

B

C

Bachelor Thesis

Datenanalyse PEP

S

Bachelor Thesis

Datenanalyse PEP

by

Jonas Mayer

in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science
in Applied Computer Science

at the Hochschule Konstanz University of Applied Sciences,

Student Number: 305630

Date of Submission: TODO

Supervisor: **Prof. Dr. Doris Bohnet**
Second Examiner:

An electronic version of this thesis is available at <https://github.com/jonez187/bachelorarbeit-htwg-latex>.

Abstract

Hier Abstract schreiben

Inhaltsverzeichnis

1	Einleitung	1
2	Theoretische Grundlagen	3
2.1	PEP-Ecopassport	3
2.2	PDF-Datenextraktion (?)	4
2.3	Statistische Methoden	4
2.3.1	4
3	Pipeline und Datenbasis	5
3.1	Pipeline	5
3.1.1	PEP Recherche	5
3.1.2	Vorhandene Pipeline	5
3.1.3	Pipeline-Versuch1.	5
3.1.4	Docling und LLM basierte Pipeline.	6
3.2	Datenbasis	6
3.2.1	Normalisierung	6
3.2.2	Datenbereinigung	6

1

Einleitung

Hier Einleitung schreiben
TESTSTETS

2

Theoretische Grundlagen

Hier werden die theoretischen Grundlagen für die vorliegende Arbeit gelegt. Ausgangspunkt ist die Beobachtung, dass Smart-Home-/IoT-Produkte Umweltwirkungen nicht nur in der Nutzungsphase (Stromverbrauch), sondern ebenso durch Materialzusammensetzung, Fertigung, Distribution und Entsorgung verursachen. Für die standardisierte Berichterstattung solcher Wirkungen existieren deklarative Formate wie die PEP Eco-passports, die Indikatoren entlang des Lebenszyklus ausweisen. Damit diese Angaben für quantitative Analysen nutzbar werden, sind konsistente Begriffe, Einheiten und Moduldefinitionen ebenso erforderlich wie ein Verständnis zentraler statistischer Verfahren zur Muster- und Zusammenhangsanalyse. Dieses Kapitel führt daher zunächst in Struktur und Inhalte von PEP-Deklarationen ein und skizziert anschließend die methodischen Bausteine (u. a. Lineare Regression und PCA), die in den folgenden Kapiteln zur Reduktion von Variablen, zur Erklärung von Indikatorvarianz und zur Ableitung einer praxistauglichen Heuristik für Produkte ohne PEP eingesetzt werden.

2.1. PEP-Ecopassport

Was ist PEP-Ecopassport, was steht drin, was ist interessant für mich

2.2. PDF-Datenextraktion (?)

Theoretische Grundlagen PDFs und Daten (Aufbau Textlayer usw), gpt Modell für extraktion (Vor-/Nachteile)

2.3. Statistische Methoden

Welche statistischen Methoden werden eingesetzt und worauf basieren sie?

2.3.1.

3

Pipeline und Datenbasis

Im folgenden Kapitel wird die Pipeline zur Datenextraktion aus PEP Ecopassport PDFs sowie die daraus gewonnene Datenbasis beschrieben.

3.1. Pipeline

Beschreibung vorheriger Pipeline, Probleme und Endresultat

3.1.1. PEP Recherche

Probleme mit der PEP Recherche und wie ich sie deshalb durchgeführt habe

3.1.2. Vorhandene Pipeline

Lennys Pipeline und Probleme

3.1.3. Pipeline-Versuch1

Lennys Pipeline versucht upzgrade. Hier werden vor allem die Probleme gezeigt

3.1.4. Docling und LLM basierte Pipeline

Docling beschreiben, neue gpt-api usw.

3.2. Datenbasis

Wie viele PDFs, wie ichs noch weiter flachgezogen habe, Probleme usw

3.2.1. Normalisierung

Mapping Material, Strommodell usw

3.2.2. Datenbereinigung

Wie Daten bereinigt (zb zu leere PDFs aussortiert), Ausreißer manuell analysiert, Werte nachtragen