Anforderungskatalog zur Definition des Datascience Life Cycles

Benötigte Informationen:

-Definition des Daten Managements (Datenformate, Datenstruktur, Datendichte, Datenablage, verwendete IT-Tools und deren Versionierung, Zugriffsschema)

-Definition der Git/GitHub und One-Drive-Struktur als zentrale Datenaustauscheinheiten

-Definition von ETL-Pipeline (Data Flow Diagramm, Datensemantik, Transformationsschema)

-Definition des Scrapping-Verfahrens, Labelingschema der Daten)

-Definition der Cloud-Struktur (Relationen der Datenbank,

-Definition der stat. Kennzahlen in der Versuchsplanung (Art des Versuchsplanes, Definition der Variablen, Standardabweichungen, Fehlerfortpflanzungen, stat. Power, FDS, Korrelationsgüte, p-Wert, AIC, BIC)

-Definition der Qualitätskriterien für die KI-Algorithmen (Stichprobenumfang, Güte der Voraussage, Güte des Models, Geschwindigkeit der Code-Ausführung)

-Anforderungen an Visualisierung der Daten (Ausführungszeit vs. Datenmenge, eindimensional vs. verschachtelt)