BSDSAP:

Entwicklung eines KI-gesteuerten autonomen Roboters für die Navigation in einer häuslichen Umgebung

Projektbericht

eingereicht bei

Dr. Martin Prause

AKAD

von

Justin Stange-Heiduk

Hengstrücken 132

37520 Osterode am Harz

Telefon: 015233817587

Studiengang: Data Science

5. Fachsemester

Matrikelnummer: 8149363

Datum: 09.07.2025

[Abbildungsverzeichnis iii](#_Toc202942117)

[1. Einleitung 1](#_Toc202942118)

[1.1 Motivation und Problemstellung 1](#_Toc202942119)

[1.2 Zielsetzung der Arbeit 1](#_Toc202942120)

[1.3 Aufbau der Arbeit 1](#_Toc202942121)

[2. Methodisches Vorgehen 1](#_Toc202942122)

[2.1 CRISP-DM als Strukturmodell 1](#_Toc202942123)

[2.2 Überblick über verwandte Arbeiten (Related Work) 1](#_Toc202942124)

[2.2.1 Deep Learning für PdM (Zhong & Wang) 1](#_Toc202942125)

[2.2.2 Survival Analysis in der Instandhaltung (Carpentier et al.) 1](#_Toc202942126)

[2.2.3 Weitere Ansätze im Vergleich (z. B. GNN, Transformer, klassische ML) 1](#_Toc202942127)

[3. Business Understanding 1](#_Toc202942128)

[3.1 Zielsetzungen aus Sicht von OEMs und Werkstätten 1](#_Toc202942129)

[3.2 Herausforderungen: Datenverfügbarkeit, Blackbox-Modelle, wirtschaftlicher Nutzen 1](#_Toc202942130)

[3.3 Der SCANIA Component X Datensatz im Forschungskontext 1](#_Toc202942131)

[4 Data Understanding 1](#_Toc202942132)

[4.1 Datenquellen und Struktur (Operation, Repair, Spezifikation) 1](#_Toc202942133)

[4.2 Merkmalsübersicht: Sensorik, Histogramme, Zielgrößen 1](#_Toc202942134)

[4.3 Explorative Analyse und Besonderheiten (Imbalance, Zensierung) 1](#_Toc202942135)

[5 Data Preparation 1](#_Toc202942136)

[5.1 Feature Engineering und Preprocessing für Bi-LSTM 1](#_Toc202942137)

[5.2 Datensatzaufbereitung für XGBoost + AFT (inkl. Zensierung) 1](#_Toc202942138)

[5.3 Trainings-, Validierungs- und Testsplits 1](#_Toc202942139)

[5.4 Umgang mit fehlenden Werten und Normalisierung 1](#_Toc202942140)

[6 Modellierung 2](#_Toc202942141)

[6.1 Auswahl der Modellierungsstrategien 2](#_Toc202942142)

[6.1.1 Regression mit Bi-LSTM + Attention 2](#_Toc202942143)

[6.1.2 Survival Analysis mit XGBoost + Accelerated Failure Time 2](#_Toc202942144)

[6.2 Implementierung: Frameworks, Hardware, Laufzeitumgebung 2](#_Toc202942145)

[6.3 Hyperparameter, Trainingsstrategie und Modellarchitekturen 2](#_Toc202942146)

[7 Evaluation 2](#_Toc202942147)

[7.1 Regressionsmetriken (MAE, RMSE) 2](#_Toc202942148)

[7.2 Survival-Metriken (Concordance Index, ggf. Brier Score) 2](#_Toc202942149)

[7.3 Explainability (XAI) 2](#_Toc202942150)

[7.3.1 SHAP für XGBoost 2](#_Toc202942151)

[7.3.2 Attention-Analyse für Bi-LSTM 2](#_Toc202942152)

[7.4 Vergleich und Interpretation der Modelle 2](#_Toc202942153)

[7.5 Grenzen der gewählten Ansätze 2](#_Toc202942154)

[8 Fazit und Ausblick 2](#_Toc202942155)

[8.1 Zusammenfassung der Ergebnisse 2](#_Toc202942156)

[8.2 Bewertung der Modellansätze 2](#_Toc202942157)

[8.3 Mögliche Weiterentwicklungen (z. B. GNN, Multitask, Transfer Learning) 2](#_Toc202942158)

[Anhang A.1 iv](#_Toc202942159)

[Literaturverzeichnis v](#_Toc202942160)

[Eidesstattliche Versicherung vi](#_Toc202942161)

# Abbildungsverzeichnis

**Es konnten keine Einträge für ein Abbildungsverzeichnis gefunden werden.**

# Abkürzungsverzeichnis

# 1. Einleitung

## 1.1 Motivation und Problemstellung

## 1.2 Zielsetzung der Arbeit

## 1.3 Aufbau der Arbeit

# 2. Methodisches Vorgehen

## 2.1 CRISP-DM als Strukturmodell

## 2.2 Überblick über verwandte Arbeiten (Related Work)

### 2.2.1 Deep Learning für PdM (Zhong & Wang)

### 2.2.2 Survival Analysis in der Instandhaltung (Carpentier et al.)

### 2.2.3 Weitere Ansätze im Vergleich (z. B. GNN, Transformer, klassische ML)

# 3. Business Understanding

## 3.1 Zielsetzungen aus Sicht von OEMs und Werkstätten

## 3.2 Herausforderungen: Datenverfügbarkeit, Blackbox-Modelle, wirtschaftlicher Nutzen

## 3.3 Der SCANIA Component X Datensatz im Forschungskontext

# 4. Data Understanding

## 4.1 Datenquellen und Struktur (Operation, Repair, Spezifikation)

## 4.2 Merkmalsübersicht: Sensorik, Histogramme, Zielgrößen

## 4.3 Explorative Analyse und Besonderheiten (Imbalance, Zensierung)

# 5. Data Preparation

## 5.1 Feature Engineering und Preprocessing für Bi-LSTM

## 5.2 Datensatzaufbereitung für XGBoost + AFT (inkl. Zensierung)

## 5.3 Trainings-, Validierungs- und Testsplits

## 5.4 Umgang mit fehlenden Werten und Normalisierung

# 6. Modellierung

## 6.1 Auswahl der Modellierungsstrategien

## 6.1.1 Regression mit Bi-LSTM + Attention

## 6.1.2 Survival Analysis mit XGBoost + Accelerated Failure Time

## 6.2 Implementierung: Frameworks, Hardware, Laufzeitumgebung

## 6.3 Hyperparameter, Trainingsstrategie und Modellarchitekturen

# 7. Evaluation

## 7.1 Regressionsmetriken (MAE, RMSE)

## 7.2 Survival-Metriken (Concordance Index, ggf. Brier Score)

## 7.3 Explainability (XAI)

### 7.3.1 SHAP für XGBoost

### 7.3.2 Attention-Analyse für Bi-LSTM

## 7.4 Vergleich und Interpretation der Modelle

## 7.5 Grenzen der gewählten Ansätze

# 8. Deployment

## 8.1 Zielgruppe und Anwendungsszenario 8.2 Visualisierung und Nutzung der Vorhersagen (z. B. PowerBI) 8.3 Beispielhafte Nutzung: Werkstatt, Flottenmanagement, OEM 8.4 Grenzen und offene Punkte für echte Produktion

# 9. Fazit und Ausblick

## 9.1 Zusammenfassung der Ergebnisse

## 9.2 Bewertung der Modellansätze

## 9.3 Mögliche Weiterentwicklungen (z. B. GNN, Multitask, Transfer Learning)

# Anhang A.1

# Literaturverzeichnis

# Eidesstattliche Versicherung

Ich versichere, dass ich das beiliegende Assignment selbstständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie alle wörtlich oder sinngemäß übernommenen Stellen in der Arbeit gekennzeichnet habe.