**Ablauf Auto-IA Workshop IJK Hannover**

**==============================**

**Lernziele:**

    - Dictionary Analyse:

        - fremdes Diktionär benutzen

        - eigenes Diktionär aufbauen

    - Topic Modeling:

        - Modell berechnen

        - Modell evaluieren

        - Modell interpretieren

     - Visualsierung von Ergebnissen

     - Validierung von Ergebnissen

     - ggf. Text-Klassifikation mit maschinellem Lernen

**Mögliche Fragestellungen für den Corona-Datensatz:**

- Wie verläuft die Konjunktur von Lockerungen und Lockdown in der Berichterstattung?

- In welchen Themen sind eher Männer bzw. Frauen repräesentiert? (Dictionary mit männl. / weibl. vornamen)

- Welche Länder stehen im Verlauf der Pandemie im Fokus der Berichterstattung? (Dictionary mit Ländernamen)

- Wie verändert sich das Themenvokabular, wenn Artikel entweder Hendrik Streek oder Christan Drosten zitieren? (STM mit  topical content covariate)

- Wie wird die die Regierungsarbeit von Jens Spahn im Verlauf der Pandemie öffentlich bewertet? (Sentiment dictionary vs. Klassifikation), <https://rdrr.io/github/kbenoit/quanteda.dictionaries/man/data_dictionary_Rauh.html>

- Alternativlosigkeitsrhetorik

**Breakout Räume:**

- entlang thematischer Filterung mit Topic Models?

- entlang Methoden?

    - dictionaries

    - classification

    - topic modeling

    - text2vec?

- Labeling-Tool vorbreiten, wenn Klassifikation gemacht werden soll?

**Tag 1: Einführung + vorbereitete Übungen**

    - 09:00 - 10:00: Vorstellungsrunde, Intro VS

    - 10:15 - 11:15: VL Preprocessing + Frequenzanalyse + Topic Model Intro

    - 11:30 - 12:30: erste Diskussion zu möglichen Fragen an das Korpus

    - MIttagspause

    - 13:30 - 14:30 Übung Preprocessing

    - 14:45 - 15:45 Übung Frequenzanalyse

    - 16:00 - 17:00 Fortsetzung Übung + Diskussion

**Tag 2: Weitere Übungen + Transfer auf neue Daten**

     - 09:00 - 10:00: Übung Topic Modeling,

     - 10:15 - 11:15: Übung Topic Model Evaluation

     - 11:30 - 12:30: Corona-Daten einlesen + erste Exploration

     - Mittagspause

     - 13:30 - 14:30: Gruppeneinteilung, Klärung Fragestellung/Operationalisierung

     - 14:45 - 15:45: Gruppenarbeit "Preprocessing"

     - 16:00 - 17:00: Präsentation des Zwischenstand + nächste Schritte im Plenum

**Tag 3: Arbeit mit eigenen Daten**

     - 09:00 - 10:00: Übung nach Bedarf (z.B. SVM-Klassifikation)

     - 10:15 - 12:30: Gruppenarbeit "Auswertung I"

     - Mittagspause

     - 13:30 - 14:30: Präsentation des Zwischenstand + nächste Schritte im Plenum

     - 14:45 - 17:00: Gruppenarbeit "Auswertung II", ggf. gemeinsames Datenlabeln

**Tag 4: Arbeit mit eigenen Daten**

     - 09:00 - 10:00: Übung nach Bedarf (z.B. SVM-Klassifikation)

     - 10:15 - 12:30: Gruppenarbeit "Auswertung III"

     - Mittagspause

     - 13:30 - 14:30: Gruppenarbeit "Validierung"

     - 14:45 - 15:45: Vorbereitung Abschlusspräsentation

     - 16:00 - 17:00: Abschlusspräsentation