UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Computer Science
Innovation and Technology Management Curriculum

Manfredo Estuardo Aceituno Pérez

Multi criteria decision process model for postal service in combination with crowd delivery

Master's Thesis (20 ECTS)

Supervisor(s):
Uku Tulev, Msc
Eduard Ševtšenko, PhD

Tartu 2025

Multi criteria decision process model for postal service in combination with crowd deliverers

Abstract:

Routing problems in logistics planning are crucial for companies, as they can lead to economic challenges and unstructured supply chain processes. Many professionals in logistics focus on optimizing models to solve these issues, which affect both internal and external processes of organizations, directly and indirectly. This thesis aims to conduct a business analysis of an international postal and logistics company in the Baltics, focusing on how implementing a crowdsourced delivery (CSD) service for parcel orders could optimize their logistics operations. The first part of the thesis includes comprehensive literature on route optimization, an analysis of different VRP methods, the second part proposes solutions for crowdsourced logistics (CSL). The third part evaluates the delivery business model, the viability of integrating CSD, and the fourth part presents a review of its impact on logistic operations, including the benefits and disadvantages of this alternative.

The sections three and four were simulated by randomly generated standardized transactional data, where we tested unsupervised machine learning clustering (ML) models to infer the optimal parcel allocation and route sequences, bridging the gap between parcel allocation, route planning and delivery execution. These methods provided a basis for the parcel delivery company, addressing the nuances of potentially implementing partial CSD into their operations model, and provide as well, a reference for further research on integrating crowdsourcing into logistics business models.

Keywords: Operational Research (OR), Vehicle Routing Problems (VRP), Logistics and transportation optimization, Crowdsourced delivery (*CSD*), Business analysis, Predictive modeling.

CERCS: P160 Statistics, operation research, programming, actuarial mathematics

Mitmekriteeriumiline otsustusmudel postiteenuse jaoks rahvakulleri kaasamisel

Kokkuvõte:

Marsruutimisprobleemid logistikaplaneerimisel on ettevõtetele võtmetähtsusega, kuna need võivad põhjustada majanduslikke raskusi ja tarneahela struktureerimatust. Paljud logistikavaldkonna spetsialistid keskenduvad mudelite optimeerimisele nende probleemide lahendamiseks, mis mõjutavad organisatsioonide sise- ja välisprotsesse nii otseselt kui ka kaudselt. Käesoleva lõputöö eesmärk on analüüsida äritegevust ühes rahvusvahelises postiteenuste ja logistika ettevõttes Balti riikides, keskendudes sellele, kuidas rahvakulleri teenuse (CSD – *crowdsourced delivery*) rakendamine võiks optimeerida nende logistikaoperatsioone. Uurimus koosneb neljast osast: esimene osa annab põhjalikku kirjanduse ülevaate marsruudioptimeerimise kohta; teine osa sisaldab erinevate sõidukimarsruudi probleemide (VRP – *vehicle routing problems*) analüüsi; kolmas osa hindab kohaletoimetamise ärimudelit ja CSD teostatavust; neljas osa annab ülevaate CSD mõjust logistikaprotsessidele ja toob välja selle lahenduse eelised ja puudused.

Uurimistöö kolmas ja neljas osa rakendavad simulatsioone, toetudes juhuslikult genereeritud standardiseeritud tehinguandmetele. Testisime järelevalveta masinõppe (ML – *machine learning*) klasterdamise mudeleid, et teha kindlaks optimaalne pakkide jaotamine ja marsruutide järjestus, luues seeläbi silla pakkide jaotamise, marsruudiplaneerimise ja kohaletoimetamise vahel.

Töö annab sisendi pakiveoettevõttele, käsitledes osalise CSD võimalikku rakendamist tegevusmudelis ja pakub ühtlasi lähtepunkti edasiseks uurimistööks rahvakulleri lähenemise integreerimisel logistikamudelitesse.

Võtmesõnad: Operatsioonanalüüs (OR – *operational research*), sõidukimarsruudi probleemid (VRP), logistika ja transpordi optimeerimine, rahvakuller (CSD), ärianalüüs, ennustav modelleerimine.

CERCS: P160 Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika.

Acknowledgements

I would like to express my sincere thanks to my family for their constant emotional and affectionate support throughout my studies. To my mother, Julieta Pérez de Aceituno, my father, Manfredo Aceituno Barrera, and my sister, Ligia Maria Aceituno, thank you for always being there for me and for your constant encouragement.

I want to thank my grandparents, Mario Pérez, Sila Reyes and Mimí Barrera, as well as the Díaz Aceituno family, Ileana, Jose Mario, María Mercedes, Jose Antonio, and Juan Silverio for their enthusiasm from the moment I told them I was going to study abroad a master's degree. Their excitement and encouragement meant a lot to me.

I am very grateful to Sander and Johanna Tammist, and their children, Säde, Silver, and Evar, who welcomed me to Estonia from the very beginning. They have not only been great friends, but I consider them my Estonian family.

I would also like to thank my Professor, Eduard Ševtšenko, and my thesis advisor, Uku Tulev, for trusting in my skills and giving me the opportunity to work on this thesis which falls within the research areas of their Digital Supply Chain Research Group. I especially appreciated the helpful feedback and productive discussions with Uku Tulev, he was very detail oriented, with a strong focus on mathematical methods and literature.

Finally, I would like to thank Aleksei Vahrušev, who, representing the company Omniva, generously gave his time, expertise, and insights around the subject of logistics, which were essential for the development of this thesis.

Manfredo Aceituno Pérez
Tartu, 2025

Contents

i. Nomenclature 1

1. Introduction 2

    1.1 Scope 2

    1.2 Objective of research 3

    1.3 Research questions 3

    1.4 Research method 4

    1.5 Thesis contribution 4

    1.6 Thesis structure 4

2. Systematic Literature Review 6

    2.1 Literature Sources 6

    2.2 Search Query 6

    2.3 Inclusion and Exclusion Criteria 7

    2.4 Paper Selection 8

    2.5 Data Extraction Strategy 9

    2.6 Quantitative Overview 10

3. Last mile Logistics Literature 12

    3.1 Crowdsourcing Logistics Overview 12

    3.2 Crowdsourcing Operations Framework 13

    3.3 Studies on Parcel Delivery 17

4. Methodology and Implementation 21

    4.1 Theoretical principles of Path Planning Algorithms 21

    4.2 Unsupervised Machine Learning Algorithms 24

    4.3 Parcel Segregation 28

    4.4 Conditions for Scenario Simulation 31

5. Model Development 33

    5.1 Data processing: Delimiting the geographical region 33

    5.2 Randomly simulated: Delivery points and Parcel attributes 35

    5.3 Data Normalization and Application of Clustering algorithms 38

    5.4 Implementation of Unsupervised ML Clustering Algorithms 38

        5.4.1 K-means Clustering algorithm 39

        5.4.2 DBSCAN Clustering algorithm 41

5.4.3 Methodology for Parcel segregation to Crowd couriers 44
5.4.4 Simulation of scenarios 46
6. Benefits and Drawbacks of crowdsourced parcel delivery 50
7. Conclusion 51
7.1 Future Work 52
7.2 Omniva's addendum on the thesis research 53
List of References 55
Appendix 61
1. License 61

List of Tables

Table 1. Search strings 7
Table 2. Exclusion/inclusion criteria utilized. 7
Table 3. Paper selection 9
Table 4. Information extraction form 10
Table 5. Scenario setting overview 32
Table 6. Sample of randomly generated drop off points 36
Table 7. Comparative clustering results 44
Table 8. From Parcel Machine to Drop off points simulation 47
Table 9. From Postal Office to Drop off points simulation 48
Table 10. From Departure Depot to Drop off points simulation 49
Table 11. Advantages and disadvantages classification 50

List of Figures

Figure 1. Research framework to provide answers to the main research question 3
Figure 2. Systematic Literature Review Process 4
Figure 3. Distribution of papers published per year 11
Figure 4. Number of Papers per Database 11
Figure 5. General current state process of last mile logistics operations 14
Figure 6. General Last Mile Logistics Operations Framework 15
Figure 7. Occasional Couriers from Parcel Machines to Customers’ Doorsteps Framework . 16
Figure 8. Occasional Couriers from Postal Offices to Customers’ Doorsteps Framework 16
Figure 9. Occasional Couriers from Centralized Hub to Drop off Points. 17
Figure 10. Map from OpenStreetMap representing a) starting point in a road network, and b) starting point in a graph/edges network. 22
Figure 11. Simplified example of the shortest path approach for a VRP 23
Figure 12. Flowchart of the clustering algorithms for the proposed scenarios 27
Figure 13. Comparison overview of crowdsourcing scenarios 32
Figure 14. Estonia PUDO’s 35
Figure 15. Tartu County PUDO’s 35
Figure 16. Elbow method for K-Means 39
Figure 17. Drop off points K-Means clustering 41
Figure 18. Elbow method adaptation for DBSCAN 42
Figure 19. Drop off points DBSCAN clustering 44
Figure 20. Parcel dimensions 45

List of Code Blocks

Code block 1. PUDO's dataset filtering 34
Code block 2. Tartu County, Estonia. PUDO's segmentation 34
Code block 3. Randomly generated set of parcel attributes 36
Code block 4. W score coefficients 37
Code block 5. Unscaled data 38
Code block 6. Normalized (scaled) data 38
Code block 7. Davies-Bouldin iterative score method 40
Code block 8. Unsupervised ML K-Means algorithm method 41
Code block 9. Silhouette Score 42
Code block 10. Parcel segregation: parameters for OC's 45
Code block 11. Parcel segregation: filtering candidates for OC's 46

i. Nomenclature

Abbreviation	Definition
B2C	Business to Customer
CSL	Crowdsourced Logistics
DD	Departure Depot
EC	Exclusion Criteria
IC	Inclusion Criteria
LMD	Last Mile Delivery
ML	Machine Learning
MRQ	Main Research Question
OC	Occasional couriers
OR	Operational Research
PM	Parcel Machine
PO	Postal Office
PUDO	Pick up & Drop off points
RQ	Research Question
SRL	Systematic Literature Review
TCO	Total Cost of Ownership
TDI	Tembi Delivery Index
VRP	Vehicle Routing Problem

1. Introduction

With the rapid development and possibilities that technology offers to simplify and improve service delivery, various companies have taken advantage of the potential and innovation of the shared logistics model - crowdsourced logistics (CSL) - to improve not only their performance in terms of direct and indirect costs, delivery performance expectations, but also the use of the existing network of local couriers to effectively meet demand needs. Crowdsourcing implemented in transport models is one of the main services that have emerged with a focus on LMD, especially because the process of moving goods from a distribution center to the final customer is often costly in terms of proximity to the planned route, the speed at which the service can be completed, availability, or vehicle capacity in relation to delivery zones[1]. Originally, companies or startups that emerged with the goal of providing LMD services focused mostly on rides or food delivery. Startups like Lyft or Uber for rides, as well as Instacart, Ubereats, and Onfleet for food and other goods delivery in the United States, Glovo in Spain, or Wolt and Starship in Estonia and Finland, all focus on food delivery. Nevertheless, rather than food, in the Baltics, the constant demand for parcels, instant delivery, and increasing customer expectations are presenting unforeseen opportunities for logistics service providers and the fleet freight system. Among the players in the delivery market, new service models based on LMD are becoming increasingly popular, enabling third-party applications and individual couriers who affiliate and travel using their own means of transport, such as scooters, bikes, on foot, or public transport, who, through compensation, could offer a more efficient and instant delivery in nearby areas, besides the current self-service parcel lockers. Therefore, we believe that one challenge this crowdsourced system may be viable to bring lower operating costs for companies, compared to the traditional logistics models [2] and provide a better LMD service, not only through distance savings but through transportation time.

1.1 Scope

Our thesis focal point is on experimental route optimization. The study considers transactional simulated data based on the parcel criteria delivery business model from a postal and logistics service provider operating in the Baltics who manages goods of different categories such as documents, toys, electronics, clothing, etc. The postal service provider currently is operating by picking and delivering goods with its own fleet and parcel lockers across the Baltics, nevertheless, this thesis does not explicitly aims to develop a software system for transportation, but focuses more on modeling different scenarios for the delivery operations of a logistics company in a bounded region of the Baltics, applying stochastic and machine learning methods to address the acceptability of occasional couriers (OC's), and providing a decision-making framework where transportation tasks can be carried out by a wide array of participants.

1.2 Objective of research

The crowdsourcing concept empowers individuals, either on foot or using their own vehicles such as cars, scooters, or bicycles, to pick up and deliver packages for the company's clients. This approach is intended to provide a logistics parcel delivery company with a better understanding of the outcomes, benefits and drawbacks of complementing their current reliance on their own fleet with crowdsourced delivery.

1.3 Research questions

In this section, we define our main research questions and break them down into smaller research questions to refine our main objectives for this thesis.

MRQ: What are the key factors influencing the implementation of a crowdsourced shipping approach for LMD in postal logistics services?

The thesis main question breaks down into the following focus:

- **RQ1:** What is the framework of reference for postal delivery operations?
- **RQ2:** What condition(s) can be formulated for identifying parcel candidates for crowdsourcing?
- **RQ3:** How can unsupervised Machine Learning algorithms be implemented to cluster and allocate parcels for crowdsourcing delivery in simulated scenarios?
- **RQ 4:** What are the benefits and drawbacks in the context of the crowdsourcing modality for a postal and logistics company?

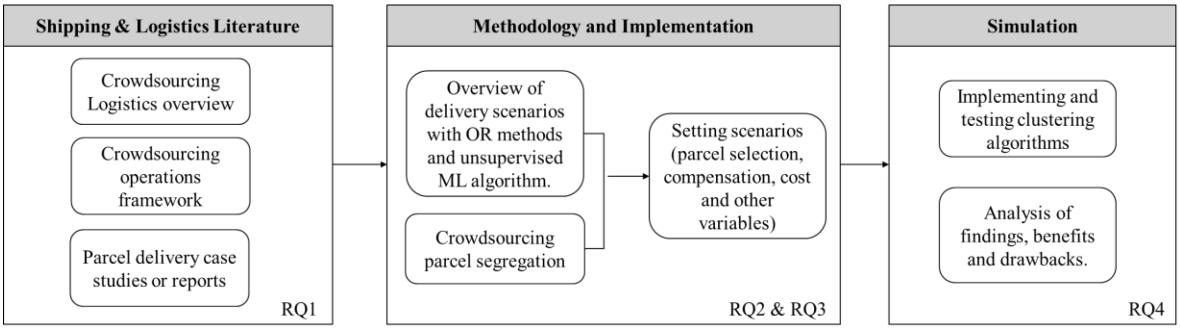


Figure 1. Research framework to provide answers to the main research question

1.4 Research method

We conducted Systematic Literature Review (SLR) using Kitchenham *et al* (2007). [6] approach. Primarily focusing on the methods and case studies that explore (RQ1) the reference framework for postal delivery operations, (RQ2) the conditions that determine crowdsourcing allocation, (RQ3) what are the modeling techniques and its performance, and (RQ4) the range of advantages and challenges in crowdsourced logistics operations. The main objective of the SLR was to explore different aspects of crowdsourcing in the Last Mile Delivery. Thus, proposed recommendations and present conclusions on how crowdsourcing cases improve the logistics operations for a postal delivery company.

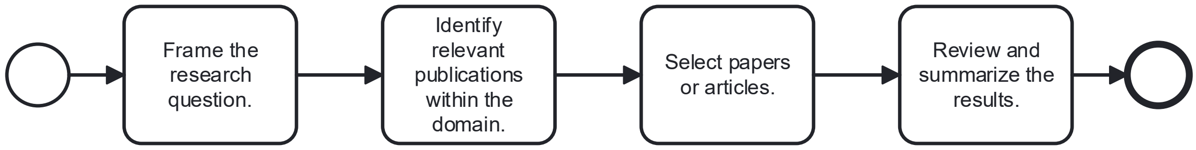


Figure 2. Systematic Literature Review Process

1.5 Thesis contribution

Methodologically, the thesis topic makes two major contributions:

- Presenting a comprehensive theoretical review of the LMD in Vehicle Routing Problems, and a business analysis that involves crowdsourcing for parcel relay between crowdsources and end customers (B2C).

- While several studies take all parcel demand for a crowdsourcing service, it is still a gap in understanding which parcels are the better candidates for this model. Therefore, applying existing Operational Research (OR) mathematical methods, and developing, training, and testing a Machine Learning (ML) model using real PUDO's data from a Postal & Logistics Company, we aim to infer optimal recommendations and bridge the gap between route planning, delivery execution, and expected costs for a crowdsourced delivery service, allowing for a simulation of small-scale situational cases and the comparison of solutions for larger and more realistic parcel delivery data.

1.6 Thesis structure

The thesis is organized into seven chapters as follows: Chapter 2 introduces the SLR about the topic. It describes how we did the search for papers and previous research documentation around the subject and presents a general overview of the data extraction strategy and a quantitative overview of the sources and publications. Chapter 3 focuses on logistics literature, framework for logistics operations and presents studies that have already been explored on parcels delivery. Chapter 4 introduces vehicle routing methods for transportation we used, describes what the unsupervised ML algorithms are, and the conditions that were set to build scenarios for parcel allocation and delivery. Chapter 5 introduces the decomposition strategy for the development of models, data processing, generation of random standardized data for simulations and implementation of ML clustering algorithms. Chapter 6 introduces a comparative chart of a crowdsourcing business model. In Chapter 7, we give the main conclusions, and the final chapter ends with recommendations we would like to propose for further investigation.

2. Systematic Literature Review

We performed the systematic literature review (SLR) method of crowdsourced shipping and its potential to innovate the postal logistics service in LMD. In order to analyze the existing knowledge regarding crowdsourcing in postal logistics, we followed the steps defined by Kitchenham *et al*. (2007) [6] to ensure a structured review process and synthesizing the findings according to the research questions (RQs), **RQ1** frameworks and models used in postal delivery systems, specifically focusing on operations related to LMD. **RQ2** the criteria and conditions that help identify parcels suitable for crowdsourced delivery. This required exploring studies that described parcel characteristics, geographic factors, customer preferences, and other methods used for parcel selection. **RQ3** analyzing the types of unsupervised ML algorithms used in postal delivery, their accuracy, and their impact on the overall performance implementation for crowdsourcing-based postal logistics, **RQ4** understanding the operational, financial, and strategic advantages of adopting crowdsourcing in LMD, as well as the challenges and limitations associated with it.

2.1 Literature Sources

To identify potentially relevant papers, the databases were selected based on the coverage of publications within the field logistics and route optimization. For the literature sources we identified that the suitable electronic databases were Google Scholar, Emerald Insight, Springer Link, Scopus, ACM Digital Library, IEEE Xplore, and we also included References to pertinent works related to the research questions; these sources included articles or conference reports. And finally, we conducted backward referencing (snowballing) [6] to identify additional relevant papers.

2.2 Search Query

Exploring in the database directory involved selecting the most pertinent literature based on searching key terms and queries as presented in Table 1, these were used for each platform: *Crowdsourcing delivery*, *Last-mile*, *Parcel delivery*, *Route optimization*, V*ehicle routing problem*, *Machine learning*, *Crowdsourcing path planning*, *Ride sharing logistics*, and *Crowdsourcing logistics pricing optimization*.

For the search string, we used the operators AND and OR, we also used synonyms such as *postal delivery* for *parcel delivery* to potentially find papers that described the scope accordingly. We combined the keywords in the databases to make sure we found as many relevant publications as possible about crowdsourcing in postal logistics. We noted that the term “*crowdsourced delivery*” was not consistently used, therefore, we chose to include "*crowdsourced shipping*" to make the search category broader.

Table 1. Search strings

Search String Queries	Databases
"Crowdsourcing delivery" AND "Last-mile"	Google Scholar, Springer Link, Scopus, IEEE Xplore
"Parcel delivery" OR "Route optimization" AND “Heuristics”	Google Scholar, Springer Link, Scopus, IEEE Xplore, Emerald Insight
"Vehicle routing problem" AND "Crowdsourcing" OR "Last-mile"	Springer Link, Scopus, ACM Digital Library, IEEE Xplore
"Crowdsourcing path planning" AND "Optimization"	Springer Link, Scopus, IEEE Xplore, ACM Digital
"Ride sharing logistics" OR "Crowdsourcing" AND "Delivery"	Google Scholar, Springer Link, IEEE Xplore, ACM Digital Library
"Crowdsourcing logistics pricing optimization" AND "Route planning"	IEEE Xplore, Google Scholar, Scopus, Springer Link
"Crowdsourcing logistics" AND "Vehicle routing problem" OR "Last-mile"	Springer Link, IEEE Xplore, ACM Digital Library, Scopus
"Machine learning" OR "Last-mile delivery" AND "Crowdsourcing path planning"	Google Scholar, Springer Link, Scopus, ACM Digital Library, IEEE Xplore, Emerald insight

2.3 Inclusion and Exclusion Criteria

We defined the exclusion and inclusion criteria based on Table 2.

Table 2. Exclusion/inclusion criteria utilized.

Inclusion Criteria	Exclusion Criteria
The paper discusses crowdsourcing delivery or last-mile delivery.	The paper is not related to crowdsourcing or last-mile delivery.
The paper focuses on route optimization or vehicle routing problem.	The paper does not provide any information about route optimization or logistics.
The study uses data analysis, optimization heuristic models, or machine learning in the context of delivery logistics or path planning.	The paper does not use any data, models, or relevant methods for analyzing logistics with machine learning or heuristic methods.

The paper presents research or case studies from real-world or practical applications.	The paper is too theoretical and does not discuss practical examples or case studies.
The paper is digitally accessible.	The paper has limited/closed access.
The paper language is English.	The paper is not written in English.

2.4 Paper Selection

The selection of papers began with searching for keywords using the queries described in Section 2.2. Searching in the digital libraries and databases mentioned in Section 2.1 resulted in two hundred and seventy relevant documents for this thesis. The queries focused on crowdsourcing in postal delivery and limited the results to papers covering crowdsourcing for LMD and route optimization with stochastic or machine learning approaches. The papers were analyzed manually by reading the title, abstract, and introduction to ensure they matched the scope of the research.

The first paper selection used the inclusion and exclusion criteria, as shown in Table 2. During the screening process, five papers were removed as duplicates. Additionally, twenty one papers focusing on crowdsourcing path planning were excluded during the abstract review, despite being a keyword in the initial search. This was because these papers examined path planning as hierarchical tasks assignment among workers, and in other cases modeling indoor paths for emergency disasters, both main outcomes had different scope than crowdsourced postal delivery in last-mile logistics. Some papers related to parcel delivery and route optimization were also excluded because they covered drone assisted topics that were not related to courier postal delivery.

The remaining sixty eight papers that passed these filters were those for which access was granted. And eight additional papers were obtained through snowballing. For the source News or Magazine Articles, in IC.6 (Inclusion Criteria 6.), six publications were relevant, but only four where in English, the other two were in Estonian, nevertheless we decided to include them as they were significant. At the end, forty four papers, journals or previous research publications were considered for the study of this thesis.

Two filters were applied to narrow down the final papers, which then underwent quality assessment (see Section 2.5). The first filter was general, as described above, while the second filter was more specific, using both inclusion/exclusion criteria and quality assessments for documents that passed the initial selection criteria.

The final set of papers consisted of forty five papers were found most relevant to the research. The reason for the reduction in the number of papers was that many of them fell outside the scope of this study, as most focused on food delivery business models or software applications for vehicle-sharing systems, none of which concentrated on the operational research for postal delivery or machine learning for LMD scenarios through crowdsourcing.

Table 3. Paper selection

	Selection Criteria	Science Direct	Google Scholar	Springer Link	ACM Digital Library	Emerald Insight	IEEE Xplore	Snowball	News Articles/ Magazines	Total (After Filtering)
	Papers initially selected	60	52	30	20	15	40	35	18	270
	Inclusion Criteria									
IC 1.	Relevant to crowdsourcing postal delivery	50	40	25	15	10	30	28	15	213
IC 2.	Focuses on route optimization or vehicle routing problems	40	30	20	12	8	25	22	12	169
IC 3.	Uses data analysis, optimization heuristics, or machine learning in logistics	35	25	19	10	7	20	18	10	144
IC 4.	Includes real-world research or case studies	25	20	15	8	5	15	12	8	108
IC 5.	Digitally accessible	20	15	10	5	3	10	10	6	79
IC 6.	In English**	23	12	8	3	2	8	8	4	68
	Exclusion Criteria									
EC 1.	Irrelevant to postal delivery or last-mile logistics	8	4	5	1	1	2	0	0	21
EC 2.	Duplicates or non-relevant papers	2	0	1	1	0	1	0	0	5
	Final papers after filtering	**13**	**8**	**3**	**1**	**1**	**5**	**8**	**6**	**45**

**IC 6.: In the category of News/Magazines Articles two publications were in Estonian and considered because of its relevance for this thesis.

2.5 Data Extraction Strategy

We planned to collect the data we need to answer every research question, see Table 4. First, we gathered basic information from all the papers, like the title, author, study domain, where it was published, and the year. After that, we focused on the data we needed to answer the research questions. For **RQ1**, we looked at the information that evidence and let to understand the framework for postal delivery operations, *e.g.*, the delivery process workflow and the factors that are important for the system. For **RQ2**, we collected data to help to figure out what makes a parcel a good candidate for crowdsourcing, *e.g.*, the conditions or rules to be considered to identify which shipping could be delivered through crowdsourcing. For **RQ3**, we focused on gathering data about how well Machine Learning algorithms can be adjusted to predict the costs and allocation of parcels for crowdsourcing in a fixed context. Finally, for **RQ4**, we collected information about the benefits and drawbacks of using crowdsourcing for postal and logistics companies, *e.g.*, specific situations where crowdsourcing is most useful and when it might not work well.

Table 4. Information extraction form

Information	Description
	Paper identification data
DOI	Digital Object Identifier (ID of the paper).
Title	Title of the paper.
Author(s)	Authors of the paper.
Year	Year of publication of the paper.
Publication	Site where the paper was published.
	Study factors
Crowdsourcing in LMD	The purpose of using crowdsourcing in logistics.
Framework for Postal delivery	Description of the framework used for postal delivery operations (RQ1).
Parcel candidate conditions	Conditions for identifying parcels suitable for crowdsourcing (RQ2).
Mathematical and Machine learning models	Details and implementation of Heuristic and Machine Learning models (RQ3).
Benefits of crowdsourcing	Benefits of using crowdsourcing for postal and logistics (RQ4).
Disadvantages of crowdsourcing	Inefficiencies of using crowdsourcing for postal and logistics (RQ4).
Context of crowdsourcing	Scenarios and business models where crowdsourcing is used (RQ4).

2.6 Quantitative Overview

The distribution of the papers over the year of publication is shown in the Figure. 3. Among the identified papers, the earliest paper in which the concept of crowdsourcing was written is from 2006. Since then, it seems that studies on crowdsourced shipping and postal logistics have been published mainly from the year 2015. However, we found that it was a trending topic in academic research around 2019 to 2022, where twenty three papers that are referenced to our thesis were published in that period time. All the identified publications were peer-reviewed, forty-four were selected from digital research databases and they fall in the categories of conference papers, journals, case studies, books or publications form consultancy firms.

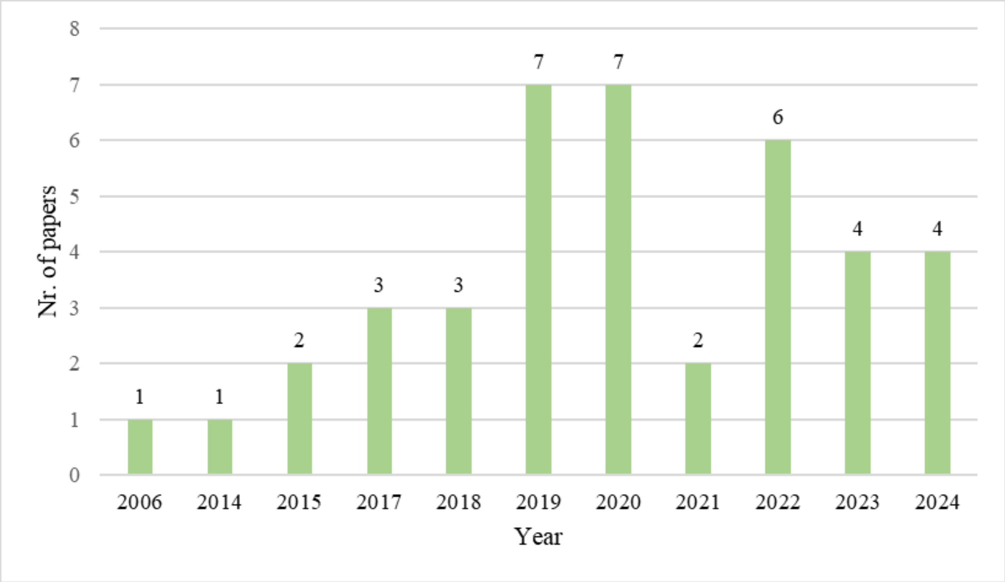


Figure 3. Distribution of papers published per year

Figure. 4 shows the distribution of papers across different search databases, including Scopus, IEEE Xplore, Google Scholar and others mentioned in section 2.1. This distribution helps us understand the spread of research on crowdsourced shipping and postal logistics across different platforms and sources used for this thesis.

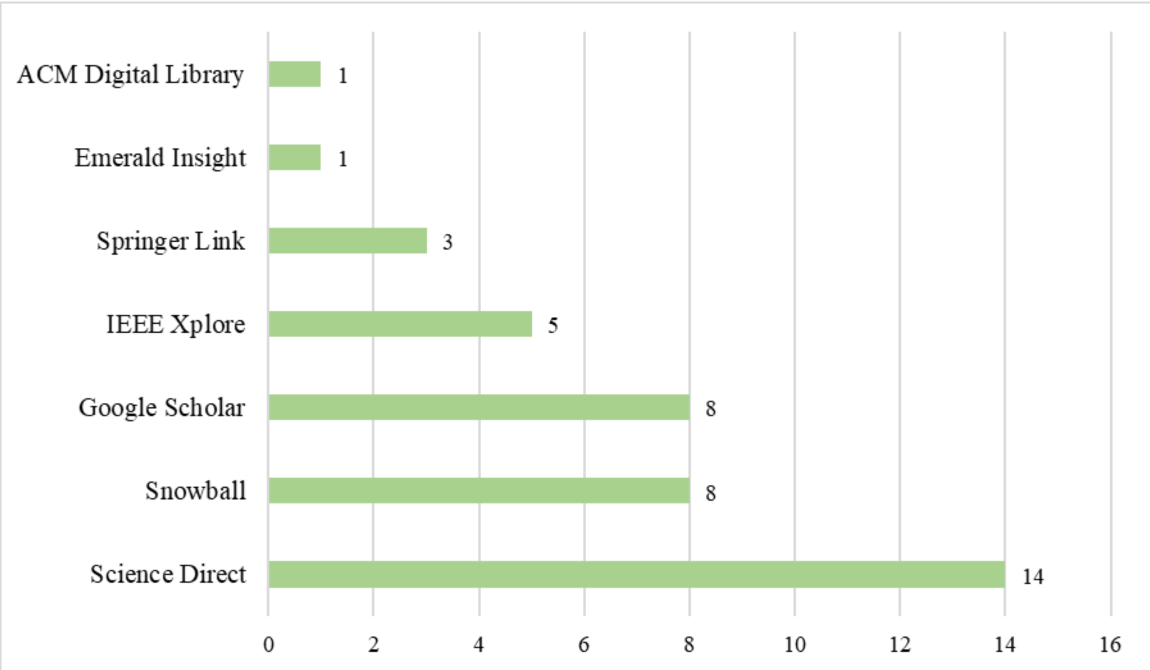


Figure 4. Number of Papers per Database

3. Last mile Logistics Literature

3.1 Crowdsourcing Logistics Overview

Today, the topic of Crowdsourced Logistics (CSL) has become important as a more efficient alternative to improving the delivery of different goods. This means improving quality, time, ease of implementation, reliability, operational strategy, low cost, etc., both for the company that produces a product, the logistics provider, and especially for the final customer who wants to satisfy an immediate need and at the best possible price.

The term "crowdsourcing" is actually a combination of the words "crowd" and "sourcing," as described by Howe [9], who coined the concept through examples of how the digital ecosystem, companies, consumers, and other agents join the value chain even if they do not have the expertise or are not part of it in the first place. "Crowd" is defined as a group of people, and "outsourcing" refers to third parties who act as employees of a company without being part of it, carrying out tasks, activities, or processes.

In Howe's words, companies grew up in the Internet age and were designed to take advantage of the networked world. But now the productive potential of millions of connected enthusiasts is attracting the attention of traditional businesses. Therefore, paraphrasing Howe's idea, crowdsourcing is a way to connect with people outside the company who are part of the network and could help develop and speed up market access to businesses. In the logistics industry, crowdsourcing has opened many possibilities for delivery or relying on external agents for fleet shipping. This approach has become popular for executing Last Mile Delivery (LMD) in local areas, especially because of the benefits it brings to the collection and delivery of packages, as we aim to study with this thesis.

To organize a crowdsourcing delivery approach, it is essential to have an information and communication system that links the organization with the large community of independent collaborators and the final customer. This thesis will not focus on specifying all the details of what an IT system involves, nor the type or design of it, but we will emphasize the premise that it is a necessary factor for crowdsourcing logistics to manage and control the postal delivery process confidentially and securely. It also helps measure the quality of the service provided by the network of collaborators, whom we call occasional couriers (OC's), keep their motivation and participation high, and, on the other hand, provide information to the customer, such as price, location, and status of their package, as well as other communication features.

This helps not only maintain the company's reputation and the quality of the service but also ensures that the crowdsourcing model is successful.

Therefore, "crowd logistics designates the outsourcing of logistics services to a mass of actors, whereby the coordination is supported by a technical infrastructure. The aim of Crowd Logistics is to achieve economic benefits for all stake- and shareholders." [8] The system platform is made available and accessible through mobile phones, tablet, or web browser with the purpose of coordinating the demand and supply of transport services. Additionally, it also handles management processes and invoice processing tasks, bringing economic advantages for the main business and among the participants.

The goal of CSL is to complete LMD, which is the process where a good is moved from the warehouse shelf, also known as the distribution center, to the vehicle that goes directly to the customer's doorstep. The goal of the LMD is to speed up delivery times and reduce the access gap to unpopulated towns, crowded, or less frequent areas of delivery routes.

Last Mile Delivery can be broadly categorized into two types [4]. First, a) delivery that goes directly from the distribution center to the customer, or to a transfer node and then to the customer, or second, b) partial crowdsourced delivery, where the delivery does not fully depend on occasional couriers (OC's), but instead there is a fleet of trucks that permanently make deliveries. As the segregation of parcels, routes, costs, schedules, and other variables are fixed, the delivery service is balanced with the goal of minimizing total costs [10].

In both cases, LMD, the crowdsourcing participants or OC's, rely on walking, cycling, or also use other types of vehicles, usually two-wheeled, like a scooter or motorcycle, the general idea is that they tend to be closer to the customer and although they might have certain limitations, such as collection-delivery distance and load capacity, compared to truck carriers, they tend to improve many scenarios of the logistics system.

3.2 Crowdsourcing Operations Framework

As we mentioned, crowdsourcing logistics models are generally divided into two main types. This thesis focuses on the second model, which *Kafle, N. et al* [10] calls partial crowdsourced delivery, where the aim is to find the optimal scenarios for segregating outlier parcels that occasional couriers (OC's) can collect and carry out the Last Mile Delivery process.

Given the overview of LMD, the general model of last mile logistics operations process is represented in Figure 5., which focuses on the operational level where the main courier could be a fleet truck owned or subcontracted by the postal service company. The case study in Estonia includes intermediate delivery points like parcel machines or letter boxes within the current logistic operational framework; these points are crucial not only for route planning but also to improve delivery conditions given factors such as space limitations in urban areas, covering customers who can't be reached to doorstep, or under the last mile concept, load consolidation for optimal truck capacity use and cost per kilometer.

The general last mile logistics operation framework in Figure 5. consists of the following stages: 1) the parcels are consolidated at departing depots, which can be post offices functioning as pick up and drop off warehouses or distribution centers, 2) the truck carrier collects the parcels, 3) the truck carrier follows the route plan to the delivery points, 4) the truck carrier could perform a relay where parcel is taken and transferred before its destination, 5) the truck carrier completes the delivery (at the parcel machines, letter boxes, or in some cases at the customer's doorstep), 6) the truck driver returns to the departing depot.

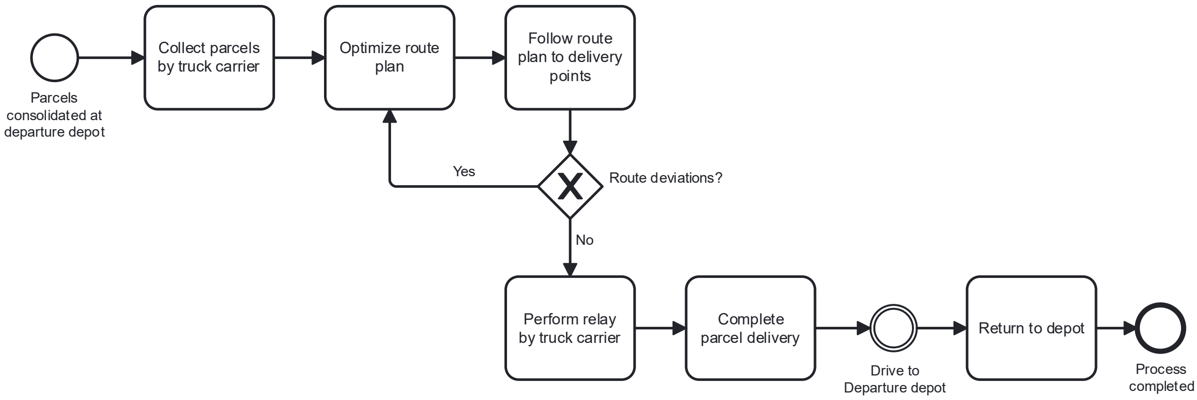


Figure 5. General current state process of last mile logistics operations

Based on the process model in Figure 5, we noted that most of the literature focuses on making more efficient one of the intermediate or final activities right before the final delivery of the parcel to the customer [1][11][20][21][22][23][24]. Therefore, the approach we suggest in this study looks at three frameworks described in Figures 7, 8, and 9. These are scenarios that can be in line with the current shipping process and will allow us to analyze, and understand new results in a systematic way where crowdsourcing can be implemented as consequence of the research and approach to the LMD problem for the Postal & Logistics Company.

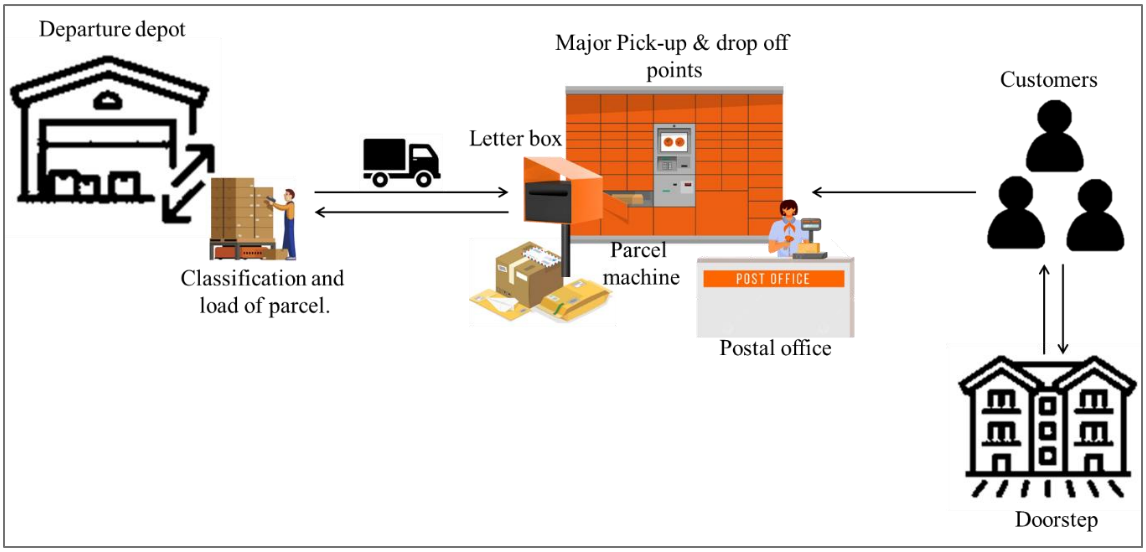


Figure 6. General Last Mile Logistics Operations Framework

Figure 6. represents the process model of Figure 5. of the current state of LMD for the study case covered in this thesis. The current process utilizes PM's and PO's as one of their main resources to shorten the LMD and facilitate customers the pick-up and drop off of parcels.

LMD is a logistics field that is constantly adapting to new technologies, methods or changes in demand, that is the reason why it is always subject to gradual optimization and iterative improvements. Even though truck deliveries cannot be completely replaced and present inefficiencies or external limitations defined by factors like the country's geography, road infrastructure, access routes, and traffic hours, we propose three crowdsourcing operation frameworks, in Figure 7. Occasional Couriers from Parcel Machines to Customers' Doorsteps, Figure 8. Occasional Couriers from Postal Offices to Customers' Doorsteps, and Figure 9. Occasional Couriers from Centralized Hub to Drop off Points. These scenarios could take advantage of OC's and iteratively optimize the LMD for parcels that cannot be delivered by the pure truck courier but could be segmented and gradually improve the direct delivery operation; that is what we analyze.

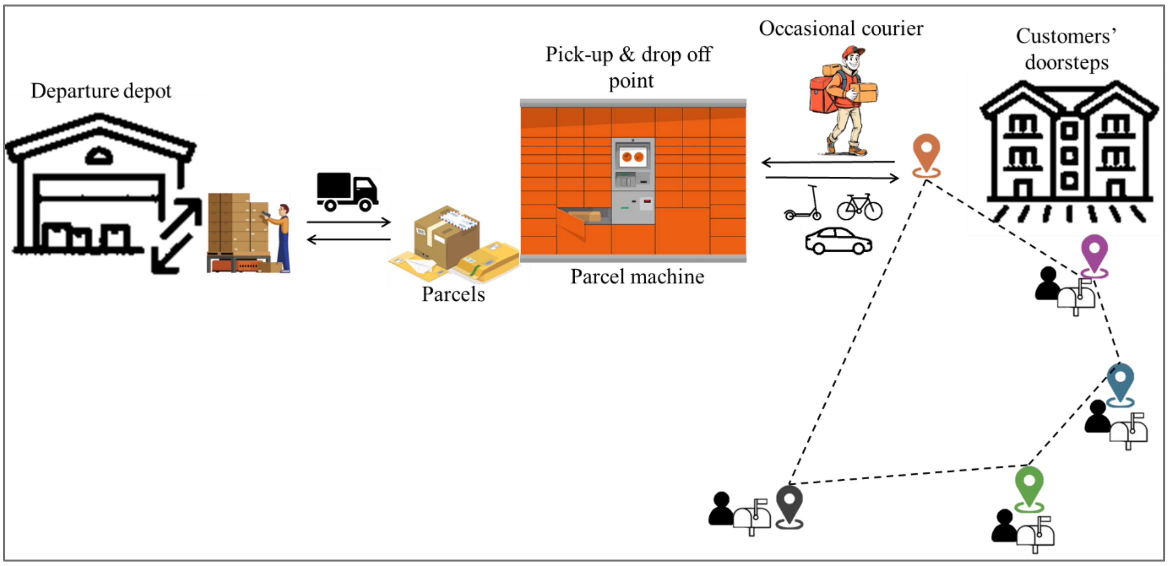


Figure 7. Occasional Couriers from Parcel Machines to Customers' Doorsteps Framework

Figure 7 consists of a scenario where the (OC collects one or more parcels from the parcel machine (PM) and delivers them to the customers' doorstep. This process can only be initiated under the assumption that the truck carrier has already collected the parcels at the Departure depot (DD) and made the drop off at the PM. The route and means of transportation are determined by the OC under the premise that it must protect the parcel and improve the operation and service level in the new LMD.

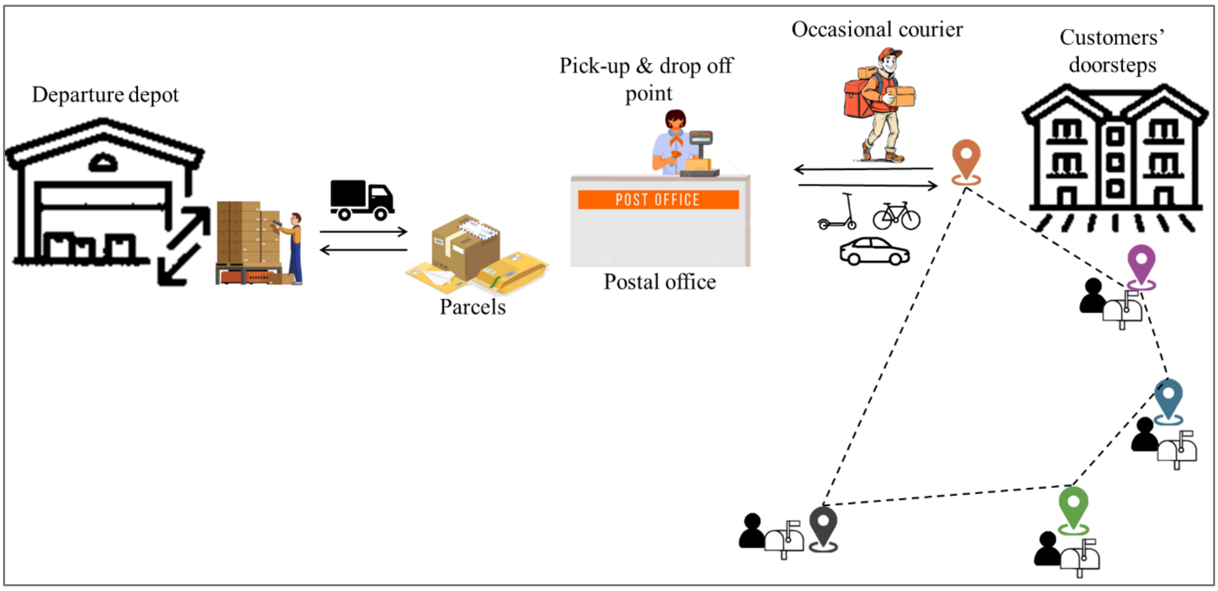


Figure 8. Occasional Couriers from Postal Offices to Customers' Doorsteps Framework

Figure 8 consists of a scenario where the OC collects one or more parcels at the PO and delivers them to the customers' doorstep. This process can only be initiated under the assumption that the truck carrier has already collected the parcels at the DD and made the drop off at the PO's.

The route and means of transportation are determined by the OC under the premise that it must protect the parcel and improve the operation and service level in the new Last Mile Delivery.

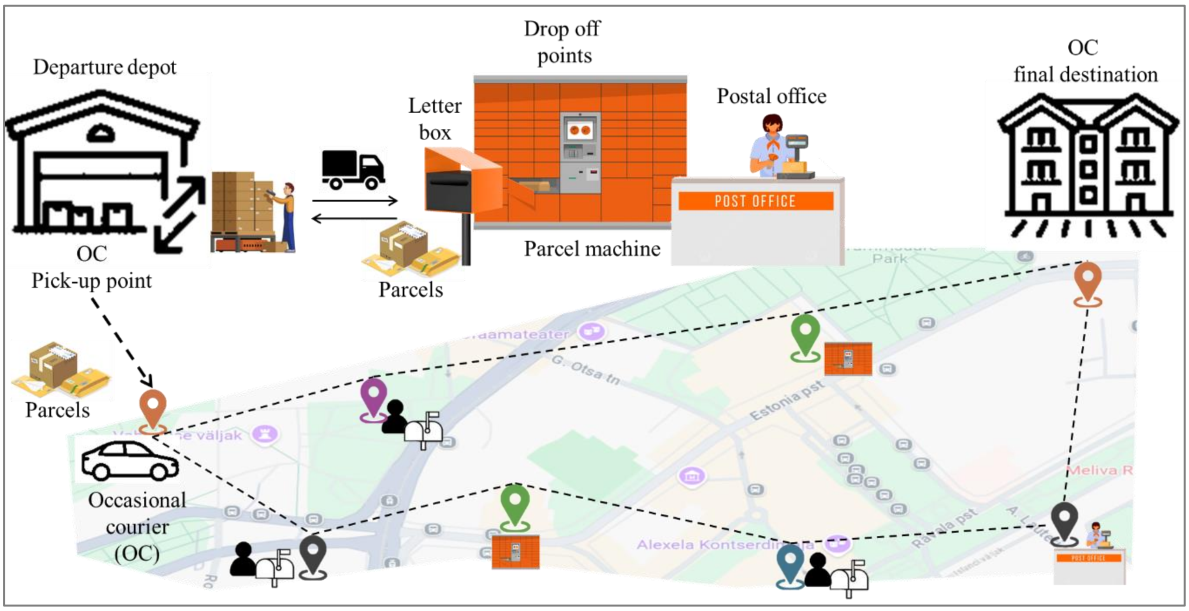


Figure 9. Occasional Couriers from Centralized Hub to Drop off Points.

Figure 9 consists of a scenario where the OC collects one or more parcels directly at the DD and as he drives along the route towards his destination, the parcel delivery operation is carried out at any of the drop off points (PM, PO, or customer doorstep) during the journey. The route and quantity of parcels to deliver are determined by the OC, but its means of transportation must be a four door passenger vehicle (*e.g.*: car, SUV, minivan, or pickup truck) under the premise that the parcel will be protected, and the operation and service level in the new LMD improved. In this scenario, the process for the fleet of truck carriers that provide service to the logistics company continues operating without any other changes.

3.3 Studies on Parcel Delivery

As we have described in Figure 5, the parcel delivery process follows a standard procedure that can be improved to reduce the last mile gap iteratively as technology and innovations add value and aim to reduce costs associated with the tasks and actors involved. One of the most positive aspects of crowdsourcing for shipping or delivery services is its similarity to traditional courier or transportation services [7][56], where a crowd of people exclusively carry out the collection, transportation, and delivery process safely for a company, but not necessarily as a fixed or contracted service with other implied costs and conditions.

A crowdsourcing process for parcel delivery could be not uniform or on the other hand, more favorable in certain conditions depending on the type of business, products, and other factors. For this reason, we will provide relevant case studies and exploration models that contribute to the understanding and analysis of the topic addressed in this thesis.

In 2015, the e-commerce company Amazon added crowdsourcing to its logistics model to optimize its Last Mile Delivery through the Amazon Flex platform [11]. Independent drivers (or gig workers, as they are called) could subscribe for free to the platform and choose a time that worked best for them to pick up and deliver parcels on the same day using their own vehicle. Eventually, the number of independent driver subscribers allowed Amazon to make deliveries even within a time window of hours. The Amazon Flex model works within the United States and started adding features like customer chat, so the independent driver could give updates on the delivery or receive instructions from the person receiving the parcel. In addition, Amazon Flex implemented incentives like discounts for vehicle maintenance workshops or insurance, for example, depending on whether the crowdsource earned these rewards.

Another case of crowdsourcing delivery is that of the supermarket chain Walmart [11], who in 2018 started with the pilot and first tests to test if independent drivers could deliver groceries to their customers. The program was called Spark Delivery [13], and its operation is quite similar to Amazon Flex. Even though during the launch of the Spark Delivery program, Walmart stores were an average of about 16 km from their customers in Arkansas, they implemented their own program to tackle the Last Mile Delivery. Initially Walmart tried to make partnerships with UberEats or Lyft, but today the program is still running under their own business model with the name Spark Driver [12].

The previous are delivery models that work in the United States of America. But there are also European companies like PiggyBaggy or DHL Express who have implemented pilots and initiatives to improve Last Mile Delivery (LMD).

For example, in 2013 [14], DHL executed the crowdsourcing delivery concept in Stockholm, Sweden, with the platform MyExpress. Even though the concept was a test that later joined other digital solutions like MyDHL+, today they still have new methods to make their LMD better relying on different methods and stages [15]. More recently, in 2022 DHL Express started using self-driving cars in Tallinn, Estonia to reduce their Last Mile Delivery costs lower [16]. Therefore, instead of being dependent only on their own fleet or considering occasional couriers (OC’s) as happened before with the platform MyExpress, the report indicates that DHL tested self-driving cars by sending packages between the three of their offices located in Tallinn. They have a partnership with Clevon, a self-driving cars company. One thing they found was that the key to an efficient LMD is to deliver on short routes. This is because after delivering the packages, there is a problem of driving back to the DD without any parcels.

Another study of crowdsourcing delivery is PiggyBaggy in Finland. The paper by Paloheimo *et al*. [17] shows the background and results of adding crowdsourcing to the book delivery service of the public library in the city of Jyväskylä, Finland. They used crowdsourcing deliveries where the residents or readers themselves could help in the delivery process while travelling along their daily commute. PiggyBaggy started with the ridesharing idea to make pick up and drop off faster and to lower the costs of LMD using crowdsourcing. Their service is based on a process where people handle and transport each other’s packages along their daily commute or shopping trips. Besides the case that Palohemio *et al*. described, PiggyBaggy joined with e-commerce and other delivery companies to make their solution scalable. It is not only cheap and efficient but also aims to be sustainable. [18]

The Estonian E-commerce Association (Eesti E-kaubanduse Liit) published in March 2024 [25] and February 2025 [26] the results of the Tembi Delivery Index (TDI) [24], which provides relevant information about the pricing of LMD frameworks in which parcel delivery companies operate. As Tembi, an E-commerce data analytics company, states on its platform, the “Tembi Delivery Index tracks the average price that e-commerce businesses charge private customers for delivery."[27] As a case study for LMD in parcel delivery, the TDI in logistics companies, e-commerce retailers, and consumers provide an accurate and recent basis of the average pricing dynamics of delivery across the three most common methods (Home delivery, Parcel lockers, and PUDO) in Nordic and Baltic countries. [27][28][29]

According to the Estonian E-commerce Association [25], in February 2024, Estonians ordered 1.16 million packages from parcel machines, which is 17.5% and 176,000 packages more than in February 2023. And in January 2025 [26] started with decent growth in e-commerce. Estonians ordered almost 15% more parcels from parcel machines in January than in the same month of 2024, which can clearly be seen as an opportunity to take advantage of the current framework in which parcel delivery operations (Figure 6), and to explore the outcomes with the proposed scenarios for this thesis (Figures 7, 8, and 9).

Each of the cases described above shows not only factual aspects of customer behavior towards the increasing demand of parcel delivery but also supports the logistics literature and study cases as we aim exploring potential opportunities based on crowdsourcing parcel delivery under certain operational scenarios. Also, they provide context on how the different people involved in the process, OC's, customers, e-commerce, stores, and transport companies, work together proving that LMD is not a static solution, but a concept in constant adaptation to small niches of a company procedures since there are new views and ideas that continue shaping crowdsourcing in logistics.

4. Methodology and Implementation

In this section, we are going to present the proposed methodology for each of the scenarios in Figures 7, 8, and 9. First, we will provide more details about the background of the problem and how each scenario would try to optimize the LMD in logistics operations for parcel delivery. Therefore, we will review the mathematical methods that OR presents for routing problems, as well as ML algorithms as a novel approach for parcel segregation. But also, we will examine, through academic literature and case studies, what are the incentives that could motivate OC's to participate in a crowdsourcing model like the ones we are proposing in this thesis.

4.1 Theoretical principles of Path Planning Algorithms

If we explore the concept of heuristics, as Kahneman [31] technically defined it is "a simple procedure that helps find adequate, though often imperfect, answers to difficult questions."

Herbert A. Simon *et al.* also wrote about heuristics and perhaps were some of the first people to attempt to program a computer to simulate decision-making. Simon *et al.* in their publication *Heuristic Problem Solving: The Next Advance In Operations Research* [32], referred to heuristics as "being used by both humans and intelligent machines: Digital computers can perform certain heuristic problem-solving tasks for which no algorithms are available… In doing so, they use processes that are closely parallel to human problem-solving processes."

Oxford English Dictionary defines heuristics as: "Computing of a program: that solves problems or makes decisions by trial and error or through empirically derived rules (often used to obtain approximations when more formal or exact methods are too slow or complex)." [33]

On the other hand, Gevaers *et al.* [30] defines Last Mile Logistics as "the final leg in a business-to-consumer (B2C) delivery service and indicate that the efficiency of the last mile can be determined by five fundamental factors: the level of consumer service, security and delivery type, the geographical area, the degree of market penetration and density, the vehicle fleet and technology employed, and the environmental impact.

Therefore, as we have described, the last mile in transportation faces challenges in delivery optimization, but these complexities can be addressed computationally, and it is at that intersection where heuristics are fundamental to finding satisfactory solutions that are aligned with practical scenarios. A classic example of optimization that represents heuristics in transportation and delivery is the Travelling Salesman Problem (TSP) [34][35]. In this problem, a certain number of given cities must be visited, but the main goal of this problem is to find the shortest possible path, and although there is no single method that is the most acceptable or optimal, combining clustering algorithms approaches with other tools and restrictions such as type of vehicle, speed limits, average traffic light's duration, distance, occasional courier compensation and cargo weight or volume, to approximate to the scenario we want to simulate and achieve by find the optimal strategy. In principle, there are several heuristic algorithms that are employed to optimize VRP's and path planning such as Bellman Ford, [37], A* (A star) [38] or Dijkstra's [39], although there is not a unique method that guarantees to provide a global optimal result for every VRP, they aim to feasible solutions for small scale scenarios.

In principle, these VRP algorithms are based on nodes (circle shape), edges (line) and a weight (value) that represents the distance between the nodes. The weight is oftentimes called the distance or cost, and it is the result of the variables impacting the start and unvisited nodes

a) Road network map	b) Nodes and edges network on map

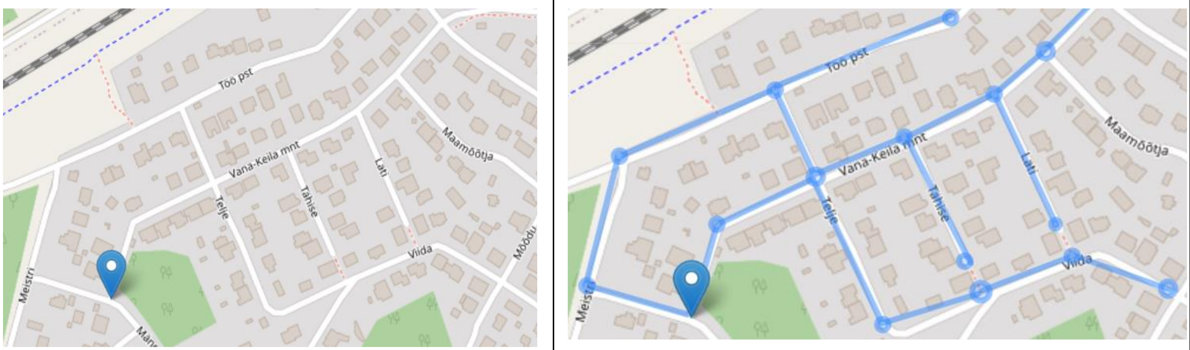


Figure 10. Map from OpenStreetMap representing a) starting point in a road network, and b) starting point in a graph/edges network.

The goal behind the heuristic VRP algorithms through the graph's theory basis, is to reduce the weight given between the nodes and edges network. We present a simplified example in Figure 11, to describe the algorithm mathematically.

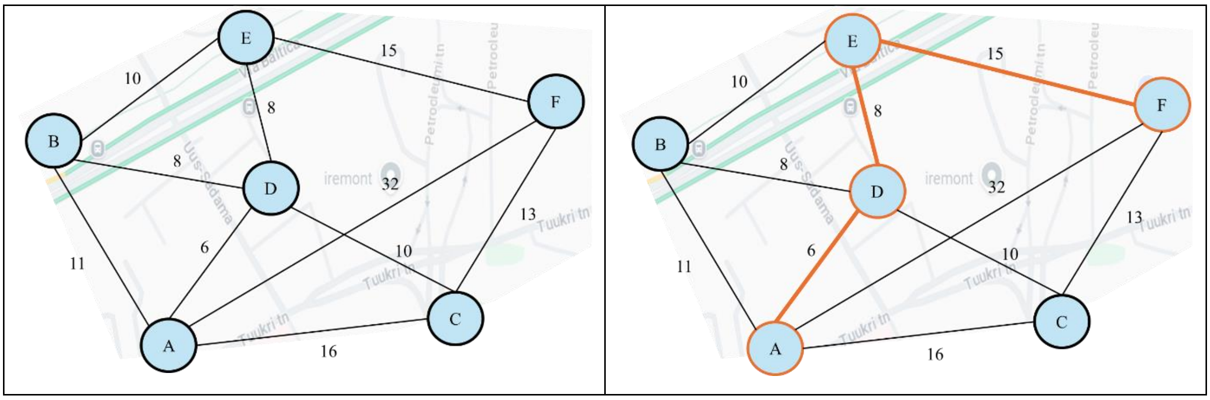


Figure 11. Simplified example of the shortest path approach for a VRP

Graph ***G*** is defined to be a set ***V*** (vertices) elements called also nodes, and a set ***E*** of directed arc segments called edges, and each edge E has a cost associated with it.

Case statement: Given a weighted graph $G = (V, E)$ where: $V = \{A, B, C, D, E, F\}$ represents the set of nodes (locations).

E is the set of edges representing connections between nodes, with associated weights representing costs.

A is the starting node (departure depot).

F is the destination node (occasional courier destination).

The objective is to find the optimal shortest path from **nodes A** to **F**, allowing the occasional courier for visits to intermediate nodes $\{B, C, D, E\}$ for parcel deliveries.

Variables:

$\boldsymbol{d(u, v)}$ is the distance or cost (weight) of the edge between nodes u and v, where $(u, v) \in E$.

P is a sequence of nodes representing a path from A to F, $P = \{V_0, V_1, \dots, V_n\}$, where $V_0 = A$ and $V_n = F$.

$\boldsymbol{w(P)}$ is the total weight (sum of the trajectory or total cost of the trajectory) of path P, calculated as:

$w(P) = \sum d(V_i, V_{i+1})$ for $i = 0$ to $n - 1$, where the objective is to minimize $w(P)$.

Constraints:

P must start at node A and end at node F.

P may include any subset of the intermediate nodes $\{B, C, D, E\}$.

In this case, we are implementing the Dijkstra's algorithm to evaluate all possible routes, and determine the optimal ones:

Route 1: A-D-E-F ($w(P_1) = 29$)

Route 2: A-C-F ($w(P_2) = 29$)

Route 3: A-D-C-F ($w(P_3) = 29$)

Route 4: A-B-E-F ($w(P_4) = 36$)

As a result, the three routes have the same $w(P_n) = 29$, nevertheless the optimal paths for the occasional courier (OC) could rely on **a)** covering the greatest number of nodes before reaching the destination F; for that scenario either Route 1 or Route 3 fulfills the criteria. Another condition could be **b)** taking the path where there are one or more nodes with delivery priorities. For instance, if the delivery at node C is a priority, then the assigned path should be Route 3. Assuming that there are at least two OC's available at the same time span and heading to a similar destination F, then we could: **c)** assign Route 1 and Route 2, to deliver the parcels in $\{D, E\}$ and $\{C\}$ respectively and deliver at $\{B\}$ with own courier fleet. In Figure 11, we are not yet explicitly considering the compensation for the occasional couriers (OC's), nor the weight and size of the parcels they can transport. These factors, which could influence the optimization process, are simplified in the graph example. Nevertheless, through this simplified approach, the typical route problem is depicted, and three possible optimal solutions (a, b, or c) were provided by the method.

4.2 Unsupervised Machine Learning Algorithms

Unsupervised learning is an essential type of machine learning algorithm. Its focus is on categorizing or grouping datasets that are not classified under any specific category. Its approach is innovative because these models, without human intervention, can find patterns or relationships between the data. These algorithms are classified into three types: clustering, association rule learning, and dimensionality reduction. [41] [42]. In the case of crowdsourcing, machine learning algorithms for data classification are a way to optimize and segregate the starting and destination location for parcel collections and deliveries, the occasional courier's ability to carry out the transportation, the distance to be covered, and the incentive payment, based on certain similarity patterns. [40]. Given the methods and applications in the field of unsupervised learning in ML, we chose K-Means and DBSCAN as the most viable algorithms for this thesis. Other authors broadly classify the unsupervised clustering algorithms into at least nine different categories [57]. For our thesis we are focusing on the partition-based and density-based categories, which are the ones in which K-Means and DBSCAN fall respectively. K-Means as a partition-based algorithm offers an advantage on computational time to iterate over the data, this means that it is efficient with computer resources and large datasets. It has some limitations such as being sensitive to outliers and the need of methods to estimate its testing optimal parameters. Nevertheless, it is helpful to find quantitative and qualitative insights of data that might have similarities that are difficult to observe and group. [57]

On the other hand, DBSCAN was selected for our thesis because it is sensitive to data density based on two parameters as we explained in further sections: the neighborhood radius (eps) and the minimum number of points (MinPts). DBSCAN requires a moderate level of time and resources for iterate and it is useful for clustering data with arbitrary shapes, which is significant for our testing, for example, it could be the case that around a pickup location, many drop-off points might be densely located in a specific area or direction, not necessarily surrounding the pickup point, but still are connected and DBSCAN is sensible to identify and cluster those cases. However, we also considered some of its disadvantages, such as its sensitivity to the initial parameters settings and the computational resources needed for processing. [57]

In general, not only both are computationally efficient for data partitioning or clustering, but both can discover patterns related to crowdsourcing[43][45][53]. Even though the data is diverse and unstructured, the algorithms can identify the patterns. K-means, also known as the weighted kernel K-means [45], is a centroid-based clustering algorithm. Its function is to group the data into a set number of K groups, where each K-centroid receives a different set of data based on its proximity to the analyzed feature. For example, if the data were parcels, it would group them based on weight, dimensions, location, or delivery distance. It's important to note that the K-value for this algorithm impacts the accuracy of the results, which is why it is recommended to initialize the centroid until the clustering doesn’t change much after each iteration. [40]. However, it should be considered that this type of algorithm can be affected by the distribution of the data if there are many outliers around a K-centroid. The selection of an appropriate K can be determined, but it is possible that the same model with this algorithm may not be able to predict the variability of data generated daily in the parcel delivery operation [43].

On the other hand, the DBSCAN algorithm, whose name comes from the words Density-Based Spatial Clustering of Applications with Noise, [44][46] is based on grouping data according to the density or frequency of each data point. It has the characteristic of identifying clusters that may follow irregular patterns, making it effective for handling outliers. Unlike K-Means, it does not require a K-value for centroids. For the purposes of this thesis, we determined that this algorithm works better for classifying parcels based on their location for PUDO points. The key idea is that for each point of a cluster the neighborhood of a given radius must contain at least a minimum number of three points according to the MinPts rule[1], nevertheless, the more points, the more precise is the model to identify clusters or noise.

The choice of K-means and DBSCAN for this thesis is due to the ability of both algorithms to have low susceptibility to classifying randomly distributed and unstructured data, as occurs in a parcel delivery process. Both approaches allow for adaptive classification given the behavior of the operation and the variability in parcel delivery, where the quantity and type of parcels, PUDO locations, and the assignment of occasional couriers are constantly changing.

Next, we show the flowcharts of both algorithms. Our goal is to visually explain how they work, how we use them, their main differences and what results to expect. Figure 12 makes it simple to understand K-Means and DBSCAN as both were used in this thesis.

[1] A. M. Sefidian, *How to determine epsilon and MinPts parameters of DBSCAN clustering*, Sefidian.com (Dec 18th, 2022). Retrieved on Apr. 22nd 2025 from: https://www.sefidian.com/2022/12/18/how-to-determine-epsilon-and-minpts-parameters-of-dbscan-clustering/

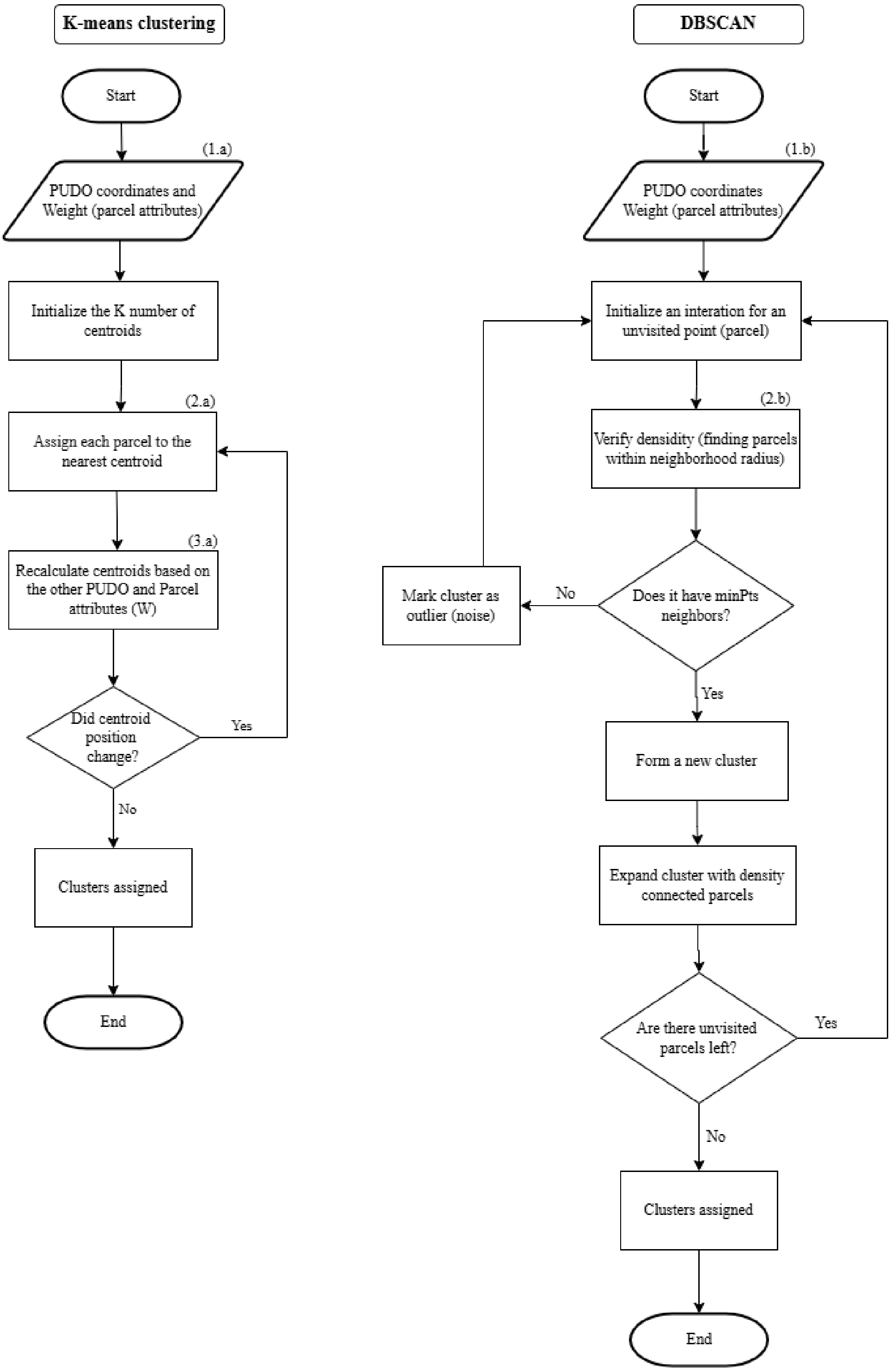


Figure 12. Flowchart of the clustering algorithms for the proposed scenarios

To describe the key stages of both algorithms and their use in this thesis, we will now mathematically detail how they are implemented.

For the **K-Means (PUDO and parcels weight)** algorithm:

Where X_i, is spatial coordinate that includes the parcel weights such as:

$X_i = (coordinates, W_i)$ (1.a)

This allows us to obtain the magnitude of each edge taking into account both, the geographic proximity and the weight of W_i.

$d(p, C_{k1}) = \sqrt{(X_{i1} - C_{k1})^2 + (X_{i2} - C_{k2})^2 + (W_i - W_k)^2}$ (2.a) is the Euclidean distance.

$C_k = \frac{1}{|C_k|} \sum_{X_i \epsilon C_k} X_i = \left(\frac{1}{n}\sum X, \frac{1}{n}\sum Y, \frac{1}{n}\sum W\right)$ (3.a) represents the centroids reallocation after clustering points in a random cluster K.

For the **DBSCAN (PUDO and parcels weight)** algorithm:

Where X_i, is spatial coordinate that includes the parcel weights, as

$X_i = (coordinates, W_i)$ (1.b)

Like K-means, this starting point allows us to obtain the magnitude of each edge by considering both the geographic proximity and the weight of W_i from the starting or collection point, and X_j represents the delivery points with their weight W_i.

$N_\varepsilon(X_i) = \{X_j \epsilon x \mid d(X_i, X_j) \leq \epsilon\}$ (2.b) evaluations of the nearest neighborhood threshold for every X_i. If the distance between X_i y X_j is less than or equal to the threshold radius ϵ, then X_j belongs to the neighborhood X_i.

Where $d(X_i, X_j) = \sqrt{(X_{i,lat} - X_{j,lat})^2 + (X_{i,lon} - X_{j,lon})^2 + (W_i - W_j)^2}$

If there are enough neighboring points ($\geq minPts$), then X_i becomes the center point of a new cluster of possible valid deliveries from that exit point.

The value of W, which is crucial for the parcel allocation in the clustering process, will be further detailed in Chapter 4.3 Parcel Segregation. The formulation and justification of W will be introduced in the upcoming section.

4.3 Parcel Segregation

In this section we present a viable proposal to respond to RQ2 and complement the solution that will be addressed later for RQ3. In this chapter our focus will be on defining the rules that allow us to classify the parcels, not only according to their intrinsic characteristics but also with respect to their location around the pickup or drop off point. We begin by presenting the general properties of a parcel and how these should be integrated with other variables of a pick-up and drop-off process.

As we have indicated, the DML can always be incrementally innovated, but at the same time, as indicated by Gevaers *et al.* [30] there are five factors that define delivery efficiency: level of consumer service, security & type of delivery, geographical area & market density, fleet & technology, and environment. Of these factors, we consider the first four to be the determinants as a starting framework for exploring the parcel segregation method in this thesis: level consumer service, since it considers lead time; security & type of delivery since it considers the factors associated with the handling, transportation and trust of the consumer; geographical area and market density, because the parcel segregation process must take into account the pool of deliveries that will be assigned to the occasional couriers in order to maintain a constant participation and improvement of the LMD; and finally, fleet & technology, where the type of vehicle of the occasional courier must be considered in such a way that it does not compromise costs and overall efficiency when traveling the route, satisfying the combination of mentioned variables. Since the segregation parcel method aims to optimize the LMD under the crowdsourcing model, the delivery can't rely entirely on occasional couriers, therefore we present a standard function that will allow us to simulate and evaluate different scenarios. We describe below the function we have proposed for this thesis:

The function W (1) represents a compound value derived from multiple parcel-related attributes. Rather than being a simple physical weight, W refers to a delivery score that integrates characteristics and constraints combined through a formula designed to reflect the logistical effort required for delivery. In other words, the lower the value of W, the more convenient we believe it is to assign a parcel to the crowdsourcing model.

$$W = \alpha \cdot D + \beta \cdot H + \gamma \cdot G + \delta \cdot C - \theta \cdot Inc \qquad (1)$$

Where D is the level of service to the consumer represented as distance between PUDO points. H is the handling complexity, which considers the physical weight, volume, and fragility factor of the parcel. G is the geographical area demand density, where a low density would be equivalent to a higher complexity to assign an occasional courier. C is the cost factor per fleet kilometer. Inc is the incentive for the occasional courier, and lastly $\alpha, \beta, \gamma, \delta, \theta$ are adjustable weights to control the relative importance of each factor in the simulation of the proposed scenarios.

Function D (2) is determined by the Euclidean distance, between the PUDO points, very similar to how the K-means algorithm (2.a) or DBSCAN (2.b) uses it, but for D it is calculated only on the coordinates of origin and destination.

$$D = \sqrt{(Lat_o - Lat_d)^2 + (Lon_o - Lon_d)^2} \qquad (2)$$

Where Lat_o is Latitude of origin, Lat_d is Latitude of destination, Lon_o is longitude of origin and Lon_d is longitude destination.

Function H (3) is determined by three variables that describe the characteristics of each parcel. The weight given in kilograms (kg), the volume in cubic meters (m^3) and fragility weighed with a subjective scale between 1 (not very fragile) and 5 (very fragile).

$$H = \lambda_1 \cdot ParcelWeight + \lambda_2 \cdot ParcelVolume + \lambda_3 \cdot ParcelFragility \qquad (3)$$

Function G (4) is determined by the density ratio determined by the number of deliveries per square kilometer, this factor allows the generation of areas with higher demand.

$$G = \frac{1}{\#deliveries\ per\ km^2} \qquad (4)$$

Function C (5) is determined by the costs associated with the fleet and the type of vehicle. These are weighted according to the impact of each constant. Where $\frac{FleetCost}{km}$ is a given constant of the combination of the TCO [49][50] and η_1 is the weighting factor, and $FleetTypeScore$ represents parametric values associated with different modes of transport, such as business fleet, or those used by an occasional courier such as bicycle, own vehicle and walking, for example.

$$C = \eta_1 \cdot \frac{FleetCost}{km} + FleetTypeScore \qquad (5)$$

Function Inc (6) is determined by the adjustable coefficients ρ_i based on importance rank, multiplied by the $Volume$ and $Distance$ where the parcel must be delivered by the OC. The incentive is not a cost, but a variable of benefit for the OC as effort perceived, therefore it is subtracted from W.

$$Inc = \rho_1 \cdot ParcelVolume + \rho_2 \cdot DeliveryDistance \qquad (6)$$

In general, the approach allows us to randomly model different types of W, thus obtaining a series of combinations to evaluate and answer under which conditions parcels are assigned to an occasional courier and which should maintain their current flow in the delivery process.

Two very important aspects regarding segregation are: a) not to compromise the overall operation of a parcel delivery system, and b) to maintain the participation of the OC’s, otherwise the availability of crowdsourcing will decrease.

By incorporating W for parcel segregation into the clustering algorithms, the model accounts not only for spatial proximity but also for delivery complexity results. This dual consideration ensures a more realistic simulation of parcel assignment.

4.4 Conditions for Scenario Simulation

In this section we present the conditions for scenario simulation to test the thesis of the outcomes of parcel delivery by implementing crowdsourcing. In Chapter 3.2 we described the theoretical process separately and how each of them is going to be set. Nevertheless, model and obtain valuable information, we will assume some numbers for coefficients, based on desk research [51] [52] and expert interviews[2].

General assumptions:

In order to standardize the models, we will not set a specific time window to make collections and deliveries. In addition, it is also not considered that the parcels will be returned or rejected by the client. The location where the experiment will be conducted is in Tartu County, since it is one of the areas that has elements that satisfy the methodology and all three scenarios (DD, PM and PO). Given the limitations of computational processing and generation of data combinations, the simulation will be done on a random sample of twenty-nine PUDO's which is equivalent to 50% of the total in Tartu. It is also assumed that the deliveries will be B2C, and, moreover, the parcel will be equivalent to one unit, and not to a package with different consolidated parcels. Finally, no seasonal demand is established, weather or traffic, and as for specific assumptions such as vehicle type or other coefficients, clarification will be made in each scenario.

The Table 5 Provides an overview of Figure 13 Where three scenarios’ settings are established and will be simulated in chapter 5.

[2] Presentation of applied research - online meeting with Omniva representatives, April 1, 2025. Meeting minute.

Table 5. Scenario setting overview

	Description
Scenario 1	A scenario whereby the LMD is improved by allocating parcels from the PM to an OC that will deliver within a small radius range of ~6 km. The execution of the LMD is assumed by using a cargo bicycle vehicle.
Scenario 2	A scenario whereby the LMD is improved by allocating parcels from the PO to an OC that will deliver within a radius range of ~6 km. The execution of the LMD is assumed by using a cargo bike vehicle.
Scenario 3	A scenario whereby the LMD is improved by allocating parcels from the DD to an OC that will deliver within a radius range higher between ~6 km to 12 km. The execution of the LMD is assumed by using 4WD vehicle.

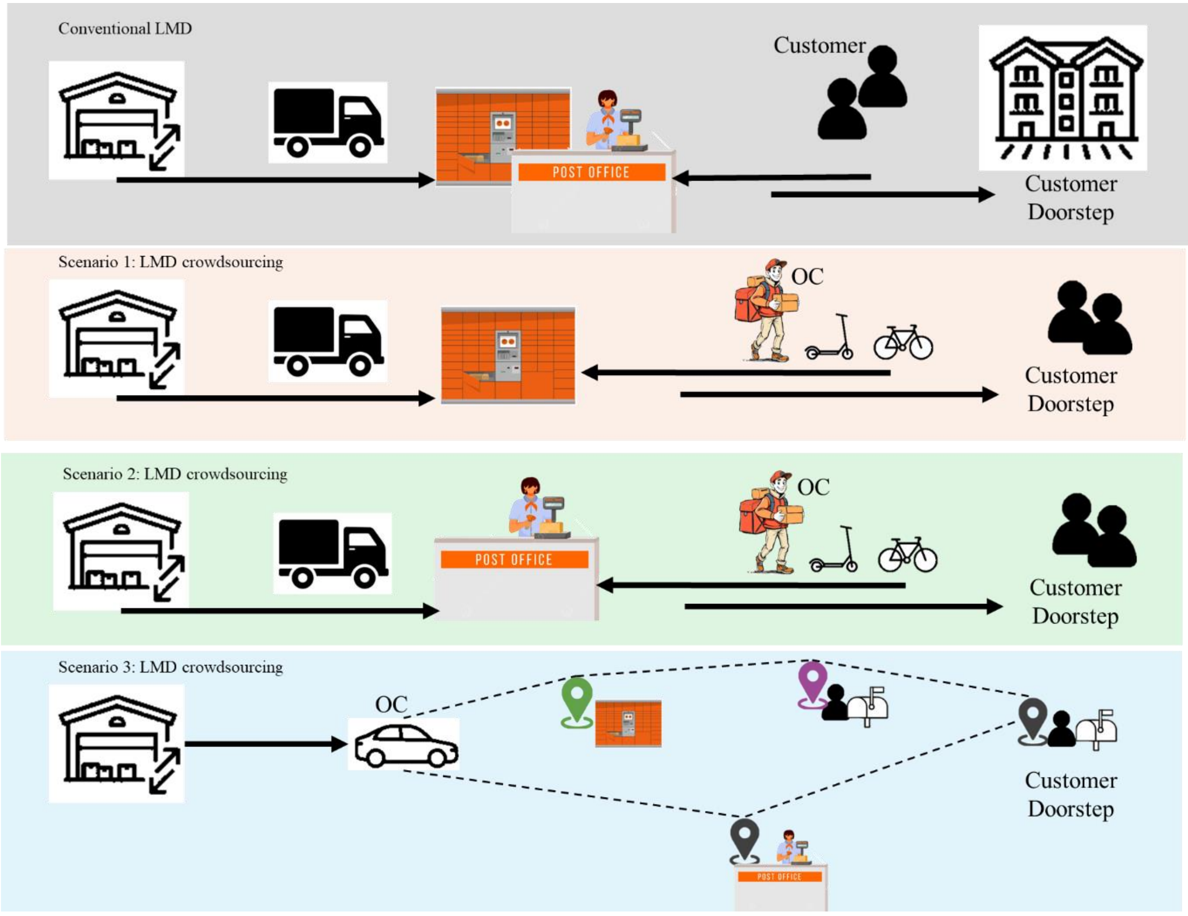


Figure 13. Comparison overview of crowdsourcing scenarios

5. Model Development

For the development of the methodology, construction and computational representation of the parcel segregation model, of we followed the decomposition strategy [36], where the general problem was broken into small situational cases or subproblems, and each of them were solved either in parallel or sequentially. Based on this strategy, we divided the general problem into four subproblems, the first decomposition is to delimit the geographical region, in this case we selected Tartu County in Estonia, since it is a populated area and there are PM's, PO's and a DD which are the pickup points that satisfy the proposed scenarios for this thesis. The objective in this stage is to geolocate the PUDO points within the county. The second decomposition is to randomly generate delivery points that represent customers to be served. The third decomposition is to use ML algorithms to cluster the customers and parcels with the PUDO points in a two-dimensional plane. The fourth and last decomposition is to find the set of optimal parcel allocation and routes sequence based on the resultant W between nodes, and since the PUDO points are based in geographic coordinates, we transformed them by an approximate value into cartesian coordinates[3] to measure distances in kilometers.

5.1 Data processing: Delimiting the geographical region

The data processing was done using Python programming language, version 3.8.18. The choice is based on several advantages for the development of this thesis. First, Python has a high-level syntax, which means it is clear and easy to read, almost like every day human language. This helps to understand it better and makes it easier to reproduce. Second, Python has a wide list of libraries that are specialized in data processing, ML algorithms, operations research, math operations, graph creation, and map visualization. This makes it very useful for our thesis. And finally, its portability and compatibility guarantee that the code can be transferred or run on different operating systems[4]. [58]

Starting with the data processing, the dataset we used as a starting point contains the geographic coordinates of the pickup points located in the Baltic countries. However, as shown in Code block 1, we only processed the points that belong to Estonia, according to the scope of this thesis and we obtained the following results: a total of 466 combined PMs and POs, and 3 DDs.

[3] Howard Veregin, "How Big is a Degree?" University of Wisconsin–Madison, State Cartographer's Office, January 21, 2022, https://www.sco.wisc.edu/2022/01/21/how-big-is-a-degree (retrieved April 6, 2025).
[4] The computer code used to analyze the data and make simulations is available in a private online repository. Access can be requested by email to: aceitunomanfredo@gmail.com

Code block 1. PUDO's dataset filtering

```
# Creating a dataframe for Pick up & Drop off (PUDO) points:
# Parcel Machines & Postal Offices within the territory of Estonia
pudo_points = df[(df['COUNTRY'] == 'EE') & (df['PUDO_POINT'].isin(['Parcel
Machine', 'Postal Office']))]

# (Terminals) Departure depots dataframe
depot_points = df[(df['COUNTRY'] == 'EE') &
(df['PUDO_POINT'].isin(['Terminal']))]
```

Later, as shown in Code block 2, we did a segmentation of the county within the territory of Estonia that would be the focus of the simulation. The selection of Tartu County is based, as mentioned before, on two main reasons: first, because it includes the PUDO points, which represent the logistic nodes used as starting points in the parcel delivery model. Second, because it is a less congested area compared to Tallinn, which is important for parcel distribution using crowdsourcing. In Tartu, the distance between urban, peri-urban, and rural sectors is manageable, not only for running the simulation but also for applying the crowdsourcing delivery model studied in this thesis. With Code block 2, we found that Tartu County has a total of 58 PUDO points (PM, PO & DD combined).

Code block 2. Tartu County, Estonia. PUDO's segmentation

```
# Filtering PUDO points by Tartu Maaakond (Tartu County)
tartu_df = df[df['COUNTY'].str.contains('Tartu', na = False,
regex=False)][['COUNTY','LONGITUDE_COORDINATE', 'LATITUDE_COORDINATE',
'PUDO_POINT']]
```

Figure 14 shows the PUDOs in the territory of Estonia, where the POs are marked in blue, the PMs in orange, and the DDs with a black balloon icon. Figure 15 shows the PUDOs in Tartu County. In this last figure, the distribution of the pickup points is more noticeable, most of them are PMs, but there are 2 POs, and 1 DD.

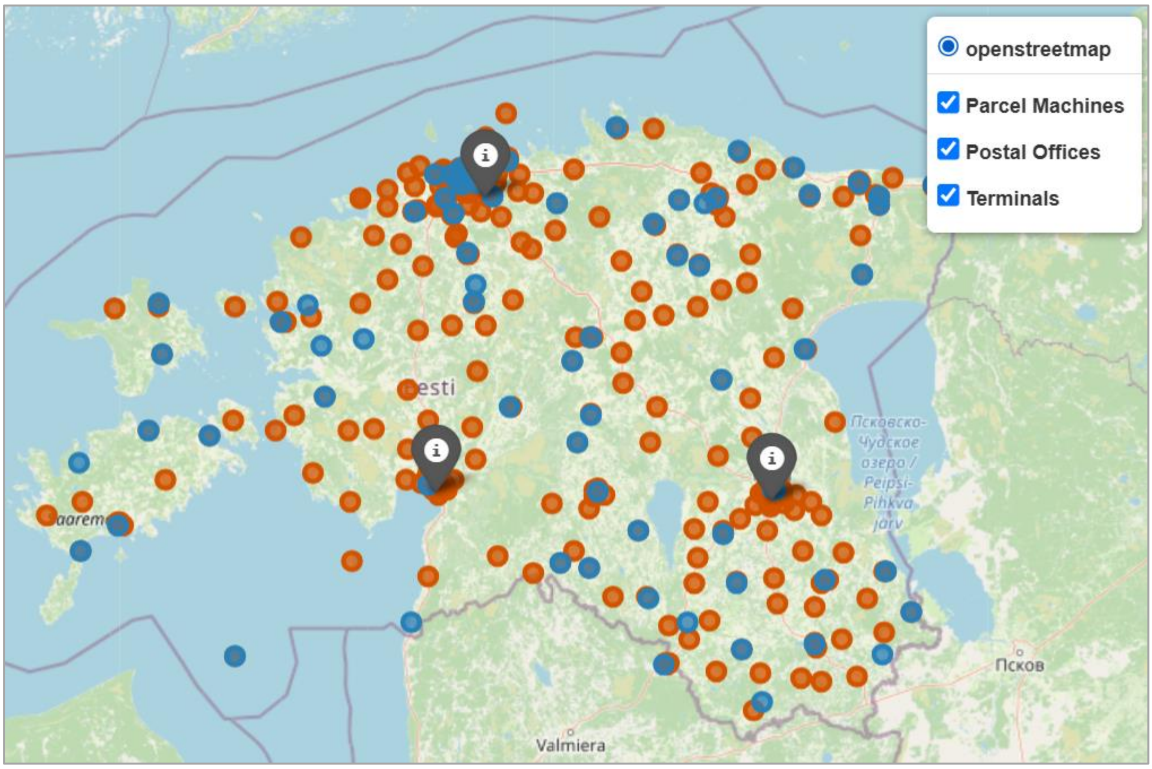


Figure 14. Estonia PUDO’s

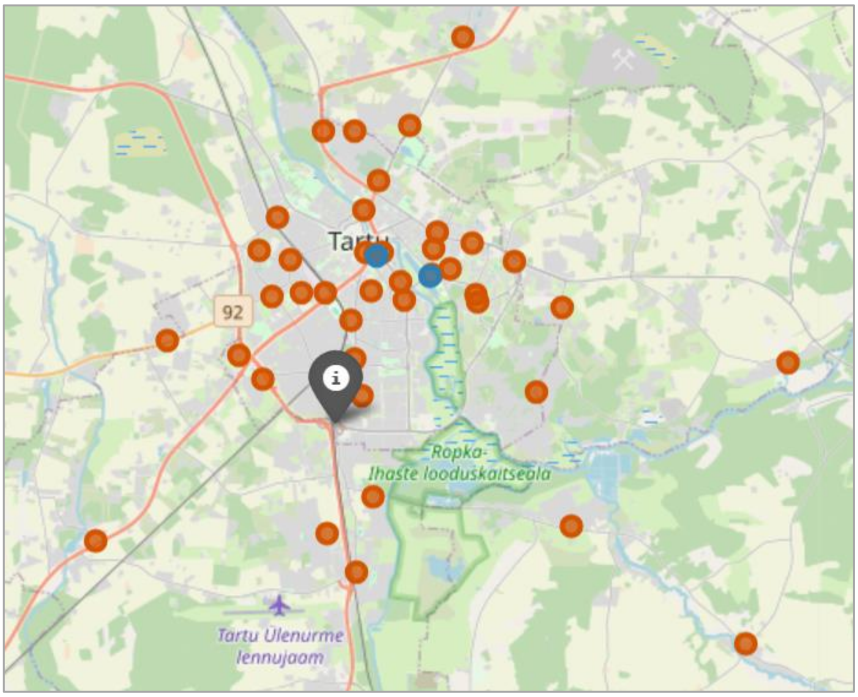


Figure 15. Tartu County PUDO’s

5.2 Randomly simulated: Delivery points and Parcel attributes

Around each pickup point, we decided to randomly generate between 4 and 30 drop-off points. This was done to simulate densely populated areas with high demand and to handle inefficiencies that delivery operations face in real-world contexts. To complement the idea behind the random range of drop-off points: if the number of parcels around a pickup point is high, it means there could be more stops for the occasional courier. However, due to some limitations like volume, distance, or load capacity (among others), this setup allows us to realistically apply a formula for parcel segmentation, as we will show later.

As shown in Table 6, a single pickup point can have several drop-offs, each with its own DROP_ID, which is a unique value that represents each customer.

Table 6. Sample of randomly generated drop off points

	PICKUP_LAT	PICKUP_LON	DROPOFF_LAT	DROPOFF_LON	PUDO_TYPE	DROP_ID
0	58.221024	26.406515	58.240000	26.487585	Postal Office	A1
1	58.221024	26.406515	58.183187	26.356458	Postal Office	A2
2	58.221024	26.406515	58.248602	26.444548	Postal Office	A3
3	58.221024	26.406515	58.203670	26.463281	Postal Office	A4
4	58.221024	26.406515	58.207512	26.370784	Postal Office	A5
5	58.221024	26.406515	58.217514	26.369030	Postal Office	A6
6	58.221024	26.406515	58.272863	26.406997	Postal Office	A7
7	58.221024	26.406515	58.181247	26.337805	Postal Office	A8
8	58.221024	26.406515	58.261044	26.409891	Postal Office	A9
9	58.221024	26.406515	58.202403	26.400966	Postal Office	A10

Code block 3 contains the set of parameters that were randomly assigned to each DROP_ID parcel. These parameters—such as rates, volume, dimensions, and weight—were based on publicly available documentation about the current parcel delivery model [51] [52] but adapted to a crowdsourcing scenario. For example, the simulated weight ranges go from 0.2 to 20 kilograms, and the volume ranges from 0.01 to 0.2 cubic meters. This way, we make sure that the random values fit a parcel that would be around 40 to 50 cm per side, but still, that doesn't always mean it's a good option for an occasional courier (OC). Other factors like the distance to the drop-off point or the incentive score for the OC are also considered.

Code block 3. Randomly generated set of parcel attributes

```
# Parcel weight between 0.2kg and 20kg
dropoff_df['Weight_kg'] = np.random.uniform(0.2, 20.0, len(dropoff_df))

# Volume between 0.01m3 and 0.2m3
dropoff_df['Volume_m3'] = np.random.uniform(0.01, 0.2, len(dropoff_df))

# Fragile parcel factor
```

```
dropoff_df['Fragility_Factor'] = np.random.uniform(1.0, 3.0,
len(dropoff_df))

# Fleet cost per km (Euros)
dropoff_df['Fleet_Cost_per_km'] = np.random.uniform(0.3, 0.7,
len(dropoff_df))

# Distance between 3km and 8km
dropoff_df['Distance_km'] = np.random.uniform(3.0, 8.0, len(dropoff_df))

# Incentive for OC (the bigger the parcel and the far, higher incentive)
dropoff_df['Incentive'] = (
    0.5 * dropoff_df['Weight_kg'] +
    0.2 * dropoff_df['Volume_m3'] * 100 +
    0.1 * dropoff_df['Distance_km']
```

Finally, in Code block 4, to get the composite score W of each parcel, we defined the weight coefficients to control the relative importance of each factor in the simulation. The weight coefficients range from 1 to 3, from low to high importance. The reason for each weighting is as follows: Alpha has a value of 1, because even if many parcels reach the maximum weight of 20 kg, this is still manageable for an occasional courier (OC). Beta, the volume factor, was given a weight of 2, as it compensates for the parcel's size. On the other hand, Gamma represents the fragility of the parcel. This factor was weighed as 1.5, since it reflects the need for balanced handling by the OC. Also, the delivery company considers an insurance fee based on the shipping and delivery process [51] [52]. Delta, which stands for cost per km, was weighted with a value of 1, to evaluate the system simulation under direct compensation for travel effort. And Epsilon, the incentive factor, was also given a weight of 1 to keep it neutral and avoid overstating the role of incentives in the simulation.

Code block 4. W score coefficients

```
# Coefficients for testing W (1 to 3 / [Low-Med-High])
alpha = 1.0 # Parcel weight factor
beta = 2.0 # Parcel Volume factor
gamma = 1.5 # Parcel fragility factor
delta = 1.0 # Cost per km factor
epsilon = 1.0 # Occasional courier incentive factor
```

5.3 Data Normalization and Application of Clustering algorithms

With simulated data already in a random distribution that matches real ratios for a parcel delivery process, it is now possible to start with the clustering algorithms. However, this requires a normalization process, which involves adjusting the scales of the drop-off points and W to a common scale. Code block 5 and Code block 6 show a sample of the input and the transformation, where the mean is 0 and the standard deviation is +/-1.

Code block 5. Unscaled data

```
# Drop off coordinates (2D) and W score (1D)
X = dropoff_df[['DROPOFF_LAT', 'DROPOFF_LON', 'W']].values

…output sample:
array([[58.24000045, 26.487585  ,  7.93798303],
       [58.18318708, 26.35645823, 14.58496963],
       [58.24860174, 26.44454784, 12.62377426],
       [58.2036697 , 26.4632811 ,  8.39617573]])
```

Code block 6. Normalized (scaled) data

```
Scaling or normalizing the data. Std = +/-1 and Mean = 0
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

…output sample:
array([[-1.62620141, -1.35578329, -0.36840413],
       [-2.30688325, -2.09185932,  1.54334815],
       [-1.52314913, -1.59737106,  0.97928507],
       [-2.0614807 , -1.4922125 , -0.23662247]])
```

$X_{norm} = \frac{X-\mu}{\sigma}$ (1) Z-score normalization is the formula behind *scaler.fit_transform(X)* function, where X is the original value of the variable, μ is the mean variable and σ is the standard deviation.[5]

5.4 Implementation of Unsupervised ML Clustering Algorithms

Next, we describe the implementation of the clustering algorithms used to analyze parcel delivery simulation through crowdsourcing with occasional couriers. The algorithms used to find meaningful patterns and segment the geographic areas, as well as the delivery characteristics, were K-Means and DBSCAN.

[5]ScikitLearn, StandardScaler, (Accessed: March 22, 2025) available at: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

The choice of these algorithms was based on their complementarity and their ability to address different aspects of the data structure, such as proximity and the formation of homogeneous groups. They also allow us to identify potential outliers or groups with irregular shapes. However, we also provide a comparison of the effects of the clusters obtained with each algorithm.

As explained in section 4.2, K-Means and DBSCAN are unsupervised clustering algorithms. K-Means, which is centroid-based, was applied to identify partitions, making it easier to segment the study area where parcels are similar in terms of drop-off points and W scores. On the other hand, DBSCAN, which is density-based, was implemented to discover clustering patterns with irregular shapes and to identify areas of high concentration. It does this without the need to predefine the number of clusters, allowing us to identify possible noise presence.

Before applying each algorithm, we carried out a parameter tuning process, using a combination of numerical and graphical methods. The detailed justification of the selected parameters for each algorithm will be presented in the following subsections, showing the process that led to their choice.

5.4.1 K-means Clustering algorithm

The K-Means algorithm requires certain parameters that can be estimated through approximation using iterative or graphical mathematical methods. These values are calculated from the normalized variables. For this thesis, we applied the method known as the Elbow graph [53][1], as shown in Figure 16.

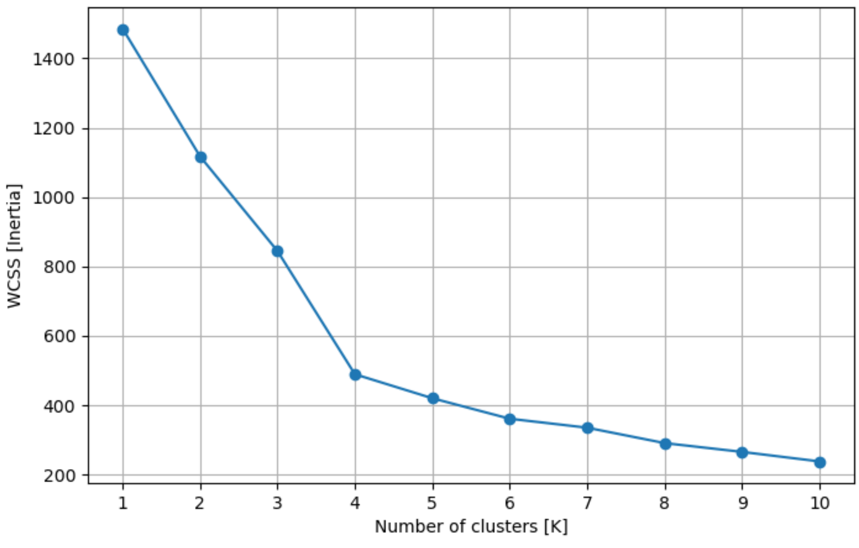


Figure 16. Elbow method for K-Means

Where WCSS stands for Within-Cluster Sum of Squares, also known as Inertia. WCSS is a metric used to evaluate how dense or compact the formed clusters are. A low WCSS value indicates that the points within the cluster are close to its centroid. On the other hand, the optimal value of K is the one where there is a downward trend in WCSS. This point of inflection is known as the "Elbow". For our simulation, the chosen value of K is 4 clusters. A value greater than 4 would lead to overfitting, as it would create noise with many groups with data specific to small patterns. But a value lower than 4 would force a generalized grouping of the data distribution.

Unlike the Elbow method, there are also numerical methods. For this thesis, we used the Davies-Bouldin index[57], as shown in Code block 7, which provides a quantitative value for a range of possible K clusters. With both approaches, we complement the choice of the optimal K value and compare the graphical method with the iterative-mathematical one.

The interpretation of the Davies-Bouldin index is that the closer index is to zero, the more compact the clusters are, meaning they have low internal dispersion. Conversely, the further the index moves from zero, the more separated the clusters are from each other, with large distances between the centroids.

Code block 7. Davies-Bouldin iterative score method

```
# Davies Boulding score method to calculate the optimal cluster number
from sklearn.metrics import davies_bouldin_score

range_n_clusters = [3, 4, 5, 6, 7]

…output:
For n_clusters = 3, the Davies-Bouldin index is: 1.00
For n_clusters = 4, the Davies-Bouldin index is: 0.76
For n_clusters = 5, the Davies-Bouldin index is: 0.95
For n_clusters = 6, the Davies-Bouldin index is: 0.95
For n_clusters = 7, the Davies-Bouldin index is: 0.96
```

The Python script in Code block 8. is ran in the background to perform the cluster algorithm iteration and calculate the optimal groups of drop off points which will be used for the parcel segregation in the route planning. As is seen in Figure 17, there are four clusters as the parameter was established.

Code block 8. Unsupervised ML K-Means algorithm method

```
# K-means cluster with K = 4 based on the previous optimal calculation
methods
kmeans = KMeans(n_clusters=4, random_state=42)
dropoff_df['KMeans_Cluster'] = kmeans.fit_predict(X_scaled)
```

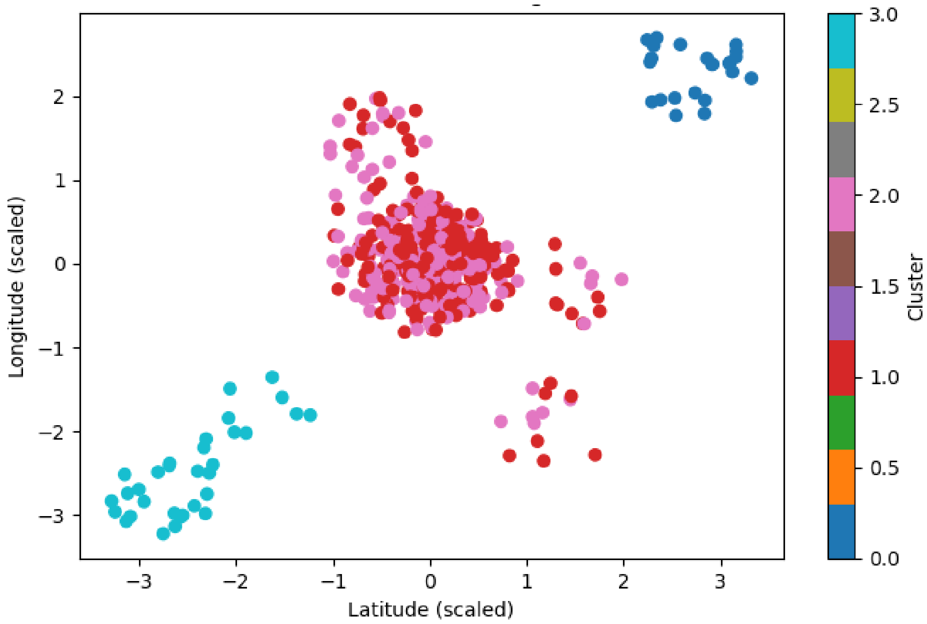


Figure 17. Drop off points K-Means clustering

5.4.2 DBSCAN Clustering algorithm

In the same way, the unsupervised algorithm DBSCAN requires the configuration of certain parameters to identify the optimal setup, according to the distribution of the data that was normalized using the function *scaler.fit_transform()*. DBSCAN doesn't require the number of clusters to be known, in other words, the clustering number is not predefined. However, it is still possible to make both graphical and numerical estimations of the parameters that define its behavior.

As shown in Figure 18, we adapted the idea of the Elbow Method to estimate an appropriate value for the epsilon (eps) parameter. To adapt the method, we first set minPts (the minimum number of points required to form a dense cluster) and related it to the distance shown on the Y-axis (eps). The idea is to find, by approximation, a point where there is a sharp change in the slope. This turning point helps us determine the value of eps.

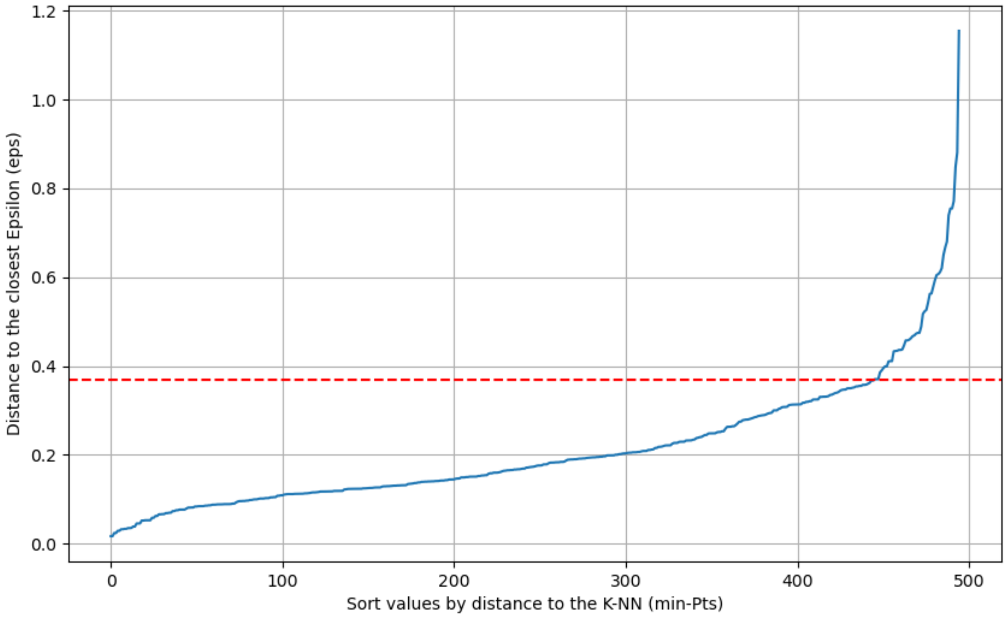


Figure 18. Elbow method adaptation for DBSCAN

By implementing iterative numerical methods, as shown in Code block 9, we can identify more precisely the result of setting different values for eps and minPts in the algorithm. This allows us to anticipate the number of clusters and the expected noise. The Silhouette Score[6] [7] [57], is interpreted as follows: higher scores indicate an optimal fitting value of epsilon, meaning that the clustering has well-categorized and well-segregated groups. This numerical method excludes the noise (labeled as -1) during the calculation process. Both graphically and numerically, we obtain complementary and more direct metrics to configure the algorithm.

Code block 9. Silhouette Score

```
# Silhouette_score numerical method to determine eps and optimal DBSCAN
algorithm parameters
from sklearn.cluster import DBSCAN
from sklearn.metrics import make_scorer, silhouette_score # DBSCAN
clustering metric
from sklearn.model_selection import GridSearchCV

…output:
Hyperparameters obtained from the systematic iterative approach:
eps=0.4, min_samples=3: Clusters=13, Noise=66, Silhouette=0.08
eps=0.4, min_samples=4: Clusters=9, Noise=82, Silhouette=0.11
eps=0.4, min_samples=5: Clusters=5, Noise=106, Silhouette=0.16
```

[6] Scikit-learn. Silhouette_score. Retrieved April 4, 2025, from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html

[7] Scikit-learn. Demo of DBSCAN clustering algorithm. Retrieved April 4, 2025, from https://scikit-learn.org/stable/auto_examples/cluster/plot_dbscan.html#sphx-glr-auto-examples-cluster-plot-dbscan-py

```
eps=0.5, min_samples=3: Clusters=13, Noise=52, Silhouette=0.09
eps=0.5, min_samples=4: Clusters=9, Noise=71, Silhouette=0.13
eps=0.5, min_samples=5: Clusters=5, Noise=92, Silhouette=0.18
eps=0.5, min_samples=3: Clusters=12, Noise=38, Silhouette=0.11
eps=0.5, min_samples=4: Clusters=10, Noise=46, Silhouette=0.14
eps=0.5, min_samples=5: Clusters=7, Noise=70, Silhouette=0.24
eps=0.6, min_samples=3: Clusters=9, Noise=32, Silhouette=0.21
eps=0.6, min_samples=4: Clusters=8, Noise=39, Silhouette=0.20
eps=0.6, min_samples=5: Clusters=7, Noise=59, Silhouette=0.15
eps=0.6, min_samples=3: Clusters=8, Noise=18, Silhouette=0.32
eps=0.6, min_samples=4: Clusters=6, Noise=30, Silhouette=0.34
eps=0.6, min_samples=5: Clusters=5, Noise=41, Silhouette=0.37
eps=0.6, min_samples=3: Clusters=8, Noise=10, Silhouette=0.33
eps=0.6, min_samples=4: Clusters=7, Noise=22, Silhouette=0.35
eps=0.6, min_samples=5: Clusters=6, Noise=34, Silhouette=0.28
eps=0.7, min_samples=3: Clusters=8, Noise=7, Silhouette=0.33
eps=0.7, min_samples=4: Clusters=7, Noise=16, Silhouette=0.35
eps=0.7, min_samples=5: Clusters=7, Noise=23, Silhouette=0.28
eps=0.8, min_samples=3: Clusters=8, Noise=6, Silhouette=0.34
>>>> eps=0.8, min_samples=4: Clusters=7, Noise=12, Silhouette=0.37 <<<<
eps=0.8, min_samples=5: Clusters=6, Noise=20, Silhouette=0.36
```

Based on the iterative results, the optimal parameters for the DBSCAN algorithm are eps = 0.8 and min_samples = 4, since this combination returns a higher Silhouette Score with a low amount of noise, which indicates that the classification will include a higher proportion of data within a meaningful number of clusters. The DBSCAN clustering graph can be seen in Figure 19.

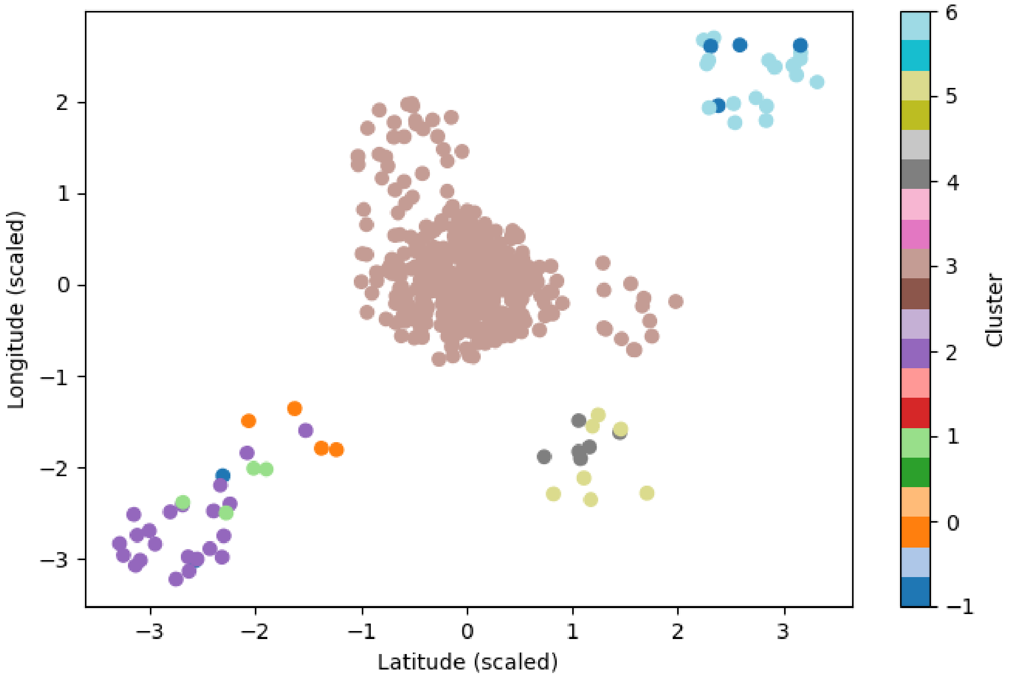


Figure 19. Drop off points DBSCAN clustering

Once the clustering process was completed, we were able to assign each drop-off point to a specific pickup point, and at the same time, compare the results of both algorithms, as shown in Table 7. K-Means grouped the data into four clusters, while DBSCAN adjusted to a more granular clustering due to the density in some areas of data distribution. Even though the parameters were adjusted using numerical and graphical methods, K-Means assigned all parcels to belong to a cluster of candidates, while DBSCAN classified six drop-off points as noise. These points can be analyzed more carefully and either manually reclassified or excluded from the segregation for OCs. On the other hand, out of the 495 parcels, only 260 were assigned to a cluster using K-Means, while 443 were grouped using DBSCAN.

Table 7. Comparative clustering results

	Model	Total Clusters	Noise Points (Percentile 15)	Avg W (crowd candidates)	Std W (crowd candidates)	Crowdsourcing Candidates	Avg W (All parcels)	Dropoff Total parcels
0	KMeans	4	0	6.761762	2.598208	260	9.21889	495
1	DBSCAN	7	6	9.094320	3.486878	443	9.21889	495

5.4.3 Methodology for Parcel segregation to Crowd couriers

The International Postal and Logistics Company suggested keeping a minimum threshold of 10% for the parcels that should be handled by occasional couriers, in order to ensure the viability and continuity of the service under this scheme. Therefore, as shown in Code block 10, we set a threshold of 15% for the simulation, which is slightly higher than the suggested value. Due to the randomness of the drop-off points, we selected parcels whose delivery radius falls within a range of 2 km to 7 km from the pickup point. However, this distance may increase as drop-off points are added along the same route for the occasional courier (OC). The fragility threshold was set to 1.5 on a normalized scale from 1 to 3, based on the assumption that the parcels in the simulation are easy to handle and carry for the OC. Regarding the weight threshold, we set it at 10 kg, considering handling, fragility, and volume, and also to include the scenario of collection and delivery from the DD, since larger or heavier parcels usually depart from the terminals. As for the volume, the maxium threshold was set at $0.08m^3$ to ensure that the parcel does not exceed the volume of a Type L, as shown in Figure 20. Finally, we applied a filter to select parcels with a pickup location at a PM, PO, or DD, to cover the different scenarios addressed in this thesis.

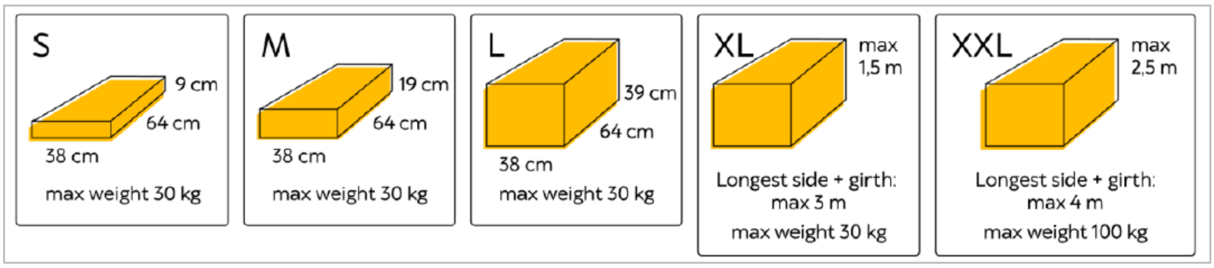


Figure 20. Parcel dimensions[8]

Code block 10. Parcel segregation: parameters for OC's

```
# Filter for W (less or equal to the 15% of the parcels)
w_threshold = dropoff_df['W'].quantile(0.15)
w_filter = dropoff_df['W'] <= w_threshold

# PUDO distance range between 2km to 7km
distance_filter = (dropoff_df['Distance_km'] >= 2) &
(dropoff_df['Distance_km'] <= 7)

# Fragility factor of the parcel
fragility_threshold = 1.5 # Mininimum subjective value for the experiment
because the parcel companies have an insurance coverage fee.
fragility_filter = dropoff_df['Fragility_Factor'] <= fragility_threshold

# Weight factor of the parcel
weight_threshold = 10 # in kg
weight_filter = dropoff_df['Weight_kg'] <= weight_threshold


# Volume of the parcel
volume_threshold = 0.08 # in m3
volume_filter = dropoff_df['Volume_m3'] <= volume_threshold

# PUDO Type (PM or PO)
pudo_pm_po_filter = dropoff_df['PUDO_TYPE'].isin(['Postal Office', 'Parcel
Machine', 'Terminal'])
```

From the total parcels distributed among the pickup points, 53 parcels can be assigned to OCs in the simulation of the shortest delivery route with an occasional courier.

[8] Omniva. (2024, November 15). *PRICE LIST OF PARCEL SERVICES FOR PRIVATE CUSTOMERS*, Parcel dimensions figure. Parcel services price list. Retrieved April 4, 2025, from https://www.omniva.ee/en/price-lists-private/

Code block 11. Parcel segregation: filtering candidates for OC's

```
# Parcel candidates segregation based on filter criteria
parcel_candidates_df = dropoff_df[w_filter & distance_filter &
                                   (volume_filter | weight_filter |
fragility_filter) & pudo_pm_po_filter]

… output:
Total candidates for crowdsourcing: 53
```

5.4.4 Simulation of scenarios

To simulate the scenarios computationally, we used the integrated development environment (IDE) Visual Studio Code and the Jupyter notebook extension v2025.3.0, since it allows programming the solution in blocks, just as we defined it according to the decomposition strategy [36]. The equipment used to run generate the data and simulation was a Lenovo computer, with an Intel Core i5-10210U CPU @ 1.60GHz 2.10 GHz and 8.00 GB of RAM. As a general rule, from the total segregated parcel candidates, we randomly selected one pickup point per proposed scenario, making sure each had at least 2 drop-off points. The following tables in this section show the simulations and their outcomes. Finally, in Chapter 6, we present the benefits and drawbacks of crowdsourcing, and we end the thesis in Chapter 7 with conclusions.

Table 8. From Parcel Machine to Drop off points simulation

City center of Tartu, Estonia

PM coordinates: Lat, Lon (58.377780, 26.724674)

Route list

PICKUP_LAT	PICKUP_LON	DROPOFF_LAT	DROPOFF_LON	PUDO_TYPE	DROP_ID
58.37778	26.724674	58.357202	26.715354	Parcel Machine	S1
58.37778	26.724674	58.376658	26.650007	Parcel Machine	S2
58.37778	26.724674	58.397117	26.696716	Parcel Machine	S11
58.37778	26.724674	58.357118	26.718556	Parcel Machine	S12
58.37778	26.724674	58.329512	26.690374	Parcel Machine	S18

```
# Constraint Programming (CP) and Routing solver engine in OR tools for shortest path

from ortools.constraint_solver import pywrapcp, routing_enums_pb2
# Euclidean distance library
from geopy.distance import geodesic

…output:
Optimal route for pickup (58.37778, 26.724674) is:
Start: Parcel Machine
Drop-off order: ['S12', 'S1', 'S18', 'S2', 'S11']
End: Parcel Machine
```

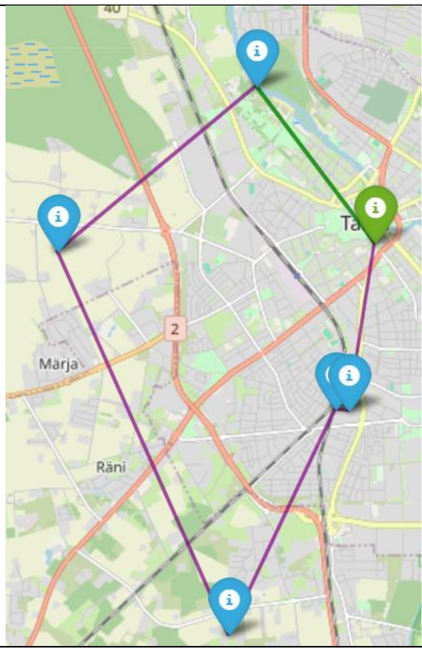


Table 9. From Postal Office to Drop off points simulation

Elva. Tartu, Estonia

PO coordinates: Lat, Lon (58.221024, 26.406515)

Route list

PICKUP_LAT	PICKUP_LON	DROPOFF_LAT	DROPOFF_LON	PUDO_TYPE	DROP_ID
58.221024	26.406515	58.207512	26.370784	Postal Office	A5
58.221024	26.406515	58.217514	26.369030	Postal Office	A6

```
# Constraint Programming (CP) and Routing solver engine in OR tools for shortest path

from ortools.constraint_solver import pywrapcp, routing_enums_pb2
# Euclidean distance library
from geopy.distance import geodesic

…output:
Optimal route for pickup (58.221024, 26.406515) is:
Start: Postal Office
Drop-off order: ['A6', 'A5']
End: Postal Office
```

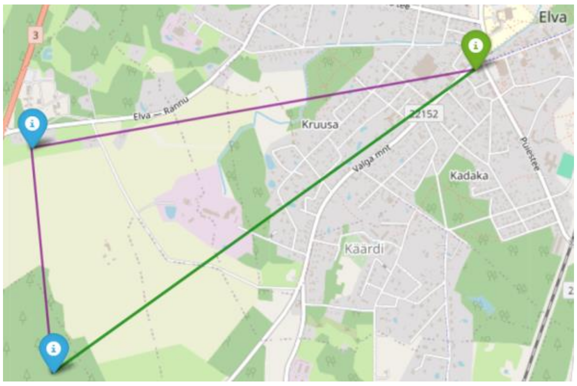


Table 10. From Departure Depot to Drop off points simulation

Võru 167 Eesti Post Terminal. Tartu, Estonia

DD coordinates: Lat, Lon (58.344923, 26.71343)

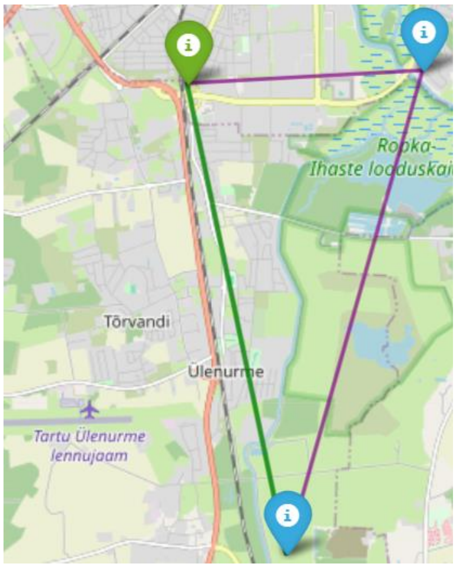
Route list

PICKUP_LAT	PICKUP_LON	DROPOFF_LAT	DROPOFF_LON	PUDO_TYPE	DROP_ID
58.344923	26.71343	58.292635	26.734545	Terminal	B4
58.344923	26.71343	58.346407	26.763538	Terminal	B9

```
# Constraint Programming (CP) and
Routing solver engine in OR tools for
shortest path

from ortools.constraint_solver import
pywrapcp, routing_enums_pb2
# Euclidean distance library
from geopy.distance import geodesic

…output:
Optimal route for pickup (58.344923,
26.71343) is:
Start: Terminal
Drop-off order: ['B9', 'B4']
End: Terminal
```



Each Table, 8, 9 and 10 divide the simulation into three sections. The first section contains the data and reference of the pickup point. The second section includes the list of deliveries with the PUDO’s coordinates as well as the DROP_ID; these are not in optimal order. And finally, the third section includes the optimal sequence of each drop-off point in which the parcels should be delivered and the graphical representation on a map, where the blue markers are drop-off points, and the green ones are pickup points. The purple line shows the route the OC must follow through the different drop-off points starting from the pickup point, and the green line shows the return route back to the pickup point from the last drop-off point.

6.Benefits and Drawbacks of crowdsourced parcel delivery

Although the crowdsourcing model can solve some problems in LMD, the partial decentralization of the logistics operation for collection and delivery also brings inherent challenges not only for the OCs regarding formalizing a partnership program, maintaining security controls when including a third-party courier for package handling, conditions and performance ranking, incentives, or fair assignments based on vehicle, distance, or load, but also requiring a reevaluation of customer success policies, staffing size, rate adjustments, service level agreements, IT infrastructure setup, and the reconfiguration of operational planning for different areas where the service is provided, to name just a few of the aspects to consider.[54][55]

Depending on the region, type of vehicle, delivery windows, and other dynamics that a partially crowdsourcing-dependent parcel delivery model might have, Table 11 separates the advantages and disadvantages in the operation and management of crowdsourcing parcel delivery.

Table 11. Advantages and disadvantages classification

Advantages	**Disadvantages**
Scalability by providing the immediate capacity for outsourcing OC's based on demand without hiring permanent staff.	Variability in service standards, less consistency with delivery protocols and procedures.
Time window flexibility by having OC's in various times or seasons, including weekends or peak hours.	OC's legal liabilities, labor rights and loyalty for the flexible nature of crowdsourcing delivery work.
Staff cost reduction and lower fleet costs, since the OC's use their own vehicle.	Insurance coverage, regular background check, or inappropriate vehicle conditions for handling the package.
Geographical coverage by facilitating direct deliveries in a broader network.	Integration of a technological platform with existing logistics system, coordinating tasks for independent OC's and maintaining communication for an effective customer service, resolving issues and complaints.
Potential for innovation and agility with approaches such as alliances with other businesses where customers get goods from (supermarkets, electronic or accessories shops, and miscellaneous stores), and expanding the delivery business.	Customers trust, brand image for the logistics company and partnerships, risks of theft, loss or damage of goods and parcels, higher pricing.

7. Conclusion

In this thesis, we explored the crowdsourcing model for a parcel delivery process, generating random data with industry-standard ranges in logistics to simulate the delivery operation with an occasional courier. We also created a model as one of the parameters for the unsupervised ML clustering algorithms, with the focus of segregating parcels for OCs in LMD. Four research questions that cover the main question are answered and presented below:

RQ1: What is the framework of reference for postal delivery operations? We found that the logistics operation follows a reference framework based on maintaining a flow of collection and delivery. It is important to note that none of the studies we referenced in our thesis had a substantial difference in terms of the stages of a logistics process. However, certain researchers focused more on the mode of transportation or costs. Nevertheless, we can state that the reference framework begins with collection from various origin points, followed by classification and processing of shipments to logistics centers. Then, the shipments are directed via transportation that follows a logical order of routes, ultimately delivering the shipment to the recipient.

RQ2: What condition(s) can be formulated for identifying parcel candidates for crowdsourcing? Each logistics service provider could set conditions to identify candidate parcels for crowdsourcing. However, as we found in Chapter 4, there are factors that define delivery efficiency: level of consumer service, security & type of delivery, geographical area & market density, fleet & technology, and environment. Based on these factors, we proposed and formulated the W score, which is a function that weighs intrinsic and extrinsic variables of the parcel. This allowed us to use it as a measurement standard and as a parameter for parcel matching of possible candidates for crowdsourcing. Its interpretation follows this rule: the lower the value of W, the more convenient we believe it is to assign a parcel to the crowdsourcing model.

RQ3: How can unsupervised Machine Learning algorithms be implemented to cluster and allocate parcels for crowdsourcing deliver in simulated scenarios? We conducted an analysis of the clustering algorithms and applied two different types, K-Means and DBSCAN, with the purpose of evaluating their effectiveness in classifying parcels based on their delivery destinations. The results indicated that the use of both algorithms allows for segmentation by W score and geography. However, their application was limited to static scenarios, where the shipping demand and delivery time-window remained constant. We also found the relevance of optimization parameters before running the algorithm on the data set. To validate both algorithms, we present in Chapter 5, the clustering figures, numerical methods, and a comparative table with the results. Regarding the methodology, we demonstrated that both unsupervised Machine Learning clustering algorithms can be applied under predefined assumptions about the study variables, and although K-Means forces clustering while DBSCAN identifies possible outliers, both contribute on a small scale and with considerable success in the context of optimizing parcel allocation for crowdsourced occasional couriers.

RQ4. What are the benefits and drawbacks in the context of the crowdsourcing modality for a postal and logistics company? Based on the general findings described in Chapter 6, we confirm that the implementation of crowdsourcing as an alternative to gradually improve LMD in a segment of parcel delivery service is positive in terms of agility, feasibility, and as an opportunity to innovate logistics operations. However, there are also drawbacks that could weigh more for different companies and markets, which could lead to criteria such as partial implementation, as demonstrated in the simulation of this thesis. In this way, the service does not rely solely on occasional couriers, but the participation of both demand and supply remains active, allowing the logistics company to adapt the availability of crowdsourcing according to the evolving needs.

7.1 Future Work

For future research and potential improvements, the recommendation is to implement a simplified model like the one we have presented, but adaptable to variables such as operation windows, seasonality, or delivery time calculations based on a denser layer of nodes to optimize the detail of route patterns. Although computational resources for processing mathematical operations tend to be limited, and our thesis relied on geospatial and operations research tools, both of which are open-source, we believe it would be interesting to incorporate a map graph layer and multiple occasional couriers for a specific region and execute a more realistic simulation of demand and delivery with crowdsourcing.

7.2 Omniva's addendum on the thesis research

Tallin, May 2025

Addendum: Industry Validation of the thesis "Multi criteria decision process model for postal service in combination with crowd delivery."

Industry Validation of Thesis research by **Manfredo Estuardo Aceituno Pérez.** Addendum submitted by: **Aleksei Vahrušev**, Group Product Manager, Omniva Logistics Product Team

As the industry stakeholder who provided the original problem statement and practical data for this thesis, I would like to formally express my appreciation for the research conducted by **Manfredo Estuardo Aceituno Pérez** under the supervision of **Uku Tulev** and **Eduard Ševtšenko**.

Manfredo's thesis explores the integration of crowdsourced delivery (CSD) with existing postal fleet operations. This is not only a relevant and timely topic but also directly addresses one of the most pressing strategic challenges we face at Omniva: scaling our fleet dynamically in response to volume spikes, particularly during the December–January peak season. Currently, our in-house fleet is provisioned for this peak, but for the remaining 10 months, we experience significantly lower volume — roughly 1/3 — resulting in underutilization and elevated cost per stop.

The thesis lays a clear foundation using DBSCAN and K-Means clustering algorithms to simulate parcel assignment and route clustering scenarios, which help infer when crowdsourcing is viable. The ML implementations are basic but effective — easy to follow, and pragmatic — and they demonstrate that even simple clustering techniques can enable dynamic segmentation of delivery loads. This validates our own internal experiments and strengthens the case for operationalizing such methods.

In parallel to this academic work, our product team conducted live experiments using the open-source VROOM optimization API (https://github.com/VROOM-Project/vroom), simulating various courier profiles and cost models. We set up a vehicle type to mimic crowd-sourced drivers (cost: €3/km, no fixed cost) versus internal couriers (cost: €0.30/km + €100/day fixed cost). The optimization algorithm frequently routed lower-volume deliveries to the crowd driver when location and cost conditions made it the optimal choice — supporting the same logic and feasibility conclusions Manfredo presented.

We fully agree with the thesis conclusion that crowd-based courier sourcing is a viable and cost-effective strategy when integrated intelligently. Furthermore, Manfredo's simulation models serve as an excellent baseline to iterate upon. Our next steps include extending these concepts into real-time, production-grade systems — integrating dynamic vehicle capacity, SLAs, time windows, and multi-objective cost functions — to route tasks based on thresholds like cost per stop.

Ultimately, Manfredo's thesis has proven both theoretically relevant and practically viable, and we commend his effort to bridge academic modeling with operational logistics constraints. We see clear potential to build on this work and pursue live pilots in 2025.

We also express interest in further developing practical experiments in collaboration with Manfredo Estuardo Aceituno Pérez at Omniva and jointly preparing a submission for the EIS Applied Research Grant Programme (https://eis.ee/en/grants/programme-for-applied-research/). Building on Manfredo's thesis and the opportunity to expand it into a full dynamic logistics model, we envision creating an automatic crowdsourcing engine driven by a bottom-up cost estimation layer — assigning cost factors to stops based on multilayered operational conditions. This will help Omniva achieve its financial goal of reducing operational costs and maintaining competitiveness against international logistics players in the Baltics. Furthermore, it opens the opportunity to explore an asset-light SaaS model to deliver services in regions without a physical Omniva workforce — allowing us to license our logistics optimization system and help other postal companies transition to modern, cost-efficient logistics infrastructure.

Aleksei Vahrušev
Group Product Manager – Baltic Logistics Product Team

omniva

DELIVERING
HAPPINESS

List of References

[1.] De Maio, A., Musmanno, R., & Vocaturo, F. (2023). Balancing risks and monetary savings when the crowd is involved in pickups and deliveries. In S. Terzi, K. Madani, O. Gusikhin, & H. Panetto (Eds.), *Innovative intelligent industrial production and logistics. IN4PL 2023. Communications in computer and information science* (Vol. 1886). Springer, Cham. https://doi.org/10.1007/978-3-031-49339-3_7

[2.] Punel, A., & Stathopoulos, A. (2017). Modeling the acceptability of crowdsourced goods deliveries: Role of context and experience effects. *Transportation Research Part E: Logistics and Transportation Review*, *105*, 18–38. https://doi.org/10.1016/j.tre.2017.06.007

[3.] Blohm, I., Zogaj, S., Bretschneider, U., & Leimeister, J. M. (2018). How to manage crowdsourcing platforms effectively? California Management Review, 60(2), 1 122-149. https://doi.org/10.1177/0008125617738255

[4.] Huang, K., & Ardiansyah, M. N. (2019). A decision model for last-mile delivery planning with crowdsourcing integration. *Computers & Industrial Engineering*, *135*, 898-912. https://doi.org/10.1016/j.cie.2019.06.059

[5.] Kitchenham, B. A., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (Technical Report EBSE 2007-001). Keele University.

[6.] Okoli C, Schabram K. 2010. A guide to conducting a systematic literature review of information systems research. SSRN Electronic Journal 10 (26)

[7.] Pourrahmani, E., & Jaller, M. (2021). Crowdshipping in last mile deliveries: Operational challenges and research opportunities. Sustainable Production and Consumption, 27, 101063. https://doi.org/10.1016/j.seps.2021.101063

[8.] Mehmann, J., Frehe, V., & Teuteberg, F. (2015). Crowd Logistics − A Literature Review and Maturity Model. In W. Kersten, T. Blecker, & C. M. Ringle (Eds.), *Innovations and Strategies for Logistics and Supply Chains: Technologies, Business Models and Risk Management. Proceedings of the Hamburg International Conference of Logistics (HICL)* (Vol. 20, pp. 117-145). epubli GmbH.

[9.] Howe, J. (2006, June). The rise of crowdsourcing. Wired, 14(6). Retrieved March 3, 2025 from http://www.wired.com/wired/archive/14.06/crowds_pr.html

[10.] Kafle, N., Zou, B., & Lin, J. (2017). Design and modeling of a crowdsource-enabled system for urban parcel relay and delivery. *Transportation Research Part B: Methodological*, [1] *96*, 186-205. https://doi.org/10.1016/j.trb.2016.12.022

[11.] Alnaggar, A., Gzara, F., & Bookbinder, J. H. (2021). Crowdsourced delivery: A review of platforms and academic literature. *Omega*, *98*, 102139. https://doi.org/10.1016/j.omega.2019.102139

[12.] Walmart Inc. (2025). Frequently Asked Questions, getting started. Spark Driver. Retrieved March 16, 2025, from https://www.sparkdriverapp.com/en_us/faqs.html

[13.] Bose, N. (2018, September 5). Walmart trials grocery delivery to rival Amazon Flex. *Reuters*. Retrieved March 17, 2025, from https://www.reuters.com/article/us-walmart-delivery/walmart-trials-grocery-delivery-to-rival-amazon-flex-idUSKCN1LL1Q1/

[14.] LogisticsMatter. (n.d.). DHL crowd sources deliveries in Stockholm with MyWays. Retrieved March 15, 2025, from https://logisticsmatter.com/dhl-crowd-sources-deliveries-in-stockholm-with-myways/

[15.] Simmons, R., & Discover content team. (2023, June 7). 4 ways to improve your last-mile delivery performance. DHL. Retrieved March 17, 2025, from https://www.dhl.com/discover/en-global/logistics-advice/import-export-advice/last-mile-solutions

[16.] Clevon. (2023). *DHL Express Estonia and Clevon Deliver the Future of Last Mile Delivery Today* [Case study]. Clevon. Retrieved March 16, 2025, from https://clevon.com/wp-content/uploads/2023/09/Clevon-DHL-case-study.pdf

[17.] Paloheimo, H., Lettenmeier, M., & Waris, H. (2016). Transport reduction by crowdsourced deliveries – a library case in Finland. *Journal of Cleaner Production*, *132*, 240-251. https://doi.org/10.1016/j.jclepro.2015.04.103

[18.] European Commission. (2014, October 1). *PiggyBaggy – the Uber of parcel delivery*. CORDIS. Retrieved March 17, 2025, from https://cordis.europa.eu/project/id/652511/reporting

[19.] Rohmer, S., & Gendron, B. (2020, April). *A Guide to Parcel Lockers in Last Mile Distribution - Highlighting Challenges and Opportunities from an OR Perspective*. CIRRELT. Retrieved from https://www.cirrelt.ca/documentstravail/cirrelt-2020-11.pdf

[20.] Skiver, Ryan L and Godfrey, Michael, Crowdserving: A Last Mile Delivery Method for Brick-and-Mortar Retailers (2017). Global Journal of Business Research, v. 11 (2) p. 67-77, Available at SSRN: https://ssrn.com/abstract=3043353

[21.] J. Li, M. Noto and Y. Zhang, "Optimization Model of Takeout-Delivery Process Based on Concept of Crowdsourcing," *2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS)*, Hachijo Island, Japan, 2020, pp. 1-6, doi: 10.1109/SCISISIS50064.2020.9322744.

[22.] Sawik, B. Optimizing Last-Mile Delivery: A Multi-Criteria Approach with Automated Smart Lockers, Capillary Distribution and Crowdshipping. Logistics 2024, 8, 52. https://doi.org/10.3390/logistics8020052

[23.] Wicaksono, S., Lin, X., & Tavasszy, L. A. (2022). Market potential of bicycle crowdshipping: A two-sided acceptance analysis. Research in Transportation Business & Management, 45 (Part A), 100660. https://doi.org/10.1016/j.rtbm.2021.100660

[24.] dos Santos, A. G., Viana, A., & Pedroso, J. P. (2022). 2-echelon last mile delivery with lockers and occasional couriers. Transportation Research Part E: Logistics and Transportation Review, 162, 102714. https://doi.org/10.1016/j.tre.2022.102714

[25.] Väät, T. (2024, 18. märts). Veebruarikuu pakiautomaatide mahtude monitooring. Eesti E-kaubanduse Liit. Retrieved on March 3, 2025 from https://www.e-kaubanduseliit.ee/uudised/veebruarikuu-pakiautomaatide-mahtude-monitooring

[26.] Väät, T. (2025, 17. veebruar). Jaanuarikuu pakiautomaatide mahtude monitooring. Eesti E-kaubanduse Liit. Retrieved on March 3, 2025 from https://www.e-kaubanduseliit.ee/uudised/jaanuarikuu-pakiautomaatide-mahtude-monitooring-2

[27.] Bugaj, M. (2024, March 4). Tembi Delivery Index. Tembi blog. https://www.tembi.io/blog/tembi-delivery-index

[28.] Veskimeister, A. (2024, May). Delivery Process - Tembi Delivery Index - April 2024 [Graphical report]. Retrieved on March 3, 2025, from: LinkedIn. https://www.linkedin.com/posts/parcellockercentral_tembi-delivery-index-april-2024-activity-7196728685704425474-KMwQ/

[29.] Veskimeister, A. (2024, December). *Tembi Delivery Index - Parcel Lockers & Home Delivery, October 2024* [Graphical report]. Retrieved on March 3, 2025, from: LinkedIn. https://www.linkedin.com/posts/veskimeister_i-asked-chat-gpr-to-analyse-the-tembi-delivery-activity-7257383459948556288-lh6b/

[30.] Gevaers, R., Van de Voorde, E., & Vanelslander, T. (2014). Cost modelling and simulation of last-mile characteristics in an innovative B2C supply chain environment with implications on urban areas and cities. *Procedia - Social and Behavioral Sciences, 125*, 398-411. https://doi.org/10.1016/j.sbspro.2014.01.1483

[31.] Kahneman, D. (2011). Answering an easier question. In Thinking, fast and slow. Farrar, Straus and Giroux. Retrieved on March 20, 2025, from Internet Archive: https://ia800603.us.archive.org/10/items/DanielKahnemanThinkingFastAndSlow/Daniel%20Kahneman-Thinking%2C%20Fast%20and%20Slow%20%20pdf

[32.] Simon, H. A., & Newell, A. (1958). Heuristic problem solving: The next advance in operations research. *Operations Research* 6(1): 1-10. Retrieved on March 20, 2025, from Carnegie Mellon University: https://iiif.library.cmu.edu/file/Simon_box00064_fld04874_bdl0001_doc0001/Simon_box00064_fld04874_bdl0001_doc0001.pdf

[33.] Oxford University Press. (n.d.). Heuristic, adj., 2. In *Oxford English dictionary*. Retrieved March 21, 2025, from https://doi.org/10.1093/OED/7584413402

[34.] Menger, K. (1930). Das Botenproblem. En Ergebnisse eines mathematischen Kolloquiums (Vol. 2, pp. 11-12)

[35.] Flood, M. M. (1956). The Traveling-Salesman Problem. *Operations Research, 4*(1), 61–75. http://www.jstor.org/stable/167517

[36.] Boyd, S., Xiao, L., & Mutapcic, A. (2003, October 1). *Notes on decomposition methods*. Notes for EE392o, Stanford University, Autumn, 2003. Retrieved March 22, 2025, from https://web.stanford.edu/class/ee392o/decomposition.pdf

[37.] R. Bellman, "On a routing problem," Quarterly of applied mathematics, vol. 16, no. 1, pp. 87–90, 1958. Retrieved on March 23, 2025, from https://www.ams.org/journals/qam/1958-16-01/S0033-569X-1958-0102435-2/S0033-569X-1958-0102435-2.pdf

[38.] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," IEEE transactions on Systems Science and Cybernetics, vol. 4, no. 2, pp. 100–107, 1968. Retrieved on March 23, 2025, from https://ieeexplore.ieee.org/document/4082128

[39.] E. W. Dijkstra et al., "A note on two problems in connexion with graphs," Numerische mathematik, vol. 1, no. 1, pp. 269–271, 1959. Retrieved on March 23, 2025, from https://link.springer.com/article/10.1007/BF01386390

[40.] Ma, H., He, Y., Huang, M., Wen, Y., Cheng, Y., & Jin, Y. (2019). Application of K-means Clustering Algorithms in Optimizing Logistics Distribution Routes. In *The 2019 6th International Conference on Systems and Informatics (ICSAI 2019)*. IEEE Xplore. Retrieved February 28, 2025

[41.] Pykes, K. (2024, January 12). Introduction to unsupervised learning. *DataCamp*. Retrieved March 25, 2025, from https://www.datacamp.com/blog/introduction-to-unsupervised-learning

[42.] Geeks For Geeks. (2025, January 15). What is Unsupervised Learning? *GeeksforGeeks*. Retrieved March 25, 2025, from https://www.geeksforgeeks.org/unsupervised-learning/

[43.] Ramírez-Villamil, A., Montoya-Torres, J. R., Jaegler, A., Cuevas-Torres, J. M., Cortés-Murcia, D. L., & Guerrero, W. J. (2022). Integrating Clustering Methodologies and Routing Optimization Algorithms for Last-Mile Parcel Delivery. In *Computational Logistics: 13th International Conference, ICCL 2022, Barcelona, Spain, September 21–23, 2022, Proceedings* [1] (pp. 275–287). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-16579-5_19

[44.] Scikit Learn Developers. (n.d.). Demo of DBSCAN clustering algorithm. *Scikit-learn*. Retrieved March 28, 2025, from https://scikit-learn.org/stable/auto_examples/cluster/plot_dbscan.html

[45.] Rodriguez MZ, Comin CH, Casanova D, Bruno OM, Amancio DR, Costa LdF, et al. (2019) Clustering algorithms: A comparative approach. PLoS ONE 14(1): e0210236. https://doi.org/10.1371/journal.pone.0210236

[46.] Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *KDD-96 Proceedings*. AAAI Press. Retrieved March 28, 2025, from https://file.biolab.si/papers/1996-DBSCAN-KDD.pdf

[47.] Yi, J., Yan, H., Wang, H., Yuan, J., & Li, Y. (2024). Learning to estimate package delivery time in mixed imbalanced delivery and pickup logistics services. Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '24). https://doi.org/10.1145/3678717.3691266

[48.] Zhang, M., & Cheah, L. (2023). Prioritizing Outlier Parcels for Public Transport-Based Crowdshipping in Urban Logistics. *Transportation Research Record*, *2678*(3), 601-612. https://doi.org/10.1177/03611981231182429 (Original work published 2024)

[49.] Ernst & Young. (2015, April 8). *Own or lease: Are you making the right choice for your truck fleet?* SlideShare. https://es.slideshare.net/slideshow/ernst-young-total-cost-of-ownership/46773346

[50.] Robinson, J. (2024, November). *How to calculate the total cost of ownership for your fleet*. Fleetio. https://www.fleetio.com/blog/calculating-total-cost-of-ownership-for-fleet

[51.] Omniva. Price list of business customer parcel services from 10.04.2025. Omniva Price Lists. Retrieved April 4, 2025, from https://www.omniva.ee/en/business/price-lists/

[52.] SmartPosti. (n.d.). Parcel prices: Parcel locker price list for private customers. Retrieved April 4, 2025, from https://www.smartposti.ee/en/sending/parcel-prices#domestic-parcel

[53.] M. Haonan, Y. C. He, M. Huang, Y. Wen, Y. Cheng and Y. Jin, "Application of K-means Clustering Algorithms in Optimizing Logistics Distribution Routes," *2019 6th International Conference on Systems and Informatics (ICSAI)*, Shanghai, China, 2019, pp. 1466-1470, doi: 10.1109/ICSAI48974.2019.9010223. Retrieved February 28, 2025

[54.] Boysen, N., Fedtke, S., & Schwerdfeger, S. (2021). Last-mile delivery concepts: a survey from an operational research perspective. *OR Spectrum*, [1] *43*(1), 1–58. https://doi.org/10.1007/s00291-020-00607-8

[55.] Buldeo Rai, H., Cetinkaya, A., Verlinde, S., & Macharis, C. (2020). How are consumers using collection points? Evidence from Brussels. *Transportation Research Procedia*, [1] *46*, 53–60. https://doi.org/10.1016/j.trpro.2020.03.163

[56.] van Duin, J. H. R., Wiegmans, B. W., van Arem, B., & van Amstel, Y. (2020). From home delivery to parcel lockers: a case study in Amsterdam. *Transportation Research Procedia*, *46*, 37–44. [1] https://doi.org/10.1016/j.trpro.2020.03.161

[57.] Xu, D., Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Ann. Data. Sci.* **2**, 165–193 (2015). https://doi.org/10.1007/s40745-015-0040-1

[58.] Aceituno, M. E. (2025). *CSD_parcel_allocation_Thesis_modelDevelopment* (Version 1.0lib.80425). https://github.com/viable-data-analyst/

Appendix

1. License

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Manfredo Estuardo Aceituno Pérez ,

(author´s name)

1. grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the digital archives of the University of Tartu until the expiry of the term of copyright, my thesis

   Multi criteria decision process model for postal service in combination with crowd delivery. ,

   (title of thesis)

   supervised by Uku Tulev and Eduard Ševtšenko ;

   (supervisor´s name)

2. grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright;

3. am aware of the fact that the author retains the rights specified in points 1 and 2;

4. confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Manfredo Estuardo Aceituno Pérez

06/05/2025