

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

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Theoretische Grundlagen</b>	<b>3</b>
2.1	PEP-Ecopassport . . . . .	3
2.1.1	PEP-Standard . . . . .	3
2.1.2	Aufbau typischer PEP-Dokumente . . . . .	4
2.2	PDF-Datenextraktion (?) . . . . .	7
2.3	Statistische Methoden . . . . .	7
<b>3</b>	<b>Pipeline und Datenbasis</b>	<b>9</b>
3.1	Pipeline . . . . .	9
3.1.1	PEP Recherche . . . . .	9
3.1.2	Vorhandene Pipeline . . . . .	9
3.1.3	Pipeline-Versuch1 . . . . .	9
3.1.4	Docling und LLM basierte Pipeline . . . . .	10
3.2	Datenbasis . . . . .	10
3.2.1	Normalisierung . . . . .	10
3.2.2	Datenbereinigung . . . . .	10
	<b>Literatur</b>	<b>11</b>





# 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

#### 2.1.1. PEP-Standard

Der *PEP Ecopassport*<sup>®</sup> ist ein international anerkanntes Programm zur Erstellung standardisierter Umweltproduktdeklarationen für elektrische, elektronische sowie Heizungs-

, Lüftungs-, Klima- und Kälteprodukte (HVAC). Träger des Programms ist die *P.E.P. Association*, eine gemeinnützige Organisation, deren Ziel es ist, ein gemeinsames und verlässliches Referenzsystem für Umweltinformationen dieser Produktkategorien bereitzustellen. Das Programm versteht sich als Branchenspezialisierung innerhalb des Rahmens der *Environmental Product Declarations (EPD)* gemäß ISO 14025 und der Lebenszyklusnormen nach ISO 14040, und basiert somit auf international festgelegten Normen. [PEP]

Ein *PEP Ecopassport* ist somit eine *Typ III-Umweltdeklaration* im Sinne der ISO 14025. Diese Deklarationen basieren auf quantitativen Ergebnissen einer Lebenszyklusanalyse (*Life Cycle Assessment, LCA*) und dienen der vergleichenden Bewertung von Produkten mit identischer Funktion. Die Datenerhebung und Berechnung erfolgt nach vordefinierten Parametern, die in sogenannten *Product Category Rules (PCR)* und bei Bedarf in *Product Specific Rules (PSR)* festgelegt sind. Jede PEP-Deklaration unterliegt einer unabhängigen Überprüfung der angewandten Methodik und der zugrunde liegenden LCA-Daten. [PEP]

Das Programm zielt auf Transparenz und Vergleichbarkeit ab. Hersteller erhalten ein einheitliches Verfahren, um ökologische Leistungskennwerte ihrer Produkte objektiv und nachvollziehbar zu kommunizieren. Für Anwender, Beschaffer und Energieberater stellen die PEP-Daten eine verlässliche Grundlage für ökologische Bewertungen und Beschaffungsentscheidungen dar.

Die Teilnahme am PEP-Programm ist freiwillig, gewinnt jedoch in der Praxis an Bedeutung, da Umweltproduktdeklarationen zunehmend als Nachweis oder Auswahlkriterium in Ausschreibungen und Produktbewertungen herangezogen werden. Eine gesetzliche Verpflichtung zur Erstellung besteht bislang nur in Einzelfällen, beispielsweise in Frankreich, wenn ein Hersteller aktiv mit Umweltvorteilen wirbt.

Das PEP-Programm unterscheidet sich klar von unternehmensbezogenen Treibhausgas-Bilanzierungen: Es erfasst ausschließlich produktspezifische Umweltwirkungen entlang des Lebenszyklus und folgt dabei den methodischen Vorgaben der ISO 14040-Reihe. Für umfassende *GHG-Assessments* auf Organisationsebene sind PEP-Daten daher nicht geeignet.

### 2.1.2. Aufbau typischer PEP-Dokumente

Ein vollständiges PEP umfasst typischerweise etwa zehn Seiten und gliedert sich in mehrere inhaltlich definierte Abschnitte.

**Titel- und Metadatenblatt** Das Deckblatt enthält grundlegende Angaben zum Produkt (Name, Version, Sprache, Hersteller), zum Veröffentlichungs- und Revisionsdatum

sowie zum Status der Erklärung (z. B. *in review* oder *verified*). Darüber hinaus sind Kontaktinformationen, Firmenadresse und Registrierungsnummer enthalten.

**Allgemeine Produktinformationen** Dieser Abschnitt beschreibt die funktionale Einheit (*functional unit*) und die Referenzlebensdauer, hier meist 10 bis 20 Jahre. Weiterhin werden die Produktfunktion, Anwendungsbereiche und gegebenenfalls weitere Varianten aufgeführt.

**Materialzusammensetzung** Die Zusammensetzung des Produkts wird tabellarisch nach Hauptgruppen ausgewiesen, z. B. Kunststoffe, Metalle und weitere Materialien (Elektronik, Sonstiges). Im Beispiel des ABB EQ Meter entfallen 28 % auf Kunststoffe, 53 % auf Metalle und 19 % auf weitere Komponenten. Diese Angaben ermöglichen eine spätere Aggregation der Stoffanteile in harmonisierten Datenstrukturen.

**Szenarien und Lebenszyklusphasen** PEP-Dokumente sind entlang der Phasen des Produktlebenszyklus strukturiert, die den Vorgaben der EN 15804 entsprechen:

- **Herstellung (A1–A3):** Produktion und Vormaterialien, modelliert mit landesspezifischem Strommix (z. B. italienischer Grid Mix).
- **Distribution (A4):** Transport vom Werk zum Markt; häufig standardisierte Annahmen (z. B. 1 000 km Schiff, 3 300 km Lkw).
- **Installation (A5):** Montage, meist nur Verpackungsabfälle berücksichtigt.
- **Nutzungsphase (B):** Betrieb des Geräts mit angegebenem Energieverbrauch, z. B. 126 kWh über 20 Jahre, basierend auf europäischem Netzstrommix.
- **End-of-Life (C1–C4):** Entsorgungsszenario gemäß PCR-Vorgaben (Recycling-, Deponie-, Transportanteile).
- **Optionale Phase (D):** Rückgewinnung und Wiederverwendung außerhalb des Systemgrenzenmodells.

In der weiteren Datenaufbereitung werden diese Phasen zu den Kategorien *manufacturing*, *distribution*, *installation*, *use* und *end\_of\_life* zusammengefasst.

**Umweltindikatoren** Die Umweltwirkungen werden für jede Lebenszyklusphase sowie als Gesamtwert angegeben. Die für diese Arbeit relevanten Indikatoren sind in der Tabelle 2.1 aufgeführt.

Tabelle 2.1: Umweltindikatoren

Indikator	Beschreibung
Acidification	Versauerung von Böden und Gewässern durch säurebildende Emissionen
Climate Change (Fossil)	Treibhauspotenzial durch fossile CO <sub>2</sub> -Emissionen
Climate Change (Land Use and Land Use Change)	Treibhauspotenzial infolge von Landnutzungsänderungen (LULUC)
Climate Change (Total)	Gesamtes Treibhauspotenzial aus allen Quellen
Eutrophication (Freshwater)	Nährstoffanreicherung in Binnengewässern
Eutrophication (Marine)	Nährstoffanreicherung in marinen Ökosystemen
Eutrophication (Terrestrial)	Nährstoffanreicherung in terrestrischen Ökosystemen
Hazardous Waste Disposed	Entsorgung gefährlicher Abfälle
Ozone Depletion	Abbau der stratosphärischen Ozonschicht durch FCKW-Emissionen
Photochemical Ozone Formation (Human Health)	Bildung von bodennahem Ozon (Sommersmog)
Radioactive Waste Disposed	Entsorgung radioaktiver Abfälle
Resource Use (Fossils)	Nutzung fossiler Energieressourcen
Resource Use (Minerals and Metals)	Verbrauch abiotischer Ressourcen (Metalle und Mineralien)
Water Use	Entnahme und Verbrauch von Frischwasser

**Verifikations- und Anhangsangaben** Im abschließenden Teil werden die angewendeten Regelwerke und Datenquellen genannt, z. B. *PCR-ed3-EN-2015\_04\_02* und *PSR-0005-ed2-EN-2016\_03\_29*, die eingesetzte Software (z. B. SimaPro 9.3 mit Ecoinvent 3.8) sowie die Verifizierungsstelle und deren Akkreditierungsnummer. Darüber hinaus enthält dieser Abschnitt Angaben zum *Materialaufbau* und zum verwendeten *Energiemodell*. Die Materialzusammensetzung wird in der Regel als prozentuale Massenanteile nach Hauptgruppen (Kunststoffe, Metalle, Elektronik, Sonstiges) dargestellt, teils in Tabellenform, teils grafisch als Kreisdiagramm. Das Energiemodell beschreibt die angenommenen Strommixe und Spannungsniveaus je Lebenszyklusphase, beispielsweise den nationalen Grid Mix für die Herstellung und den europäischen Durchschnittsmix für die Nutzungsphase.

Obwohl der inhaltliche Mindestumfang und die zu berichtenden Umweltindikatoren durch die zugrundeliegenden ISO- und PCR-Vorgaben festgelegt sind, besteht keine feste formale Struktur. Das Layout, die grafische Aufbereitung und die Anordnung der Tabellen können je nach Hersteller, Software und Version variieren. So enthalten einige PEPs tabellarische Aufstellungen sämtlicher Indikatoren, während andere ergänzend oder teilweise ausschließlich Diagramme und grafische Vergleichsdarstellungen beinhalten.

**Einordnung für die Datenanalyse** Die standardisierte, aber formal heterogene Gliederung der PEPs ermöglicht eine eindeutige Zuordnung der Informationen zu den Ziel Datenfeldern der harmonisierten JSON-Datenbank. *functional\_unit*, *lifetime* und *manufacturing\_country* bilden die Metadatenebene, *material\_composition[]* enthält die Stoffanteile, *indicators{}* die Umweltindikatoren pro Phase, und *energy\_model[]* die modellierten Energiemixe je Lebenszyklusphase. Damit stellen PEP-PDFs eine normbasierte, inhaltlich konsistente, aber formal variable Datenquelle für die automatisierte Extraktion und statistische Auswertung dar.

## 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?





# 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

# Literatur

[PEP] Association P.E.P. "PEP-Ecopassport". In: *PEP-Ecopassport Webseite* ().