[Achtung: Verwenden Sie einen Sperrvermerk nur in sehr gut begründeten Fällen]

[evtl. Sperrvermerk]

Die vorliegende Arbeit ist bis zum [DATUM] für die öffentliche Nutzung zu sperren. Veröffentlichung, Vervielfältigung und Einsichtnahme sind ohne meine ausdrückliche Genehmigung nicht gestattet. Der Titel der Arbeit sowie das Kurzreferat/Abstract dürfen veröffentlicht werden.

Dornbirn, Unterschrift Verfasser*in

FHV
Vorarlberg University
of Applied Sciences

Classification of GPS Track Data Using AI Methods

A Case Study of Waste Collection Vehicles

Bachelor thesis
for obtaining the academic degree

Bachelor of Science in Engineering (BSc)

Vorarlberg University of Applied Sciences
Computer Science - Software and Information Engineering

Supervised by
Dipl.-Ing. Dr. techn. Ralph Hoch

Submitted by
Matthias Hefel
Dornbirn, May 2025

Dedication

Dedicated to my younger self, who never stopped chasing his dream and never will!

And to my parents, who supported me throughout this journey.

Thank you.

Kurzreferat

Klassifizierung von GPS-Spurdaten mit Unterstützung von KI-Methoden am Beispiel von Abfallsammelfahrzeugen

In der Abfallwirtschaft ist die strategische Tourenplanung ein wichtiger Prozess, in dem durch optimale Gebietsaufteilung eine maximal effiziente Fuhrparkauslastung bei möglichst geringen Kosten ermittelt wird. Dies geschieht in Entsorgungsbetrieben sowohl für bestehende Auftragsgebiete, als auch bei der Kalkulation von neuen Ausschreibungen. Vor Allem bei Regionen, in denen keine Erfahrungswerte vorliegen müssen für eine robuste Tourenplanung zahlreiche unscharfe Annahmen getroffen und manchmal auch Schätzungen vorgenommen werden. Um diese Unsicherheiten durch die Analyse von geographischen Strukturen zur verringern soll eine Technologie in die bestehende Tourenplanungssoftware der Firma integriert werden, die folgende Aufgabenstellung automatisiert lösen kann: Anhand von bestehenden GPS-Aufzeichnungen sollen strukturelle Eigenschaften der jeweilige Sammelgebiete numerisch bewertet und klassifiziert werden. Gleichermaßen sollen anhand von geographischen (und möglichst frei verfügbaren Strukturdaten) aus noch unbekannten Gebieten erhoben werden können um diese auf die selbe Art und Weise klassifizieren zu können. Dadurch entsteht einerseits eine Referenzdatenmenge (von bestehenden Sammeltouren) und eine Vergleichsdatenmenge (aus den neuen Ausschreibungsgebieten). Dort wo die Klassifizierungsdaten übereinstimmen, kann davon ausgegangen werden, dass die planungsrelevanten Kennzahlen aus bestehenden Auftragsgebieten ohne gewagte Annahmen einfach übernommen werden können. Die Klassifizierung von GPS-Daten und geographischen Strukturdaten soll mit Hilfe von künstlicher Intelligenz automatisiert erstellt werden können. Auch die Überlegung, welche geographischen Strukturdaten denn überhaupt aussagekräftig sind um einen Vergleich anzustreben, sollen ggf. mit Hilfe von KI Technologien erfolgen.

Das Ziel der praktischen Arbeit ist es einen Sandbox-Service zu implementieren, der von der bestehenden Software der infeo aufgerufen und mit Daten befüllt werden kann um so "auf Knopfdruck" Klassifizierungen und Vergleiche von GPS-Daten und Ausschreibungs-Strukturdaten zu erstellen. Die Anwender:innen haben dadurch die Möglichkeit für neue Ausschreibungen entsprechend passende Planungsparameter aus ihren bestehenden Auftragsgebieten zu berechnen und somit die Unsicherheiten bei der Ausschreibungskalkulation deutlich zu reduzieren.

GPS-Datenklassifizierung, Abfallwirtschaft, Künstliche Intelligenz, Geografische Datenanalyse, Maschinelles Lernen, Automatisierung

Abstract

Classification of GPS Track Data Using AI Methods: A Case Study of Waste Collection Vehicles

In waste management, strategic route planning is a crucial process where optimal fleet utilization is determined through the efficient division of service areas, with the goal of minimizing costs. This process is applied by waste disposal companies both for existing service areas and when calculating bids for new tenders. Especially in regions where there is no prior experience, numerous uncertain assumptions and estimates must be made for robust route planning. To reduce these uncertainties through the analysis of geographical structures, a technology will be integrated into the company's existing route planning software, which can automatically solve the following task: Based on existing GPS records, the structural characteristics of the respective collection areas should be numerically evaluated and classified. Additionally, geographical structural data (preferably from freely available sources) from unknown areas should be collected and classified in the same way. This approach will create both a reference data set (from existing collection routes) and a comparison data set (from new tender areas). Where the classification data match, it can be assumed that planning-relevant parameters from existing service areas can be applied to the new areas without risky assumptions. The classification of GPS data and geographical structural data should be automated using artificial intelligence. Furthermore, the consideration of which geographical structural data are meaningful for comparison should, if necessary, also be supported by AI technologies.

The practical goal of this work is to implement a sandbox service that can be called and populated with data by the existing software of infeo, enabling the creation of classifications and comparisons of GPS data and tender structural data "at the push of a button." This will provide users with the ability to calculate appropriate planning parameters from their existing service areas for new tenders, thereby significantly reducing uncertainties in bid calculations.

GPS Data Classification, Waste Management, Artificial Intelligence, Geographic Data Analysis, Machine Learning, Automation

Preface

[Preface Text]

Contents

List of Figures

List of Tables

List of Abbreviations

GPS Global Positioning System

AI Artificial Intelligence

ML Machine Learning

API Application Programming Interface

CSV Comma-Seperated Values

1 Introduction

"The world's most valuable resource is no longer oil, but data" [1]

In today's digital age, where electronic devices are a part of everyones daily lives, increasing amounts of data are being generating every day, and this trend shows no signs of slowing down. [2] With this increase in data, businesses ranging across all industries recognize the importance of leveraging it for decision-making and operational efficiency. This has lead to a growing demand for technolgies that can gather insights from data and integrate seamlessly into strategic processes.

One industry in which data-driven decision-making is becoming increasingly important is the waste management industry.

1.1 Problem Statement

Companies operating in the waste collection business have trouble calculating accurate bids for new service areas when expaning their field of business. They often have to make assumptions and rough estimates on several paremters concerning the operation cost in new service areas. A data driven estimation can help create more accurate and less risky assessments for unknown collection locations. This can help reduce uncertainties and improve the accuracy of bid calculations.

1.2 Motivation

Notes: GPS Data is one of the most informativ data and can lead to many insights, which infeo is interested in.

1.3 Solution Approach

1.4 Structure of the Work

2 Background and Related Work

2.1 Technical Background

2.2 Related Work

2.2.1 Comparison of GPS-Routes

3 Problem Definition and Solution Approach

3.1 Description of the Dataset

3.1.1 Overview

The dataset used is a collection of GPS tracking data collected by wastecollection vehicles from various wastecollection businesses and provided by infeo GmbH. It represents real-world data collected during regular wastecollection operation in the DACH region.

3.1.2 Source and Collection Method

The data was obtained by the onboard tracking systems installed by infeo GmbH, which collects GPS coordinates in regular intervals during regular operation. Each tracking represents a complete wastecollection route taken and includes metadata aswell as a list of GPS coordinates.

3.1.3 Structure of the Data

Each dataset entry represents a single recorded route refered to as *tracking* and contains metadata aswell as a time ordered list of gps coordinates.

Each tracking contains the following fields:

Table 3.1: Structure of a Tracking Entry

Field	Type	Description
`id`	Integer	Unique identifier of the tracking entry.
`name`	String	Name of the tracking (randomized for anonymization) identification.
`description`	String	Route metadata, often includes internal codes.
`recorded`	DateTime	Start date and time of the tracking.
`length`	Float	Total length of the route in kilometers.
`duration`	Integer	Total duration of the tracking in nanoseconds.
`vehicleId`	Integer / Null	ID of the vehicle (nullified for anonymization).
`tourId`	Integer / Null	ID of the associated tour (nullified for anonymization).
`isExported`	Boolean	Flag indicating if the tracking was exported.
`editState`	Integer	Edit state used by the system.

Each GPS point contains the following fields:

Table 3.2: Structure of a GPS Point Entry

Field	Type	Description
`id`	Integer	Unique identifier of the GPS point.
`time`	DateTime	Timestamp of when the point was recorded.
`latitude`	Float	Latitude coordinate.
`longitude`	Float	Longitude coordinate.
`speed`	Float	Instantaneous speed at the time (in km/h).
`heading`	Float	Direction of movement in degrees.
`sequence`	Integer	Position of the point in the tracking sequence.
`metaTag`	Integer	Custom metadata tag.
`metaValue`	String	Value associated with the metadata tag.
`pointBaseType`	Integer	Internal point type used by the system.

3.1.4 Size and Coverage

Formatvorlage für den Fließtext.

3.1.5 Limitations

Missing Values: GPS gaps etc, useless trackings etc.

3.2 Big Picture

Formatvorlage für den Fließtext.

3.3 Dataset Analysis

Formatvorlage für den Fließtext.

3.3.1 Sample Analysis

A small, manually selected sample of 8 tracking routes was selected for initial exploratory data analysis. Each route was inspected on the AWM-Map-Tool and then categorized into on of the four area type labels: RURAL, SUBURBAN, TOWN or URBAN. Each Label is represented by 2 tracking routes in the sample data to ensure a balanced representation.

Feature extraction was performed to calculate route-level metrics such as length, duration, bounding box area, point density, number of stops and average distance between points.

The goal of this sample is to explore patterns, validate assumptions, and identify features useful for future automatic classification.

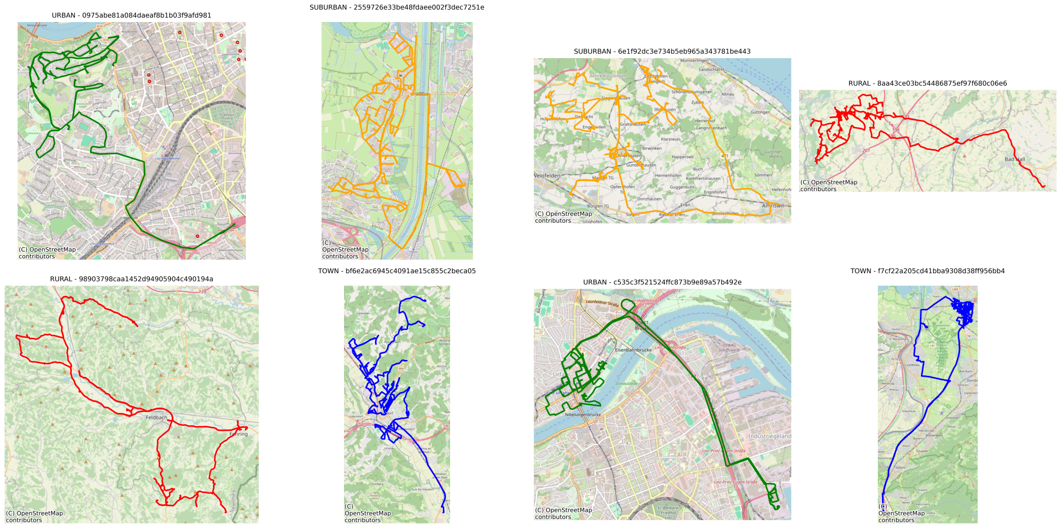

Figure 3.1: Mapgrid for selected sample trackings

The mapgrid shows spatial layout of each tracking route overlayed on a visual map. Each subplot corresponds to a single tracking and is colorcoded with the assigned lable (RURAL=Red, SUBURBAN=Orange, TOWN=Blue, URBAN=Green). This visual representation clearly shows the difference in route shapes and sizes between the four different area types.

Urban:

- Routes are geographically compact and highly localized.
- Movement appears dense with short travel distances between stops.
- Often confined to a small cluster of city blocks or neighborhoods.

Town:

- Coverage is slightly more dispersed than urban routes.
- Still relatively compact but less tightly packed.
- Serves a central area and nearby residential surroundings.

Suburban:

- Routes extend farther and cover wider areas than town routes.
- Show transitional behavior between urban and rural structures.
- Less dense stop distribution, indicating more spaced-out residential zones.

Rural:

- Routes are long and span large geographical areas.
- Stops are widely spaced, often connecting small, isolated settlements.
- The shape and path vary significantly, often following main roads between distant collection points.

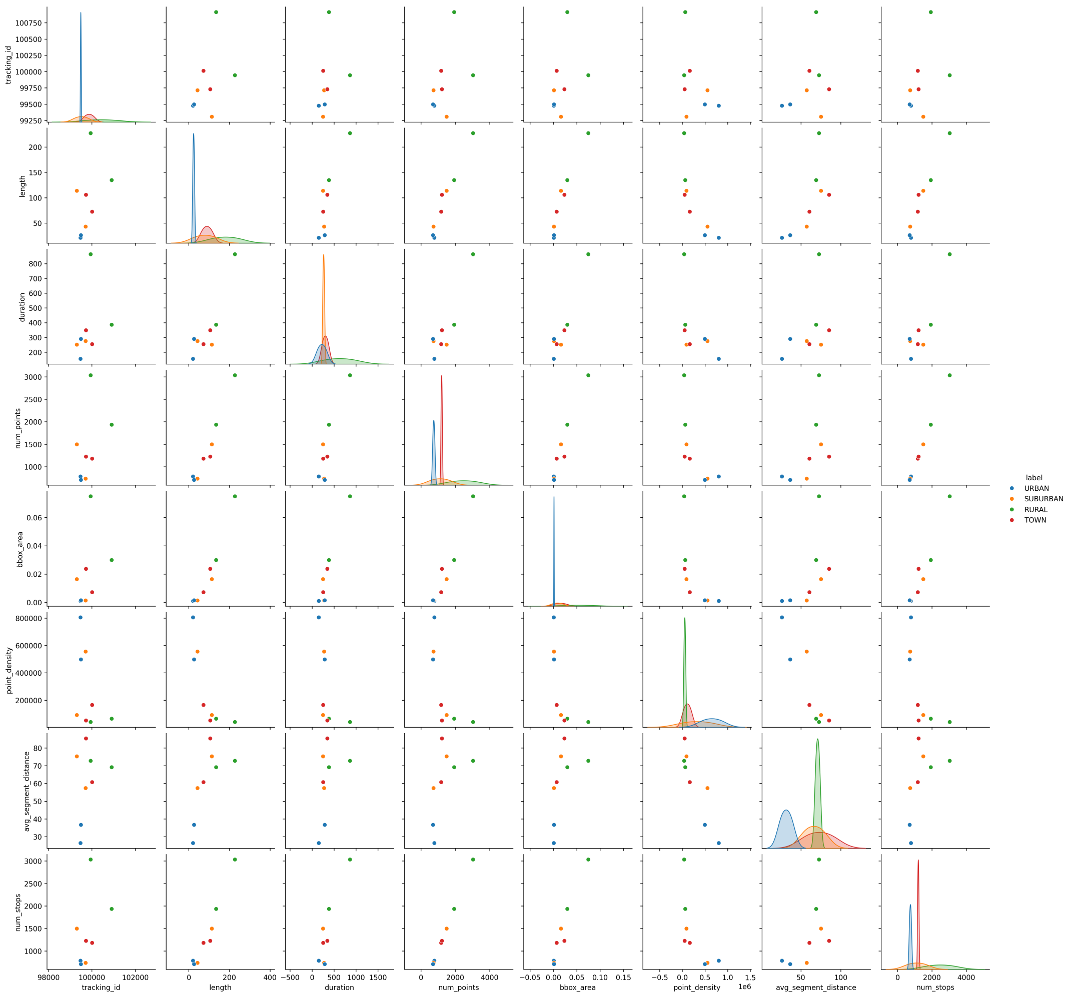

Figure 3.2: Pairplot of selected GPS route features grouped by area label

The pairplot compares the relationships between the extracted features (eg., length, duration, number of points, bounding box area, point density, average segmengt distance and number of stops). Rural and urban trackings show a clear difference in multiple features. Point density and average segment distance are especially good discriminators. Suburban and town trackings show more variance and occasionally overlap with eachother, giving a not so clear distinction.

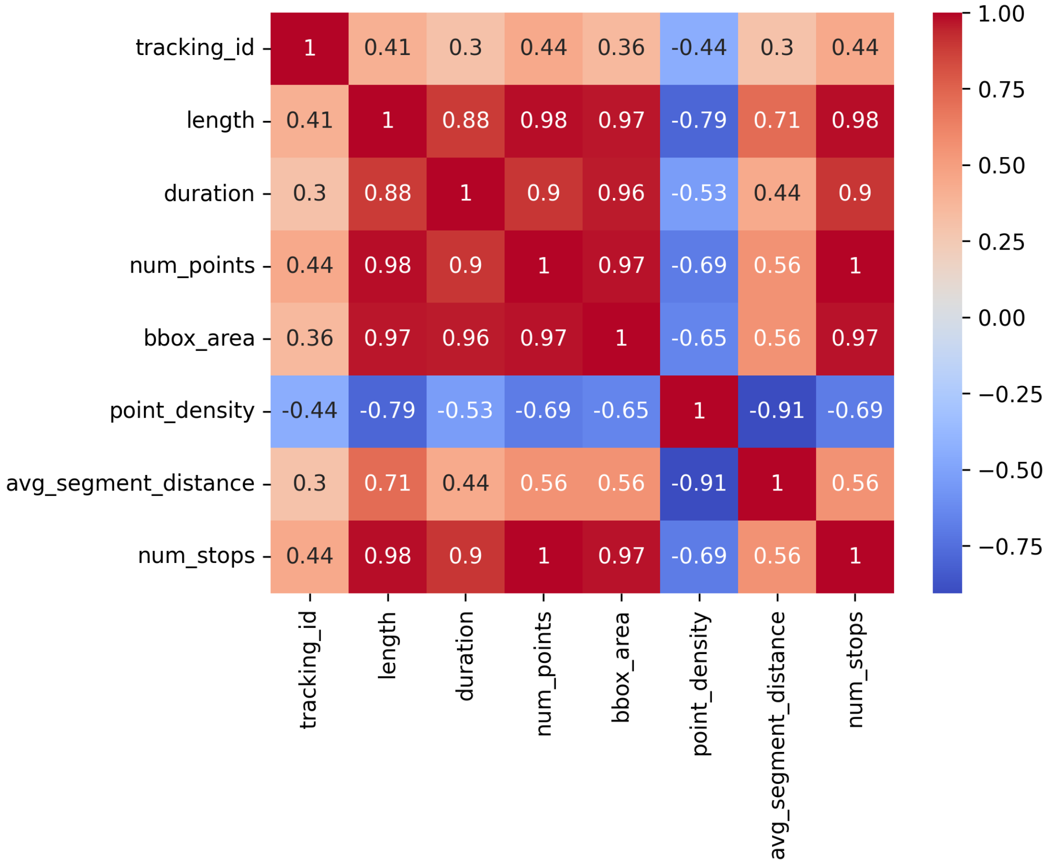

Figure 3.3: Correlation Matrix of selected GPS route features grouped by area label

The Correlation Matrix highlights the strong correlations between features in the selected sample. A high posive correlation between length, duration and number of points can be observerd. Additionally the point density has a strong negative correlation with the tracking length, number of points and average segment distance.

This correlation is to be expected and confirms the consistency of the data, since the points (GPS coordinates) are recorded in uniform intervals. This plot also indicates the possible redundance of features such as number of points and number of stops.

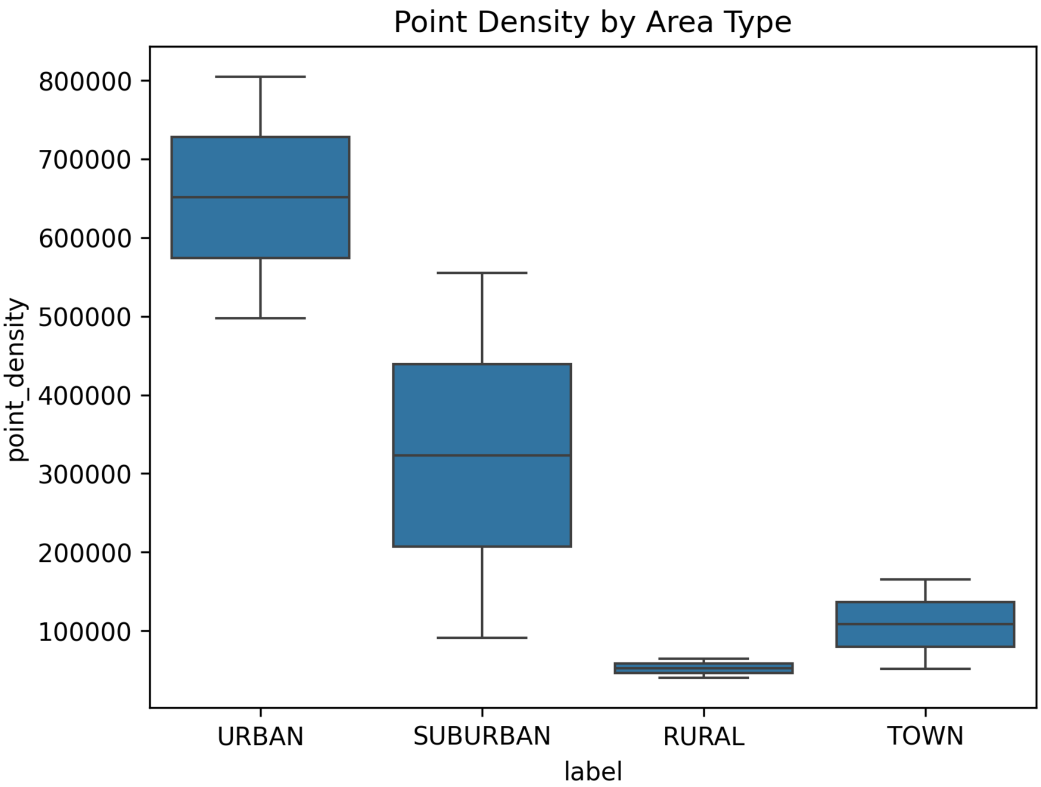

Figure 3.4: Boxplot of point density grouped by area label

The point density boxplots for each label shows a clear difference between the labels. Urban trackings appear to have the highest density and rural trackings the lowest, with suburban and town trackings falling in the middle with a wider variability.

This validates that the point density is a strong feature for classifying trackings to the four labels.

3.4 Solution Approach

Formatvorlage für den Fließtext.

4 Implementation

4.1 Implementation of the Big Picture

Formatvorlage für den Fließtext.

4.2 Integration with existing systems

Formatvorlage für den Fließtext.

5 Evaluation and Discussion

5.1 Definition of the data sets used for the evaluation

Formatvorlage für den Fließtext.

5.2 Evaluation of the results

Formatvorlage für den Fließtext.

5.3 Reflection on the results

Formatvorlage für den Fließtext.

6 Conclusion

6.1 Future Directions

Formatvorlage für den Fließtext.

6.2 Limitations

Formatvorlage für den Fließtext.

Bibliography

[1] "The world's most valuable resource is no longer oil, but data," *The Economist*, ISSN: 0013-0613. [Online]. Available: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data (visited on 04/06/2025).

[2] T. Petroc, *Data growth worldwide 2010-2028*, en. [Online]. Available: https://www.statista.com/statistics/871513/worldwide-data-created/ (visited on 04/11/2025).

[evtl. Anhang]

Formatvorlage für den Fließtext.

6.3 Use of AI tools

Table 6.1: Use of AI tools during the creation process

Working step	**AI used**	**AI tool(s)**	**Experiences / recommendations / irritations**
Find a topic idea	no	-	-
Narrow down topic / Formulate question	no	-	-
Find sources	yes	OpenAI GPT-4o	Helped identify relevant keywords and topic clusters.
Explain terms	yes	OpenAI GPT-4o	Useful for quick definitions and simple explanations.
Design text structure	yes	OpenAI GPT-4o	Good for outlining sections. Required manual adjustments.
Have content read aloud	no	-	-
Translate content	no	-	-
Dictate content	no	-	-
Paraphrase content, summarise	yes	OpenAI GPT-4o	Helpful to generate concise summaries of long text.
Write introduction	yes	OpenAI GPT-4o	Provided inspiration but required rewording.
Write main chapter	yes	OpenAI GPT-4o	Used for structuring and rephrasing, not full writing.
Write a summary	yes	OpenAI GPT-4o	Used to condense main points effectively.
Obtain text feedback	yes	OpenAI GPT-4o	Used to review tone, clarity, and consistency.
Revise text statement	yes	OpenAI GPT-4o	Helpful for reformulating statements.
Revise text formulation	yes	OpenAI GPT-4o	Polished sentences and improved flow.
Correct text formally	yes	OpenAI GPT-4o	Assisted with grammar and punctuation checking.

Affidavit

I hereby declare in lieu of oath that I have written this Bachelor thesis independently and without the use of aids other than those specified. The passages taken directly or indirectly from other sources directly or indirectly from other sources are marked as such. The thesis has not been neither in the same nor in a similar form to any other examination authority nor has it been published.

Dornbirn, on 15. May 2025 Matthias Hefel