Theorie:

* CRISPDM
  + Phasen und Beschreibung (und Arbeitsschritte)

Praxis:

* Lineare Regression:
  + Grundformel
  + Methode der kleinsten Quadrate
  + Evaluation der Modelle
    - RMSE
    - Korrelation von Attributen -> Einfluss auf bestes Modell
    - Zusammenhang Modellkomplexität <-> RMSE (Overfitting z.B.)
* Entscheidungsbäume
  + Warheits-/Konfussionsmatrix
  + Kostenmatrix

Wahrheitsmatrix

|  |  |  |
| --- | --- | --- |
|  | True Positive | True Negative |
| Pred. Positiv | True Pos. | False Pos. |
| Pred. Negativ | False Neg. | True Negativ |

Präzision Positive: TP/(TP+FP)

Recall Positive: TP/(TP+FN) (Sensitivität)

Recall Negative: TN/(TN+FP) (Spezifität)

Baseline-Modell: Modell ohne Intelligenz als Referenzwert (z.B. alle als Positive einstufen)

Clustern nach Länge & Breite nach k-NN (NN=Next Neighbors)

Aufgabe: gegeben ist Koordinatensystem mit Messpunkten, Vorhersage der Klasse eines Clusters mit k Datenpunkten, ausgehend von n Ausgangspunkten.

Suche nach den k nächstgelegenen Datenpunkten, zählen der Klassenzugehörigkeit, dann Klasseneinteilung des Clusters nach am stärksten vertretener Klasse. Datenpunkte können zu mehreren Clustern gehören.