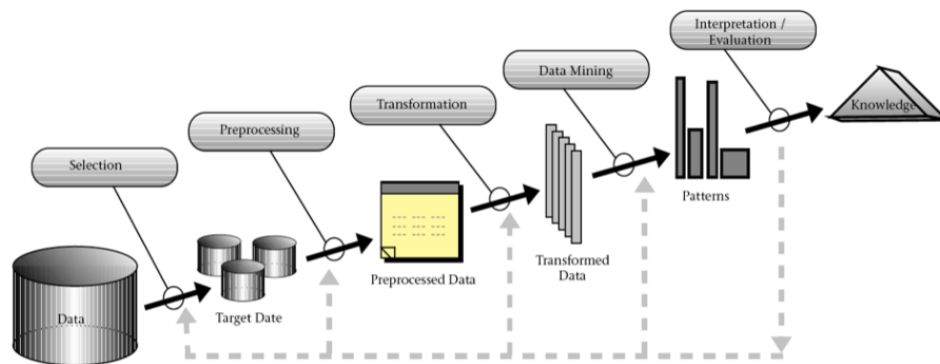
List of Tables

Chapter 1

Application Area and Goals

1.1 Introduction

- Problem Statement and idea behind the project

- General introduction similar to Project outline

1.2 Theoretical framework

- keep it small

- roughly 1 Page

Chapter 2

Data Selection

- In slides named: "structure and size of data"

- min. 1 Page

- Selection:

    - What data is available?
    - What do I know about the provenance of the data?
    - What do I know about the quality of the data?

- Exploration

    - Get an intitial understanding of the data
    - Calculate basic summarization statistics
    - Visualize the data
    - Identify data problems such as outliers, missing values, duplicate records

Chapter 3

Preprocessing and Transformation

- Transform data into a representation that is suitable for the chosen data mining methods

    - number of dimensions

    - scales of attributes (nominal, ordinal, numeric)

    - amount of data (determines hardware requirements)

- Methods

    - Aggregation, sampling

    - Dimensionality reduction / feature subset selection

    - Attribute transformation / text to term vector

    - Discretization and binarization

- Good data preparation is key to producing valid and reliable models

- Data preparation estimated to take 70-80% of the time and effort of a data mining project!

3

Chapter 4

Data Mining

- Input: Preprocessed Data

- Output: Model / Patterns

1. Apply data mining method

2. Evaluate resulting model / patterns

3. Iterate:

    - Experiment with different parameter settings
    - Experiment with different alternative methods Improve preprocessing and feature generation Combine different methods

4.1 Algorithms

Chapter 5

Interpretation / Evaluation

- Output of Data Mining

    - Patterns
    - Models

- In the end, we want to derive value from that, e.g.,

    - gain knowledge
    - make better decisions
    - increase revenue