

**Kompensationsarbeit im Modul Software Entwicklung**

Master:

“Data Science & Intelligent Analytics”

Betreuerin :

Huber Stephan, MA

Autor:

Jochen Hollich | 1810837475

Datum

10.12.2019

Inhalt

**Es wurden keine Einträge für das Inhaltsverzeichnis gefunden.**

# Konzept

Bei der Evaluierung von Klassifikationsalgorithmen gibt es eine Vielzahl an Möglichkeiten, welche im Folgenden aufgelistet und beschrieben werden. Häufig muss jedoch die Qualität des Klassifikationsmodelels gegen die Random Baseline evaluiert werden. Die Random Baseline ist das Resultat aus dem „nackten Raten“. Diese Resultate hängen davon ab inwiefern in einem Datensatz welche Klasse wie oft vorkommt.

# Begriffe

Ballanziert = Wenn in einem Datensatz bspw 10 Klassen vorkommen, und jede einzelne Klasse n /10 vorkommt ist der Datensatz balanziert

Unballanziert = Nicht Ballanziert, möglichkeit bei NN

# Random Baseline

RB = Zufallswahrscheinlichkeit für das gesamte Model

Besonders bei MulticlassKlassifikation.

BSP MNIST(Datensatz mit 10 unterschiedlichen Klassen, jede Klasse tritt gleich häufig auf, wenn wir nun einfach immer „nicht 5“ prognositizieren, so sind wir in 90% der Fälle richtig)

# Random Baseline according to Zero

Weitere Entwicklung der RB, jedoch wird nur die Klasse mit dem höchsten Anteil und nicht die Summe gewählt

# Accuracy

Kann als Treffergenauigkeit beschrieben werden, Soll in der Modellierung Maximiert werden

# LOG-LOSS

# ROC

# AUC

# CONFUSION MATRIX

## Precesion/ Recall

## F1-Score