# Dokumentation zu den Programme in Java Data Mining mit Entscheidungsbäumen

Inhalt

[Dokumentation zu den Programme in Java Data Mining mit Entscheidungsbäumen 1](#_Toc73560813)

[1. Einordnung des Themas 2](#_Toc73560814)

[2. Problematiken und Herangehensweise 2](#_Toc73560815)

[3. Kurze Zusammenfassung der Dokumentation 2](#_Toc73560816)

[4. Beschreibung der Daten sowie deren Herleitung und Beziehungen 3](#_Toc73560817)

[5. Beschreibung der Programme und Hilfs Funktionen 3](#_Toc73560818)

[1. Input/Output 3](#_Toc73560819)

[1.1. CSVReader 3](#_Toc73560820)

[1.2. XMLWriter 3](#_Toc73560821)

[6. Kurze Reflexion der Problematiken und Umsetzung sowie der Designentscheidungen. 3](#_Toc73560822)

## Einordnung des Themas

Dieses Programm ist in der Lage ein eine CSV-Datei einzulesen und auf Grundlage der eigelesenen Daten einen Entscheidungsbaum zu generieren. Hierbei wird der ID3 Algorithmus aus der Vorlesung genutzt, der durch Rekursion den Baum rekursiv generiert. Anschließend kann man den Baum als XML-Datei ausgeben lassen oder den Baum zur weiteren Informationsverarbeitung nutzen… NOT FINAL

## Problematiken und Herangehensweise

Auf Grundlage der Reorganisation hatten wir anfängliche Probleme mit der Organisation sowie der Synchronisation der Daten. Dieses Problem wurde dann schlussendlich von Herrn Moll, welcher die Nutzung von GitHub vorschlug, gelöst. Dies erlaubte es uns dann schnell und synchron and den Aufgaben im Team zu arbeiten.

Nach den anfänglichen Synchronisation Probleme, haben wir öfters Backtracking betreiben müssen, weil wir Fehler in Programmen hatten, die in späteren Iterationen komplett umgeschrieben werden mussten, diese Fehlerbehebung hat uns schlussendlich viel Zeit gekostet, uns aber auch die Möglichkeit gegeben uns näher mit verschieden Prozessen zu beschäftigen und diese zu verinnerlichen.

Es wurde versucht das Programm so effizient wie möglich zu haten, hierbei wurden Bibliotheken genutzt, die in Relation der Aufgaben am besten geeignet waren. Wir konnten somit effizient die Daten durch ein Buffer Reader einlesen. Der Umgang mit großen Daten wurde durch Stream abgewickelt, die es erlaubten sehr schnell über die Daten zu gehen.

## Kurze Zusammenfassung der Dokumentation

… NOT FINAL

## Beschreibung der Daten sowie deren Herleitung und Beziehungen

Zur vereinfachten Darstellung und Strukturierung, werden die Programme in Relation zu ihrer Anwendung gruppiert und beschrieben. … NOT FINAL

## Beschreibung der Programme und Hilfs Funktionen

### Input/Output

Hierbei handelt es sich im Allgemeinen um Programme die Daten ein- und auslesen. Diese Programme verarbeiten die Daten im ersten Schritt auf und geben sie zum Schluss zurück bzw. lesen Sie ein.

### CSVReader

… NOT FINAL

### XMLWriter

… NOT FINAL

## Kurze Reflexion der Problematiken und Umsetzung sowie der Designentscheidungen.