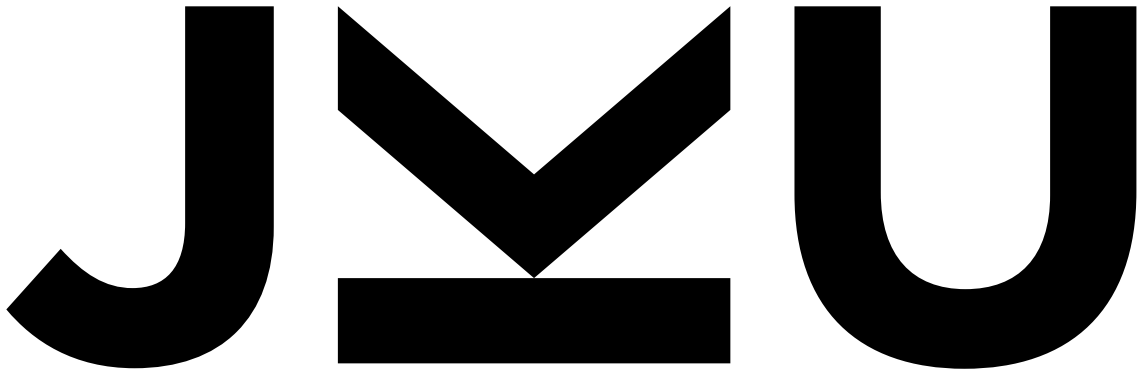
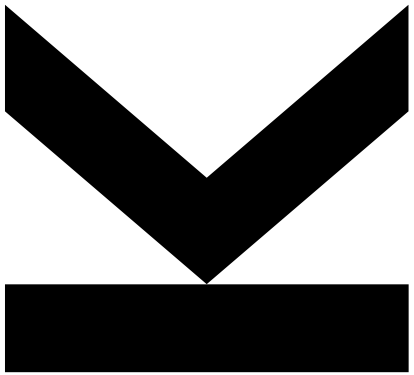


JOHANNES KEPLER
UNIVERSITY LINZ

Master Thesis – Design and Implementation of a web-based User Interface for the guided Assessment of Reliability of Classification Results using the Perturbation Approach



Speaker: Pascal Badzura
Supervisors: Assoz.-Prof. Mag. Dr. Christoph Schütz; Simon Staudinger, MSc
Master‘s Thesis Seminar (239.705)

JOHANNES KEPLER UNIVERSITY LINZ

Agenda

1. Motivation
2. Reference Process
3. Perturbation Approach
   1. Perturbation Options
   2. Example - Perturbation Table
4. Knowledge Graph
5. Methodical Approach
6. Goals
7. Status & Plan

Motivation

Motivation

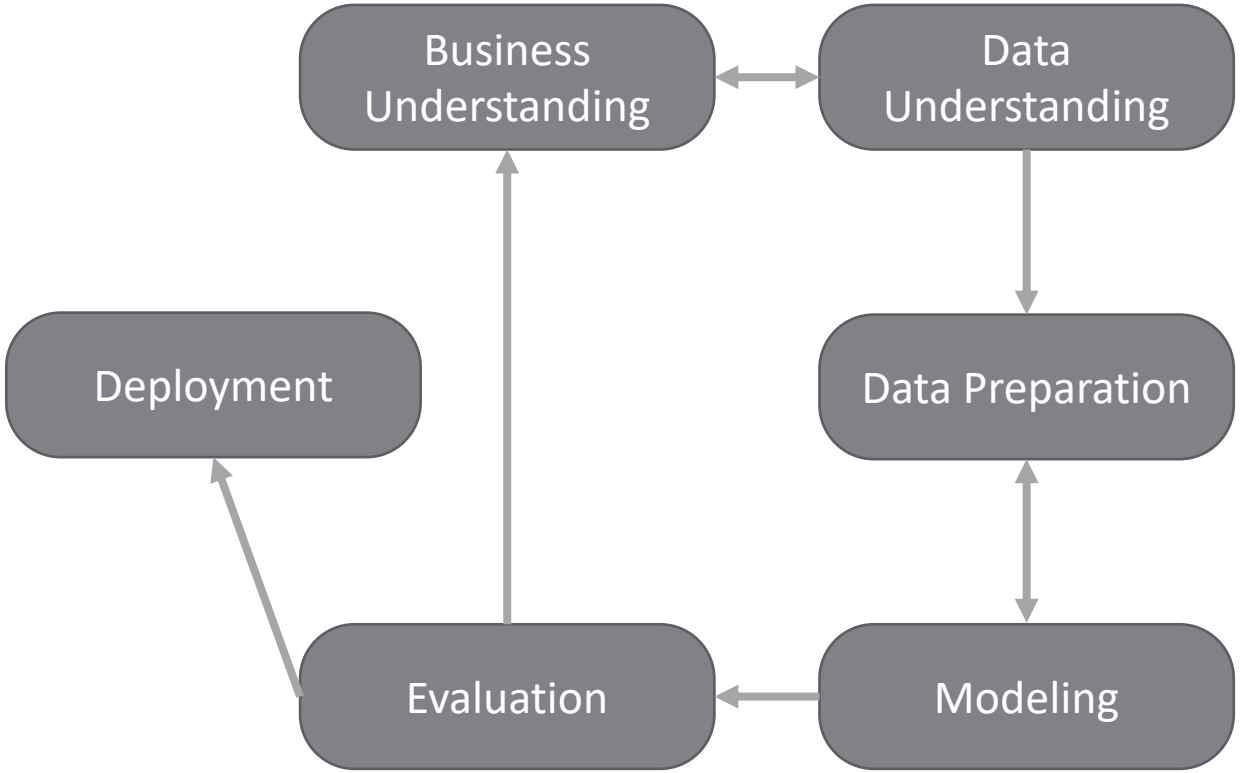
Organizations develop and use predictive classification models

How reliable are my results?

Reference Process for assessing the Reliability of **individual** Classification results

Reference Process

Reference Process



- CRISP DM
  - Methodology for data mining and machine learning projects
- Assessing the reliability of classification results
  - Along each CRISP-DM step, knowledge is gathered and saved in a knowledge graph

Perturbation Approach

Perturbation Options

- Create new feature values with perturbation options

Influencing factors

- Scale
- Volatility
- Restrictions
- Binning
- Missing values

Cardinal	Percent
	Range
	Sensor
	Fixed Amount
Ordinal	In Order
	All Values
Nominal	All Values

EXAMPLE – PERTURBATION TABLE

age	job	marital	education	default	balance	housing	...	prediction
20	technician	single	secondary	no	2143.00	yes	...	no
20	technician	single	secondary	no	2164.43	yes	...	no
20	technician	single	secondary	no	2121.57	yes	...	no
20	technician	single	secondary	no	2185.86	yes	...	no
20	technician	single	secondary	no	2100.14	yes	...	no
20	technician	single	secondary	no	2207.29	yes	...	no
20	technician	single	secondary	no	2078.71	yes	...	no
...	...	...	...	...	...	...	...	...
20	technician	married	secondary	no	2143.00	yes	...	no
20	technician	divorced	secondary	no	2143.00	yes	...	no
...	...	...	...	...	...	...	...	...
20	technician	single	secondary	yes	2143.00	yes	...	yes
20	technician	single	secondary	yes	2164.43	yes	...	yes
20	technician	single	secondary	yes	2121.57	yes	...	yes
...	...	...	...	...	...	...	...	...

Predict classification case → Generate perturbation case → Receive prediction for perturbed case

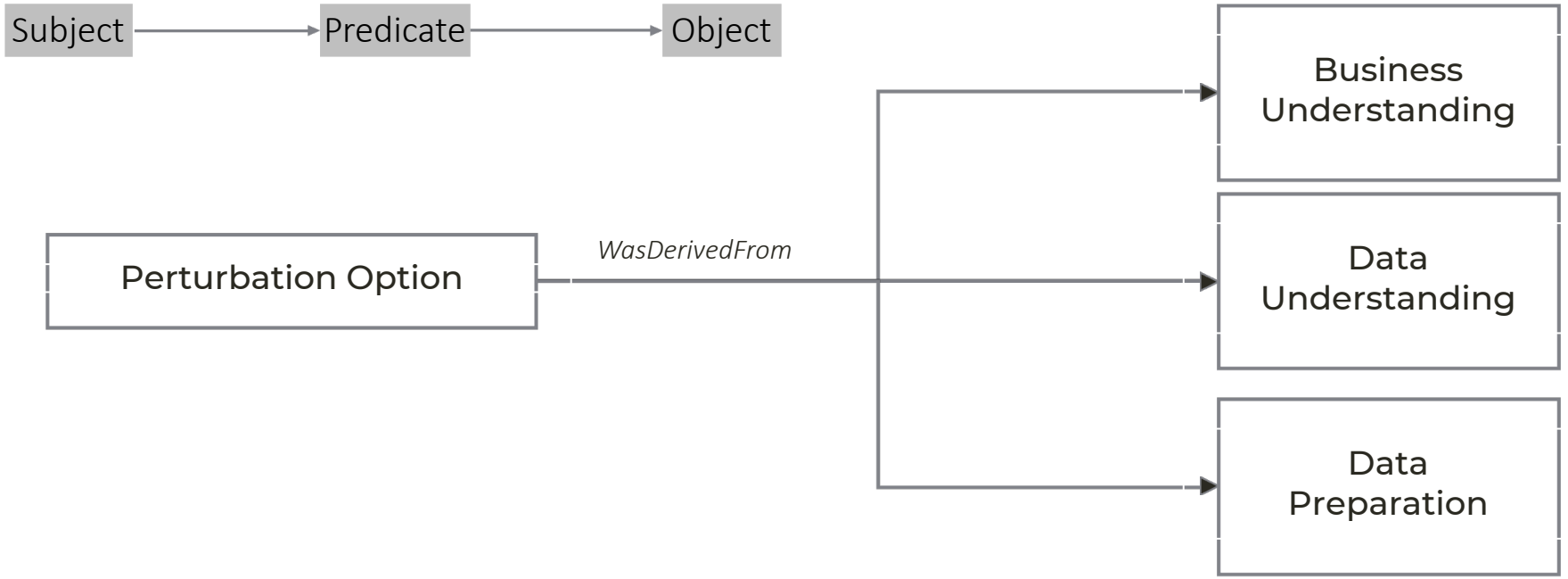
Source: Staudinger, S., Schütz C., Schrefl M.: A Reference Process for Assessing the Reliability of Predictive Analytics Results

Knowledge Graph

KNOWLEDGE GRAPH

- What is a Knowledge Graph?

  Type of data model that represents entities and the relationships between them

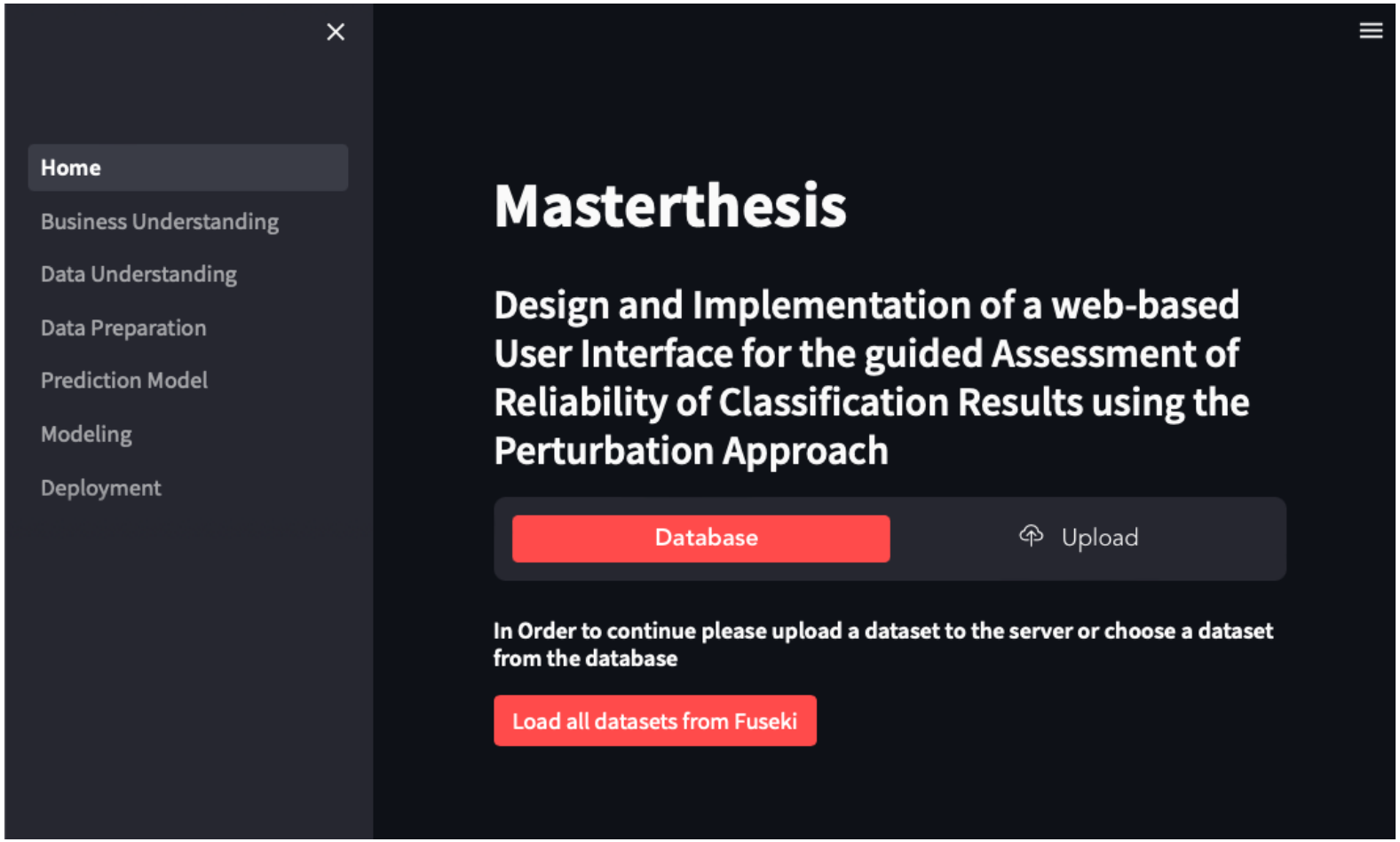


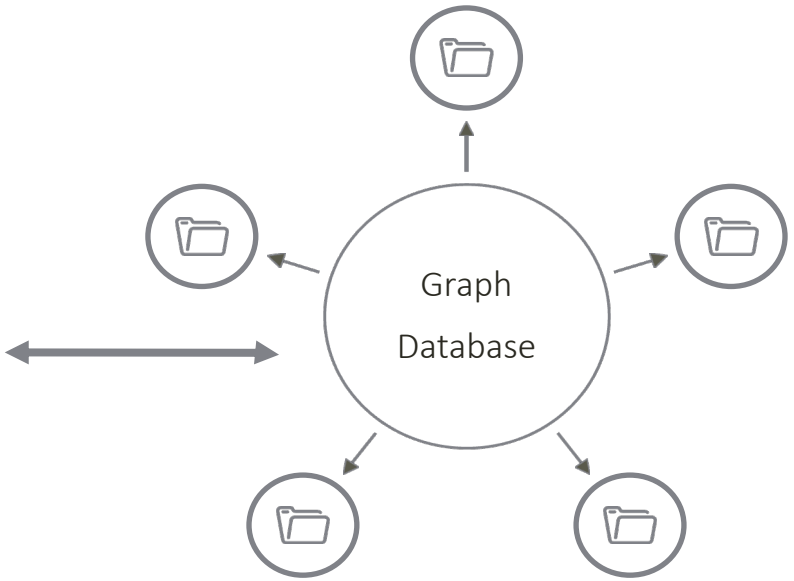
Goals

- Design and documentation
- Recommendation
- Quality management / Decision Audit

Methodical Approach

METHODICAL APPROACH



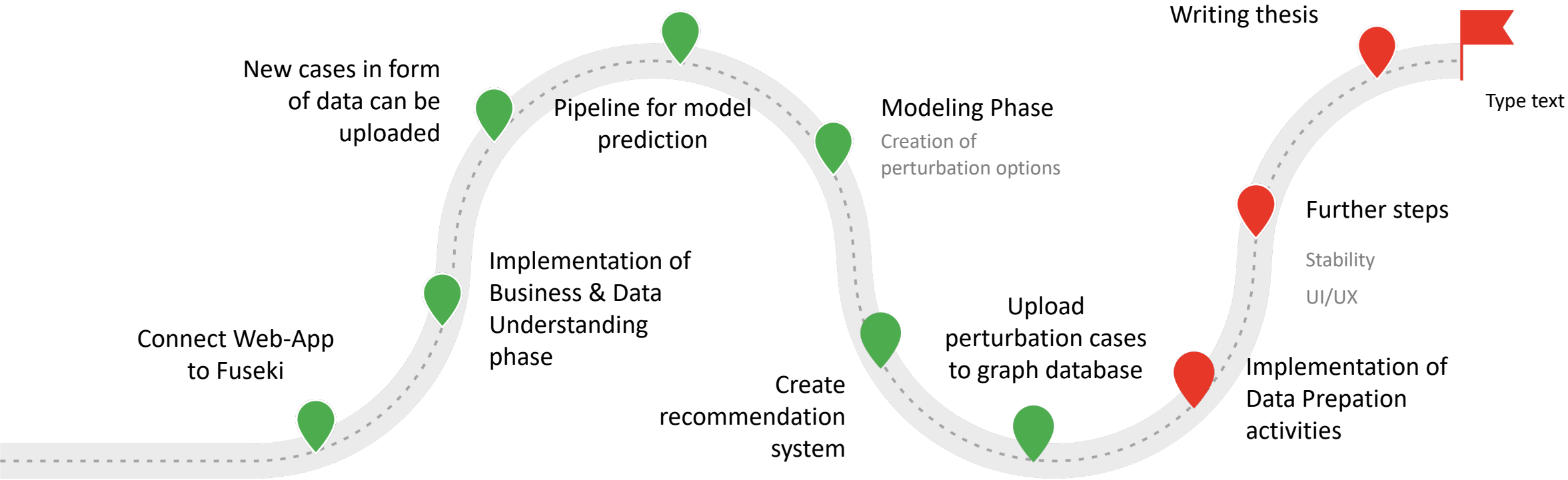


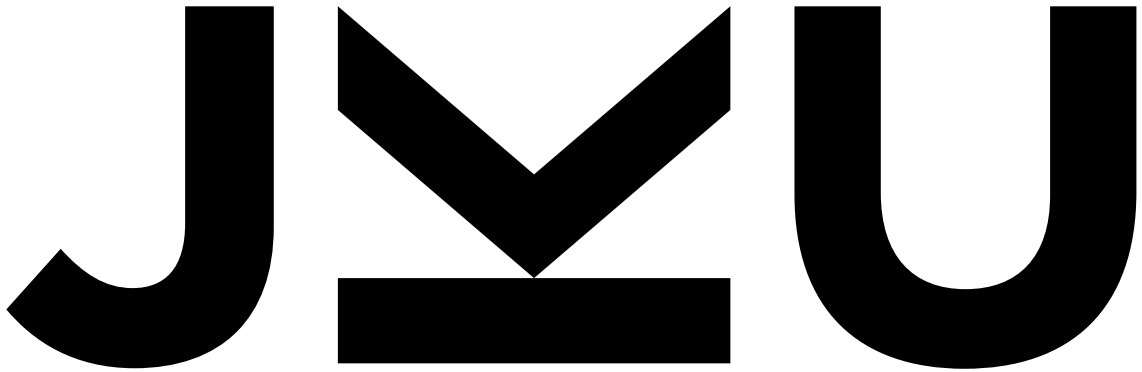
Goals

Goals

1. Generic web app which maps the entire process
2. Automatically generate and store knowledge graph to store historical decisions
3. Recommendation system

CURRENT STATUS AND PLAN





JOHANNES KEPLER
UNIVERSITY LINZ