When Trash Brings Trash

Modeling the Wicked Problem of naastplaatsing in Rotterdam

When Trash Brings Trash

MSc Thesis

Author: Claudio Prosperini
First Supervisor: Dr.ir. Sander Van Cranenburgh
Second Supervisor: Dr.ir. Petra Heijnen
External Supervisor: Daan Van Den Elzen
Programme: Engineering and Policy Analysis
University: Technische Universiteit Delft
Date: March 30th 2026

Executive Summary

This thesis investigates whether the phenomenon of naastplaatsing, the improper disposal of municipal solid waste next to municipal containers, can be modeled to understand its root causes and to improve the operational response of the Departmente of Waste Management (Stadsbeheer) of the Municipality of Rotterdam. Combining stakeholder mapping, multivariate data analysis (MVDA), machine learning (ML), and heuristic routing, the study provides a comprehensive analysis of one of Rotterdam's most persistent urban challenges.

Figure 1: Naastplaatsing in Rotterdam (Gemeente Rotterdam, J. Boute, 2023)

Rotterdam faces a rapidly growing challenge with naastplaatsing. Residents increasingly describe waste accumulations around containers as a major nuisance, contributing to unhealthy residues, degraded streetscapes, and declining confidence in municipal services. NietRnaast (nRn), the city's dedicated response team, handles hundreds of thousands of incidents annually; these are all recorded as administrative actions detailing the type of waste detected (if any), the amount of waste, the time of the intervention, and sometimes a picture of the container(s) with waste. Still, structural pressures (population density, container malfunctions) and behavioral dynamics ("waste attracts waste") continue to reinforce the phenomenon.

A theoretical framework of naastplaatsing is constructed drawing on expert interviews, field observations, and prior research. The framework identifies two core mechanisms that

shape naastplaatsing: operational triggers, such as container malfunctions, overflow, and service delays, and behavioral feedback loops, most notably the “trash brings trash” effect, which means that visible waste further normalizes improper disposal. A stakeholder mapping exercise reveals a structural misalignment in Stadsbeheer’s decision arena. The actors who generate the problem (perpetrators) hold high behavioral influence but low interest in solving it. Conversely, high-interest actors like residents, container adopters, and NietRnaast workers exhibit limited power. Strategic decision-makers (the alderman, Stadsbeheer managers) must balance a delicate trade-off between operational efficiency and behavioral deterrence, making naastplaatsing a wicked problem.

The analysis of the current NietRnaast setup reveals that certain functionalities of the tools utilized by drivers and pickers in their daily operations have detrimental effects on their efficiency, as well as negative implications on data quality. NietRnaast drivers have too many categories from which they can choose from when describing their administrative actions, leading them to often default to the more general categories out of convenience. Furthermore, their access to picture mode is tied to a report of actual naastplaatsing, meaning that they are not able to photograph containers that are clean. This has caused drivers to misreport the situation of some containers.

Figure 2: Naastplaatsing in Rotterdam (Gemeente Rotterdam, J. Boute, 2023)

A statistical analysis of factors at the neighborhood level validates some mechanisms of the theoretical framework. Operational impediments, especially container malfunctions, show a consistent and positive correlation with naastplaatsing rates. Demographic composition (notably the share of non-EU residents) also correlates positively. Other hypothesized drivers (prosperity, structural load, social cohesion) appear weaker or highly context-dependent. MVDA helps distinguish robust predictors from weaker ones; still, its linear structure limits its ability to capture behavioral feedback loops and completely explain the theoretical framework.

Tree-based ML models are trained to predict the naastplaatsing risk at the container cluster level, achieving high predictive accuracy ($R^2 \approx 0.93$–0.95). SHAP-based interpretability reveals that naastplaatsing is strongly self-reinforcing: clusters that have experienced it recently are far more likely to experience it again. Operational conditions (malfunctions, waste type, fill dynamics) and temporal patterns also shape risk. These insights extend the MVDA, showing that naastplaatsing emerges from the interaction of historical behavior, operational pressures, temporal patterns, and container characteristics.

A heuristic routing study shows that the collection process can be made substantially more efficient. Alternative routes are calculated with 5 different algorithms (heuristics), which in some cases use the prediction outputs of the ML models to inform their choice in the selection of the cluster sequence. Their evaluation is carried out over performance metrics that reflect the trade-off identified by the strategic stakeholders between collection efficiency and quality. Purely distance-based heuristics reduce travel time but ignore behavioral relevance; conversely, purely risk-based heuristics prioritize problematic containers at the cost of longer routes. Hybrid heuristics balance both metrics, with the ratio heuristic consistently outperforms actual NietRnaast routes on both distance and collection penalty; this demonstrates that efficiency and service quality need not be in conflict.

The research identifies three key areas for policy intervention:

- **Routing that Works**
  ML predictions and routing tools, co-designed with NietRnaast drivers, should be integrated in the departments workflow to increase the efficiency of the collection of naastplaatsing. Optimized routes can help with reducing costs, and ML-derived risk scores can reduce the amount of container clusters to visit.

- **Behavior that Shifts**
  A selectively fast and selectively slow strategy can address hotspots quickly in order to reduce the multiplying effect of naastplaatsing, while simultaneously reducing convenience for repeat offenders in low-risk, low-visibility areas. Targeted Handhaving presence enhances deterrence.

- **Mind the Gap**
  A structural disconnect exists between strategic planning and operational reality. Improving feedback loops, including drivers in tool design, and embedding collaborative workflows are prerequisites for the successful addressing of a wicked problem like naastplaatsing.

While Stadsbeheer cannot directly resolve the socio-economic drivers behind naastplaatsing, it can shape the environmental conditions that determine whether those drivers manifest. By combining stakeholder insight, MVDA, interpretable ML, and routing optimization, this thesis shows that naastplaatsing is modelable, predictable, and operationally actionable. The challenge ahead is the organizational integration of these insights into a proactive and inclusive waste management strategy.

Acknowledgements

Writing this thesis has been a challenging and rewarding journey. I learned what it means to conduct research, developing a project from beginning to end in (mostly) uncharted waters. This has been an invaluable opportunity for personal growth, and a great send-off for my academic career.

First and foremost, I would like to thank all the members of my graduation committee. I am extremely grateful to Sander, my first supervisor and chair, for coordinating this committee, providing critical, relevant, and timely feedback, and keeping me on the right track throughout this whole process. The advice you gave me in July, to wait for the Municipality to embark on my thesis, was the best you could have given me. I am also greatly thankful to Petra, my second supervisor, for all of her contributions to this thesis, especially the routing methodologies; we got there in the end, even with the Convex Insertion heuristic. Last but not least, I am deeply grateful to Daan, my external supervisor. Your stakeholder perspective on this project has been invaluable, and your active involvement has given me plenty of motivation to move forward. I am glad we will keep working together in the near future; hopefully, there are great things to come.

A special thanks goes to the people of Stadsbeheer that worked with and around me during these months. To anyone who helped me gather data, took part in the interviews, or played some table soccer with to blow some steam off: thank you for being part of this process. I am glad to now call you my colleagues.

I could not have undertaken this journey without the support of my family. You have put up with me at my lowest points, when I was doubtful and dejected. I am so glad I get to share this milestone with you all; I know some of you have been waiting for this day even longer than I have. To my girlfriend in particular, I want to say thank you for making it on the day of the defense.

Finally, thanks should also go to all my friends scattered between Italy and the Netherlands. If any of you ever got an earful about my thesis and naastplaatsing: I'm sorry, and thank you for bearing with me. I promise not to talk about this again for at least a little while.

Contents

Chapter 1

Introduction

Industrialization and population growth have resulted in unprecedented levels of solid waste production worldwide. According to Kaza et al. (2018), the amount of solid waste generated in municipal areas is expected to grow from 2.01 billion tons in 2016 to 3.4 billion tons by 2050. Around the world, waste generation volumes fluctuate widely between 0.11 and 4.54 kgs per capita per day. Such variation is generally correlated with income levels and urbanization rates. Traditional waste collection systems are often built on fixed schedules and manual monitoring, which can lead to undesirable outcomes: trash containers overfilling, service vehicles running unoptimized routes, and illegal dumping in underserved areas.

As cities become increasingly data-rich, digital technologies and geo-referenced information provide new opportunities for understanding and optimizing waste-related challenges. However, many urban waste problems remain insufficiently understood, in particular those driven not only by infrastructural constraints, but also by behavioral dynamics.

1.1 Naastplaatsing

Naastplaatsing is one of the most pressing challenges facing the Stadsbeheer department of the Municipality of Rotterdam. The term refers to the improper disposal of waste next to any one of the over 3000 municipal containers in the city of Rotterdam. The phenomenon corresponds to what national waste organizations classify as illegal dumping: the placement of bags, boxes, bulky items, or other household waste beside a container rather than inside it (“Bijplaatsingen”, 2023). According to the NVRD Household Waste Benchmark, 92% of (participating) Dutch municipalities are confronted with the issue of naastplaatsing to varying degrees (“Bijplaatsingen”, 2023).

Figure 1.1: Naastplaatsing in Rotterdam (Gemeente Rotterdam, J. Boute, 2023)

This concerningly growing phenomenon is a recognized issue in the city of Rotterdam, with the percentage of residents considering naastplaatsing a relevant nuisance increasing from 37% in 2014 to 56% in 2020 (Epskamp & De Vries, 2020). In 2018 NietRnaast, a joint operation between the 3 operative departments of Stadsbeheer (Reining, Inzameling, and Handhaving), was created with the objective of addressing naastplaatsing with a bottom-up approach. A number of crews canvass municipal containers and clear them of any waste that is improperly disposed of; initially this was restricted to a select area (the Delfshaven wijk), and it later spread to most neighborhoods in the city. The NietRnaast crews record data on each of the administrative actions they carry out on the municipal containers since 2021.

This effort to combat naastplaatsing imposes additional operational costs on waste disposal municipal services. Furthermore, each case of naastplaasting contributes to eroding public trust in waste systems, often discouraging proper waste usage among the rest of the population: ”waste attracts waste”, an application of the Broken Window behavioral theory (“Het veelkoppige afvalmonster: waarom boetes alleen de stad niet schoon krijgen”, 2025). Local reporting illustrates the extent of resident frustration, with news outlets repeatedly documenting overflowing containers and torn bags, resulting in a rapid deterioration of public spaces. In Kralingen-Crooswijk, residents described naastplaatsing as a “domino effect”, with delays in emptying containers triggering further dumping and subsequent hygiene problems (“Rotterdammers zijn afvaloverlast beu”, 2019). Similar frustrations surface across the city, where residents frequently call for a more involved Handhaving department and more reliable container emptying by Inzameling.

Figure 1.2: Naastplaatsing in Rotterdam (Gemeente Rotterdam, P. Gooris, 2023)

Even interventions intended to reduce naastplaatsing can provoke new tensions. In 2024, Stadsbeheer removed underground containers on Josephlaan in the Oude Westen neighborhood. This was the result of persistent naastplaatsing complaints, with neighboring containers expected to absorb the waste. Residents expressed strong concerns about accessibility, communication, and the likelihood of increased dumping in surrounding streets ("Containers verdwijnen uit Josephlaan: zo gooien mensen het ernaast en komen er ratten", 2024). Situations like this demonstrate how infrastructural adjustments can impact both the operational burden and behavioral responses, sometimes with unintended effects.

Persistent naastplaatsing is also reflected in citywide statistics. In neighborhoods such as Hillesluis, Carnisse, and Middelland, surveys indicate that more than 80% of residents experience nuisance from waste next to containers ("Hillesluis ervaart afvaloverlast het ergst", 2021). Municipal counselors have raised questions about business waste misuse, insufficient paper container capacity, and the capacity of Stadsbeheer to deal with full or broken containers. Calls for the deployment of sensors for improved monitoring reflect the operational pressures associated with managing overflowing or blocked containers.

National Dutch media also portrays naastplaatsing as a growing and expensive issue. A 2023 investigation by Pointer found that large cities in the Netherlands spend millions annually cleaning up naastplaatsing, with Rotterdam's NietRnaast crews alone addressing more than 660,000 incidents in a single year ("Vuil naast containers kost steden miljoenen, maar blijft zich opstapelen", 2023). Despite these investments, cities report that naastplaatsing continues to rise. Officials in multiple municipalities describe the issue as a "persistent and difficult-to-solve problem."

In response, Rotterdam has experimented with behavioral interventions. One recent example includes video projections of rats near chronic hotspots, meant as a deterrent for resident in an attempt to draw residents' attention to the hygiene consequences of leaving bags next to containers ("Gemeente waarschuwt voor afval naast de container door beelprojectie", 2026). These actions reflect the broader principle that "waste attracts waste," and that both environmental cues and operational conditions shape resident behaviour.

1.2 Knowledge Gap

The issue of illegal dumping is tackled in a multitude of ways, and behavioral research is also beginning to utilize quantitative analysis to support its claims. However, while a variety of analysis techniques have been employed to analyze illegal dumping, there is no perfect match for the specific issue of naastplaatsing in Rotterdam. More often than not, the research is not focusing on a purely urban context; when it does, illegal dumping is not entirely comparable to naastplaatsing due to the characteristics of the phenomenon. Placing waste "next to" bins is not the same as actively looking for secluded, under-monitored areas where to dispose of inconvenient waste.

The phenomenon of naastplaatsing, is not as extensively researched as illegal dumping is. The goal of this study is to utilize the many tools enabled by the availability of data not just to predict, but also to understand why the problem of naastplaatsing is growing by identifying its key drivers. Furthermore, little work has addressed how predictive outputs can be linked to practical municipal interventions, which is among the key goals of this thesis.

1.3 Research Objectives

The objective of this research is to model the phenomenon of naastplaatsing. This thesis' research question is thus:

Can the phenomenon of naastplaatsing in Rotterdam be modeled to understand its root causes and efficientize the response to address it?

The goals of this comprehensive analysis are articulated as follows:

- Identify the most relevant stakeholders involved in the decision arena of naastplaatsing and their differing levels of involvement and power.
- Understand the factors that influence the occurrence of naastplaatsing in different areas of Rotterdam.
- Predict the likelihood of naastplaatsing happening at the level of individual containers based on their characteristics historical record of naastplaatsing instances.
- Formulate an approach to optimize the allocation of cleaning resources and improve the efficiency of the naastplaatsing collection process.

These research objectives are addressed by the following sub-questions:

1) Who are the most relevant stakeholders in the decision arena of the naastplaatsing problem, and what are their objectives?
A theoretical framework of the naastplaatsing problem is formulated with domain knowledge, which is acquired by combining multiple streams of information: relevant literature and research, informal interviews with Stadsbeheer experts, and on-field experience with NietRnaast operators. The insights from the theoretical framework are to build a unifying representation of the decision arena, the analysis of which can further inform the current understanding of the naastplaatsing phenomenon.

2) Can MVDA techniques be utilized to validate the findings and hypotheses of the naastplaatsing theoretical framework?
To evaluate the insights gained with the qualitative analysis of naastplaatsing a statistical model at the neighborhood level of it is developed. The specific goal is to understand if the predictors that have been found to be relevant through the framework are similarly impactful according to the data. Data analysis techniques like PCA and OLS-R are utilized on datasets that contain demographic and socio-economic information on the neighborhoods of Rotterdam.

3) Can ML algorithms be employed to model naastplaatsing in Rotterdam to a satisfying level of prediction accuracy and result explainability?
Heterogeneous data from multiple sources is utilized to train a variety of models, similarly to how Geng et al. (2024) compare different ML algorithms, starting from a simple baseline. Tree models, such as Decision Trees, Random Forest, and Gradient Boosting regressors, are particularly well suited for modeling naastplaatsing since they can capture the non-linear relationships characteristic of waste management processes. These models also integrate with explainability techniques like feature importance ranking and SHAP

(SHapley Additive exPlanation), which are applied to the trained ML models in order to evaluate them based on performance and interpretability.

4) Can the collection process of naastplaatsing be made more efficient?
The routes that nRn crews follow for the collection of improperly disposed waste are not dictated by any algorithm. The Municipality of Rotterdam has access to the fill levels through its WasteVision (WV) system, and the outputs of the best performing ML model could inform the routing algorithm to distribute cleaning shift in different areas of Rotterdam more efficiently. The ML model assigns to each waste container a naastplaatsing risk score, which is used to build routes based on simple, interpretable heuristics to define the sequence of visited nodes. The main goal of this step is comparing these newly defined against each other, and whenever possible with the actual routes taken by NietRnaast drivers on a given day.

1.4 Thesis outline

Following this chapter's introduction, this thesis is organized as follows. Chapter 2 consists of a literature review of the various techniques utilized to leverage data in the field of MSWM, with a focus on MVDA, ML, and routing. Chapter 3 introduces the background of the study, bringing together insights obtained from interviews, relevant academic and municipal literature, and on-field observations. This chapter also describes the datasets used for the analyses and defines the theoretical framework that supports this thesis. Chapter 4 describes the methods applied in each stage of the research; Chapter 5 presents the results derived from these methods. Chapter 6 discusses how the research objectives in light of the findings, illustrates policy recommendations, addresses the study's limitations, and proposes directions for future work. Finally, Chapter 7 concludes the thesis by summarizing its main contributions.

Chapter 2

Literature Review

This chapter aims to present a literature review on the applications of various analysis techniques in the field of Municipal Solid Waste Management (MSWM). The results of the review are separated conceptually, with the themes spanning state of the art reviews, ML studies, routing applications, and behavioral analysis. A section is dedicated to studies pertaining to the city of Rotterdam. The chapter concludes with a summary table, which will serve to compare the various sources examined and highlight the knowledge gap that this study aims to address.

2.1 Core Concepts

2.1.1 AI and IoT in Urban Waste Management

There are several reviews on the application of AI technologies within the context of Smart Cities and MSWM. These mainly focus on the application of various analysis techniques to the increasing amount of data available through the Internet of Things (IoT), and the advancements in remote sensing technology.

The study from Bohm et al. (2013) provides a complete overview of the Multivariate Data Analysis (MVDA) methods employted in MSWM. These techniques mainly focus on exploratory data analysis, factor classification and prediction of parameters. The highlighted methods are Principal Component Analysis (PCA) and Partial Least Squares Regression (PLS-R), which are respectively used to reduce the dimensionality of data matrixes with many predictors and to develop quick prediction models. The review from Xia et al. (2022) covers the application of various ML algorithms through the entire process of MSWM, from collection to final disposal, underscoring how live data can be used to increase the efficiency of the whole process. Javed et al. (2023) present an overview of XAI techniques used in smart cities, underscoring the technologies that are essential towards its development. Big Data, IoT and Computer Vision are among the most relevant to ML. Fang et al. (2023) provide a general overview of different uses of ML models in waste management, from prediction of waste generation to the detection of illegal dumpings. Fraternali et al. (2024) focus on remote sensing to detect and characterize illegal waste disposal sites: Deep Learning (DL) models for satellite images are among the applications of ML for this specific topic.

These surveys attest to the established use of MVDA techniques in waste management, as well as a rapid growth in the use of ML algorithms in recent years. Of those, the most commonly implemented models are Artificial Neural Networks (ANN), which were utilized in more than half of the articles examined by Xia et al. (2022). Convolutional Neural Networks (CNN) are also DL models that have rose in usage for the analysis and classification of images. These models are of the black-box kind, which makes them hard to interpret and explain. Explainable AI models represent an opportunity to move beyond this issue, but in the field of MSWM such research is still rare.

2.1.2 Statistical and Spatial Analysis in Waste Management

Statistical and geospatial analysis are instrumental in understanding the complex mechanisms of waste management. MVDA methods are essential for the analysis of huge data structures characterized by hidden patterns and spatial heterogeneity.

Mikhailov et al. (2007) utilize PCA to analyze landfill composition: the reduction in dimensionality of the predictor matrix allowed researchers to identify the specific areas within the landfills. Similarly, a study on illegal dumping sites in India (Mondal & Mandal, 2024) employed PCA as a factor analysis technique, in order to reduce the amount of independent variables (like population density, road density, pond density) into a few principal components that could capture a satisfactory amount of the dataset's variance (in this case, 76%). Furthermore, Mondal and Mandal (2024) utilized an Average Nearest Neighborhood approach to investigate the clustering of illegal dumping sites. This was used to determine one of the system's independent variables, dumpsite frequency.

Similarly, Liu et al. (2019) conducted a study on urban solid waste generation in 30 provinces of China, which were divided into 3 economic consumption groups based on k-means clustering analysis. After this distinction, the researchers built a Multiple Linear Regression (MLR) model, to explore the relationship between a dependent variable and multiple predictors. Syafrudin et al. (2023) make use of a OLS-R model to perform an exploratory factor analysis of the predictors in their research on illegal waste dumping. This allowed to identify altitude and population density as relevant predictors, with positive and negative correlation respectively. Variance Inflation Factor (VIF) analysis was also employed in the preliminary phases of some of these studies (Mondal & Mandal, 2024; Syafrudin et al., 2023), to identify any potential problems of collinearity between the predictors that were being considered for the analysis.

The statistical and spatial models and techniques mentioned here allow not only to predict the volume of waste generation and the occurrences of illegal dumping, but also to identify and consider socio-economic and demographic factors.These can be used to device specific waste management strategies for different regions, areas, or neighborhoods

2.1.3 Machine Learning in Waste Management

A growing body of contemporary display the integration of various ML techniques with MSWM, and in particular the phenomenon of illegal dumping.

Glanville and Chang (2015) utilize Binary Logistic Regression (BLR) to classify the variables that influence the spatial distribution of illegal dumping in Queensland, Australia. The most significant predictive factors were population density, primary use of the land, and an assortment of distances to the nearest facilities (roads, landfills).

Johnson et al. (2017) use a Gradient Boosting Regression Tree (GRBT) to predict weekly municipal solid waster generation in New York City, using historic data as the training set. The model was chosen based on its ability to outperform more traditional statistical methods (ARIMA), as well as its interpretability. In the case of this short term prediction, internal features, such as average weekly temperature, were more significant than external ones, like demographic and socio-economic data.

Thakker et al. (2020) propose an application of XAI to a black-box, DL model used to recognize varying levels of gully blockage, which are believed to cause urban floodings. The opaqueness of the CNN used in this study was addressed by including expert knowledge in the classification process through formal frameworks (ontologies and reasoners). The explainable version of the model actually saw an increase of the overall predictive performance.

Sakti et al. (2023) use satellite imagery from Sentinel-2 to identify illegal dumping sites around a river in Indonesia. A Random Forest (RF) algorithm was used to classify ground coverage between buildings, detritus, water, vegetation, and dirt in the images provided by the satellite. Du et al. (2023) analyze the road characteristics associated with illegal dumping in low demographic density areas. They employ Kernel Density Estimation (KDE) to analyze existing illegal dumping grounds and a RF algorithm to analyze areas classified as low, normal or high risk. The main drivers behind illegal dumping appear to be distance from urbanized areas, as well as natural features that reduce visibility, like bushes and trees.

Geng et al. (2024) use Positive-Unlabeled (PU) learning for predicting the illegal dumping of hazardous waste. PU learning is used to overcome a lack of labeled historic data, which is fundamental for regular supervised ML techniques. The study compares 5 ML algorithms: K-Nearest-Neighbors (KNN), RF, Multi-Layer Perceptron (MLP), Logistic Regression (LR), and Decision Tree (DT). PU-KNN and PU-RF were the models that displayed the better performance and robustness overall.

Adeleke and Jen (2025) tackle the prediction of MSW generation through a multi-stage ML framework, utilizing Principal Component Analysis (PCA), K-Means Clustering, and SHAP values as an XAI technique to increase explainability. The latter allows to rank features in order of predictive impact, highliting refuse-removal access as the most significant predictor. Humidity levels, average household income, and population density also demonstrated a positive correlation to MSW generation.

This sample of sources shows an intensive application of ML and XAI techniques to waste management and illegal dumping; however, the context is often peri-urban, rural, or industrial. Furthermore, the practice of illegal dumping researched in these studies is quite distinct from naastplatsing, which happens in densely populated areas and in close proximity to the apt containers for disposal. The use of ML models in municipal decision-making raises important concerns. Sophisticated models function as black boxes, making it difficult to understand how specific predictions are generated. These inscrutable systems may be affected by wrongly generated biases, which can have serious consequences when it comes to allocating public resources. This undermines trust in end-users, which may result in less uptake by stakeholders. The emergence of Explainable AI (XAI) presents an opportunity: by making black-box models more interpretable, their opacity can be left behind. XAI has already been recognized as a promising, growing field (Adadi & Berrada, 2018). Model explainability, however, has been associated with a trade-off of the model's performance (Arrieta et al., 2019; Gunning et al., 2019; Holzinger, 2018): the most accurate models are the least transparent. Still, this notion has been challenged in numerous studies (Crook et al., 2023; Herm et al., 2023; Rudin, 2019) , stating that the relationship between accuracy and explainability is not so linear. As stated by Leichtmann et al. (2023), in contexts where decisions made by AI can have significant impact, explainability is a desirable quality for a ML model.

2.1.4 Routing Heuristics in Waste Management

Efficient routing is a fundamental element of modern waste management. The complexity of real-world constraints, such as vehicle capacity, dynamic container fill levels, and time dynamics, make exact optimization impractical, especially for large scale problems. As a result, a wide range of heuristic algorithms have been applied to real waste management problems, from basic route construction heuristics (nearest neighbor, Clarke & Wright savings) strategies integrating Genetic Algorithms and ML techniques.

Hemmelmayr et al. (2013) explores the Periodic Vehicle Routing Problem with Intermediate Facilities (PVRP-IF) for the collection of different waste fractions at given fixed points. The trucks collecting this waste unload it at locations different from their starting depot, making these points intermediate facilities (IFs). To optimize the route position of these IFs two approaches are attempted: an exact insertion of the IF node based on the shortest detour between two fixed nodes, and a greedy procedure that only inserts an IF node when capacity is about to be exceeded. For real-world instances, the PVRP-IF algorithm achieved an average 25% reduction in routing costs compared to current manual planning.

A study (Louati et al., 2019) to develop an efficient waste collection in Sfax, Tunisia modeled real world characteristics like varying waste amounts and collection vehicle capacities. Researchers created a Python script in ArcGIS that employs a custom version of the Dijkstra algorithm for the spatial routing, with the added constraint of the current vehicle's waste fill-level to determine its next move. The new method reduced total travel distance and operational time compared to the manual routing process.

The research of Yazdani et al. (2021) addresses construction and demolition waste collection in Sydney, Australia. The proposed approach is based on a simheuristic model that combines stochastics and deterministic components to solve the problem. The routing solutions are generated by a Genetic Algorithm (GA) pipeline, based on an initial selection of random solutions to a deterministic formulation of the routing problem. The quality of GA solutions is then evaluated stochastically, by accounting for the uncertainty of travel time between locations.

The paper by Shi et al. (2020) models the collection of household waste in an urban setting as multi-depot vehicle routing problem (MDVRP); the optimization goal is to minimize the total transportation distance. Researchers use a Sector Combination Approach to generate an initial set of solutions, which are then refined using the Merge-Head and Drop Tail strategy. The abundance of starting solutions ensures that the model does not get stuck in a local optimum.

A study by Dereci and Karabekmez (2022) evaluates various heuristics to optimize municipal solid waste collection in Istanbul, Turkey. The chosen formulation is a Capacitated Vehicle Routing Problem (CVRP), with the constraint of visiting each location exactly once. Among the evaluated heuristics, the Path Cheapest Arc strategy combined with Tabu Seach produced the best results for distance minimization. More frequent collection was found to improve citizen satisfaction, offsetting a marginal increase in travel distance as a tradeoff.

Zhao et al. (2024) embed a Reinforcement Learning (RL) strategy in a routing model aiming to optimize waste collection by minimizing carbon emissions and operational costs. The authors propose a Q-learning based hyperheuristic particle swarm optimization (QLHPSO), which selects the most effective low-level heuristic (LLH) from a pool of eight options. This model reacts rapidly to disruptions, such as new service requests and vehicle breakdowns; each vehicle is allowed to have multiple trips to collect waste, with the aim of reacting to the environmental changes timely.

Overall, these studies attest that heuristics are consistently utilized to address the challenges of waste collection routing. Heuristics provide the flexibility necessary to perform comparative analyses to evaluate a wide array of solutions for real world scenarios.

2.1.5 Behavioral Drivers of Illegal Dumping

Literature discussing the behavioral drivers behind illegal dumping is relevant to this thesis project, as the results of the ML can only validate an already existing theoretical framework of the phenomenon.

The previously mentioned study by Glanville and Chang (2015) identified the key drivers associated with illegal dumping on the Sunshine Coast of Australia. Crucially, while the most significant factors were environmental (distance to the nearest street, landfill, and roadside amenities), the BLR analysis determined that socio-demographic factors, like population density or average median income, were not as relevant.

Similarly, Du et al. (2023) point towards natural factors, like vegetation coverage, or traffic volume and road type. However, social factors are also taken into account: isolated areas are also classified as high risk. Furthermore, proximity to shops and businesses, especially those that treat furniture or function as charities, is also considered as a risk factor for illegal dumping.

The hybrid study by Hohl et al. (2023) explores more in depth the social predictors of illegal dumping, incorporating stakeholders' perspectives in the analysis. Economic decline, lack of surveillance, high concentration of vacant buildings, and ”Dumping Culture” perception are among the factors social factors that this study associates with illegal dumping. This is in line with the ”Broken Window” theory, which suggests that signs of public disorder discourage the public from upholding standards of civil behavior.

2.1.6 Illegal Dumping in the City of Rotterdam

A study by Merkelbach et al. (2021) evaluates the effectiveness of enriching the standard policy of door-to-door canvassing to reduce illegal dumping with commitment-nudging. Households were asked to put a ”Keep the Outdoors Clean” sticker on their doors, and information boards were placed right next to waste containers. These nudges proved to be highly effective in reducing illegal dumping by two thirds immediately after the test, as well as at the 2 month follow-up. The authors suggest that these measures were likely aided by strong pre-existing levels of social cohesion, which were only emphasized by the nudging strategies.

A research paper on naastplatsing by the Municipality of Rotterdam (Epskamp & De Vries, 2020) combined multiple data streams (administrative documents, employee surveys, citizen reports of illegal disposal) to offer a bird's eye view of the main factors contributing to naastplatsing. Among these, 5 key indicators were identified:

- weekends and festivities. A combination of the reduced number of disposal workers and the increased amount of people at home during the day contributes to containers filling, which can lead to naastplatsing.
- high volume areas. Some bins are utilized by a significantly greater than average amount of citizens; the increased probability of improper disposal points towards

an inadequate capacity of the containers.

- social cohesion. In neighborhoods with low social cohesion individuals feel less responsible for their surroundings, and neighbors don't look out for problematic behavior. Moving is also quite frequent in these areas, which is often responsible for large waste to be improperly placed next to containers.
- technical difficulties. More often than not, naastplatsing can be caused by the opening of the container malfunctioning. Disposal of carton boxes can also obstruct bins even when they're not full, due to the high-volume this type of waste can occupy.
- socio-economic factors. Frequent naastplatsing has been tied with areas with lower incomes, higher percentages of students and/or young adults, and a prevalence of small, cheap households.

2.2 Conclusion of Literature Review

The reviewed literature shows that a variety of analysis techniques have increasingly been applied in the context of MSWM. Statistical and MVDA methods like PCA, PLS and MLR are proven in handling large datasets with multiple predictors, allowing for the reduction of dimensionality and evaluation of socio-economic indicators. The integration of spatial-based models enables customized analysis and solutions on clusters and other geographical aggregations. Recent studies highlight a growing interest for the application of ML techniques to MSWM, with a focus on the prediction of the volume of generated waste and the location of illegal dumping sites. These models can outperform the traditional statistical methods, and their deployment is supported by the vast amounts of data that smart cities are able to collect through the development of technologies relating to the IoT. The literature also shows that heuristic algorithms are increasingly implemented for solving complex routing problems in the waste management domain. Their flexibility allows ad-hoc specification to various operational constraints found in municipal solid waste collection. Behavioral and qualitative literature provides a theoretical understanding for the interpretation of qualitative results, underscoring the importance of socio-demographic factors and the characteristics of neighborhoods.
Table 2.2 synthesizes these contributions by categorizing the reviewed studies according to their methodological focus and research domain.

Category	**Nr**	**Methods**
Review	7	Conceptual analysis, literature review
Statistical Analysis	3	MLR, PCA, PLS, GLMM, clustering
Geospatial Analysis	3	OLS, GWR
Machine Learning	7	RF, GBRT, ANN, SHAP, LIME
Routing	6	CVRP, heuristics, GA, RL
Behavioral and Qualitative Analysis	3	Interviews, experiments

Table 2.1: Synthesis of literature review by topic

Still, there exists a knowledge gap in the literature for what concerns the analysis of naastplaatsing in dense urban contexts. The majority of the studies on illegal dumping focuses on rural, industrial, or peri-urban contexts. Fly-tipping is inherenlty different from illegal dumping: often the main issue of the latter is to find a secluded area, while naastplaatsing always happens next to municipal containers.

Furthermore, very few studies attempt to combine a theoretical framework of understanding to the qualitative results of a MVDA or ML based analysis. This study aims to utilize both quantitative and qualitative analysis to both understand naastplaatsing and predict it, with the goal of identifying the key factors causing it and supporting targeted intervention towards it.

Chapter 3

Background

3.1 Overview

This thesis sets out to understand, explain, and model the phenomenon of naastplaatsing through a combination of various analysis techniques, with the goal of improving the efficiency and the effectiveness of the municipal response to this issue.

Naastplaatsing is a complex, wicked problem that requires expert knowledge to be properly conceptualized. The first step of this research is to acquire such expertise through a combination of:

- analysis of existing relevant literature, like the naastplaatsing report (Epskamp & De Vries, 2020) from the Research and Business Intelligence (OBI) department.
- expert interviews, during which individuals familiar with naastplaatsing and the municipal process s to address it are asked to share their experience.
- on-field experience, directly experiencing the operations of the NietRnaast crew on a regular day of work.

The information acquired through these methods is synthesized in a conceptual framework representing how different factors interact with naastplaatsing and each other. A selection of these interactions is translated into eight hypotheses, which will be validated and utilized in the other stages of this study. Furthermore, the gained knowledge will be used to identify the most relevant actors involved with naastplaatsing. This representation is a problem arena, where the actors are positioned in a 2-dimensional plane based on how much power they hold over naastplaatsing, and how much they are interested in it.

This chapter is divided into three main sections. First, the datasets utilized are introduced and described. Following this, domain knowledge is presented by delving into the three aforementioned streams of information. Finally, the developed theoretical framework is described and discussed, alongside the eight hypotheses derived from it.

3.2 Datasets

3.2.1 NietRnaast

The most utilized dataset in this study is the one comprised of all the administrative actions carried out by the NietRnaast (nRn) crews during their activity. The first recorded actions date back to the beginning of 2021, and the final cutoff date is the 31st of December 2025. A total of 3907855 administrative actions is recorded in the nRn dataset. The number of actions differs between the years of activity, with the following distribution: 65515 for 2021, 718822 for 2022, 944937 for 2023, 968459 for 2024, and 1210122 for 2025. The lower amount of datapoints for the first year is due to the (then) recent adoption of the data collection process, which makes the first year of this distribution less useful for this analysis.

Each row of the nRn dataset represents one of the administrative actions carried out by the crews, and it details its relevant characteristics. Among these, the "action" column qualifies the type of intervention that was necessary; "Niets aangetroffen" means that the visited cluster was clean, while any other type of waste is labeled accordingly ("Karton", "Grofvuil", "Klein afval", etc...). This column also provides information on the status of the cluster at the time of the intervention, with entries reading "Klemmers" or "Volle container" indicating malfunctions or full containers. Container clusters are geolocated through latitude/longitude columns, and the "wijk_NietRnaast" column shows in which neighborhood it is.

3.2.2 Central Bureau voor de Statistiek

The data retrieved from the Central Bureau voor de Statistiek (CBS) dataset contains socio-economic variables and environment density proxies. The level of aggregation is neighborhood. The temporal coverage of this dataset is complete from 2021 through 2023, while most of the selected indicators are incomplete from 2024 onwards.

The key fields in this dataset are the following:

- workforce_participation (percentage of the neighborhood residents)
- avg_income_per_resident (average)
- low/mid/high education (percentage of the neighborhood residents)
- address_per_km2

3.2.3 Onderzoek 010

The Onderzoek010 dataset contains neighborhood-level information on the composition of the population, as well as social cohesion indicators in the form of responses to the Wijk Profiel survey assessing neighborhood sentiment.

Age groups are recorded in brackets of 5 years gaps, engineered into 5 groups: 0/14 years old, 15/29 years old, 30/44 years old, 45/64 years old, 65+ years old. The population's origins are recorded in the following brackets: Dutch, Europe (excluding NL), Turkey, Morocco, Suriname, Indonesia, Cape Verde, Caribbean, Africa, Asia, America and Oceania. For this analysis, these are reduced to Dutch, EU, non-EU.

Population data is available up to 2024, with predicted projections for 2025. Wijk Profiel surveys are biennial, with the latest available in 2024. The data for 2023 was imputed through an arithmetic average of the cohesion items of the years 2022 and 2024.

3.2.4 Waste Vision

The Wastevision (WV) dataset contains information on most of the containers managed by Stadsbeheer. The most relevant are:

- container_number (identifying the container univocally)
- container_site (the number of the cluster of which it is a part of)
- container_type (different levels)
- capacity (expressed in m3 or litres)
- fraction (type of waste serviced by the container)
- fills_in_days (average number of days between a container being emptied and it reaching maximum capacity)
- collection_type (scheduled or fixed)

GFE (vegetable, fruit, and food scraps) containers are not part of this system, which is used to map Inzameling collection routes based on fill-level of the containers.

3.2.5 GIS

This is a miscellaneous dataset containing the coordinates of various types of resources in Rotterdam. Hotels, restaurants, and cafes are points of interest due to their high volume of waste generation; public waste disposal facilities (milieuparks) are also relevant, since difficulty of accessing them may encourage improper disposal of bulky waste, like furniture or household appliances.

3.3 Domain Knowledge

3.3.1 Literature Review

Existing literature on naastplaatsing is the first source of domain knowledge. It is primarily produced by institutions already involved with waste management, like the municipality of Rotterdam itself. Previous research in this field adopts an approach synthetizing qualitative and quantitative analysis, in which it is similar to this study. Such studies rely on the only data available before the NietRnaast crews started recording administrative actions, which consists primarily of reports made by either citizens or Handhaving officials. As a result, the data disproportionately reflected areas where residents were more inclined to file complaints, rather than areas where naastplaatsing itself was objectively more prevalent. This bias also affected some of the findings: for example, the observed positive correlation between naastplaatsing and containers with citizen adopters likely reflects oversampling in neighborhoods with more engaged residents, not a causal relationship. This study benefits from a more complete and operationally grounded representation of the distribution of naastplaatsing, thanks to the nRn dataset. Still, this research will make a conscious effort to reduce its exposure to biases of sampling or representation of certain areas, for instance utilizing rates and percentages instead of absolute values whenever possible. Previous work is not only used to gain domain knowledge, but also contributes to shaping the research method of this thesis.

The main body of analytical work on naastplaatsing is made of municipal "status" documents: the ones inspected in this section are the "K+V Rapportage Evaluatie Vulgraadsystem" (*Evaluatie Vulgraadsysteem Rotterdam*, 2021) report, the "Verklaringen voor naastplaatsing in Rotterdam" (Epskamp & De Vries, 2020) report. For brevity, from this points forward these reports will be referred to as the "K+V report" and the "OBI report". K+V is the consultancy firm hired by Stadsbeheer to produce their report; OBI stands for Onderzoek en Business Intelligence, the department that created the research on naastplaatsing.

Both these reports are quite dated; as such, their findings are focused on the period of 2017 to 2020, a time when the nRn initiative was in its infancy and was not recording the outcome of its interventions on container clusters. Furthermore, a significant event separates these years from the present: the COVID-19 pandemic. The K+V report in particular details how during the lockdowns residents generated more waste due to being confined in their homes and increasing their reliance on delivery services. Nationally, 73% of municipalities reported an increase of naastplaatsing during this period. Furthermore, as previously stated, these documents and their analysis rely primarily on reports of naastplaatsing, as opposed to a dataset of information collected on all areas of Rotterdam in a (more or less) fair distribution.

To summarize, the findings of these reports are significant, but should also be scrutinized based on two main concerns. First, there is the potential obsolescence of the findings due to distance in time and fundamental changes in how waste generation and management operate after the Corona disruption. Furthermore, the potential biases in the collected data due to the distribution of datapoints being highly correlated with reporting tendencies of residents and/or Gemeente officers should not be overlooked.

Naastplaatsing Growth

Rotterdam experienced a significant increase of public reports of naastplaatsing between 2017 and 2020. Data shows an increase of more than 330% for naastplaatsing reports, full-container notifications by more than 260%, and reports made about defective containers by more nearly 100%. This increase coincides with a wider adoption of the city's reporting infrastructure, with submissions coming both from residents through the BuitenBeter app and from municipal teams of the Reiniging and Handhaving departments. Non-waste reports rose by 36% over the same period, suggesting an increase in public engagement with reporting platforms.

Sensor-based monitoring

The effectiveness of sensor-based monitoring directly correlates with naastplaatsing. This is because full containers, which are what sensors are designed to address and avoid, are conducive to naastplaatsing, especially among citizens that are prone to ignoring the proper waste disposal methods at the slightest inconvenience. Furthermore, residents may misinterpret the presence of trash next to the container as an indicator of its fullness, leading to more naastplaatsing: this is the foundation of the "trash brings trash" saying.

Fill-level sensors have been subject to technical failures, which lead to preventable overflows of waste in containers, which in turn generate more naastplaatsing. Since the introduction of these sensors in 2018, steps forward have been made with reliability and monitoring; still, reliability concerns persist both in literature and presently in the opinion of interview participants. Nonetheless, the fill-level sensors have contributed to a substantial increase in the efficiency of Inzameling's operations, with the number of collection routes remaining virtually the same in 2017 and 2020 even with substantially more waste volume.

Quantitative Insights

The OBI report examines how the characteristics of containers relate to naastplaatsing occurrences.

Containers are often part of clusters, groups of 2-4 containers that share the same platform. Clusters with more containers (and thus, capacity) appear to be more exposed to naastplaatsing; this may seem counterintuitive, but it is likely related to the intentional distribution of such clusters in areas with a high-volume of waste generation. Similarly, the presence of different waste containers positively correlates with naastplaatsing counts. This is likely collinear to the number of containers in a cluster.

Mondays have the highest amount of reports, reflecting a higher volume of waste generated and disposed of during the weekends, when fewer municipal workers are available. Months with vacation days and/or periods (May through July, October through December) exhibit elevated naastplaatsing levels; December is the most exposed, between Sinterklaas, Christmas, and New Year's Eve.

Malfunctions of a container's opening, or of its fill-sensor, are shown to trigger more naastplaatsing. Even legitimate appointments to dispose of bulky waste may induce "trash brings trash" behavior, since areas that make use of this municipal service are impacted (comparatively) more by naastplaatsing.

Population density is directly proportional to waste generation volumes, and in turn to naastplaatsing counts; furthermore, more citizens using municipal containers translates to a higher number of individual occasions for improper disposal of waste. The composition of the population is also investigated in the report: areas with high proportions of young adults (18-27 years old), Central-Eastern European (CEE) migrants, and short-term residents exhibit systematically higher naastplaatsing counts.

Residential turnover is a significant predictor of naastplaatsing counts. There is a twofold explanation for this connection: frequent moving indicates a lower desire to build a connection with the neighborhood, which has been linked to unwillingness to dispose of waste properly. Furthermore, on a more practical note, the act of moving (in or out of a residence generates extraordinary amounts of waste, in the form of construction residue, leftover furniture, or packaging materials.

Areas with nine or more businesses, especially restaurants and supermarkets, show around double the amount to naastplaatsing counts compared to areas with fewer businesses. Businesses often have waste disposal contracts with Stadsbeheer, consisting of a fixed tax to be paid in exchange for a fixed amount of waste to be disposed of. Exceeding this limit means incurring in extra costs; businesses aim to avoid this by utilizing municipal containers, even though this is illegal. This contributes to higher pressure on said containers, which in turn translates into more chances to have naastplaatsing.

Qualitative Insights

The OBI report also relied on interviewing workers familiar with naastplaatsing to highlight aspects of naastplaatsing that are more elusive to quantitative analysis.

Interview participants indicated that residents place waste next to containers due to minor inconveniences, such as dirty or jammed openings. Visible naastplaatsing triggers further disposal, due to either a perceived acceptability of the behavior or an assumption of container fill level.

Residents ailing from different countries could be used to different waste disposal habits and/or cultural norms, leading to naastplaatsing being considered as acceptable or even the proper way to get rid of garbage. The challenge of identifying perpetrators contributes to low enforcement and perceived repercussions, which does not discourage residents from naastplaatsing.

It is also noted that sometimes Inzameling crews get rid of naastplaatsing without reporting its presence, since it is an obstacle to collection operations.

Synthesis

The OBI report consolidates 5 composite factors that jointly predict naastplaatsing.

- Holidays: increased waste generation and reduced staff availability.
- Container Pressure: correlation between generated waste volumes (tied to population served, container capacity, and business presence) and naastplaatsing counts.
- Low Social Cohesion: frequent moving, weak social accountability, weak neighborhood ties.
- Waste Appointment Offer: approved bulky waste attracting naastplaatsing, heightened detection for adopted containers.
- Low Prosperity: socio-economic struggles, smaller housing of lower quality.

The following main conclusions can be drawn. Naastplaatsing is a growing, complex, and multifaceted phenomenon that affects Rotterdam disproportionately: spatial hotspots and temporal clustering emerge as key characteristics. Container pressure is a crucial predictor of waste overflow, which in turn correlates positively with naastplaatsing. Moreover, behavioral and social indicators amplify structural vulnerabilities, often proxying low social cohesion among residents. Finally, municipal involvement is associated with higher naastplaatsing counts, although this is likely attributable to overreporting tendencies in target areas.

Crucially, the conclusion of the OBI report is that no single factor is universally explanatory; still, their composite interpretation offers a valuable analysis of naastplaatsing as a phenomenon.

Recent Updates

As previously stated, the insights provided by the analyzed reports are dated. Still, its findings can be presently contextualized by observing the trends in the nRn dataset when it comes to naastplaatsing rates. The following figure shows how the percentage of administrative actions that resulted in any type of naastplaatsing found has been steadily growing over the years, a trend with two main interpretations. On one hand, this shows that nRn operations have become more efficient over time, targeting containers with a higher risk of naastplaatsing more often; on the other, since the total number of actions has also been growing (700k in 2022 to 1.2 million in 2025, as shown in figure 3.1), this also means that naastplaatsing is still growing as a phenomenon.

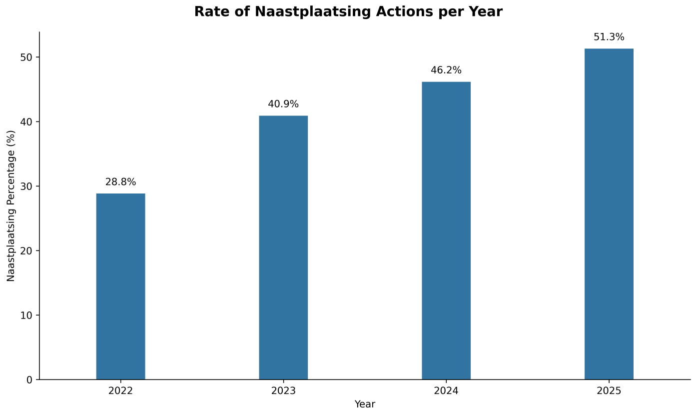


Figure 3.1: Yearly rate of administrative actions that documented naastplaatsing, 2022-2025

Regarding fill-level sensors further improvements have been made, since now most of the Inzameling routing relies on the data they provide. Still, some of the interviewed workers (Participant F, Appendix A.1) mentioned that these sensors are presently unreliable, leading to impactful disruptions of waste management services.

3.3.2 Expert Interviews

To complement the literature, unstructured and informal interviews are conducted with individuals familiar with naastplaatsing and the municipality's process to address it. These interviews are mostly open-ended, but a consistent line of questioning is presented across sessions. Participants are always asked to characterize the nature of naastplaatsing across various dimensions: prevalent types of waste, temporal patterns, hotspots. Other topics often covered include: the usage of data in operational decision-making; the effectiveness of nRn as an initiative and its interaction with the other departments of Stadsbeheer; the impact of operational and logistical factors, such as container design and placement, or collection frequency. In addition, participants are systematically asked to identify the top 3 to 5 causes of naastplaatsing according to them. Furthermore, participants fill in a power-interest matrix identifying the actors involved in naastplaatsting. These actors can be individuals (e.g. aldermen, waste collection managers), organizations (e.g. municipal departments like Reiniging or Inzameling), or groups of citizens. The actors are positioned in this space according to the degree of influence they can exert over naastplaatsing and their level of interest in addressing it.

A total of 7 interviews have been carried out across the months of November and December 2025. The professionals participating occupy various roles tied to naastplaatsing in Stadsbeheer's departments. From the NietRnaast area a manager, team leader, project leader, and product owner were involved. A manager from the Reiniging department working with NietRnaast was also interviewed, as well as a Strategic Advisor involved with containers and naastplaatsing. A Management and Group Support Trainee analyzing NietRnaast's workflow also took part in this process. The identities of these experts are not disclosed, and their contribution is referred to in the form of “Participant X”, following the interview table in Appendix A.1.

A recurring theme across all interviews is the strong correlation of container availability and waste management operational performance to naastplaatsing occurrences. Respondents often mentioned the frequency of emptying containers, their filling speed, and technical hurdles like jammed openings or collection delays as key contributors to naastplaatsing. The spatial distribution of the phenomenon also appears to be uneven: neighborhoods have widely varying rates of naastplaatsing, with areas in the Center and South of Rotterdam reported as persistent hotspots. In a more limited, yet still present, capacity, the temporal aspect is also mentioned. The first days of the week often present a more dire situation in terms of naastplaatsing due to higher waste volume generated during the weekend by people staying at home, combined with less available personnel to collect waste. Holidays (and the days immediately following them) pose similar challenges, as well as days in which extreme weather conditions impede regular waste collection operations (e.g.: the heavy snow days that occurred on the first days of January 2026).

Interviewees consistently point towards container design as one of the key determinants of naastplaatsing occurrence. Smaller openings and lower-capacity containers, usually of the HV kind (half-verdiept, half underground), make clusters more exposed to jamming and/or overflow, which in turn causes residents to leave waste next to containers. Karton waste is prevalent for this, due to the dimensions of cardboard boxes often exceeding the size of container openings. These could theoretically accommodate this type of waste in most occasions, but only if the cardboard is folded properly. This can lead to containers that appear full due to a malfunction, but in reality are not.

Municipal stakeholders also believe that naastplaatsing rates are strongly influenced by citizen behavior, in particular their relationship with social norms and each other. Multiple interviewees (Participants A, B, C, and E, Appendix A.1) distinguish between 3 categories of residents, characterized as follows:

- Type 1 residents, who are directly engaged with their community and dispose of waste properly.
- Type 2 residents, who are aware of waste disposal practices but are mostly driven by convenience. These are the residents that may walk to another container if their usual one is full or malfunctioning, but this outcome is not guaranteed.
- Type 3 residents, who are often staying in Rotterdam for a short time and are not at all connected with their surrounding area. These individuals may not speak Dutch or English, making them incapable of understanding with municipal waste disposal directives, as well as making any meaningful connection with their neighbors. These residents may also hail from parts of the world where disposing of waste directly on the road is the status quo, which does not help with reducing naastplaatsing.

The most consistent description of residents is one of lazy, convenience-driven individuals, who very often have other things on their mind and are minimally impacted by naastplaatsing and its consequences. In fact, a paradox of NietRnaast efficiency is often discussed: since collection crews operate daily, residents who dispose of their waste next to containers will often find them clean only a few hours after their dumping. Interview respondents agree on the proposition that NietRnaast work increases neighborhood cleanliness, but reduces the incentive for proper disposal behavior. This can make citizens less responsible and complacent, since the negative effects of naastplaatsing often don't persist long enough to be internalized. Furthermore, the Handhaving department is the least integrated in the work on NietRnaast; without tangible consequences, the system of disincentive through penalties loses a significant portion of its effectiveness. It must be noted that actually identifying the people responsible for a naastplaatsing occurrence is very unlikely, unless residents are caught in the act (Participants D and G, Appendix

A.1)
Respondents noted that despite the vast amount of data generated by NietRnaast operations, its impact on operations is very limited. Interviewees agreed that the existing data can be utilized to efficiently allocate resources, particularly in relation to reducing unnecessary routing through "healthy" neighborhoods and increasing visit frequency in hotspots and fast-filling locations. However, data quality is also reported as a consistent issue, with action categories often misclassified to expedite reporting, implausibly short reported cleaning timestamps, and improperly designed functionalities in the NietRnaast collection app.

During each interview, participants were asked to list the top 5 most prominent reasons for naastplaatsing. A synthesis of their responses highlighted the following drivers:

- Operational/logistic drivers, meaning container fill-speed and occurrences of overflow and/or container malfunctions. In general, the load a container has to support in terms of generated waste volume is often mentioned as the primary culprit of naastplaatsing.

- Citizen behavior, influenced by convenience-driven practices, limited motivations, insufficient understanding of disposal rules, and lack of perceived consequences of naastplaatsing.

- Low social cohesion, characteristic of temporary workers and other extra-national groups who are also often tied to high residential turnover. Lack of long term attachment to the neighborhood leads to higher naastplaatsing rates.

- "Broken Window" dynamics, the behavioral theory supporting the "trash brings trash" statement. Due to laziness or simple assumption, residents tend to normalize and participate in improper waste disposal once they see it often enough next to their usual containers.

Furthermore, interviewees were asked to fill in a Power/Interest matrix describing the decision arena of the naastplaatsing phenomenon. In here, actors are positioned according to their direct impact on the phenomenon and their involvement in its resolution. Actors can be any individual, group or organization that influences, is affected by, or participates in the processes of waste management. A synthesis of their responses is represented by the matrix listed as Figure 3.2.

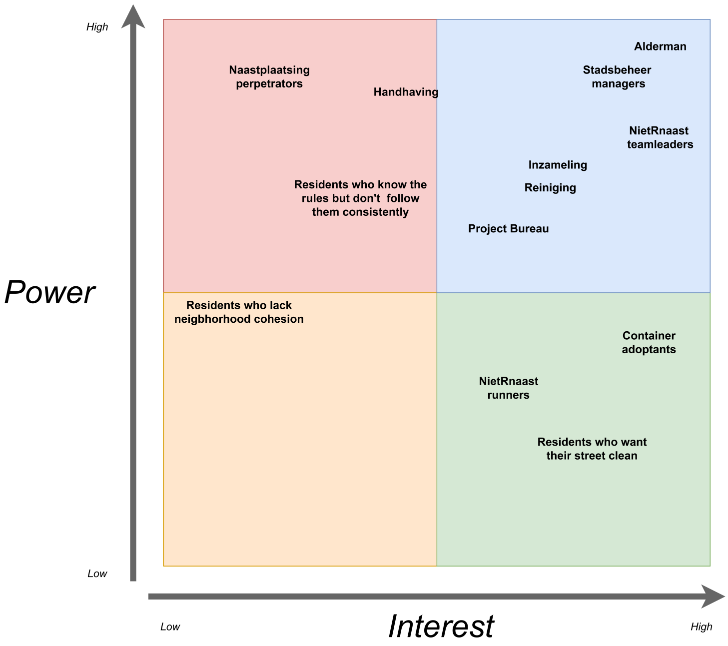


Figure 3.2: Power/Interest Matrix synthesizing expert interview responses on naastplaatsing actors

The key players are actors who can both influence decisions and care strongly about addressing naastplaatsing, and are thus positioned in the blue quadrant. The alderman holds political authority, influences enforcement emphasis, and sets the overall priorities of the waste management operation; Stadsbeheer managers execute his mandate and manage the coordination between Reiniging, Inzameling, and Handhaving. NietRnaast teamleaders oversee daily execution of the department's operations, positioning them in a uniquely powerful position when it comes to the naastplaatsing problem. Inzameling, Reiniging and Stadsbeheer workers directly shape the systems designed to respond to naastplaatsing, often interacting directly with the NietRnaast department.
The Handhaving department has the power to deter naastplaatsing through fines and presence, but may not prioritize waste issues. Lazy residents who are knowledgeable about waste management practices hold "soft power": their behavior shapes local norms and waste patterns, and while they comply when convenient, they lack interest in long-term neighborhood cleanliness. The most crucial actor in this red quadrant are of course the perpetrators: by definition, this is the category that holds the most power and has very low interest. Actors in the red quadrant are the ones that should be kept "satisfied", since their disengagement can undermine interventions. They hold behavioral power since they shape the problem, but their convenience-driven practices suggests a lack of interest in its resolution.
On the opposite side of the matrix, residents who want to keep the street clean are strongly committed, yet their power is limited to holding their neighbors accountable. Container adopters are an even more interested and powerful group, since they care about the issue enough to volunteer to monitor containers and can actually operate on them in case of malfunction through a key they are given by Stadsbeheer. NietrNaast runners are deeply engaged and knowledgeable when it comes to naastplaatsing, yet they hold limited structural influence over policy or response design.
The final quadrant contains residents who lack cohesion with the neighborhood. This

is the lowest group in terms of interest, due to their often limited time of permanence in Rotterdam. Actors in this quadrant are often characterized as the ones that require minimal management; still, it is telling that the actors that are perceived to create the most problems are not the ones who can solve it. This misalignment characterizes naastplaatsing as a truly wicked problem.

While stakeholders differ in their level of power and interest, their goals regarding naastplaatsing can broadly be divided into two perspectives: operational efficiency and behavioral impact. Actors that can be described as closer to the collection process, such as NietRnaast teamleaders, runners, and product owners, primarily focus on the efficiency and practicality of waste collection. For these stakeholders, minimizing operational metrics like travel distance is of paramount importance.

In contrast, stakeholders more involved in planning and policy design, such as project managers and strategic advisors within Stadsbeheer, tend to frame the naastplaatsing problem through a behavioral lens. The "trash brings trash" axiom is the foundational idea often discussed from this perspective. An extension of this concerns the timeframe of naastplaatsing permanence: the longer waste remains visible near containers, the more likely it is to attract additional dumping. For these actors, the key objective is therefore to act on this signifcant underlying trigger of naastplaatsing behavior, by reducing the time that naastplaatsing remains uncollected in public space, limiting its ability to spread or normalize. Similarly, residents who actively care about neighborhood cleanliness share this priority; their primary concern is the visible presence of waste in their streets.

Finally, higher-level decision makers such as Stadsbeheer managers and the alderman balance both perspectives. From a strategic standpoint, they must consider operational efficiency and budget constraints; on the other hand, responding to public perception and political pressure regarding street cleanliness is one of their top priorities. As a result, their goals simultaneously align with reducing operational costs and ensuring that naastplaatsing is removed quickly and consistently; the crux of the problem stands in balancing these two goals. These stakeholder priorities shed light on the thought process behind the operational response of Stadsbeheer to naastplaatsing, and provide this study relevant information to structure performance metrics of the routing study.

3.3.3 On-Field Experience

Domain knowledge is further enriched through on-field experience, consisting of participation in a shift of nRn cleaning. The focus of this observation are three main aspects:

- the average amount and types of naastplaatsing encountered
- the interaction between crew and the data collection process (with particular attention to data quality)
- the routing habits of the crews during their operations

Naastplaatsing is highly prevalent, appearing at an estimated 60-75% of container clusters visited across two consecutive shifts. Most occurrences consisted of relatively small amounts of waste, such as open waste bags, cardboard boxes, broken glass, and various refuse. A minority of instances amounted to severe cases, with voluminous waste like furniture, electronic appliances, and hazardous material, such as laughing gas canisters. The presence of huge and/or unsafe waste is operationally significant: nRn workers cannot deal with these instance of naastplaatsing, and are required to request the intervention of an Inzameling truck. This increases the amount of time during which a cluster has naastplaatsing, allowing the negative “trash brings trash” effect to last longer. In summary, the majority of naastplaatsing instances are easily dealt with, while a few severe ones cause the most disruption.

Until recently, the reporting app used by nRn crews to collect data allowed drivers to take pictures of the containers only if they recorded the action type as some form of naastplaatsing. This constraint, combined with the request of team leaders to occasionally take pictures of all the visited container clusters, created a data quality issue. Since the picture directive was mostly done for the Target Wijks, nRn drivers are compelled to misclassify clean clusters to activate the camera feature. This practice fuels a negative feedback loop, where already troubled neighborhoods get negative data points in spite of the real-world situation.

Drivers are presented with 17 different options for the type of action they carried out on a container cluster. These range from the different types of encountered waste (carton, small bags, furniture), to operational details about the state of the containers (full, malfunctioning), to required interventions (chemical, Handhaving officer). However, it was observed that nRn crews overwhelmingly select “Klein afval”, the most generic form of residual waste. This was not always an accurate reflection of the naastplaatsing occurrence, but rather a quick shortcut to move on to the next container cluster.

The “Klein afval” category was consistently the most chosen across the years of nRn data collection (2022-2025); however, its share has increased significantly, going from 10.29% in 2022, to 15.89% in 2023, to 19.82% in 2024, to 27.64% in 2025. This result may be a combination of the two aforementioned data quality issues: nRn drivers are compelled to select a naastplaatsing type of action, and “Klein afval” is the most generic (and already frequently selected) option available.

Additionally, nRn routing lacks optimization. Observed crews work through container clusters efficiently, but the visiting appears to be largely experience-based and intuitive. From an external perspective, this approach presents opportunities for improvement. Introducing prioritization rules, for instance on known hotspots or containers with documented chronic overflow, could reduce the negative effect of naastplaatsing attracting more waste as long as it remains next to containers.

3.4 Theoretical Framework

The theoretical framework in figure 3.3 conceptualizes naastplaatsing as the result of multiple interacting socio-economic, operational, and behavioral factors. Relationships are represented in the format factor1 → factor2, where the sign indicates whether an increase in factor1 leads to an increase (+) or decrease (-) in factor2.

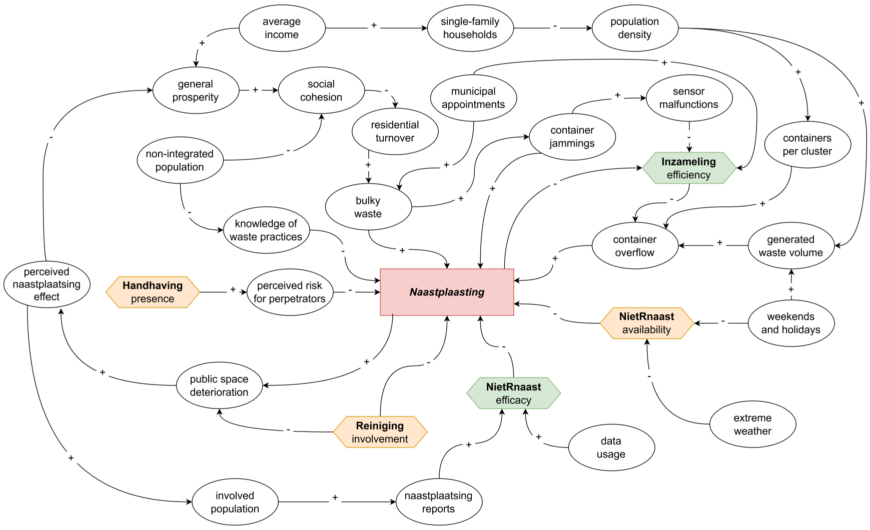


Figure 3.3: Theoretical Framework of naastplaatsing

This structure allows for identifying the feedback loops and causal chains that determine the conditions under which *naastplaatsing* emerges in Rotterdam. The following analysis of this theoretical framework focuses on these aspects:

- high impact factors, which represent the strongest drivers of naastplaatsing in the framework.
- feedback loops, which highlight the reinforcing and balancing nature of the interactions shaping the occurrence of naastplaatsing.
- municipal levers, referring to the direct impacts that Stadsbeheer can impose on naastplaatsing through its departments.
- peripheral factors not investigated by this thesis, which appear in this framework but are not empirically tested in any of the methods of this study.

3.4.1 Theoretical Framework Analysis

Factors that either trigger immediate operational disruptions around containers or alter resident behavior and expectations affect naastplaatsing in a significant way. Container jammings and overflow constitute two powerful accelerators of naastplaatsing, both through direct and chain effects. In particular, container jammings are especially harmful thanks to their impact on sensors: waste stuck in a container's opening may affect the reported fill level, leading to inefficiencies in the collection operations of Inzameling. Bulky waste is a part of this causal chain and also impacts naastplaatsing directly. Generated waste volume captures relevant dynamics of naastplaatsing in its relationship with population density and weekends/holidays: pressure on containers increases when more residents are at home, as well as when the density of inhabitants increases. Social factors also exhibit significant impact, as shown by the interactions of non-integrated residents with chains that increase naastplaatsing. The perceived effect of naastplaatsing also interacts with several causal chains, proving the relevance of behavioral factors in the challenges posed by this phenomenon.

The naastplaatsing factor is involved in several feedback loops. When waste is deposited on a containers platform, this can impede the collection operations of Inzameling; the consequent decrease in efficiency leads to higher occurrence of container overflow, which is recognized to be among the leading causes of naastplaatsing. The presence of waste also contributes to public space deterioration, which through its causal chain leads back to an increase of bulky waste. This type of sizeable waste is often associated with naastplaatsing.

Five of the presented factors represent the levers that Stadsbeheer can influence directly to address naastplaatsing. These are encoded as the effectiveness and engagement of the various Stadsbeheer departments involved in combating naastplaatsing. Their different color scheme also highlights their "economic" impact on Stadsbeheer: yellow factors will cause higher costs for waste management if they are increased, while green factors should allow saving. The role of NietRnaast is split into availability and efficacy. The latter shows the importance of data and reporting streams, which can empower their operations to address naastplaatsing in a more impactful manner. The availability factor, in combination with generated waste volume, underscores the impact of holidays and weekends, providing a theoretical underpinning to the phenomenon of post-weekend naastplaatsing surges. The Inzameling factor is one of the most central factors of the framework. It is affected by three relevant elements, and is tied directly to the high-impact container overflow variable. Outside of NietRnaast, it is the department with the most impact in terms of the operational capabilities of Stadsbeheer. Its balancing effect is also the most vulnerable to the different causal chains characterizing naastplaatsing. The Reiniging department handles city cleaning, and can be involved in addressing naastplaatsing directly and indirectly by maintaining cleanliness around containers and reducing the deterioration of the public space around containers. Handhaving is the behavioral lever at Stadsbeheer's disposal: it does not influence the system operationally, but rather it can shape the decision-making process of potential offenders. These municipal levers form the controllable backbone of the system; they operate as the entry points for policy interventions, where resource allocation can destabilize negative naastplaatsing loops and causal chains and reduce its prevalence across Rotterdam.

The framework includes several factors that are not examined by this thesis, despite their conceptual relevance. Weather conditions can impact the availability of all waste management services; however, the limited temporal resolution of the available data disqualifies them from being included in the analysis. Knowledge of waste practices is directly related to naastplaatsing, but no dataset quantifies awareness or compliance of the general population. Public space deterioration can be quantified reliably through street-image datasets, yet this falls outside the scope of this study. Modeling perceived impact of naastplaatsing requires a level of data granularity that is currently not available for this study. Municipal appointments, sensor malfunctions, and the routing logs that would be necessary to model inzameling efficiency as a dependent variable are also not available for this thesis.

3.4.2 Hypotheses Formulation

The insights on naastplaatsing described by the theoretical framework are operationalized for the research by the formulation of the following hypotheses:

1. Structural load on container (clusters) → + Naastplaatsing
The frequency and risk of naastplaatsing increase with higher container "pressure", defined as the amount of waste the container clusters are structurally exposed to. Features expressing this are container fill-speed, capacity, the amount of businesses nearby container clusters, and the average number of residents served by each cluster.

2. Malfunctions and full containers → + Naastplaatsing
Containers that frequently malfunction or reach full capacity create direct physical barriers to proper disposal, leading to more naastplaatsing. Features expressing this are the rate or count of "volle" and "klemmers" actions in the NietRnaast dataset, as well as the amount of "problematic" countainers that go beyond a certain threshold of these types of actions over a period of time.

3. Container cluster composition → ? Naastplaatsing
The mix of containers within a cluster influences naastplaatsing, though the direction of the effect is uncertain. Features expressing this are the waste fractions and the type of containers in different clusters.

4. Temporal stress → ? Naastplaatsing
Whether naastplaatsing occurs disproportionately on days with different operational pressure. Features expressing this are the rate of naastplaatsing happening on days that are outside of the post-weekend surge, so Thursday and Friday in particular, which operate at the same level as other weekdays in terms of collection frequency (compared to Saturday and Sunday) but usually display less dumping occurrences.

5. Naastplaatsing → + Naastplaatsing
The expression of the "trash brings trash" hypothesis. Features expressing this are the length of maximum and average "streaks" of consecutive interventions of container clusters that recorded naastplaatsing instances, the rate of problematic waste deposited next to containers, and the type of collection assigned to containers, scheduled or fixed. Fixed containers are visited more regularly by Inzameling, so they are believed to be less prone to overflow.

6. Social cohesion → - Naastplaatsing
Neighborhoods with stronger social cohesion experience less naastplaatsing due to shared social norms of intra-neighborhood accountability. Features expressing this are "willingness to move" and the social cohesion PCA aggregate of Wijk Profiel questions.

7. Low prosperity → + Naastplaatsing
Socio-economic disadvantage is associated with higher naastplaatsing, potentially through housing density and weaker neighborhood engagement. Features expressing this are the average income, percentage of participation to the workforce, and population shares of level of education.

8. Cultural and normative practices → ? Naastplaatsing
Demographic structure of neighborhoods may shape waste disposal norms and practices. Features expressing this are the demographic composition of the neighborhoods, in terms of age brackets and origin of the residents (distinguishing Dutch, European, and non-European citizens).

Chapter 4

Methodology

4.1 Overview

This thesis sets out to model the phenomenon of naastplaatsing through a combination of various analysis techniques, with the goal of uncovering the most relevant factors behind this issue improving the efficiency and the effectiveness of Stadsbeheer's response.
The theoretical framework formulated in the previous step sheds light on the factors that influence naastplaatsing and their interactions. This led to the formulation of a series of hypotheses, which are going to be tested through factor analysis. More specifically, Ordinary Least Squares (OLS) regression is used to estimate the relationship between a few independent variables (predictors) and a dependent (target) variable (Fiebig et al., 1992), which in this case is the naastplaatsing rate of neighborhoods ("buurts") in Rotterdam. This level of aggregation is chosen because it allows for the inclusion of socio-demographic data from the Onderzoek010 (O10) and Centraal Bureau voor de Statistiek (CBS) datasets as predictors. The validity of the aforementioned hypotheses is determined by the respective importance of one or more features that is directly related to its statement(s).
The following step moves the analysis at the level of container clusters, which are the units at which the NietRnaast crews record administrative actions. The objective at this stage is to predict the naastplaatsing risk of each container cluster. A few different Machine Learning (ML) models will be trained to make such estimates; their overall performance will be compared to determine which one is better suited to model naastplaatsing. Furthermore, a comparison of their feature importance ranking will also be used to further analyse this phenomenon. The features used in this step are not the same as the ones of the previous model: socio-demographic information is not available at this finer aggregation level, which instead allows for a more detailed analysis of the composition of clusters and their surroundings. The (validated) hypotheses will also be used to extract predictive features from the NietRnaast dataset, which can provide information on the "behavior" of container clusters when it comes to naastplaatsing.
The naastplaatsing risk calculated for each container will be used for the final operative step of this study, the routing analysis. The container clusters can be modeled as a network with nodes, edges, and weights, over which the application of the Traveling Salesman Problem (TSP) can lead to the discovery of routes to collect the improperly disposed waste. The routing analysis will consist in comparing different methods to design these routes, which will then be evaluated by analysing their performance over previously observed distributions of naastplaatsing occurrences over container clusters.

The main outcome of this study is understanding, predicting, and efficientizing the response to naastplaatsing. This goal is articulated in the following items, addressed through different research methods:

- a factor analysis of naastplaatsing, evaluating some of the hypotheses formulated with domain knowledge on neighborhood-level trends. With this outcome, we aim to answer research subquestion 2: *How can MVDA techniques be utilized to validate the findings and hypotheses of the naastplaatsing theoretical framework*?
- a predictive model of the risk of naastplaatsing for each municipal container in Rotterdam, developed with qualitative and quantitative inputs and utilized to both predict and better understand (through the comparison of predictor importance) the phenomenon. With this outcome, we aim to answer research subquestion 3: *How can ML algorithms be employed to model naastplaatsing in Rotterdam to a satisfying level of prediction accuracy and result explainability?*
- a routing approach for the collection of naastplaatsing, supported by the output of the predictive model(s) and designed to make the municipal response more efficient than the status quo. With this outcome, we aim to answer research subquestion 4: *How can the collection process of naastplaatsing be made more efficient?*

A breakdown of how each of the research design steps interacts with one another is illustrated in Figure 4.1.

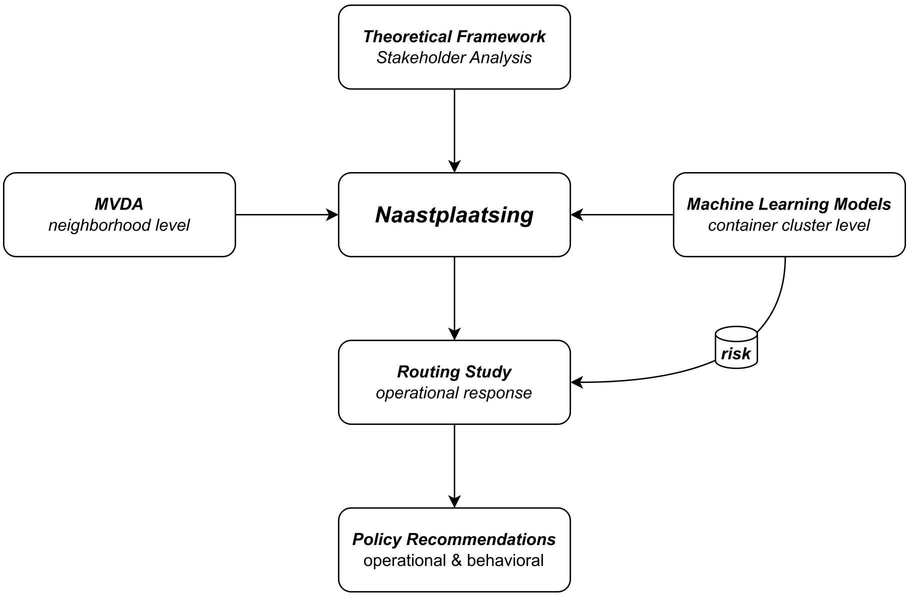


Figure 4.1: Interactions of the different methodological steps of this thesis

The following sections detail how each step of this thesis is structured. First, the OLS-R method used for Factor Analysis is described in its assumptions and relevant features involved. After this, section 4.3 provides an explanation of the tree-based ML models trained for naastplaatsing prediction at the container cluster level. Finally, the Routing Study section outlines the methods utilized to build comparable routes; a subsection is dedicated to a demonstration of how the different heuristic algorithms employed build routes, aided by a visual representation of a toy problem.

4.2 Factor Analysis

The output of the domain knowledge step is a set of hypotheses concerning the factors influencing naastplaatsing. For instance, literature and experts agree that citizens with "other concerns" are less likely to be observant of proper waste disposal. This hypothesis can be conceptualized by proxying this concern with low average income or low workforce participation and investigating its correlation with naastplaatsing rates.

Hypothesis testing is conducted at the neighbothood ("buurt") level. This is the aggregation level at which the Central Bureau voor de Statistiek (CBS) and Onderzoek010 datasets record the socio-demographic indicators involved in this analysis. The analysis is restricted to the years 2022 and 2023 due to data availability; the socio-demographic indicators are not yet available beyond 2023, while previous years lack sufficient nRn observations. The limited variation between these two years makes a panel analysis inappropriate, thus the qualitative modeling consists mainly of hypothesis testing through factor analysis.

Hypotheses are tested using Ordinary Least Squares (OLS) regression, which estimates the linear relationship between a set of features and a target variable. OLS regression is utilized for its interpretability and its suitability for hypothesis testing, with the coefficient value and statistical probability (p value) used to evaluate the formulated statements for the related features. In this case, the predictive variable is the yearly naastplaatsing rate of each neighborhood, calculated as a ratio of the interventions recorded as anything other than "Niets aangetroffen" over the total amount of administrative actions. The independent variables include:

- waste management process-related indicators from the nRn dataset, such as the rate of full containers or container malfunction actions.
- demographic variables, such as age or origin shares of the population in a given neighborhood.
- economic and social indicators, such as education levels, average income, or social cohesion metrics.

OLS regression relies on assumptions of linearity and feature independence, which makes feature engineering necessary. In particular, the data on social cohesion is recorded as a score given to multiple questions in the Wijkprofiel survey (e.g.: "Feels responsible for neighborhood", "Feels connected to neighborhood", "Satisfied with neighborhood", translated). These items are reduced to a single composite indicator using Principal Component Analysis (PCA), a dimensionality reduction technique that uses linear transformations to condense correlated variables into one (or more) component(s) representing their shared variance across a set of observations.

The resulting regression outputs are used to validate or reject the formulated hypotheses, and to provide a quantitative interpretation of the main drivers of naastplaatsing behavior at the neighborhood level.

4.3 Predictive Modeling

At the predictive stage, the analysis shifts to the container cluster level. This is because nRn crews record administrative actions at this aggregation level, recording their interventions on these groups of containers; this means that the exact container at which naastplaatsing occurred is not known. In practice, even crews cannot always determine the source container: waste is often placed in and around the shared surface between containers, making it impossible to assign responsibility to a specific one.

The target variable is still the yearly naastplaatsing rate, calculated the same way as the one at the neighborhood level; in this step, it is interpreted as a measure of naastplaatsing risk. The core assumption is that if a container is visited for 100 times over a significantly representative period of time, for example a year, and for 75 times there naastplaatsing is found, its likelihood of finding waste next to it for any given visit is $75/100 = 75\%$.

The WasteVision (WV) dataset stores information on containers like capacity (in m^3), type of waste fraction, average filling time, and others. These characteristics are aggregated to describe clusters structurally, and as features can only identify which cluster structures are prone to naastplaatsing. As such, their predictive power is limited. To improve on this, additional features are derived using expert knowledge and validated hypotheses over the available datasets.

The nRn dataset can be used to extract features that capture the naastplaatsing "behavior" of container clusters, going beyond mere structural characterization. For instance, the notion of "trash brings trash", or broken window theory, can be reflected in a variable that tracks the unbroken sequences of days with observed naastplaatsing (more plainly "streaks") for a given container cluster. Weighting naastplaatsing based on the type of waste disposed improperly, or the specific day of the week when it happens, can also provide more context and aid the predictive power of the model.

Expert knowledge also points towards the relationship between naastplaatsing occurrences and the proximity to establishments that influence the waste generation process. Points of interest include supermarkets, restaurants, hotels, and waste disposal facilities. The locations dataset contains the necessary information to determine which clusters are in the vicinity of these points of interest, and the derived predictive features can be used to ascertain the impact of these establishments over naastplaatsing rates. Following the approach used in the OBI report (Epskamp & De Vries, 2020), the radius of each container cluster is set at 100 metres.

This collection of features is used to compute a prediction of the naastplaatsing rate of each container cluster. This task is framed as a supervised regression problem. Given the heterogeneous nature of predictors, the presence of non-linear relationships, and the interaction between features, a range of machine learning models is deployed: a shallow decision tree (DT), a random forest regressor (RF), and a gradient boosting tree regressor (GB).

Decision trees are exploratory, fast models that iteratively partition the feature space with a sequence of if-then rules that minimize prediction error with each split (Gordon, 1984). The construction of the tree begins at its root node, which contains the entire training dataset. The algorithm identifies an optimal split to create the first set of child nodes; this continues through internal intermediate nodes, where the feature space is further divided one variable at a time. For each node, the regressor performs a search over all predictor variables (X) and all possible split points (S) to find the partition that most effectively reduces the error of the model. The final regions of the dataset, over which no

further splitting occurs, are called leaves. The predicted value (Y) for any observation assigned to a specific leaf is the mean of the target value for all the instances that reached it through predictor splits. Decision trees utilize metrics to distinguish leaf nodes based on their "purity"; in classification tasks, which assign prediction instances to classes (for instance, a binary 0/1), a node is considered pure if all the instances in the leaf's partition are of the same class. For regression trees, the purity of a node is calculated by the sum of squared deviations from the mean of that leaf's target value. A node is considered pure if the variance is 0, and mixed if there is a high degree of numerical variance.
This regressor is a highly interpretable, simple model. However, it is also inherently unstable and prone to overfitting. To prevent this, algorithms often apply pruning techniques to reduce the size of the decision tree by preventing the formation of branches, or removing them post-generation. Pruning rules are managed through specific hyperparameters (Probst et al., 2018), settings that determine the algorithm's structure defined before training. In the case of decision trees, the most relevant are:

- max_depth, which constrains the number of levels the tree can reach from the root to the furthest leaves. For lower depths, the decision tree becomes simpler, reducing its exposition to noise. This may reduce performance on the training set, but better generalization should positively impact test performance.
- min_samples_leaf, which ensure that each split or leaf node contains enough training examples to avoid spurious partitions.
- min_samples_split, which specifies the minimum number of samples needed to consider splitting an internal node, to prevents the algorithm from creating branches on very small subsets of the data.

Random forest regressors (Breiman, 2001) are parallel ensemble models consisting of a large number of simple decision trees. The strength of this model lies in the diversity of its components, ensured by two main randomization techniques: bagging and patching. With bagging, each tree is trained on bootstrap (randomly selected) copies of the original data. With patching, the feature splits performed at each node to build different trees are randomized to ensure better generalization.
Predictions are obtained by averaging the output of the individual models. Relevant hyperparameters for this ensemble model are again the ones that regulate the pruning rules of individual trees, like max_depth, min_samples_leaf, and min_samples_split. Furthermore, ensemble specific parameters include:

- n_estimators parameter, which determines how many trees will be a part of the random forest.
- max_features parameter, which sets the amount of predictors covered by each split in a tree, increasing diversity in the forest.

Gradient boosting tree regressors (Friedman, 2001) are sequential ensemble models. Decision trees are trained iteratively, with each model correcting the wrong predictions of its predecessor through its residuals. At each iteration, the model learns approximations of the target function that are increasingly optimized. Key hyperparameters for this model are:

- learning_rate, which controls how much each tree contributes to the ensemble.

- max_iter, which expresses the total amount of boosting stages.

Standard tree model hyperparameters max_depth and min_samples_leaf are also included. GridSearch, a systematic evaluation of model performance across a set of hyperparameter values, is used to optimize these tree regressors. Models are evaluated through both predictive accuracy, measured with loss functions, and stability across validation folds. Cross validation folds can be either composed randomly, providing a better generalization of the performance, or group-based at the neighborhood level to negate information leakage between training and test sets. The latter option is utilized for the evaluation in this step, with a total of 5 folds. The employed performance metrics are coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE). R^2 measures how much the model is able to explain the variance in the target variable, ranging from 0 to 1. This assesses the overall explanatory power of the model. RMSE is the square-root of the average squared difference between predicted and actual values. Squaring before averaging means RMSE penalizes large errors, making it sensitive to sizeable outliers. MAE is the average of the absolute differences between predicted and actual values, thus treating all errors equally.

The models are also evaluated based on how they rank their predictive features. Consistency with the insights derived from domain knowledge and qualitative analysis are considered to be as relevant as predictive accuracy. An additional interpretability method utilized i SHAP (SHapley Additional exPlanations), which decomposes a model's prediction into each feature's contributions based on cooperative game theory. The contributions of all features add up to the model's predicted value for that specific observation.

4.4 Routing Study

The final stage of the research applies the predicted naastplaatsing risk to routing optimization. Container clusters are modeled as nodes in a network, with edges representing travel distances. This routing formulation is conceptually a Traveling Salesman Problem (TSP), where the objective is to determine an efficient sequence of visits that cover a set of locations. There are, however, a few key differences with the general problem.

Crucially, while in regular applications of the TSP all the points of the network are known *a priori*, the naastplaatsing status of each container is unknown at the time of the formulation of each route; instead, only probabilistic risk probabilities are available. Capacity constraints are mostly not relevant: crews can often dispose of naastplaatsing through the on-site containers. The sole exception to this is when the waste is too big for containers, or when these are full or malfunctioning. Even still, when naastplaatsing waste is sizeable the nRn crews can request the intervention of the bigger Inzameling trucks. Finally, while visiting all nodes is an operational goal, the desired outcome is the quickest removal of any waste, in order to reduce the "trash brings trash" effect.

Given these characteristics, greedy heuristics are chosen to construct the routes in this step of the research. This type of algorithm builds solutions step by step by repeatedly choosing the locally best choice according to a predefined condition. The concatenation of local optima constructs a route that is not the absolute best solution, but is still satisfying this problem's performance metrics.

A sample of 12 neighborhoods in which NietRnaast operates the collection of naastplaatsing has been selected for this step of the study. For each of these areas, a graph representing the street network that drivers navigate to reach container clusters is built

with the OSMNX Python library. The "drive-service" network is chosen based on its ability to both faithfully represent the constraints of driving in the city of Rotterdam with a truck, while including connections between nodes that are not available to private vehicles, but are instead used by NietRnaast drivers. Each of the container cluster locations is assigned to the closest network node, and a distance matrix between these "snapped" nodes is computed to obtain the pairwise shortest-path road distances (in metres) between all cluster nodes, so that the routing can operate on a realistic distance matrix rather than simple Euclidean distances between points.
Let R be the set of remaining unvisited clusters, and $\mathrm{d(c_i,c_j)}$ the shortest-path road distance from the distance matrix D between $\mathrm{cluster_i}$ and $\mathrm{cluster_j}$. The following routing strategies are considered:

- next-best distance routing, in which the next cluster to visit is chosen based on the minimum travel distance from the current location.

$$c_{\text{next}} = \arg\min_{c \in R} \ d(\text{current},\ c)$$

- risk-based routing, in which the next cluster to visit is chosen based on the highest predicted naastplaatsing risk. This method creates "front-heavy" routes, with the objective of minimizing the time containers harbor naastplaatsing.

$$c_{\text{next}} = \arg\max_{c \in R} \ r_c$$

- k-neighbor distance-risk routing, in which the next cluster to visit is chosen based on a weighted calculation of spatial distance and predicted risk. This strategy represents a synthesis of the two previous heuristics.

$$c_{\text{next}} = \arg \max_{c \in N_k(\text{current})} \ r_c$$

$$N_k(\text{current}) = \text{KNN}(\text{current};\ R,\ d,\ k)$$

- ratio score distance-risk routing, in which the next cluster is chosen by minimizing a composite score that blends normalized road distance and predicted risk.

$$\text{score}(c) = \alpha \cdot \frac{d(\text{current},\ c)}{\max(d)} \ + \ \beta \cdot (1 - r_c)$$

$$c_{\text{next}} = \arg\min_{c \in R} \ \text{score}(c)$$

- convex hull heuristic, in which the route starts as the clusters sitting on the convex hull of cluster coordinates (i.e.: the smallest convex polygon enclosing them, like a rubber band would for some nails). The algorithm inserts new nodes by finding the shortest increment of the circuit through each convex edge and missing node combinations.

$$\text{route} = H \quad (\text{convex hull})$$

$$\Delta_{k,(a,b)} = d(a,k) + d(k,b) - d(a,b)$$

$$(a_{\text{next}},\ b_{\text{next}}) = \arg \min_{(a,b) \in \text{edges(route)}} \Delta_{k,(a,b)}$$

To compare routes constructed with heuristic algorithms between each other and with the status quo, the actual distribution of naastplaatsing occurrences for specific sets of days had to be reconstructed from the NietRnaast dataset. A distribution is a mapping of container clusters to their naastplaatsing outcome on a specific day. A visual representation of this concept is shown in figure 4.2.

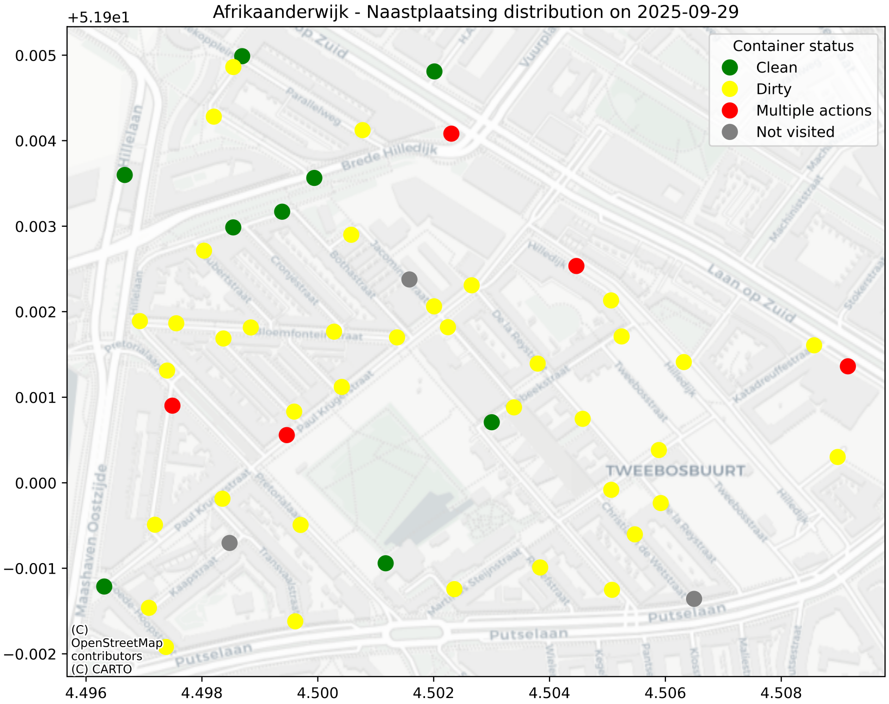


Figure 4.2: Naastplaatsing distribution in Afrikaanderwijk - 29/09/25

The classification of container clusters in the distribution is based on the number and characteristics of the actions recorded for each cluster on a given day. When a cluster is visited exactly once, its status is derived from that single action. A *Niets aangetroffen* action leads to a 0 label, which is visualized as a green dot on the distribution visualization; any other value implies that there is naastplaatsing, for which the label 1 and the color yellow is used. When a cluster appears multiple times in the daily records of the distribution, a few different scenarios are invesigated. If the actions on the same container cluster happen at the same time-stamp, this is interpreted as the crew registering several findings during a single visit. For these instances, the 2 label and the red color is used. If the time-stamps are distinct, the interpretation shifts toward follow-up visits or corrections. If the earlier action is a *Niets aangetroffen* entry, it is disregarded so that the focus remains on relevant events; a ”correction” note is added to the action’s entry, and its label and color are determined by the remaining actions. Finally, containers in the neighborhood that are not visited on the given day are labeled with a -1 and colored in grey on the map.

To construct the “actual” routes, the dataset is filtered based on a specific day and neighborhood. Once the appropriate set of actions is isolated, their registered timestamps are sorted to get the order of visited clusters in the route. This method of route extraction is built on a few constraints and assumptions:

- obtaining the actual routes is only possible for days after May 1st of 2025, since before then the recorded timestamp defaults to 00:00.

- time-stamps are recorded on a 12 hour clock without AM/PM distinction. To accomodate for this limitation, all recorded actions exhibiting time-stamps before 07:00 are shifted of 12 hours.
- actions that display a time difference of more than 60 minutes from the previous minute are considered the starting point of a different roue.

The actual route for a given neighborhood on the selected day is the longest segment of consecutive actions identified in accordance with the previous ground rules. Figure 4.3 demonstrates this process of route extraction in the Afrikanderwijk neighborhood for Monday 29th of September 2025; table 4.4 backs this visualization with the relevant datapoints.

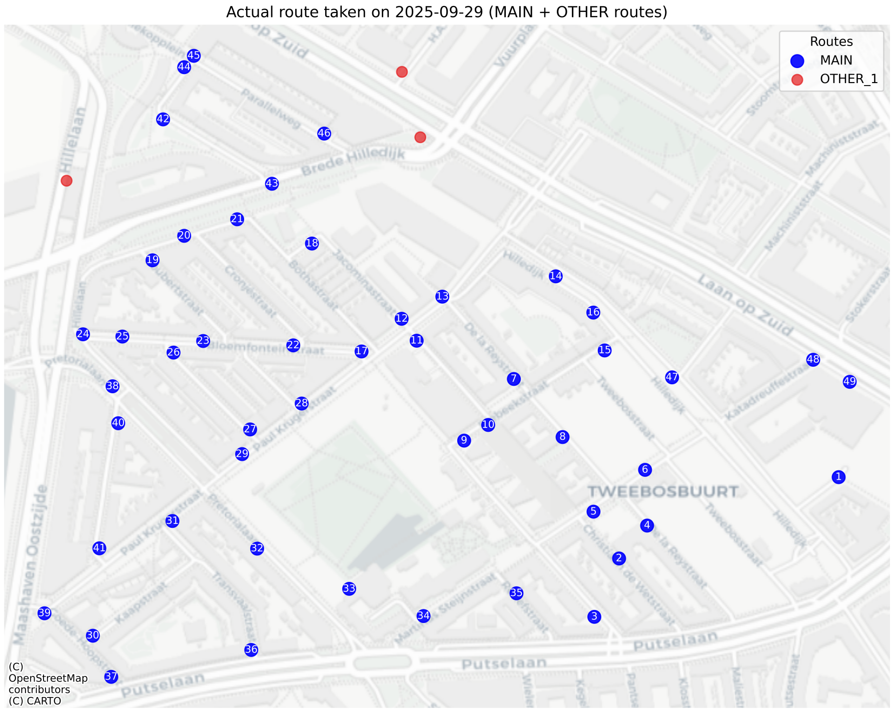


Figure 4.3: Route extracted in Afrikaanderwijk - 29/09/25

Cluster	Buurt	Lon	Lat	Time	Delta Time	Label
14985	Afrikaanderwijk	4.502305	51.904082	09:40	00:00	OTHER_1
14887	Afrikaanderwijk	4.502013	51.904809	09:40	00:00	OTHER_1
13072	Afrikaanderwijk	4.496667	51.903599	09:40	00:00	OTHER_1
14495	Afrikaanderwijk	4.508973	51.900301	17:06	07:26	MAIN
12793	Afrikaanderwijk	4.505473	51.899397	17:12	00:06	MAIN
12808	Afrikaanderwijk	4.505078	51.898748	17:13	00:01	MAIN
12796	Afrikaanderwijk	4.505919	51.899763	17:17	00:04	MAIN
12813	Afrikaanderwijk	4.505065	51.899917	17:17	00:00	MAIN
14640	Afrikaanderwijk	4.505887	51.900383	17:19	00:02	MAIN
			…			

Table 4.1: Extracted route sequence with cleaned time fields.

Only the nodes visited in the extracted distribution are utilized as the set that the heuristic routes need to visit completely, in order to ensure fairness of comparison between all results. Through these tools, the evaluation of the heuristic routes is carried out over the following metrics:

- total distance traveled, serving as a proxy for operational costs and crew workload. This is calculated by adding the distances between consecutive route nodes by locating their relative positions in the distance matrix.
- collection penalty, measuring how long naastplaatsing remained at the containers of the designated area.

The weight of each collection penalty is a product function of the position at which the container with naastplaatsing is visited (penalizing late collection) and the type of action, with heavier scoring for more disruptive types of waste or situations (refer to Appendix A.5). A full container, due to its overflow implications, carries a penalty of 20, while the more common and manageable "Klein afval" occurrences only score 1.

The performance metrics described here directly reflect the priorities expressed by the stakeholders involved in the naastplaatsing response process, as detailed in section 3.3.2. The total distance traveled captures the operational concerns of actors closest to the collection process, who are responsible for the daily feasibility and efficiency of collection routes. At the same time, the collection penalty reflects the impact of behavioral dynamics that planners and managers emphasize, and that Rotterdam residents inherently care about. The synthesis of these metrics is also a representation of the dilemma that high level management faces when dealing with naastplaatsing, where cost efficiency and quality of delivered service are difficult to balance out. Routes that produce a satisfying score in both of the metrics are expected to be more appreciated by stakeholders that need to balance both of the main requirements that emerged from the analysis of the decision arena of naastplaatsing.

4.4.1 Toy Problem

To provide a visual demonstration of how each heuristics functions, a toy problem is presented here. This consists in the random generation of a small graph, with nodes representing container clusters and edges representing the road network connections between them. Each node is given a random risk score, and each edge has a precomputed weight representing the distance between the nodes it connects. These weights are used to create a distance matrix, which is instrumental to determine the network distances between each pair of nodes; to simulate the collection penalty, a simplified version of a naastplaatsing distribution is randomly generated. This distributions is done with a binary value of 0 or 1 for containers based on whether they display naastplaatsing; the collection penalty is calculated with the same step multiplication process as the real routing study. Figure 4.4 shows the toy graph with a color grading on each node representing its risk. A full breakdown of the risk per toy node is presented in table in Appendix A.4.

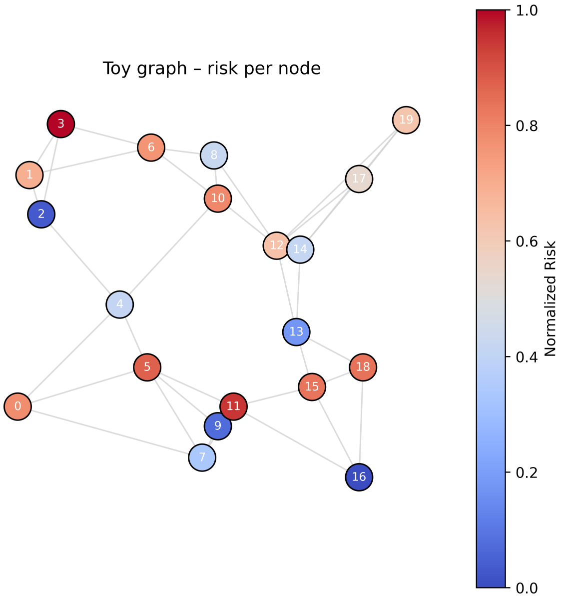


Figure 4.4: Toy node with risk score color grading

All routes start from the randomly chosen node 2. The network paths used in each route are colored in full black; multiple black lines indicate multiple usage of the same edge, which is consistent with the routing of a street network. Steps in the routes are numbered; the source node is displayed as a green dot, and the ending point as a blue square.

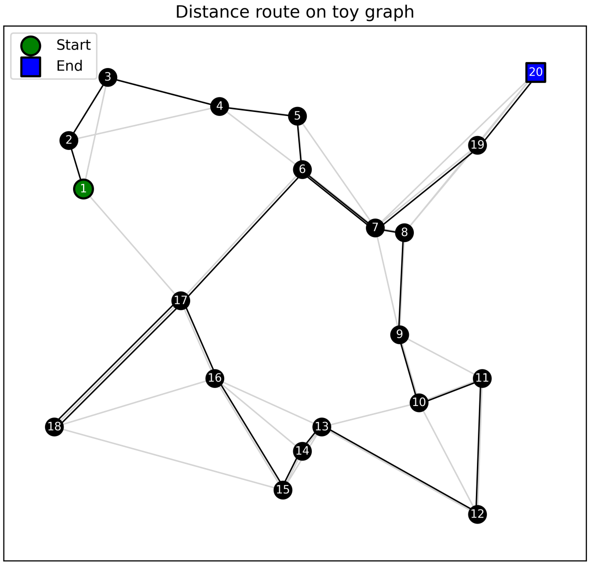


Figure 4.5: Next-best distance route on toy graph

The representation of the next-best distance route in figure 4.5 shows how this heuristic can produce satisfying results, yet still not optimize its main strength. At step 6, the algorithm chooses to go towards node 12; this choice proves to be costly towards the end of the route, where at step 17 from node 4 the only choice is to go back and forth from node 0. This useless detour could have been avoided if at step 6 the algorithm had chosen to go towards node 4 first; however, the next-best distance logic forced it to make a miopic decision, which ended up causing inefficiencies in the long run. Still, with the exception of three edges (0-4, 4-10, 10-12), the routing is quite efficient when it comes to total distance covered.

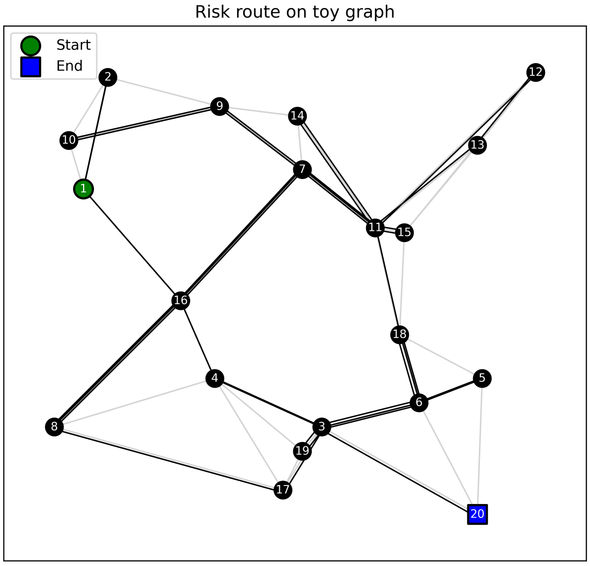


Figure 4.6: Next-best risk route on toy graph

The next-best risk heuristic produces a route (figure 4.6) where almost every edge is used more than once, indicating low distance efficiency. A simple inspection of the route node order reveals the reason for this; after the fixed start of node 2, the risks of the following nodes (3, 11, 5) are 0.975, 0.926, and 0.858. This approach is likely to produce the lowest collection penalty, but the resulting route representation is evidently inefficient when it comes to distance covered.

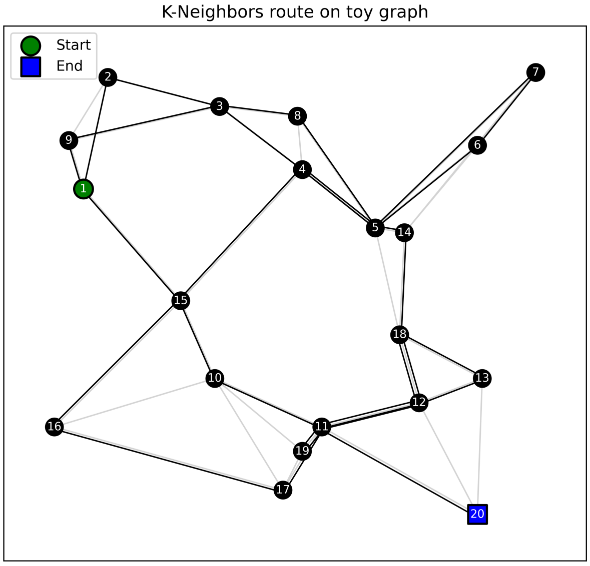


Figure 4.7: K-neighbors distance-risk route on toy graph

The k-neighbor distance-risk algorithm represented in figure 4.7 chooses the 3 ($k = \sqrt{(n = 20)/2}$) closest neighbors by distance, then picks the one with the highest risk among those. At the start of the route, the k neighbor heuristic finds itself on node 2, and its neighbors are 1, 3, and 4. Among those, node 1 has a risk of 0.697, node 3 has a risk of 0.975, and node 4 has a risk of 0.438; as a result, it chooses node 3 as the next node in the route. After this, the neighbors of node 3 are 1 (again), 6 (risk = 0.761), and 8 (risk = 0.450). Node 1 is slightly closer to 3, as it was closer to 2 as well; yet both times it is ignored, and visited at a later step (9) for a costly detour.

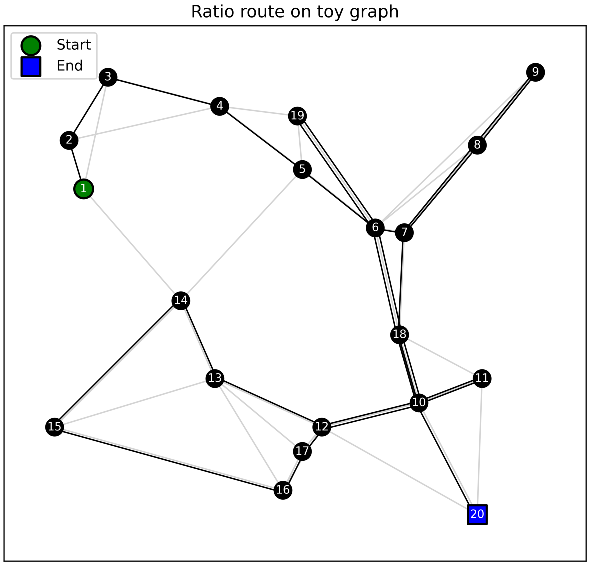


Figure 4.8: Risk-distance ratio route on toy graph

The ratio heuristic (figure 4.8) avoids this pitfall by weighing risk and distance together. This is why while having the same starting point as the k neighbor heuristic, it chooses node 1 as the next node instead of 3, because even though 3 has a higher risk, it is much farther away than 1, and the ratio of risk to distance is better for node 1. Still, the ratio heuristic is not completely agnostic to risk: at the 11th step of its route, it can choose between node 16 (risk 0.06) and node 11 (risk 0.92). Even though 16 is much closer to the current node than 11, the ratio heuristic chooses 11 because its risk is so much higher, and the ratio of risk to distance is better for 11 than for 16. The more simple distance heuristic (figure 4.5), in this position, chooses node 16 regardless, incurring in a higher penalty over node 11 since it is visited later.

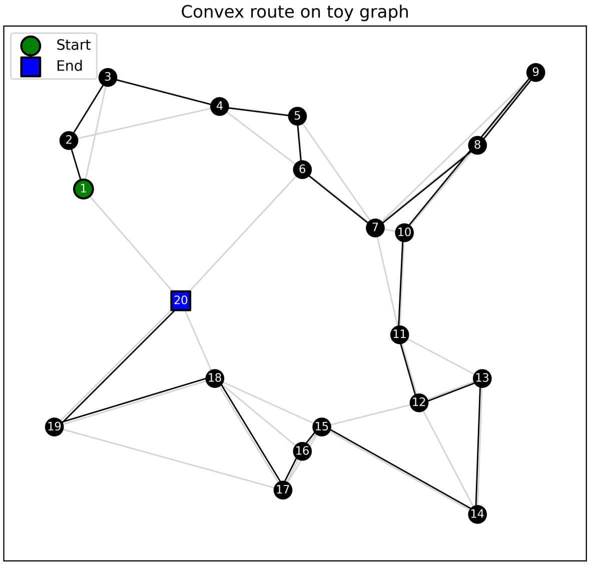


Figure 4.9: Convex insertion route on toy graph

The convex insertion heuristic produces the best-looking route (figure 4.9) distance-wise: each edges is used only once, indicating a high level of distance efficiency. The shape of the route itself also clearly resembles the convex hull of the points, demonstrating the starting point from which the logic of inserting the least expensive edges for each internal node is applied.
This toy formulation also allowed to calculate the score of each route based on the two performance criteria described in section 4.4. The results for the covered distance are detailed by figure 4.10, while the collection penalty comparison is in figure 4.11.

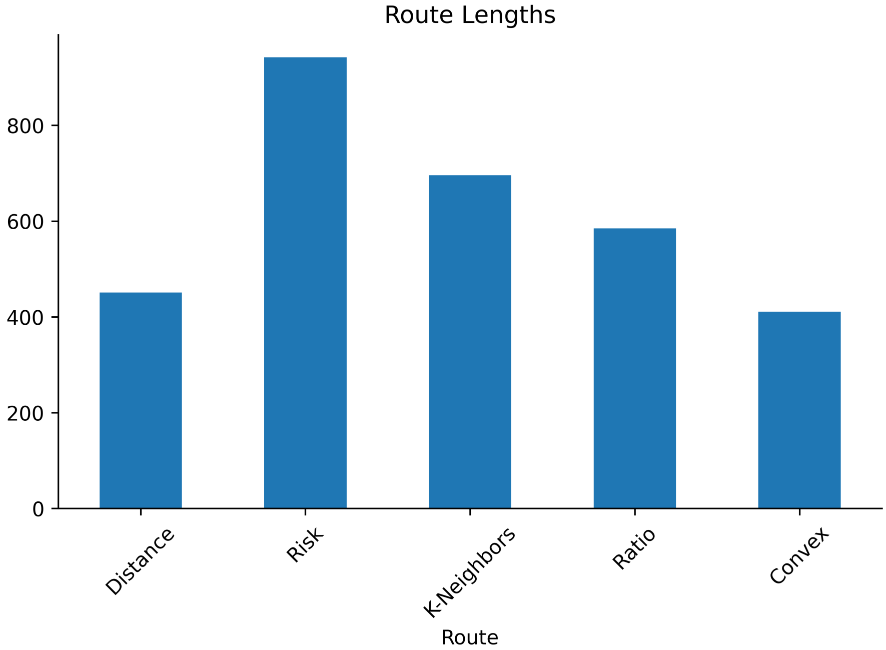


Figure 4.10: Distance covered by each route in the toy formulation

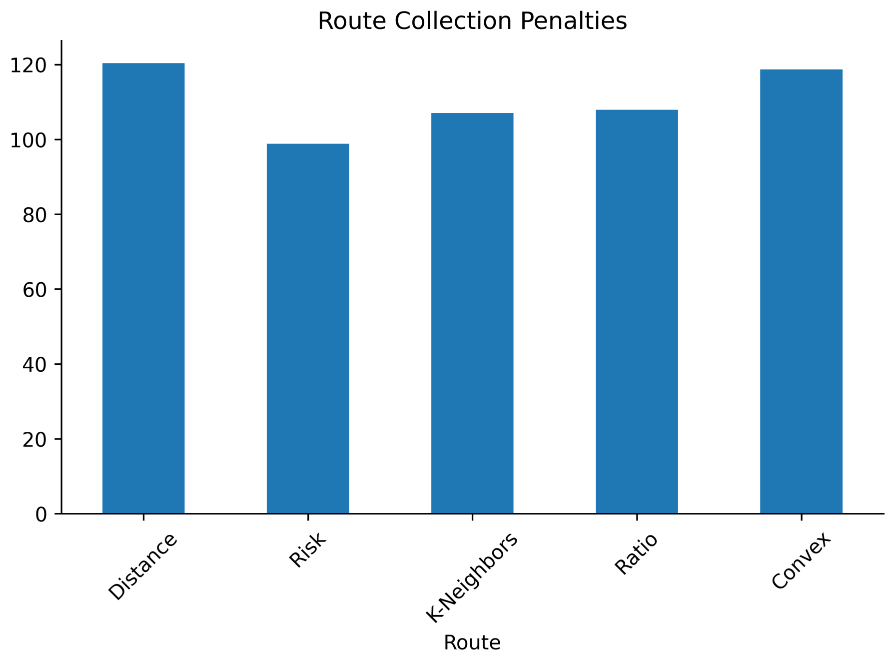


Figure 4.11: Collection penalties accumulated in each route of the toy formulation

The two distance-based heuristics, next-best distance and convex insertion, display the best performance in terms of total distance covered. The convex insertion heuristic is slightly better than next-best distance. However, their collection penalties are the highest among the heuristics, highlighting the fact these are risk-agnostic algorithms. On the other hand, the risk heuristic scores the lowest collection penalty; however, its total distance is more than double what the two distance-based algorithms managed, making it operationally infeasible. The k-neighbors risk-distance and risk-distance ratio heuristics produce the most balanced results across both performance metrics. They do not excel in any of these, but in particular the ratio route presents itself like a potentially satisfying compromise between the two metrics.

Chapter 5

Results

5.1 Overview

The following chapter details the results of each step of the analysis carried out for this study. The first section focuses on the outcomes of the factor analysis. At this stage, the variables evaluated are at the neighborhood level of granularity. The goal is to determine whether the selected socio-demographic, environmental and operational features that characterize the neighborhoods of Rotterdam display a statistical correlation with their naastplaatsing rate.

The prediction outputs of the Machine Learning (ML) models are subsequently examined. The analysis of this step is complimentary to the previous one, since it is carried out at a finer level of aggregation: container clusters. At this level, it is possible to capture the characteristics of naastplaatsing and NietRnaast operations with greater precision, and to make further analysis on the composition of container clusters when it comes to waste fractions, capacity, and types of containers. The NietRnaast dataset can be fully leveraged here, providing insights on the historical incidence of naastplaatsing on different containers: how often it happens, how often the waste next to containers is problematic, how frequently it happens consecutively, and how many times it happens outside of the usually busier days of the post-weekend surge.

Finally, the performance of the routes produced by the heuristics is reviewed over the formulated metrics. The routing study serves as a proof of concept linking analysis and prediction to real world impact: the results of the ML section are operationalized at this stage, with three of the heuristics utilizing predicted risks scores as part of their logic to choose the next container cluster to visit on their path.

5.2 Factor Analysis

Data availability partly shapes the structure of the qualitative analysis. The NietRnaast dataset become analytically relevant only from 2022 onwards, since in 2021 their data collection process was in its infancy. On the other hand, neighborhood level information on socio-demographic indicators from the CBS and Onderzoek dataset is completely up to date only up until 2023. This allows the study to utilize the available information to analyze two separate OLS regression models, one based on the data from 2022 one from 2023; the comparison of these two models sets out to identify the variables that consistently predict naastplaatsing rates across varying data conditions.

5.2.1 2022 model

An initial OLS regression was estimated using the full set of predictive variables selected from the relevant datasets (Appendix A.2). The model achieved a high R^2 score (0.83), indicating that high explanatory power of the selected predictors over the variance of the target, naastplaatsing rates. However, this implementation suffered from severe multicollinearity, reflected by a high condition number of 1545, as well as extreme VIF (Variance Inflation Factors) values among closely related variables. With 71 observations, the number of analyzed neighborhoods, and 25 predictors, the model faced an unfavorable observations-features ratio, which resulted in unstable coefficients and inflated standard errors. Analysing the correlation heatmap for the features selected for the 2022 model, it is evident that these are not linearly independent, a core assumption of OLS regression.

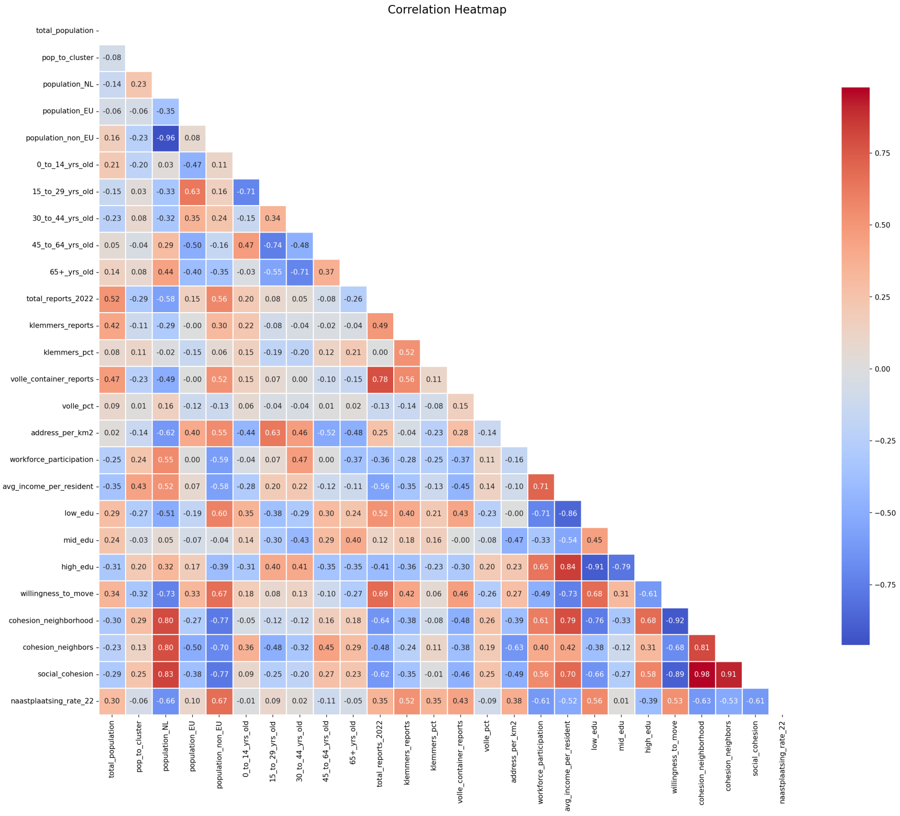


Figure 5.1: 2022 Correlation Heatmap

To improve this model's results, a reduction of the amount of predictive features is required, even at the cost of a loss in explanatory power. This subset is formulated with the goal to retain one feature per each of the hypotheses evaluated at this stage, guided by the qualitative synthesis of the domain knowledge section. More specifically, hypotheses 3 and 4 (Container cluster composition and Temporal stress) are evaluated at the Predictive Modeling step. The retained features are detailed in table 5.2.1.

Variable Name *Description*	**Hypothesis**
address_per_km2 Proxy of the pressure on container clusters	H1
klemmers_reports Reports of containers being blocked or malfunctioning	H2
volle_container_reports Reports of full containers	H2
nrn_visits_2022 Proxy for the operational focus of Stadsbeheer in specific areas during 2022	H5
social_cohesion Proxy of social cohesion, derived with PCA over the 2022 Wijk Profiel answers	H6
avg_income_per_resident Indicator of economic conditions	H7
population_non_EU Share of residents originating from outside the EU	H8
15_to_29_yrs_old Share of teenagers and young adults	H8

Table 5.1: Factor Analysis Variables

The klemmers_reports and volle_container_reports variables provide information over the operational, concrete barriers that contribute to causing naastplaatsing. The social_cohesion variable is the first principal component extracted from a PCA of the questions asked about neighborhood living in the Wijk Profiel of 2022. It alone captures over 80% of the variance of that datasets scores, so it is deemed sufficient to cover this topic. Individuals not fully integrated with the neighborhood and community are often associated with naastplaatsing, thus population_non_EU is the most relevant share to keep for this analysis. This is also on the grounds of these individuals likely lacking knowledge of either Dutch or English languages. The 15 to 29 years old age bracket includes teenagers and young adults, who are the overwhelming majority of high-school and university students. These groups have often been associated with naastplaatsing, for instance when discussing ”candy trails” between schools and supermarkets, or in the context of frequent moving in the case of university accomodation.
The VIF scores of this new set of variables are significantly improved, with only social cohesion scoring barely above 4. The R^2 score is now a bit lower than the full-feature model (from 0.83 to 0.66); however, the condition number is a much improved 5.5, meaning that there is no problematic multicollinearity among variables.

The variables that hold the most significance (p-value < 0.05) are the following:

- **klemmers_reports**: positive coefficient ($+0.431$, $p < 0.001$)

  Containers experiencing more opening malfunctions are associated with higher naastplaatsing rates, consistent with the "impediments" hypothesis.

- **population_non_EU**: positive coefficient ($+0.391$, $p = 0.005$)

  Neighborhoods with larger shares of non-EU residents show higher naastplaatsing rates, aligning with qualitative accounts about transient populations, language barriers, and weaker neighborhood attachment.

- **total_reports_2022**: negative coefficient (-0.429, $p = 0.0028$)

  Areas with more reporting activity show lower naastplaatsing rates, reflecting greater operational attention/cleanup responsiveness and/or higher civic engagement (reporting as a protective factor).

Other variables record borderline effects, with statistical significance between 0.05 and 0.12: avg_income_per_resident has a negative coefficient (-0.231), while address_per_km2 has a positive one (+0.204). Both results are in line with qualitative findings. Finally, the variables deemed to be not statistically significant ($p > 0.20$) are social_cohesion, volle_container_reports, and 15_to_29_yrs_old.
Figure 5.2 provides a visual representation of the coefficients of the variables tested on naastplaatsing rates.

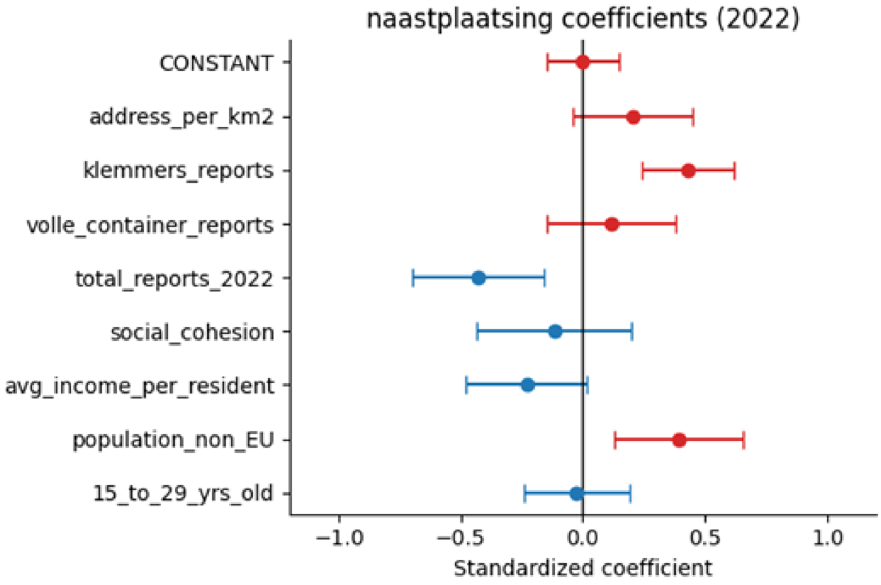
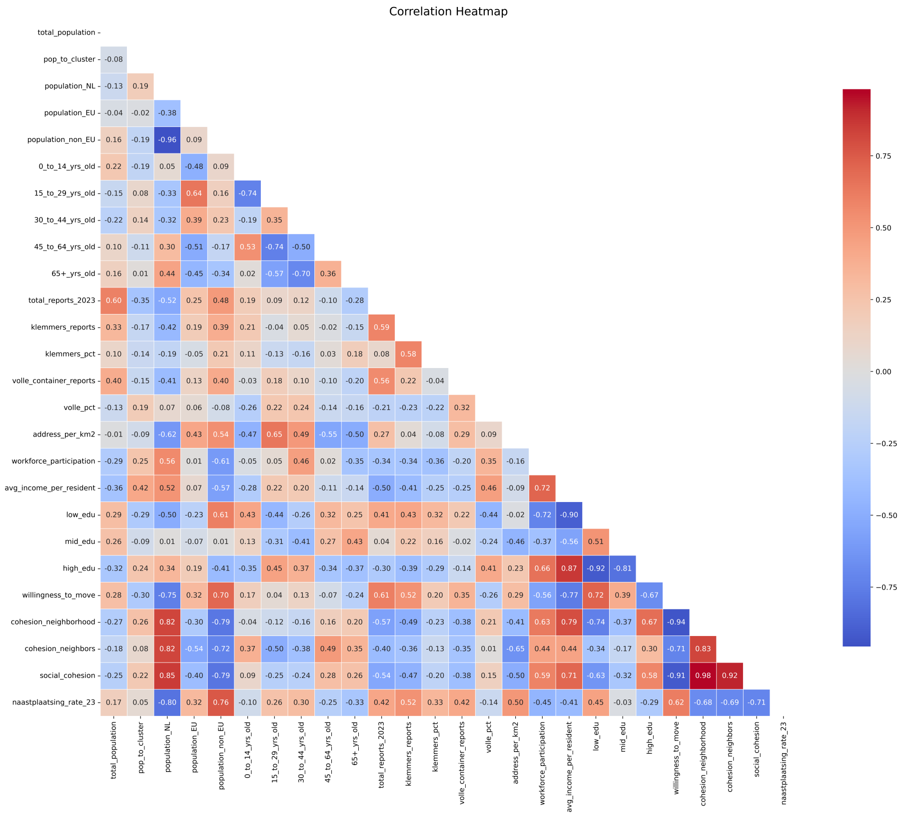


Figure 5.2: Results of the OLS regression for 2022 data with minimal predictors



	total_population	pop_to_cluster	population_NL	population_EU	population_non_EU	0_to_14_yrs_old	15_to_29_yrs_old	30_to_44_yrs_old	45_to_64_yrs_old	65+_yrs_old	total_reports_2023	klemmers_reports	klemmers_pct	volle_container_reports	volle_pct	address_per_km2	workforce_participation	avg_income_per_resident	low_edu	mid_edu	high_edu	willingness_to_move	cohesion_neighborhood	cohesion_neighbors	social_cohesion	naastplaatsing_rate_23
total_population																										
pop_to_cluster	-0.08																									
population_NL	-0.13	0.19																								
population_EU	-0.04	-0.02	-0.38																							
population_non_EU	0.16	-0.19	-0.96	0.09																						
0_to_14_yrs_old	0.22	-0.19	0.05	-0.48	0.09																					
15_to_29_yrs_old	-0.15	0.08	-0.33	0.64	0.16	-0.74																				
30_to_44_yrs_old	-0.22	0.14	-0.32	0.39	0.23	-0.19	0.35																			
45_to_64_yrs_old	0.10	-0.11	0.30	-0.51	-0.17	0.53	-0.74	-0.50																		
65+_yrs_old	0.16	0.01	0.44	-0.45	-0.34	0.02	-0.57	-0.70	0.36																	
total_reports_2023	0.60	-0.35	-0.52	0.25	0.48	0.19	0.09	0.12	-0.10	-0.28																
klemmers_reports	0.33	-0.17	-0.42	0.19	0.39	0.21	-0.04	0.05	-0.02	-0.15	0.59															
klemmers_pct	0.10	-0.14	-0.19	-0.05	0.21	0.11	-0.13	-0.16	0.03	0.18	0.08	0.58														
volle_container_reports	0.40	-0.15	-0.41	0.13	0.40	-0.03	0.18	0.10	-0.10	-0.20	0.56	0.22	-0.04													
volle_pct	-0.13	0.19	0.07	0.06	-0.08	-0.26	0.22	0.24	-0.14	-0.16	-0.21	-0.23	-0.22	0.32												
address_per_km2	-0.01	-0.09	-0.62	0.43	0.54	-0.47	0.65	0.49	-0.55	-0.50	0.27	0.04	-0.08	0.29	0.09											
workforce_participation	-0.29	0.25	0.56	0.01	-0.61	-0.05	0.05	0.46	0.02	-0.35	-0.34	-0.34	-0.36	-0.20	0.35	-0.16										
avg_income_per_resident	-0.36	0.42	0.52	0.07	-0.57	-0.28	0.22	0.20	-0.11	-0.14	-0.50	-0.41	-0.25	-0.25	0.46	-0.09	0.72									
low_edu	0.29	-0.29	-0.50	-0.23	0.61	0.43	-0.44	-0.26	0.32	0.25	0.41	0.43	0.32	0.22	-0.44	-0.02	-0.72	-0.90								
mid_edu	0.26	-0.09	0.01	-0.07	0.01	0.13	-0.31	-0.41	0.27	0.43	0.04	0.22	0.16	-0.02	-0.24	-0.46	-0.37	-0.56	0.51							
high_edu	-0.32	0.24	0.34	0.19	-0.41	-0.35	0.45	0.37	-0.34	-0.37	-0.30	-0.39	-0.29	-0.14	0.41	0.23	0.66	0.87	-0.92	-0.81						
willingness_to_move	0.28	-0.30	-0.75	0.32	0.70	0.17	0.04	0.13	-0.07	-0.24	0.61	0.52	0.20	0.35	-0.26	0.29	-0.56	-0.77	0.72	0.39	-0.67					
cohesion_neighborhood	-0.27	0.26	0.82	-0.30	-0.79	-0.04	-0.12	-0.16	0.16	0.20	-0.57	-0.49	-0.23	-0.38	0.21	-0.41	0.63	0.79	-0.74	-0.37	0.67	-0.94				
cohesion_neighbors	-0.18	0.08	0.82	-0.54	-0.72	0.37	-0.50	-0.38	0.49	0.35	-0.40	-0.36	-0.13	-0.35	0.01	-0.65	0.44	0.44	-0.34	-0.17	0.30	-0.71	0.83			
social_cohesion	-0.25	0.22	0.85	-0.40	-0.79	0.09	-0.25	-0.24	0.28	0.26	-0.54	-0.47	-0.20	-0.38	0.15	-0.50	0.59	0.71	-0.63	-0.32	0.58	-0.91	0.98	0.92		
naastplaatsing_rate_23	0.17	0.05	-0.80	0.32	0.76	-0.10	0.26	0.30	-0.25	-0.33	0.42	0.52	0.33	0.42	-0.14	0.50	-0.45	-0.41	0.45	-0.03	-0.29	0.62	-0.68	-0.69	-0.71	

Figure 5.3: 2023 Correlation Heatmap

5.2.2 2023 model

The analysis of 2023 started with fitting a model with a full set of predictors; once again, the results were a high R^2 score (0.83), and high condition number (1623), attesting to multicollinearity problems. Figure 5.3 shows the correlation heatmap of the full feature set for the 2023 naastplaatsing rate.
The same set of minimal variables per hypothesis was utilized to reduce the model, which produced the following set of statistically relevant predictors:

- **klemmers_reports**: positive coefficient (+0.364, $p = 0.00027$)
- **population_non_EU**: positive coefficient (+0.459, $p = 0.00054$)

Other variables record borderline effects, with statistical significance between 0.05 and 0.2: volle_containers_reports has a positive coefficient (+0.168) and total_reports_2023 has a negative coefficient (-0.214), and social cohesion has a negative coefficient (-0.226). Finally, addres_per_km2, ave_income_per_per_resident, and 15_to_29_yrs_old have a p value of above 0.3, making them not statistically significant.
Figure 5.4 provides a visual representation of the coefficients of the variables tested on naastplaatsing rates.

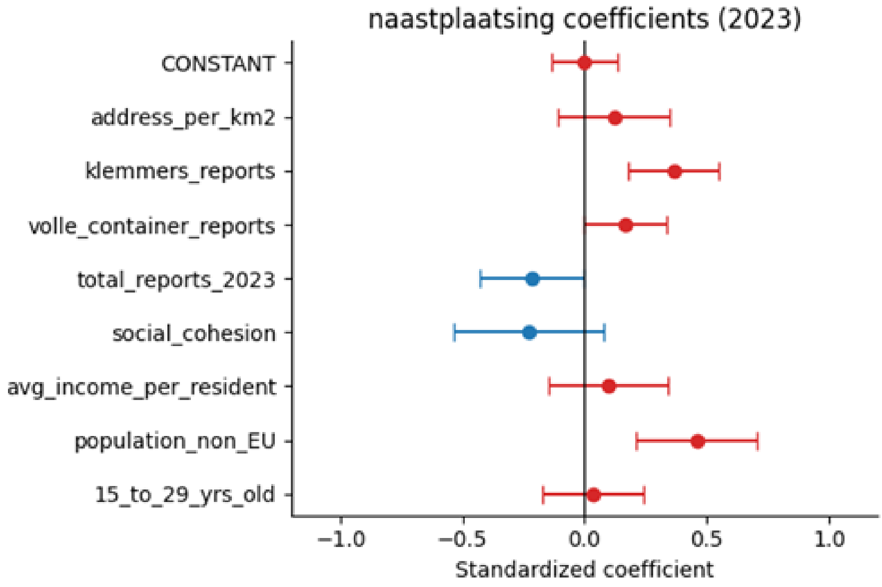


Figure 5.4: Results of the OLS regression for 2023 data with minimal predictors

5.2.3 Model Comparison

Container malfunctions and origin composition of resident population are the two signals that are consistently significant and positive across both 2022 and 2023 models. The direction of their coefficient is consistent with the findings of the domain knowledge section, which associate their increase with a similar trajectory of the target variable. These results point towards a (partial) confirmation of hypothesis 2, and a stronger confirmation of hypothesis 8.

The overflow variable starts out as non-significant in the 2022 model, shifting to quasi relevant in the 2023 model with a p value of 0.05. The quantitative results converge to an agreement with the qualitative insights gained from expert knowledge and strengthening the statement of hypothesis 2. Reporting volume, on the other hand, seems to disprove the idea that the efficiency of NietRnaast (proxied by the frequency of its visits) is tied to more naastplaatsing, since the coefficient for this variable is consistently negative and statistically significant.

The address_per_km2 variable does not hold statistical significance across both models, and its coefficient is not substantial. This would partially disprove hypothesis 1, although perhaps pressure from population level may be captured more directly by the malfunctions/overflow variables.

Social cohesion remains negatively correlated with naastplaatsing rates in both models, partially validating the statement of hypothesis 6. Its lesser statistical significance is perhaps tied to its linear dependence to other variables included in the model (namely prosperity and population composition), given that it retains the highest VIF value across both models.

The avg_income_per_resident and 15_to_29_yrs_old remain mostly statistically inconsequential in both models, pointing to either a disproval of hypotheses 7 and 8 or to the inadequacy of these proxies to represent the relationship with the target variable.

5.3 Predictive Modeling

After analysing which naastplaatsing-related characteristics of neighborhoods interact most significantly with the phenomenon, the level of granularity is reduced to the container cluster level. In this step, three tree based regression models are trained to predict the normalized naastplaatsing rate at the container cluster level over the 2025 NietRnaast actions dataset. The variables used at this stage are divided into the following categories, with a full breakdown in Appendix A.3:

- malfunctions/overflow: rates of malfunction/overflow rates, fill time of the fastest container in the cluster.
- operational involvement: total NietRnaast visits in 2025, type of collection (scheduled vs fixed), presence of container adopter.
- cluster composition: counts of waste fraction containers in the cluster, counts of container types in the cluster, capacity of containers in the cluster.
- business pressure and location access: amount of hotels, restaurants, cafes in a 100 meter radius of the container cluster, distance to the closest public waste disposal facility.
- temporal patterns: average and max streak of interventions that found trash next to the containers, the rate of problematic waste deposited next to containers, and the rate of naastplaatsing happening on days that are outside of the post-weekend surge.

For the results of the Hyperparameter tuning, refer to Appendix B.

5.3.1 Tree Models

The evaluated models are a (Shallow) Decision Tree, a Random Forest, and a Gradient Boosting regressor. Table 5.3.1 lists the performance of each of these models across the performance metrics described in section 4.3.

Model	**RMSE**	**MAE**	**R^2**
GradientBoosting	0.067	0.047	0.949
RandomForest	0.068	0.045	0.948
ShallowTree	0.080	0.053	0.928

Table 5.2: Model Performance

All models demonstrate strong predictive performance across the selected evaluation metrics, indicating that each algorithm is capable of capturing the underlying relationships between the target variable and features with low error. Among the three, the Gradient Boosting regressor performs best overall, achieving the highest R^2 value (0.949) and the lowest RMSE (0.067). and MAE (0.047). This suggests that it not only models the greatest proportion of variance in the data but also produces the smallest prediction errors in terms of average magnitude (MAE) and error sensitivity to larger deviations (RMSE). The Random Forest model follows closely with comparable error levels and only a marginally lower R^2 (0.948), indicating similarly robust performance. It edges out the Gradient Boosting model in sensitivity to outliers, as shown by its superior MAE score (0.045). The shallow decision tree yields reasonably good results (R^2 = 0.928), yet still noticeably worse than the ensemble-based models, which is expected given its limitations. Overall, the results reinforce that tree models, and in particular ensemble methods, are well suited for this predictive task.

Model	**Feature**	**Importance**
Shallow Tree	avg_streak	0.2511
	bad_waste_count	0.0269
	nietrnaast_visits_2025	0.0121
	klemmers_rate_2025	0.0035
	bad_naastplaatsing_day_share	0.0034
Random Forest	avg_streak	0.2272
	bad_waste_count	0.0306
	nietrnaast_visits_2025	0.0169
	bad_naastplaatsing_day_share	0.0088
	fastest_fill_time	0.0059
HistGB	avg_streak	0.2137
	bad_waste_count	0.0333
	nietrnaast_visits_2025	0.0237
	bad_naastplaatsing_day_share	0.0143
	klemmers_rate_2025	0.0129

Table 5.3: Top five features by SHAP importance for each tree-based model.

SHAP (SHapley Additive exPlanations) feature importance quantifies how much each input variable contributes to the model's predictions by computing Shapley values. This approach is based on cooperative game theory, evaluating all possible combinations of features to determine their marginal contribution; the additive result is the actual prediction produced by the model. This method provides a consistent, model-agnostic explanation of feature influence by decomposing each prediction into these additive contributions. Across the three machine learning models evaluated, the SHAP analysis reveals a stable set of dominant predictors: the top five features consistently include *avg_streak, bad_waste_count, nietrnaast_visits_2025, bad_naastplaatsing_day_share, klemmers_rate_2025* and *fastest_fill_speed*, although their presence and/or relative ranking varies slightly between models. Gradient Boosting and Random Forest show similar importance profiles, with features other than the *avg_streak* scoring more substantially than the shallow decision tree, which exhibits a more concentrated dependence on the strongest predictor. This shows how the ensemble models produce more nuanced results. For each of the three models, a SHAP summary plot is created. This plot provides an overview of how all features contribute to the model's predictions across the dataset. Each point represents a single observation, with its position on the x-axis showing the SHAP value: how much that feature increases or decreases the prediction output. Wider positions on the x-axis indicate significant contribution. Features are ordered vertically by overall importance, so the most influential variables appear at the top. The blue-red colour gradient indicates the original feature value, with blue indicating a low value and red a high one: this helps in identifying patterns such as whether high or low values of a feature drive predictions upward or downward. In combination, the vertical thickness of the variable observations shows the quantity of datapoints that display a certain prediction influence. This can help determine whether their effects are consistent or highly variable across observations.

Figures 5.5 and 5.6 show the summary plots for each of the ensemble models.

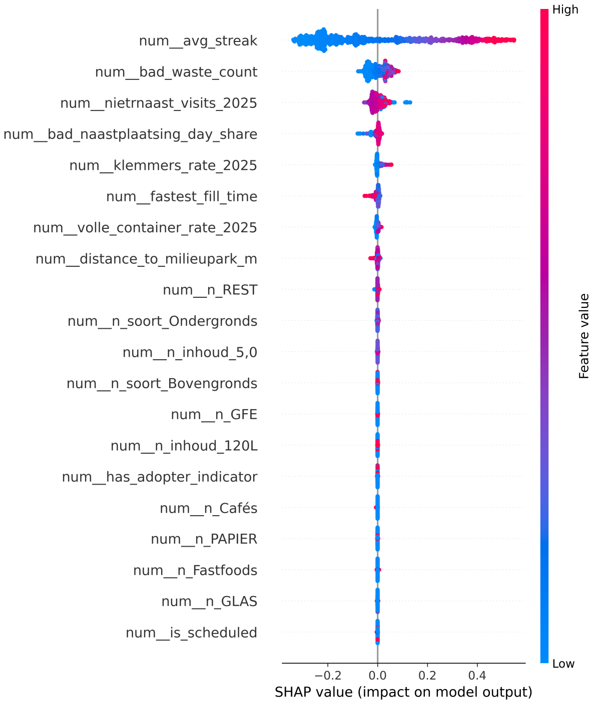


Figure 5.5: SHAP summary plot for Random Forest model

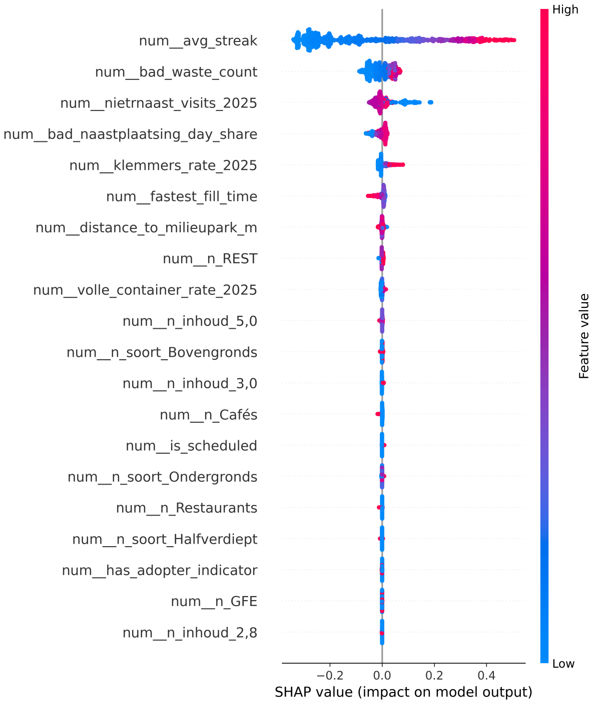


Figure 5.6: SHAP summary plot for Gradient Boosting model

The *avg_streak* variable dominates prediction, with a wide spread of red-positive and blue-negative impacts. Clusters that experience consecutive and/or frequent naastplaatsing occurrences score a higher naastplaatsing risk, meaning they are more likely to experience it again; viceversa, clusters with low streak scores tend to be less exposed to naastplaatsing. This result is consistent with the feedback loop of the "trash brings trash" effect described by literature and experts. The *avg_streak* variables represent the historic behavior of naastplaatsing on container clusters, which indicates that the history of specific container holds significant importance over its likelihood of showing naastplaatsing. The relevance of the *bad_waste_count* suggests that the type of naastplaatsing waste is strongly correlated to the risk. The constant presence of the *klemmers_rate* feature further solidifies the relevance of these incident reports; conversely, the absence of *volle_container_rate* also reinforces the comparative relevance of these two features, already observed with the results of the OLS regression discussed in section 5.2.3. The variable capturing operational attention, *nietrnaast_visits_2025*, show with its red-negative blue-positive impacts that more visits do contribute to reducing naastplaatsing rates. The temporal aspect of naastplaatsing is underscored by the result of *bad_naastplaatsing_day_share*: containers that experience naastplaatsing in a greater quantity even beyond the post-weekend surge days are overall more exposed to risk. Finally, the singular contribution of *fastest_fill_speed* in the Random Forest results can be interpreted as follows: very high (bright red) values of fastest fill speed of a container in the cluster, meaning that the fastest container is quite slow compared to the average, reduce the risk of naastplaatsing. This confirms the theoretical framework's conjecture that container with higher fill times, dealing with a lower volume of generated waste, are less exposed to naastplaatsing.seems to indicates that high-pressure containers, which have to deal with higher volumes of waste, are more exposed to naastplaatsing. Overall, the SHAP results provide interpretable evidence that these key variables drive model performance across all approaches.
The summary SHAP plot of the shallow Decision Tree regressor can be found in Appendix C.

5.3.2 Individual SHAP visualizations

SHAP waterfall plots provide a further understanding of the individual impact of different features on the final predicted target. A few examples individual datapoints from the Gradient Boosting regressor model are discussed in the following.
For the instance depicted in figure 5.7, the model assigns an extremely high probability of naastplaatsing. This is almost entirely driven by the *avg_streak* variable, which shows a significant 2.4 value; features like *klemmers_rate* and *bad_waste_count* also have very positive values, yet their effect isn't nearly as impactful. This data point is likely one of the red bright dots on the far right of the *avg_streak* row in figure 5.6. The contribution of all the other features is minimal.

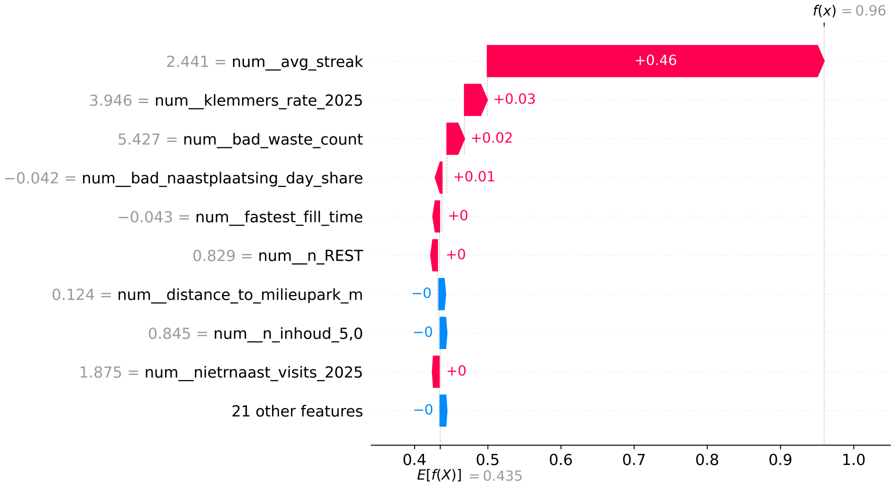


Figure 5.7: SHAP waterfall plot of a high-risk cluster

Figure 5.8 displays a low risk cluster, with a predicted score of 0.073. Similarly to the previous case, the *avg_streak* feature contributes to most of the prediction. In this instance the effect on the prediction output is the opposite, given the low value of the variable (-0.516); exceptionally low values for *bad_naastplaatsing_day_share* and *bad_waste_count* also contribute to reduce the model's output. The only positive push comes from the extraordinarily small amount of NietRnaast visits in 2025, which increases the overall risk of the cluster. This is conceptually sound, considering the idea that clusters with less operational attention can deteriorate in unexpected ways; still, the contribution is not significant to sway the final prediction meaningfully.

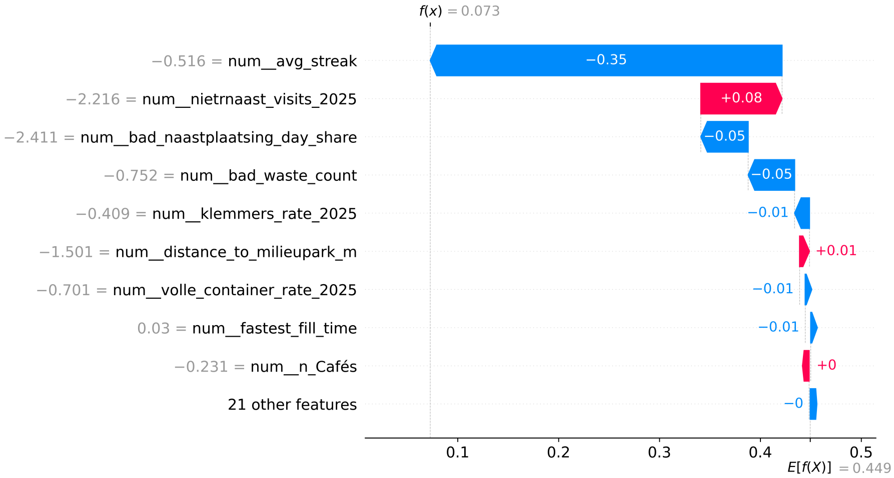


Figure 5.8: SHAP waterfall plot of a low-risk cluster

The prediction of a mid-risk cluster in figure 5.9 is less of a foregone conclusion. Here the contribution of the *avg_streak* is only ranked fourth in absolute terms, pushing the effect slightly above even if its value is actually negative. It is likely that this value is not negative *enough* for the model to use it to lower the prediction output. The three features that influence this cluster's prediction are still among the most relevant ones. The extremely high rate of *klemmers* reports increases the risk prediction, which is however reduced again by a low value of the *bad_waste_count* feature. The lack of recorded naastplaatsing incidents in the later days of the week, attested by the low value of the *bad_naastplaatsing_day_share* feature, also contributes to dropping the prediction output.

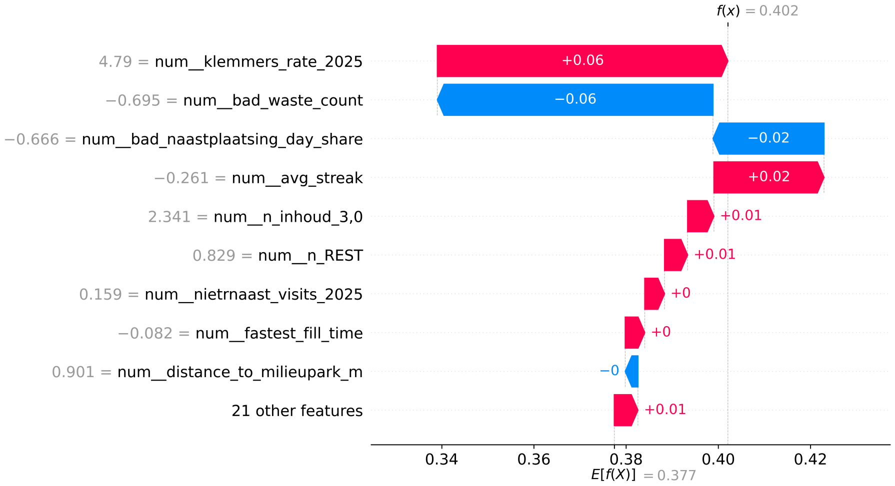


Figure 5.9: SHAP waterfall plot of a medium-risk cluster

5.4 Routing Study

The routing study represents the operationalization of the predictions computed by the ML models in the previous step of this research. Its purpose is to verify whether these predictive outputs can be translated into viable, real-world applications. In doing so, the findings of the stakeholder analysis are also assessed. Specifically, the routing methodology evaluates how different routing configurations perform when addressing naastplaatsing, using metrics that capture both sides of the trade-off between cost efficiency of the NietRnaast operation and the quality of its coverage and interventions. By testing multiple strategies from both perspectives, this routing study provides empirical evidence on whether the proposed methodological approach can balance operational efficiency with service quality in practice.

The 12 selected neighborhoods are: Afrikaanderwijk, Beverwaard, Bospolder, Delfshaven, Groot IJsselmonde, Het Lage Land, Kleinpolder, Liskwartier, Prinsenland, Tarwewijk, Tussendijken, and Zuidwijk. Figure 5.10 shows their position on the map of Rotterdam, along with a color grading indicating their global naastplaatsing rate in 2025.

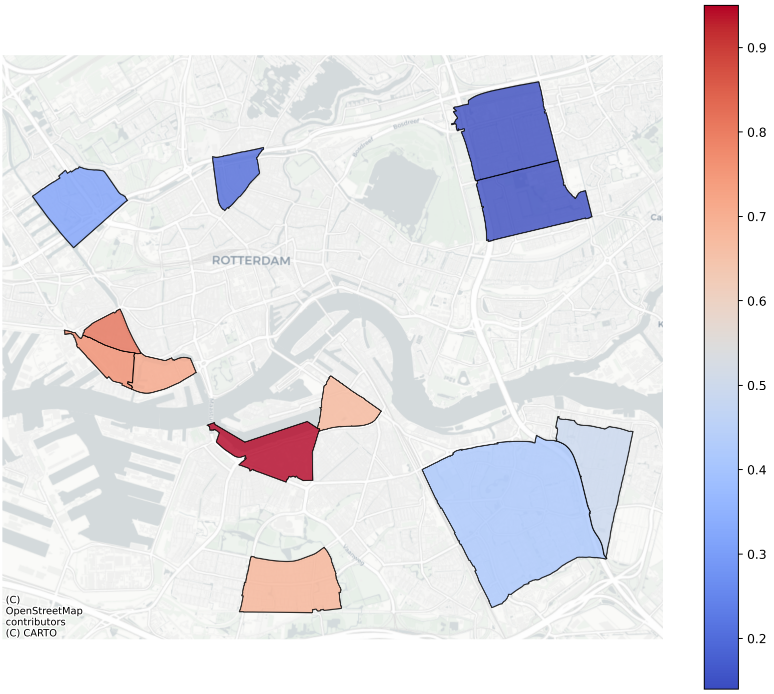


Figure 5.10: Naastplaatsing rate (2025) of neighborhoods in the routing study

An initial observation shows that the average amount of container clusters visited in different days of the week tends to vary significantly. The average amount of container clusters visits per day in the areas is broken down by the graph in figure 5.11. Saturday and Sunday display, as expected, a significantly lower amount of visits compared to the rest of the weekdays.

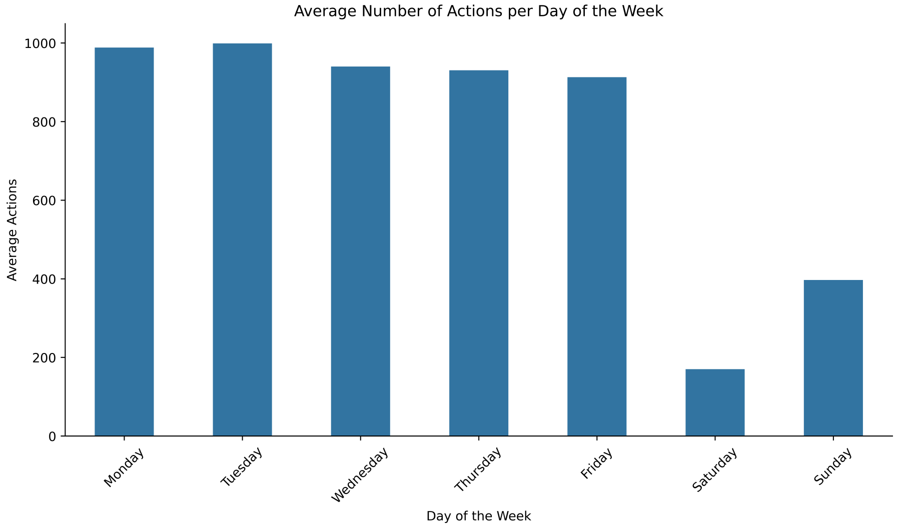


Figure 5.11: Average actions per day (2025) in routing neighborhoods

Extracted distributions and routes of actual NietRnaast operations by day are thus extracted in groups of 10 instance per area, 2 instances per each day of the week (Monday through Friday), with random sampling of months between May and December of 2025 to accomodate for the time-stamp availability constraint described in section 4.4.

The 5 heuristics utilized to build routes are next-best distance routing, risk-based routing, k-neighbor distance-risk routing, distance-risk ratio routing, and convex hull heuristic; the constructed routes are evaluated on the basis of the total distance they cover and the collection penalty they incur on for a specific naastplaatsing distribution extracted from the NietRnaast dataset. The heuristics build routes for 10 days over each of the 12 areas selected for this study; there are thus 720 total routes, 120 distributions and actual routes and 600 routes built by the 5 heuristics.

Evaluating the performance of a heuristic's results is a matter of interpretation and perspective. Different stakeholders may be interested in different things; some may prefer routes that minimize the covered distance above anything else, while others may be drawn to more balanced solutions. To provide a complete overview, the two following visualizations are presented and discussed. Figure 5.12 shows how many times each heuristic has scored the best in either of the performance metrics, while figure 5.13focuses on the comparison between each heuristic and the actual route's performance.

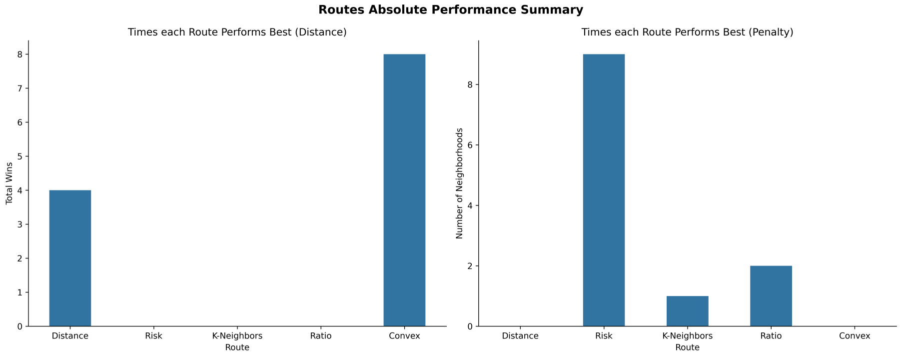


Figure 5.12: Absolute performance of heuristic routes

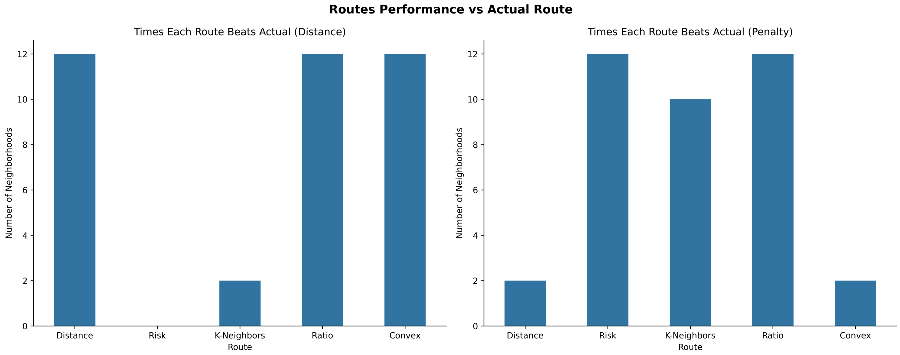


Figure 5.13: Performance of heuristic routes against actual route

These visualizations clearly display the performance trends across the 5 heursistics. The purely distance-based approaches, next-best distance and convex insertion, dominate the distance metric. Convex insertion wins 75 times and next-best distance 44, whereas all other heuristics win only once combined (the ratio heuristic). The two distance heuristics overwhelmingly outperform not only their competitors, but also the actual routes: both convex insertion and next-best distance beat the real-world route 93 times (out of 120 total) in terms of traveled distance. However, this dominance does not translate to collection penalty. Distance-based heuristics remain among the weakest performers on this metric, with only 8 (distance) and 7 (convex) absolute wins, and only 34 and 32 improvements over the actual routes respectively. This aligns with their expected behavior: being risk-agnostic methods, they minimize distance at the expense of risk-related performance. The purely risk-based heuristic displays the opposite profile. It wins 66 times in terms of collection penalty, by far the highest of any other method, and improves on the actual route 80 times. Yet it is never the best algorithm distance-wise, confirming the substantial route length inefficiency characteristic of risk prioritization. In terms of actual route comparison, it consistently represents a degradation relative to the status quo. The hybrid heuristics present a more nuanced picture. The k-neighbors method performs moderately well on collection penalty, with 22 wins and a strong 67 improvements relative to the actual routes. However, it does not win any distance competitions and outperforms the actual routes only 28 times in terms of distance, consolidating it as a mediocre option. In contrast, the ratio heuristic stands out. Although it wins only 1 absolute distance comparison, it improves upon the actual routes 88 times on the same metric, equalizing the results of the purely distance-based algorithms. It also performs strongly on collection penalty, with 17 absolute wins and 66 improvements relative to the actual route, surpassing both distance-based approaches and approaching the performance of k-neighbors. While it does not match the risk heuristic's dominance on collection penalty, it avoids the severe downside that comes with its distance-agnostic logic. These results show that the ratio heuristic is the most well-balanced method: it is the only heuristic that consistently outperforms the actual route on both metrics at scale, doing so without the extreme trade-offs present in the purely distance or risk driven strategies. For stakeholders concerned with balancing the trade-off between efficiency and operational quality, the ratio heuristic offers the most robust compromise across the competing objectives.

The average performances by route on both distance and collection penalty metrics for each of the 12 neighborhoods can be found in Appendix C.

Chapter 6

Discussion

This chapter interprets the results of the proposed methods in light of the research question and subquestions posed in section 1.3. After addressing these research objectives, policy recommendations are discussed. The Limitations section reflects on the methodological and data-related hurdles that shaped this study and impact its generalizability. The final section outlines directions for future work on the subject.

6.1 Research Objectives

This section discusses how the empirical findings answer the main research question:

Can the phenomenon of naastplaatsing in Rotterdam be modeled to understand its root causes and efficientize the response to address it?

To answer the main research question, 4 research objectives have been identified and tied to the different steps of this thesis. The results of each step provide answers to the research sub-questions, and are discussed in the following.

1) Who are the most relevant stakeholders in the decision arena of the naastplaatsing problem, and what is currently understood about it?
The key stakeholders in the decision arena are the Alderman, Stadsbeheer management, NietRnaast teamleaders and operators, Inzameling and Reiniging personnel, Handhaving officers, various groups of residents (including perpetrators, concerned residents, and container adoptants), and higher-level strategic advisors. Their objectives can be summarized by two perspectives:

- operational efficiency: minimising collection time, distance, and costs.
- behavioural impact: reducing the visibility, duration, and spread of naastplaatsing to prevent further dumping.

High-level decision makers must ultimately balance both perspectives when shaping policy and operational responses. The theoretical framework constructed in this study, a result of the combination of the three streams of information used to build domain knowledge, cements that Naastplaatsing is a complex phenomenon, driven by multiple correlated factors. The decision arena is a complex network of stakeholders, whose interests and influence shape naastplaatsing and how Stadsbeheer responds to it.

Stakeholders can be characterized as operationally or behaviorally inclined, depending on which of the two perspectives they prioritize. The most strategically relevant actors are those who must bridge the gap between these two goals: the alderman, Stadsbeheer managers, and (to a lesser extent) NietRnaast managers. Their public-facing positions require balancing the efficiency demands of daily operations with the behavioral objectives tied to naastplaatsing public perception and long-term waste norms. The stakeholder mapping of figure 3.2 also reveals a fundamental misalignment in the power–interest structure of the decision arena. The actors whose behaviour directly produces naastplaatsing possess high influence but low interest in resolving it. Perpetrators represent the most problematic configuration: they hold substantial physical power, and with their actions can shape the behavior of other actors. Yet, they have minimal intrinsic motivation to change. Similarly, "lazy but knowledgeable" residents contribute to undesirable waste patterns while lacking meaningful accountability. By contrast, many of the actors who are highly committed to maintaining clean streets—such as concerned residents, container adoptants, and NietRnaast runners hold limited structural power. They care about neighborhood cleanliness and perceive the effects of public space deterioration produced by naastplaatsing, but can only operate within narrow operational or social boundaries. This imbalance underscores why naastplaatsing persists as a wicked problem: the stakeholders who could meaningfully reduce the phenomenon are not the ones with the motivation to do so, and those who are motivated lack the authority or structural leverage needed for significant change.

2) How can MVDA techniques be utilized to validate the findings and hypotheses of the naastplaatsing theoretical framework?

The multivariate data analysis (MVDA) conducted in this study provides a meaningful, yet partial, validation of the theoretical framework. The conjectures on the impact of operational impediments and the origin composition of residents on naastplaatsing rates at the neighborhood level are consistently and robustly supported by the Ordinary Least Squares (OLS) regression yearly models at the neighborhood level. Other hypotheses and framework relationships are tentatively backed, with variables often demonstrating clear limits in their ability to proxy the real phenomenon. MVDA techniques proved to be useful in validating and refining the framework, but not sufficient on their own.

The initial set of OLS models (section 5.2), estimated separately for 2022 and 2023 at neighborhood level, shows that MVDA techniques can capture a substantial share of the variance in naastplaatsing rates. This high explanatory power, however, is tied to the full set of feature, which introduces issues of severe multicollinearity and instability in the coefficient estimates. The reduction strategy employed to address this issue is supported by the hypotheses formulated in section 3.4.2.

Operational impediments are the most robust predictors of naastplaatsing rates at the neighborhood level. The "klemmers" variable is positively correlated with the target variable and highly statistically significant in both yearly models. The predictor representing overflow incidents also inches towards significance in the 2023 model. Together, these findings provide a strong validation of H2. Resident composition matters consistently on the side of origin, rather than age. The share of non-EU residents in neighborhoods is positively correlated to naastplaatsing rates and statistically significant across both yearly models; on the other hand, the 15 to 29 years old variable chosen to represent age portion of the cultural and normative practices hypothesis is not significant in either year. At this spatial scale, H8 is only partially validated. Pressure and prosperity variables are directionally consistent with the hypothesis they represent, but their statistical signal is weak. This suggests that H1 and H7 are partially supported, but perhaps could be validated at a different aggregation level, likely a finer one. Social cohesion is negative in both yearly models, but it is not statistically significant in the 2022 and barely so in the 2023 one. This is likely because it is the variable with the most linear dependence with other features, mainly demographic composition and prosperity; furthermore, the evaluation of this metric must always be contextualized with the fact that the 2023 values are an imputation of 2022 and 2024 observations, since the data generating process is biennial. These results offer at best a tentative support of H6.

These results show that MVDA techniques can indeed be used to validate key aspects of the naastplaatsing theoretical framework. However, an important caveat must be considered: the quality of the validation depends on the adequacy of the chosen proxies and the scale of analysis. It is perhaps telling that one of the most solidly validated relationships, the operational H2, stems from the most solid and plentiful dataset available, requiring minimal transformations and proxies to achieve an empirically usable variable. The answer to sub-question 2 is nuanced: MVDA is not the enough to evaluate the framework on its own, but it is an effective instrument for distinguishing robust relationships from weaker ones.

3) Can ML algorithms be employed to model naastplaatsing in Rotterdam to a satisfying level of prediction accuracy and result explainability?

The results of the predictive analysis indicate that tree-based models are highly suitable for capturing the complex, non-linear relationships between predictors and naastplaatsing rates container cluster level. Model's performance is strong consistently across Decision Tree, Random Forest, and Gradient Boosting regressors; SHAP-based feature importance provides a transparent and interpretable decomposition of how each feature contributes to the predicted risk. These explanations strongly align with the theoretical framework and domain knowledge, confirming the usefulness of these models for understanding and forecasting naastplaatsing patterns.

The *avg_streak* variable dominates prediction, providing a strong confirmation of H5. Clusters that experience consecutive and/or frequent naastplaatsing occurrences score a higher naastplaatsing risk, meaning they are more likely to experience it again. This feedback loop is consistent with the "trash brings trash" effect described by literature and experts. The temporal dynamics captured by *bad_naastplaatsing_day_share* highlight that clusters experiencing naastplaatsing outside the typical post-weekend surge are more exposed to sustained risk. This confirms that naastplaatsing patterns vary not only in magnitude but also in timing, as stated by H4, and that persistent occurrences across the week imply vulnerability. The *nietrnaast_visits_2025* variable displays a SHAP pattern consistent with the conceptual understanding of naastplaatsing: higher visit counts are associated with lower predicted naastplaatsing, suggesting that operational attention has a mitigating effect. Earlier assumptions about visit frequency merely correlating with problematic clusters are replaced by the definition of a protective relationship between responsiveness and reduced risk. The prominence of *bad_waste_count* indicates that the type of dumped waste, especially when it is of significant volume, is closely tied to recurring naastplaatsing. Clusters consistently receiving problematic or bulky waste show higher predicted risk, confirming that waste composition acts as a structural predictor of future incidents. The strong statistical signal of *klemmers_rates* in the OLS regression is partially picked up by the Decistion Tree and the Gradient Boosting regressors, further solidifying the relevance of these operational features. The contribution of *fastest_fill_speed* confirms that clusters with slower-filling containers, and thus lower structural pressure, tend to exhibit lower risk.

Together, these findings indicate that ML algorithms can indeed model naastplaatsing with both strong predictive accuracy and high explainability. The performance results support the viability of the ML-based risk predictions, while SHAP interpretations validate key mechanisms of the theoretical framework. ML models are excellent at detecting correlation, yet they should never be burdened with explaining causation: the explainability methods allow humans in the loop with domain knowledge to contextualize their outputs. This robust approach makes these ML predictions a valuable tool for supporting strategic understanding and operational decision-making in Rotterdam's naastplaatsing response.

4) How can the collection process of naastplaatsing be made more efficient?

The routing study shows that the collection process of naastplaatsing can be made substantially more efficient, both in terms of operational costs (distance traveled) and service quality (collection penalty). While no single heuristic dominates both metrics, the distance-risk ratio heuristic consistently outperforms the actual routes across both dimensions, demonstrating that more balanced and efficient collection strategies are attainable without compromising behavioral responsiveness.

The evaluation relies on two performance metrics that reflect the perspectives prioritized by stakeholders:

- total distance traveled, which represents a proxy for workload, fuel consumption, equipment wear and tear, and route efficiency. These are the main concerns of operational teams.
- collection penalty, capturing how long naastplaatsing remains visible at each container and weighting more disruptive waste more heavily. This metric expresses behavioral, normative, and quality-of-service perspectives emphasized by planners, managers, and residents.

Together, these metrics synthesize and operationalize the dilemma highlighted in the decision arena: the need to reconcile operational feasibility with timely responses to prevent behavioral feedback loops of naastplaatsing.

Across the five heuristic strategies tested, several clear patterns emerge. The distance-based heuristics outperform the other algorithms consistency on route efficiency, but systematically fall short in the timely collection of disruptive waste, making them unfit for addressing behavioral priorities. The risk-based heuristic excels in prioritzing problematic containers, but the routes it creates are simply infeasible from an efficiency perspective. Hybrid heuristics provide a more balanced solution. The k-neighbors method improves upon the actual routes primarily in terms of collection penalty, but its distance performance remains mediocre. The ratio heuristic is instead capable of improving the actual route on both performance metrics consistently, delivering the most balanced outcome by far. Unlike the risk-based strategy, it maintains operational feasibility; unlike the distance heuristics, it remains attentive to disruptive waste that shapes resident perceptions and behavioral spillovers.

The overall insight of the routing study is that NietRnaast's current operational setup leaves substantial room for improvement, and that efficiency gains do not require choosing between operational or behavioral priorities. The ratio heuristic shows that routes can be shorter and high risk containers can be prioritized simultaneously. This balanced performance aligns closely with the goals of high-level strategic stakeholders, who must continuously negotiate the trade-off between cost efficiency and street-level service quality. Given these insights, the answer to the main research question:

Can the phenomenon of naastplaatsing in Rotterdam be modeled to understand its root causes and efficientize the response to address it?

is **Yes**. This study demonstrates that naastplaatsing in Rotterdam can be effectively modeled both to deepen understanding of its underlying mechanisms and to improve the efficiency of the operational response.

The theoretical framework developed through literature, expert interviews, and field insights, successfully characterizes naastplaatsing as a wicked problem. It identifies the two overarching goals of the stakeholders involved with this problem, the imbalance in the decision arena caused by their positioning, and the relationships between predictors and naastplaatsing that need to be investigated to better understand this phenomenon. The theoretical framewoek provides the underlying support to broaden the understanding of naastplaatsing, supporting more empirical steps like the predictive modeling when its results need causal contextualization.

MVDA confirms and refines several of the mechanisms proposed in the theoretical framework, expressed in the form of hypotheses. The results of the OLS regressions and their comparison demonstrate that statistical modeling can meaningfully distinguish between strong and weak explanatory factors. Where these traditional methods reach their limitations due to non-linear relationships and multicollinearity, ML extends the analysis by capturing more complex interactions and offering interpretable, cluster-level predictions. The tree models trained as part of the predictive modeling step provide accurate and interpretable predictions of naastplaatsing risk at the container-cluster level; SHAP analyses reveal that both historical behavioral patterns ("trash brings trash") and operational conditions (malfunctions, waste types, temporal stress) govern the prediction of risk, showing good adherence between the model's thought process and the theoretical framework of naastplaatsing developed through domain knowledge.

Finally, the routing study shows that integrating risk into operational decision-making, even in a simplified and experimental setting, can lead to concrete improvements in collection efficiency. While different routing heuristics embody different stakeholder goals and perspectives, the ratio heuristic consistently outperforms the actual routes on both distance and collection penalty, dominating in terms of the efficiency/quality trade-off.

This confirms that modeling not only can support understanding naastplaatsing, but it can also directly support the design of operational improvements, aligning with the various priorities of the stakeholders involved in naastplaatsing and its response strategy.

6.2 Policy Recommendations

The findings of this thesis show that naastplaatsing is a complex phenomenon, shaped by a multitude of interconnected factors. While some of these are in more or less direct control of Stadsbeheer, like the involvement of its departments, or the position of container clusters, many other of the upstream drivers of naastplaatsing seem to fall outside of its mandate. Some of the structural causes of this phenomenon are broader social and spatial dynamics: population mobility, housing shortages, socio-economic degradation are not factors that can be tackled through analysis and optimisation. However, Stadsbeheer can operate to reduce both the frequency and the impact of naastplaatsing by acting on the mechanisms that are within its decision-making boundaries. In the following, three broader policy approaches are discussed, detailing which area of the naastplaatsing area they are designed to tackle and how.

6.2.1 Routing that Works

From a strictly operational point of view, the routing study demonstrates that the collection process can be made more efficient. A further development of the risk-distance ratio heuristic would allow to allocate resources more efficiently, with equipment suffering from less wear and overall collection times becoming shorter. Simple routing tools should be integrated into the daily NietRnaast workflow; these must be designed in tandem with the primary users, the drivers, in order to adapt them to the operational reality of naastplaatsing collection. ML-based predictions could enable NietRnaast to reduce the size of their routes. Clusters with consistently low predicted risk can be visited less frequently, freeing up time and resources for hotspots where naastplaatsing is persistent or disruptive. This lesser attention is appropriate especially for clusters with no history of recurrent naastplaatsing, but, as discussed next, it also can serve a behavioral purpose.

6.2.2 Behavior that Shifts

Reducing frequency and reducing impact do not necessarily require the same operational intensity everywhere. The study shows that despite the flaws with some of the tooling and practices, NietRnaast operates at a high level of efficiency. As described by a few interview participants (Participants A, B, C, Appendix A.1), this can turn out to have a negative impact on naastplaatsing: waste is removed so quickly that perpetrators do not have the opportunity to experience the consequences of their own actions. This creates a perverted behavioral incentive for chronic offenders: naastplaatsing remains convenient, and its effects are short-lived. At the same time, the persistence of naastplaatsing in visible areas has a well-documented and empirically investigated effect: the "trash brings trash" theorem, effectively a waste multiplier. This creates a genuine dilemma: collecting too quickly everywhere reduces visibility but maintains convenience, but collecting too slowly increases the chance of triggering a multiplying effect. The missing piece to solve this puzzle is naastplaatsing's variability: the risk on container clusters is not uniform, and it is also predictable. Stadsbeheer should pursue a selectively fast and slow response strategy, guided by predicted risk and behavioral dynamics. Hotspots and areas where multiplying effects are prevalent should still be subject to quick collection; however, chronically problematic containers could be visited less frequently, to reduce the convenience perpetrators currenltly enjoy. This is only one of the levers that Stadsbeheer can act upon to elicit a change of behavior in the most problematic naastplaatsing stakeholder: the other is perceived risk. Coordinated presence of Handhaving near chronic hotspots can shift local norms and reduce opportunistic dumping. This does not require blanket enforcement, but targeted visibility, which could even be based on the ML-derived risk map and other time-based considerations, like the current day of the week or the proximity to holidays.

6.2.3 Mind the Gap

This study has uncovered a structural disconnect between the ”thinking” stakeholders that designs workflows and tools, and the operational level that must use them. Many of the flaws of the NietRnaast app, for instance, are integration failures more than technical challenges. This persists largely because primary users are not involved in the design process, and also because the feedback loop between these stakeholders is not working as intended. This thesis offers many different directions in which to develop the operational response to naastplaatsing; yet, any of the findings and improvements designed at this stage could be completely useless if not integrated with the people that must work with them. Stadsbeheer must work to close the gap between operational practice and strategic decision making: this means both in terms of communication between personnel, and data usage and sharing between the layers of the departments that work together to address the challenges posed by naastplaatsing. From an internal policy perspective, Stadsbeheer should embed ”collaborative design” practices into its solution and innovation cycle by establishing structured feedback loops between drivers, team leaders, and tool developers. Furthermore, the mechanisms to report and correct design choices must be formalized. The flow of information between street-level expertise and thinking-level decisions needs to be facilitated; this alignments is essential to tackle a problem as complex as naastplaatsing.

Together, these recommendations show that although Stadsbeheer cannot address some of the root causes of naastplaatsing, it can influence the mechanisms that regulate it. By integrating data-driven routing, applying behavioral insights through a selective fast–slow strategy, and ensuring that operational expertise guides strategic planning, Rotterdam can develop a more resilient and responsive system for managing naastplaatsing.

6.3 Limitations

This research aimed to comprehensively analyze the phenomenon of naastplaatsing, utilizing all the available data and techniques to produce a faithful representation of this issue and provide insights valuable to further address this challenge. Despite these efforts, there are some limitations that contextualize the boundaries of this study’s findings.

Across the multiple accessed datasets, completeness, granularity and temporal coverage vary greatly. NietRnaast data is reliable and usable only from 2022 onwards, and its timestamp feature only since May of 2025. The socio-demographic datasets (CBS and Onderzoek010) are complete only up to 2023, with social cohesion variables requiring imputation due to the biennial nature of their collection. The finest spatial granularity recorded by these datasets is neighborhood level, while the NietRnaast dataset collects actions at the container cluster level; the Waste Vision (WV)dataset, which describes features like filling speed and waste fraction, deals instead with containers. While a synthesis between container and cluster level was possible, the neighborhood-level variables were not included in the Predictive Analysis section, in order to avoid articial clustering of the observations. GFE containers are also entirely missing from the WV dataset, since they are not included in the collection routes used by the Inzameling departments; this is because this type of containers only recently started to integrate fill sensors in their management. Finally, the lack of access to GPS data lead to devising a method to approximate a representation of the current routes; while this approach is deemed sufficient for comparative purposes, having access to the real routes would improve the reliability of this study's results.

The behavior of the drivers in charge of the data generating process for the NietRnaast dataset contributes to a few data quality concerns. Drivers have been observed systematically misclassifying the type of actions they carry out on the container clusters they visit. This is done in an effort to move through their route quickly; in a haste to move on to the next container, NietRnaast drivers select the "Klein afval" option, which is the most general type of small waste residue. This category is massively overrepresented in the NietRnaast dataset, growing from 10.29% in 2022 to 27.64% in 2025. Furthermore, until recently drivers could only take a photo if they recorded an action as a form of naastplaatsing. NietRnaast team leaders sometimes ask their drivers to take pictures for all the containers they visit, meaning that the workers are forced to mislabel clean clusters as dirty, introducing noise in the dataset. This issue has particularly negative consequences in Target Wijks, areas of Rotterdam that have reported issues with naastplaatsing and waste management. Drivers servicing these areas are asked to take pictures here more often, leading to a feedback loop that negatively reinforces the "bad" reputation of these neighborhoods.

Both OLS and ML are correlational techniques, and as such do not imply causation. This is especially relevant when the evaluated predictors represent proxies of the actual concepts that the modeler is trying to evaluate. However, the strong foundation provided by the domain knowledge obtained in the first step serves as a way to shift the burden of proving causation away from the models.

Finally, it is worth restating that the routing study is comparative and not a definitive solution of the routing problem of naastplaatsing. The reconstruction of the routes from the NietRnaast dataset provides an approximation of the actual route, and the timestamp sorting may improperly sequence clusters with identical timestamps, as well as include clusters that on any specific day were visited by other drivers. The collection penalty metric is a simple proxy: if a trash bag is encountered at step 100 of the route, instead of step 1, it is not implied that its impact on the cleanliness of that area is one hundred fold. Operationally deployable routes were not the goal of this step of the study.

6.4 Future Work

The results of this thesis open several avenues for further research into naastplaatsing and its operational response.

The predictive models used in this thesis are regressors, producing a continuous risk score per container cluster. While this is effective for evaluating feature importance and designing a comparative routing study, this regression compresses all the information of the NietRnaast dataset into a single row of information per container cluster. An extension of this approach is the development of a daily classification model, where each row corresponds to a specific cluster on a specific day labeled with a 0/1 target variable for naastplaatsing occurrence. The predictive features would include cluster characteristics, similar to the ones used for this study's models; furthermore, the temporal information of each day, like the specific day of the week, the month of the year, or a holiday marker, could be used by this classifier to determine naastplaatsing occurrences. With the benefit of reliable timestamps, precise weather variables could be included in the analysis.

The hyperparameters of the mixed heuristics could also be fine-tuned to extract the maximum performance out of these metrics. In particular, the balance of factors in the risk-distance ratio heuristic and the number k-neighbors heuristic could be readjusted for different areas and days, adapting to the heterogeneity of naastplaatsing occurrence patterns.

All the evaluated heuristics are subject to a completeness constraint, in an effort to ensure comparability among each other and with the actual routes, mimicking the operational modus operandi of NietRnaast. However, a significant portion of the container clusters in Rottedam show extremely low predicted risk. A Selective Routing heuristic could determine a risk threshold used to eliminate certain clusters from select daily routing computations. Such an approach could reduce the routing distance quite significantly, and the reduction of clusters to visit could lead to shorter times to reach high priority clusters.

Chapter 7

Conclusion

The goal of this thesis was to investigate whether the phenomenon of naastplaatsing in Rotterdam could be modeled to improve the understanding of its underlying mechanisms, while also supporting a more efficient and effective operational response. The findings of this research indicate that naastplaatsing can be understood, predicted, and managed more intelligently, but only when analytical insights are integrated with organizational realities, behavioral dynamics, and the lived expertise of the people who address the problem every day.

This work's initial step was constructing a theoretical framework based on literature, expert interviews, and on-field observations of NietRnaast workers. This framework described naastplaatsing as a wicked problem, characterized by multiple interacting factors, misalignment in stakeholder power and interest, and damaging feedback loops. The mapping of stakeholders revealed that the perpetrators, those who generate the problem, hold significant behavioral power yet possess little interest in resolving the issue. Conversely, many of the actors who care deeply about neighborhood cleanliness have limited formal influence. High-level strategic stakeholders must balance cost efficiency with service quality, while operational teams focus on feasibility and practicality. This asymmetry is what limits the impact of any intervention.

The multivariate data analysis (MVDA) phase brought quantitative insights to this framework. Through yearly OLS regression models at neighborhood level, several mechanisms emerged as statistically robust: container malfunctions and overflow incidents consistently correlate with higher naastplaatsing rates, as does the demographic composition related to resident origin. Other hypothesized relationships, such as population prosperity and social cohesion, appeared weaker or more scale-dependent, suggesting that they operate in more subtle or localized ways. While MVDA cannot capture the full complexity of the phenomenon, it allowed for a distinction between stronger and weaker explanatory factors. It also showed the limits of traditional statistical techniques, especially with systems characterized by many collinear variables.

Machine Learning modeling extended the analysis into a more granular and operationally relevant domain. Tree-based models demonstrated excellent predictive performance at the container-cluster level, substantiating the idea that naastplaatsing is a patterned phenomenon. The dominance of features expressing the historical behavior of containers confirmed the self-reinforcing nature of naastplaatsing. SHAP interpretations further illustrated the roles of operational disruptions (malfunctions), temporal stress, waste type, and operational involvement of municipal waste services.

These findings deepen the theoretical framework and provide a bridge between causal understanding and the ability of ML models of reliably detecting correlation.

The routing study translated risk estimation into actionable operational application. By evaluating five heuristics it showed that NietRnaast's current operational practice can be improved significantly. While purely distance-driven strategies offer significant efficiency gains, they perform poorly on collection penalty. Conversely, purely risk-driven approaches excel at reducing visibility and its multiplier effects, but are operationally infeasible due to excessively long distances covered. The ratio heuristic emerged as the most balanced method, consistently outperforming the actual routes on both distance and collection penalty. This finding demonstrates that a hybrid routing approach is technically effective and strategically aligned with the dual priorities of cost and quality that define Stadsbeheer's mandate.

Together, these analytical components support a coherent set of policy recommendations. Stadsbeheer can improve the efficiency of its operational response by integrating routing optimization and risk prediction into daily NietRnaast operations, provided that drivers and team leaders are directly involved in the design and development of these tools. It can also influence the behavioral dynamics that reinforce naastplaatsing by selecting container clusters to service at different speeds: quick removal in hotspots to prevent spread, slower removal in chronic low-visibility locations to reduce convenience for repeat offenders. The final recommendations is to close the gap between strategic decisions and operational practice. Many of the current system's shortcomings, like the flaws in the NietRnaast app, are due to a failure of integration between tools and their primary users. Collaborative design, institutionalized feedback loops, and structured data sharing are enhancements that should be pursued.

While Stadsbeheer cannot tackle the deep socio-economic roots of naastplaatsing, it can shape the conditions under which those causes translate into visible waste. The combined use of MVDA, machine learning, and routing optimization provides a robust analytical foundation for understanding and managing naastplaatsing. The overarching call to action of this thesis is organizational: technical solutions achieve impact only when they are embedded within the people of the system they intend to improve. Addressing naastplaatsing is an operational challenge first and foremost. Still, its behavioral component requires a strategic approach; the alignment of all the expertise at the disposal of Stadsbeheer is fundamental to deal with the challenge posed by this wicked problem.

Bibliography

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adeleke, O., & Jen, T.-C. (2025). Explainable ai and machine learning-based analysis of municipal solid waste generation rate: A south african case study. *Waste Management*, *206*, 115036. https://doi.org/https://doi.org/10.1016/j.wasman.2025.115036

Arrieta, A. B., Rodríguez, N. D., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2019). Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *CoRR*, *abs/1910.10045*. http://arxiv.org/abs/1910.10045

Bijplaatsingen [Latest access: 18 March 2026]. (2023). NVRD. https://www.nvrd.nl/dossier/bijplaatsingen/

Bohm, K., Smidt, E., & Tintner, J. Application of multivariate data analyses in waste management (L. V. de Freitas & A. P. B. R. de Freitas, Eds.). In: *Multivariate analysis in management, engineering and the sciences* (L. V. de Freitas & A. P. B. R. de Freitas, Eds.). Ed. by de Freitas, L. V., & de Freitas, A. P. B. R. London: IntechOpen, 2013. Chap. 2. https://doi.org/10.5772/53975

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Containers verdwijnen uit josephlaan: Zo gooien mensen het ernaast en komen er ratten [Latest access: 18 March 2026]. (2024). Open Rotterdam. https://openrotterdam.nl/containers-verdwijnen-uit-josephlaan-zo-gooien-mensen-het-ernaast-en-komen-er-ratten/

Crook, B., Schlüter, M., & Speith, T. (2023). Revisiting the performance-explainability trade-off in explainable artificial intelligence (xai). https://arxiv.org/abs/2307.14239

Dereci, U., & Karabekmez, M. E. (2022). The applications of multiple route optimization heuristics and meta-heuristic algorithms to solid waste transportation: A case study in turkey. *Decision Analytics Journal*, *4*, 100113. https://doi.org/https://doi.org/10.1016/j.dajour.2022.100113

Du, L., Zuo, J., Vanzo, J., Chang, R., & Zillante, G. (2023). Assessing and predicting the illegal dumping risks in relation to road characteristics. *Waste Management*, *169*, 332–341. https://doi.org/10.1016/j.wasman.2023.07.031

Epskamp, M., & De Vries, C. (2020). *Verklaringen voor naastplaatsingen in rotterdam, 2018–2019*. Onderzoek en Business Intelligence. Rotterdam.

Evaluatie Vulgraadsysteem Rotterdam (No. 20332-004). (2021, December 21) (Report commissioned by Stadsbeheer Rotterdam). KplusV. Arnhem.

Fang, B., Yu, J., Chen, Z., Osman, A. I., Farghali, M., Ihara, I., Hamza, E. H., Rooney, D. W., & Yap, P.-S. (2023). Artificial intelligence for waste management in smart cities: A review. *Environmental Chemistry Letters*, *21*(4), 1959–1989. https://doi.org/10.1007/s10311-023-01604-3

Fraternali, P., Morandini, L., & Herrera González, S. L. (2024). Solid waste detection, monitoring and mapping in remote sensing images: A survey. *Waste Management*, *189*, 88–102. https://doi.org/https://doi.org/10.1016/j.wasman.2024.08.003

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gemeente waarschuwt voor afval naast de container door beelprojectie [Latest access: 18 March 2026]. (2026). De Havenloods. https://www.dehavenloods.nl/nieuws/algemeen/62750/gemeente-waarschuwt-voor-afval-naast-de-container-door-beelde

Geng, J., Ding, Y., Xie, W., Fang, W., Liu, M., Ma, Z., Yang, J., & Bi, J. (2024). An ensemble machine learning model to uncover potential sites of hazardous waste illegal dumping based on limited supervision experience. *Fundamental Research*, *4*(4), 972–978. https://doi.org/https://doi.org/10.1016/j.fmre.2023.06.010

Glanville, K., & Chang, H.-C. (2015). Mapping illegal domestic waste disposal potential to support waste management efforts in queensland, australia. *International Journal of Geographical Information Science*, *29*(6), 1042–1058. https://doi.org/10.1080/13658816.2015.1008002

Gordon, A. D. (1984). *Biometrics*, *40*(3), 874–874. Retrieved March 15, 2026, from http://www.jstor.org/stable/2530946

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). Xai—explainable artificial intelligence. *Science Robotics*, *4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Hemmelmayr, V., Doerner, K. F., Hartl, R. F., & Rath, S. (2013). A heuristic solution method for node routing based solid waste collection problems. *Journal of Heuristics*, *19*(2), 129–156. https://doi.org/10.1007/s10732-011-9188-9

Herm, L.-V., Heinrich, K., Wanner, J., & Janiesch, C. (2023). Stop ordering machine learning algorithms by their explainability! a user-centered investigation of performance and explainability. *International Journal of Information Management*, *69*, 102538. https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2022.102538

Het veelkoppige afvalmonster: waarom boetes alleen de stad niet schoon krijgen [Latest access: 18 March 2026]. (2025). AT5. Retrieved March 18, 2026, from https://www.at5.nl/artikelen/235623/het-veelkoppige-afvalmonster-waarom-boetes-alleen-de-stad-niet-schoon-krijgen

Hillesluis ervaart afvaloverlast het ergst [Latest access: 18 March 2026]. (2021). Dagblad010. https://dagblad010.nl/politiek/hillesluis-ervaart-afvaloverlast-het-ergst

Hohl, B. C., Kondo, M. C., Rupp, L. A., Sadler, R. C., Gong, C. H., Le, K., Hertlein, M., Kelly, C., & Zimmerman, M. A. (2023). Community identified characteristics related to illegal dumping; a mixed methods study to inform prevention. *Journal of Environmental Management*, *346*, 118930. https://doi.org/https://doi.org/10.1016/j.jenvman.2023.118930

Holzinger, A. From machine learning to explainable ai. In: In *2018 world symposium on digital intelligence for systems and machines (disa)*. 2018, 55–66. https://doi.org/10.1109/DISA.2018.8490530

Javed, A. R., Ahmed, W., Pandya, S., Maddikunta, P. K. R., Alazab, M., & Gadekallu, T. R. (2023). A survey of explainable artificial intelligence for smart cities. *Electronics*, *12*(4). https://doi.org/10.3390/electronics12041020

Johnson, N. E., Ianiuk, O., Cazap, D., Liu, L., Starobin, D., Dobler, G., & Ghandehari, M. (2017). Patterns of waste generation: A gradient boosting model for short-term waste prediction in new york city. *Waste Management*, *62*, 3–11. https://doi.org/https://doi.org/10.1016/j.wasman.2017.01.037

Kaza, S., Yao, L. C., Bhada-Tata, P., & Van Woerden, F. (2018). *What a waste 2.0: A global snapshot of solid waste management to 2050*. World Bank. https://doi.org/10.1596/978-1-4648-1329-0

Leichtmann, B., Humer, C., Hinterreiter, A., Streit, M., & Mara, M. (2023). Effects of explainable artificial intelligence on trust and human behavior in a high-risk decision task. *Computers in Human Behavior*, *139*, 107539. https://doi.org/https://doi.org/10.1016/j.chb.2022.107539

Liu, J., Li, Q., Gu, W., & Wang, C. (2019). The impact of consumption patterns on the generation of municipal solid waste in china: Evidences from provincial data. *International journal of environmental research and public health*, *16*, 1717. https://doi.org/10.3390/ijerph16101717

Louati, A., Son, L. H., & Chabchoub, H. (2019). Smart routing for municipal solid waste collection: A heuristic approach. *Journal of Ambient Intelligence and Humanized Computing*, *10*(5), 1865–1884. https://doi.org/10.1007/s12652-018-0778-3

Merkelbach, I., Dewies, M., & Denktas, S. (2021). Committing to keep clean: Nudging complements standard policy measures to reduce illegal urban garbage disposal in a neighborhood with high levels of social cohesion. *Frontiers in Psychology, Volume 12 - 2021*. https://doi.org/10.3389/fpsyg.2021.660410

Mikhailov, E., Tupicina, O., Bykov, D., Chertes, K., Rodionova, O., & Pomerantsev, A. (2007). Ecological assessment of landfills with multivariate analysis — a feasibility study [Special Issue: WSC-5 2006]. *Chemometrics and Intelligent Laboratory Systems*, *88*(1), 3–10. https://doi.org/https://doi.org/10.1016/j.chemolab.2006.11.006

Mondal, S., & Mandal, B. (2024). Assessment of urban solid waste management in a class ii indian city using geospatial and statistical approaches: A case study of rampurhat municipality. *Waste Management Bulletin*, *1*(4), 74–92. https://doi.org/10.1016/j.wmb.2023.09.003

Probst, P., Bischl, B., & Boulesteix, A.-L. (2018). Tunability: Importance of hyperparameters of machine learning algorithms. https://arxiv.org/abs/1802.09596

Rotterdammers zijn afvaloverlast beu [Latest access: 18 March 2026]. (2019). Open Rotterdam. https://openrotterdam.nl/rotterdammers-zijn-afvaloverlast-beu/

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. https://arxiv.org/abs/1811.10154

Sakti, A. D., Sembiring, E., Rohayani, P., Fauzan, K. N., Anggraini, T. S., Santoso, C., Patricia, V. A., Ihsan, K. T. N., Ramadan, A. H., Arjasakusuma, S., & Candra, D. S. (2023). Identification of illegally dumped plastic waste in a highly polluted river in indonesia using sentinel-2 satellite imagery. *Scientific Reports*, *13*(1), 5039. https://doi.org/10.1038/s41598-023-32087-5

Shi, Y., Lv, L., Hu, F., & Han, Q. (2020). A heuristic solution method for multi-depot vehicle routing-based waste collection problems. *Applied Sciences*, *10*(7). https://doi.org/10.3390/app10072403

Syafrudin, S., Ramadan, B. S., Budihardjo, M. A., Munawir, M., Khair, H., Rosmalina, R. T., & Ardiansyah, S. Y. (2023). Analysis of factors influencing illegal waste dumping generation using gis spatial regression methods. *Sustainability*, *15*(3). https://doi.org/10.3390/su15031926

Thakker, D., Mishra, B. K., Abdullatif, A., Mazumdar, S., & Simpson, S. (2020). Explainable artificial intelligence for developing smart cities solutions. *Smart Cities*, *3*(4), 1353–1382. https://doi.org/10.3390/smartcities3040065

Vuil naast containers kost steden miljoenen, maar blijft zich opstapelen [Latest access: 18 March 2026]. (2023). Pointer (KRO-NCRV). https://pointer.kro-ncrv.nl/vuil-naast-containers-kost-steden-miljoenen-maar-blijft-zich-opstapelen

Xia, W., Jiang, Y., Chen, X., & Zhao, R. (2022). Application of machine learning algorithms in municipal solid waste management: A mini review [PMID: 34269157]. *Waste Management & Research*, *40*(6), 609–624. https://doi.org/10.1177/0734242X211033716

Yazdani, M., Kabirifar, K., Frimpong, B. E., Shariati, M., Mirmozaffari, M., & Boskabadi, A. (2021). Improving construction and demolition waste collection service in an urban area using a simheuristic approach: A case study in sydney, australia. *Journal of Cleaner Production*, *280*, 124138. https://doi.org/https://doi.org/10.1016/j.jclepro.2020.124138

Zhao, Y., Shen, X., Chen, W., & Pan, H. (2024). Energy-efficient dynamic vehicle routing for waste collection by q-learning-based hyperheuristic particle swarm optimization. *IEEE Access*, *12*, 56196–56213. https://doi.org/10.1109/ACCESS.2024.3389749

Appendix A

Supplementary Tables

A.1 Interview Participants

Participant	**Date**	**Department**	**Role**
Participant A	04/11/25	Project Bureau	Project Leader
Participant B	11/11/25	Project Bureau	Trainee
Participant C	17/11/25	Project Bureau	Advisor
Participant D	17/11/25	NietRnaast	Product Owner
Participant E	18/11/25	Reiniging	Manager
Participant F	19/11/25	NietRnaast	Team Leader
Participant G	18/12/25	NietRnaast	Manager

Table A.1: List of Interviews Conducted

A.2 OLS Regression Variables

Variable	**Description**
total_population	Total number of residents in the neighborhood
pop_to_cluster	Ratio of residents to container clusters
population_NL	Share of residents with Dutch origin
population_EU	Share of residents with non-Dutch EU origin
population_non_EU	Share of residents with non-EU origin
0_to_14_yrs_old	Number of residents aged 0–14
15_to_29_yrs_old	Number of residents aged 15–29
30_to_44_yrs_old	Number of residents aged 30–44
45_to_64_yrs_old	Number of residents aged 45–64
65+_yrs_old	Number of elderly residents aged 65+
cohesion_neighborhood	Indicator of overall neighborhood cohesion
cohesion_neighbors	Indicator of perceived trust/contact between neighbors
social_cohesion	Combined social cohesion index
willingness_to_move	Share of residents intending to relocate soon
total_reports_2023	Total number of operational reports (NRN actions)
klemmers_reports	Number of jammed container incidents
klemmers_pct	Share of reports involving container malfunctions
volle_container_reports	Number of full container incidents
volle_pct	Share of reports involving full containers
address_per_km2	Address density per square kilometer
workforce_participation	Proportion of residents active in the labor force
avg_income_per_resident	Average income level per resident
low_edu	Number of low-education residents
mid_edu	Number of middle-education residents
high_edu	Number of highly educated residents

Table A.2: OLS Regression Variables and Descriptions

A.3 ML Variables

Feature	**Description**
klemmers_rate_2025	Rate of jammed container incidents in the cluster
volle_container_rate_2025	Rate of full-container incidents in the cluster
fastest_fill_time	Fastest fill time among containers (proxy for load/pressure)
is_scheduled	Binary indicator for a cluster part of a scheduled route
has_adopter_indicator	Whether the cluster has an assigned container adopter
nietrnaast_visits_2025	Number of NRN visits to the cluster
n_GFE	Count of organic waste containers in the cluster
n_GLAS	Count of glass containers
n_PAPIER	Count of paper containers
n_REST	Count of residual waste containers
n_TEXTIEL	Count of textile containers
n_soort_Bovengronds	Number of above-ground containers
n_soort_Halfverdiept	Number of semi-submerged containers
n_soort_Ondergronds	Number of underground containers
n_inhoud_120L	Count of 120L containers
n_inhoud_2,8	Count of 2.8 m^3 containers
n_inhoud_3,0	Count of 3.0 m^3 containers
n_inhoud_3,3	Count of 3.3 m^3 containers
n_inhoud_3,5	Count of 3.5 m^3 containers
n_inhoud_4,0	Count of 4.0 m^3 containers
n_inhoud_4,5	Count of 4.5 m^3 containers
n_inhoud_5,0	Count of 5.0 m^3 containers
n_Cafés	Number of cafés near the cluster
n_Fastfoods	Number of fast-food locations near the cluster
n_Hotel-restaurants	Number of hotel-restaurants near the cluster
n_Restaurants	Number of restaurants near the cluster
distance_to_milieupark_m	Walking distance from cluster to nearest waste facility
avg_streak	Average consecutive number of days with naastplaatsing
bad_waste_count	Number of disruptive or bulky waste occurrences
bad_naastplaatsing_day_share	Share of naastplaatsing occurring outside peak days

Table A.3: Machine Learning Features and Descriptions

A.4 Toy Cluster-Level Risk Predictions

Cluster	**Predicted Risk**
0	0.774
4	0.439
5	0.859
1	0.697
2	0.094
3	0.976
6	0.761
10	0.786
9	0.128
8	0.450
7	0.371
11	0.927
12	0.644
15	0.823
14	0.443
13	0.227
17	0.555
16	0.064
18	0.828
19	0.632

Table A.4: Example Predicted Risk Scores (Scaled 0–1)

A.5 Collection Penalty Scoring System

Action Type (NL)	**Action Type (EN)**	**Penalty Score**
Niets aangetroffen	Nothing found	0.0
Klein afval	Small waste	1.0
Karton	Cardboard	5.0
Zakken	Bags	5.0
Grofvuil	Bulky waste	5.0
Volle container	Full container	20.0
Vol cluster (alle containers)	Entire cluster full	20.0
Klemmer	Jammed container	20.0
Chemisch afval	Chemical waste	5.0
Matrassen	Mattresses	5.0
Kerstboom	Christmas tree	5.0
Witgoed	Large household appliance	10.0
PV Handhaver	Enforcement report (no waste)	0.0
Vuile container	Dirty container	0.0
Bouw- of bedrijfsafval	Construction or commercial waste	5.0
PV Reiniger	Cleaning service report	0.0
Straat verzakt	Street issue	0.0

Table A.5: Action Score Map for Collection Penalty (Dutch with English Translation)

Appendix B

Model Specifications

B.0.1 Decision Tree

Hyperparameter	Values
max_depth	3, 4, 5, 6, 8
min_samples_split	2, 5, 10, 20
min_samples_leaf	1, 2, 5, 10

Table B.1: Decision Tree Parameter Grid

B.0.2 Random Forest

Hyperparameter	Values
n_estimators	200, 400, 800
max_depth	None, 5, 10, 20
min_samples_split	2, 5, 10
min_samples_leaf	1, 2, 5
max_features	sqrt, log2, 0.7

Table B.2: Random Forest Parameter Grid

B.0.3 Gradient Boosting

Hyperparameter	Values
max_iter	200, 400, 600
learning_rate	0.03, 0.05, 0.1
max_depth	None, 3, 5, 7
min_samples_leaf	10, 20, 50
max_leaf_nodes	15, 31, 63

Table B.3: GradientBoosting Parameter Grid

B.0.4 Hyperparameters

Model	**Best Parameters**	**Best CV $\mathbf{R^2}$**
Decision Tree	max_depth = 6 min_samples_split = 5 min_samples_leaf = 1	0.902
Random Forest	n_estimators = 800 max_depth = 20 max_features = 0.7 min_samples_leaf = 1 min_samples_split = 2	0.926
Gradient Boosting	learning_rate = 0.05 max_depth = 3 max_iter = 400 max_leaf_nodes = 15 min_samples_leaf = 20	0.928

Table B.4: Best Hyperparameters from Grid Search

Appendix C

Additional Figures

C.1 SHAP Summary of Decision Tree regressor

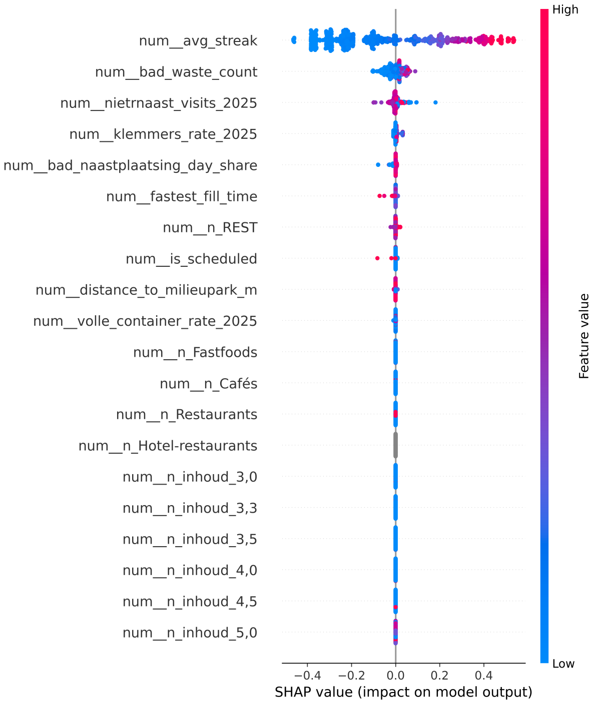


Figure C.1: SHAP summary plot for Shallow Tree model

C.2 Routing Average Performance per Neighborhood

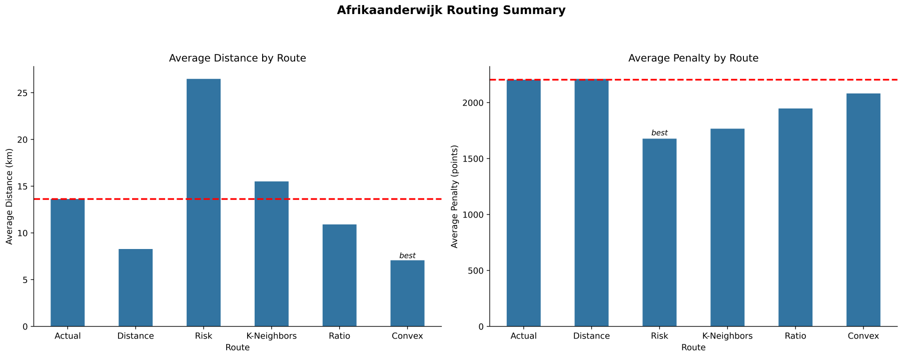


Figure C.2: Afrikaanderwijk routing summary

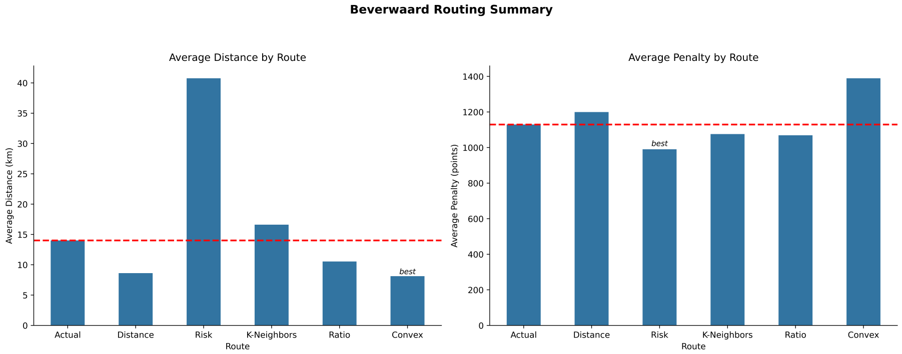


Figure C.3: Beverwaard routing summary

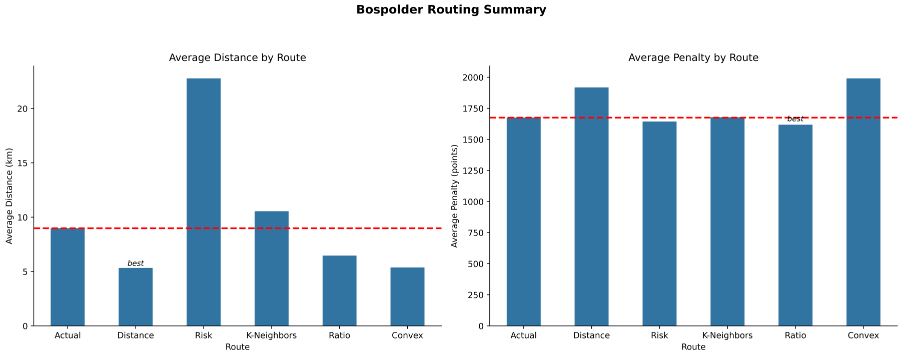


Figure C.4: Bospolder routing summary

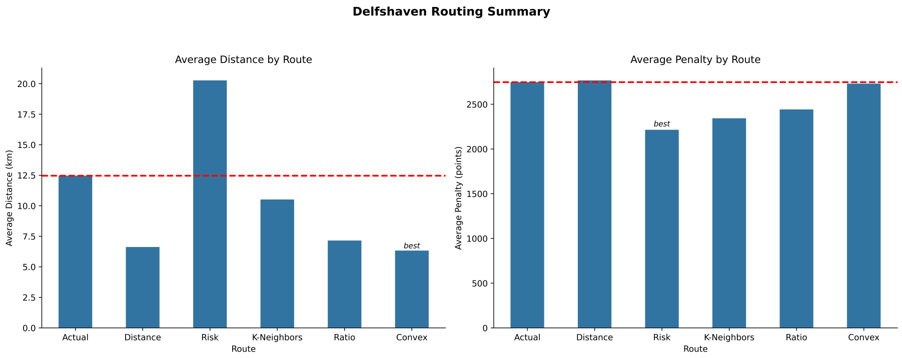


Figure C.5: Delfshaven routing summary

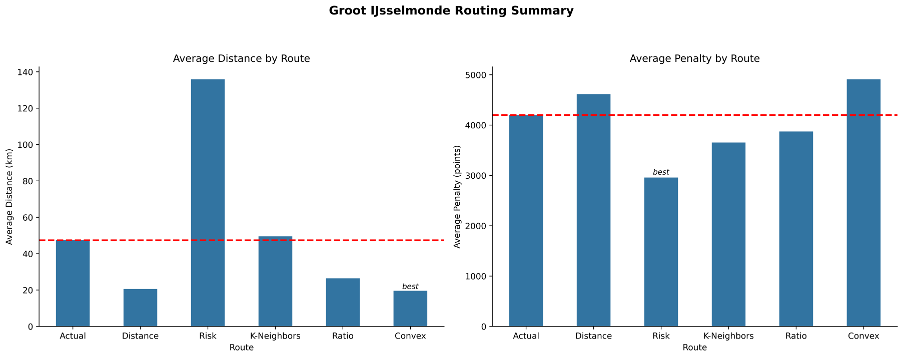


Figure C.6: Groot IJsselmonde routing summary

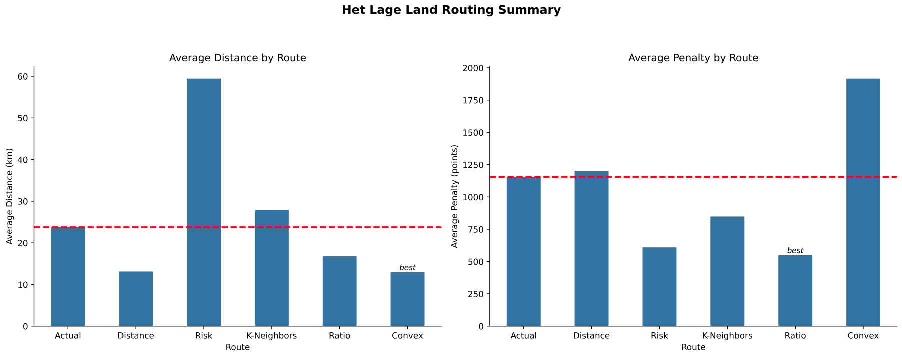


Figure C.7: Het Lage Land routing summary

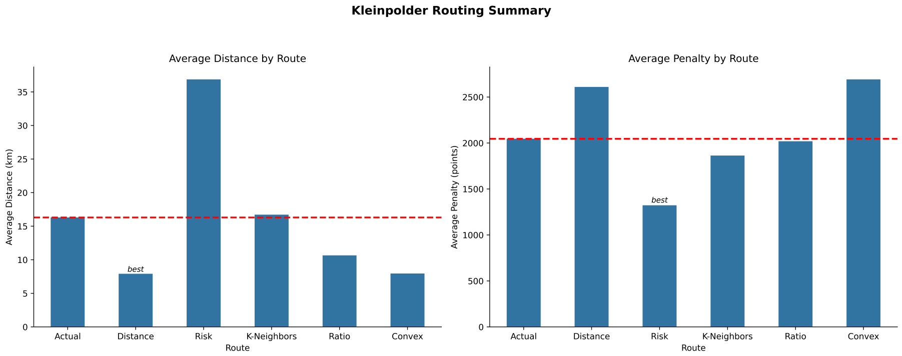


Figure C.8: Kleinpolder routing summary

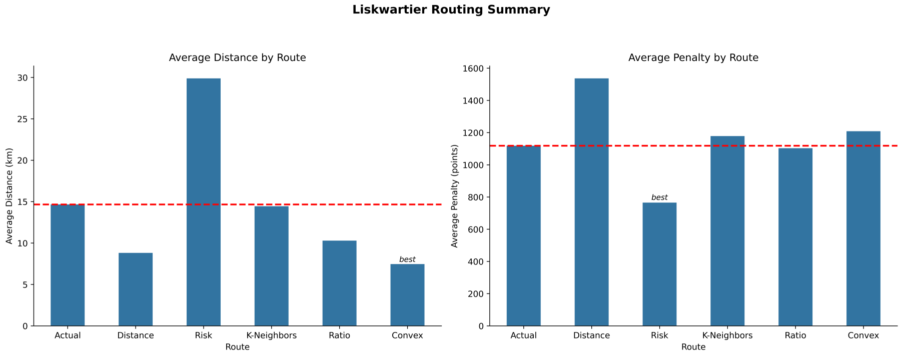


Figure C.9: Liskwartier routing summary

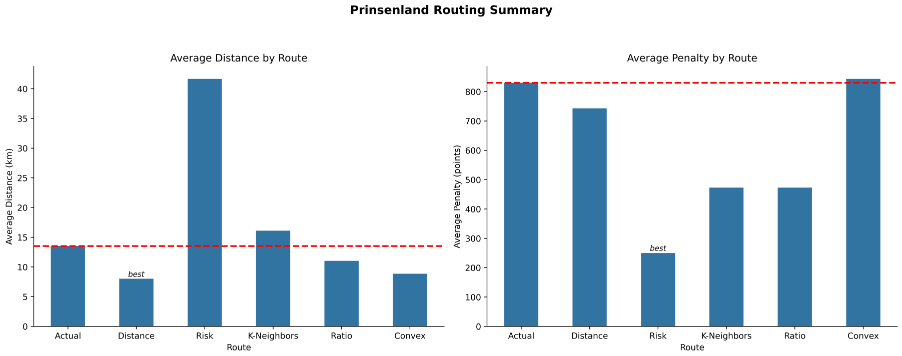


Figure C.10: Prinsenland routing summary

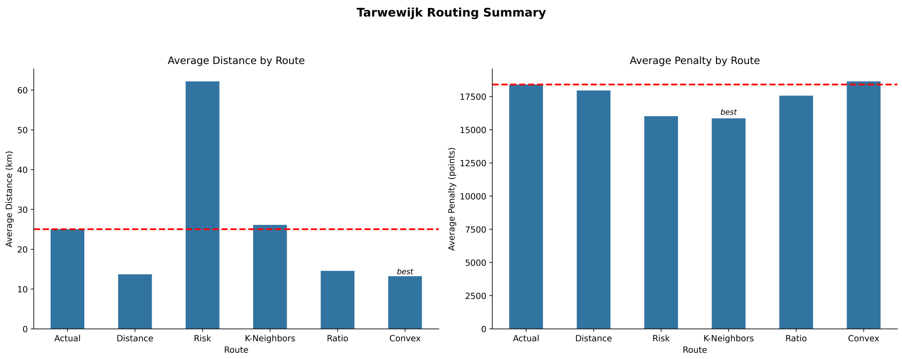


Figure C.11: Tarwewijk routing summary

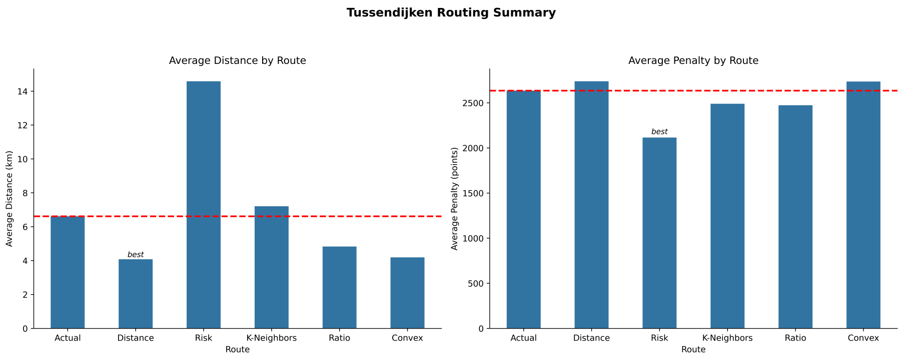


Figure C.12: Tussendijken routing summary

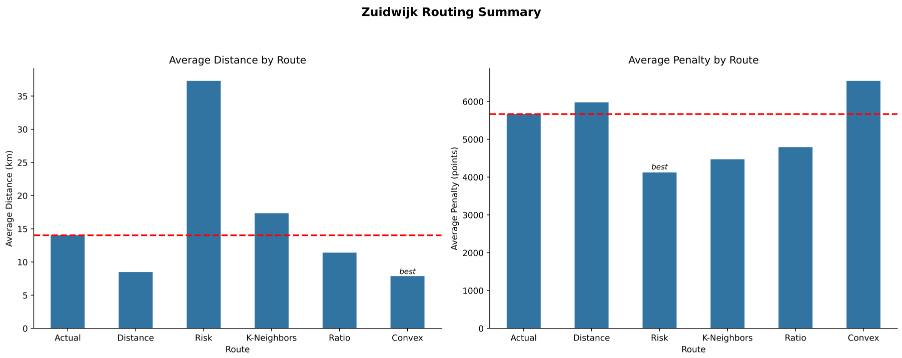


Figure C.13: Zuidwijk routing summary