UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR IT-SICHERHEIT

# Evaluation of a privacy preserving smart city simulation based on Cities:Skylines

# *Bewertung einer datenschutzfreundlichen Smart-City-Simulation auf der Grundlage von Cities:Skylines*

Bachelorarbeit

verfasst am
Institut für IT-Sicherheit

im Rahmen des Studiengangs
IT-Sicherheit
der Universität zu Lübeck

vorgelegt von
Peter Hodgkinson

ausgegeben und betreut von
Prof. Dr. Esfandiar Mohammadi

Lübeck, den 23. November 2022

## Eidesstattliche Erklärung

*Ich erkläre hiermit an Eides statt, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.*

Peter Hodgkinson

## Zusammenfassung

Nachhaltige Optimierungen und intelligente Lösungen, die unser tägliches Leben und unsere Infrastruktur betreffen, scheinen häufig im Konflikt zum Datenschutz zu stehen, da Optimierungen dieser Art oft datenbasiert sind. Insbesondere in Städten besteht Optimierungsbedarf, um den Anforderungen der Bürger*innen gerecht zu werden und um sich auf den Klimawandel vorzubereiten. In dieser Arbeit möchte ich die Möglichkeiten datenschutzfreundlicher Simulationen untersuchen, die eine Möglichkeit darstellen, dem Konflikt Privatsphäre versus Optimierung zu begegnen. Simulationen werden häufig als Instrument eingesetzt, um die Auswirkungen einer bestimmten Veränderung oder Verbesserung zu untersuchen oder um verschiedene Szenarien zu testen. Was-wäre-wenn-Fragen definieren solche Szenarien und sind eine Hauptmotivation für die Simulation von Städten. Das Computerspiel Cities:Skylines wurde als Simulationsgrundlage verwendet, um den Stromverbrauch von Gebäuden in der Stadt Lübeck zu simulieren. Ziel der Arbeit war es, diese Simulation in eine Privatsphäre erhaltende Simulation zu transformieren. Entscheidend dafür war die Entwicklung einer maschinellen Lernmethode, welche die Anpassung von Stromverbrauchswerten von Gebäuden an realistische Werte ermöglichte sowie die Integration eben jener Methode in Cities:Skylines. Die Simulation wurde in zwei Experimenten evaluiert und die Ergebnisse mit einem hypothetische Datensatz verglichen. Im ersten Experiment wurde die allgemeine Genauigkeit bewertet, während im zweiten Experiment die Leistung dieser Simulation in einem Was-wäre-wenn-Szenario untersucht wurde. Mit dieser angepassten Simulation konnte ein durchschnittlicher Fehler von 4 % im Allgemeinen und von 1 % in einem spezifischen Szenario erreicht werden. Die Ergebnisse zeigen Möglichkeiten auf, Szenarien für Städte unter Wahrung der Privatsphäre zu simulieren und demonstrieren, dass es keinen Konflikt zwischen Privatsphäre und Fortschritt geben muss.

## Abstract

Sustainable optimisations and smart solutions concerning our daily life and infrastructure often seem at odds with privacy due to the data-driven nature of these optimisations. In cities in particular, there is a need for optimisation in order to meet the demands of citizens and to prepare for a changing climate. In this work, I investigate the possibilities of privacy preserving simulations which offer a way to overcome the conflict between privacy and optimisation. Simulations are often used as a tool to evaluate the outcome of a certain change or improvement, or to test different scenarios. What-if questions define such scenarios and are a main motivation for the simulation of cities. The computer game Cities:Skylines was used as a simulation base to simulate the electricity consumption of buildings in the city of Lübeck. The aim was to transform the simulation into a simulation able to preserve privacy. A decisive step towards a privacy preserving simulation was the development and integration of a machine learning model into Cities:Skylines, so as to allow the adjustment of the electricity consumption of buildings towards more realistic values. The setup was evaluated and tested against a hypothetical ground truth in two experiments. The first experiment evaluated the general accuracy while the second experiment dealt with the performance of this setup in a what-if-scenario. It showed, that an average error of 4 % in general and of 1 % in a specific scenario could be achieved with this setup. The results indicate opportunities of simulating scenarios for cities in a privacy preserving way and demonstrate that there does not have to be a conflict between privacy and progress.

# Contents

# 1

# Introduction

Simulations are an emerging tool for analysing hypothetical scenarios. Climate change makes adequate preparation imperative, particularly in urban areas. Extreme weather events due to climate change can threaten human health and infrastructure especially in cities where billions of people live and work [19]. The smart city concept, which combines transformations in a city with digital technologies, can be a guiding principle towards a more sustainable and prepared future city. Simulations can be a valuable tool to evaluate the outcome of potential adaption or mitigation strategies to climate change in cities and therefore they are part of the smart city concept. As an example for possible scenarios, so called what-if questions like the following are a main motivation for simulations of a city:

What will happen to a power grid if all households in a certain area install a heat pump? Can the power grid withstand the higher load? What will happen if every citizen has an electric car? Is it possible to achieve a static energy base load when all cars are used as a storage?

The answers to these questions are important and it would make a difference if we had a tool to produce these answers. However, it is important to respect individual privacy. A tool which is able to answer what-if questions normally needs quite a lot of personal data to calculate the outcome of a decision based on the way people behave. Collating personal data can be problematic because political systems can change and future authorities could use the gathered data to the detriment of their citizens. Also in a less dramatic way, that kind of data could be used by companies to improve their advertising, which is exactly what is happening right now [9]. At least in the European Union, there are also many data protection regulations that must be complied with [8]. These facts make privacy preserving methods both necessary and attractive.

To conclude, we need to develop simulation tools which, for instance, evaluate various ways of adapting to climate change, but we should do it whilst still upholding individual privacy, which means that we need effective privacy preserving tools. In this thesis, I am investigating the accuracy of a privacy preserving simulation based on the computer game Cities:Skylines.

## 1.1 Contributions of this Thesis

The aim of this thesis is to extend an existing simulation, so that it can function as a privacy preserving simulation and to investigate its accuracy. I used a simulation of Lübeck (Germany) based on Cities:Skylines, a city simulation computer game. The chosen parameter to research the accuracy of the simulation was the electricity consumption of buildings. The constant guiding principle for the transformation towards a privacy preserving simulation was to what extent what-if-scenarios could be successfully designed and implemented. There were four major problems to solve.

First of all, a dataset needed to be designed which had to fulfil three conditions. First, it had to suite the existing simulation, because it was used as reference in a first experiment. Secondly, the electricity consumption values for every single building and hour should produce a realistic curve. Consistency is here more important than scaling. And lastly, everything had to be done with a focus on privacy. As part of the second challenge, a strong machine learning model had to be developed to improve and to extend the simulation. This machine learning model was used to make the electricity consumption of buildings in the simulation more realistic. Addressing the third challenge, the machine learning model had to be integrated into the simulation.

The target for the first experiment was to evaluate the accuracy of the improved simulation. For this purpose, the electricity consumption value for each building was recorded and compared against the values of the generated dataset. The Mean Average Percentage Error (MAPE) of this comparison was around 4 %.

In the second experiment, the accuracy was evaluated in further detail. The aim here was to discover the possibilities of the simulation for a what-if-scenario. The design and implementation of a what-if-scenario are part of the fourth challenge. The example under consideration here was, what would happen if every house in a certain area uses a heat pump? The investigation focused on the accuracy of the simulation at this level of detail. The electricity consumption of the selected houses with a heat pump was compared with the electricity consumption value of the same houses with no heat pump. It showed that the simulation is able to simulate a scenario in such detail. A percentage difference of 1 % of the comparison supported this result.

Both experiments were carried out with two different machine learning models. A comparison of these models showed that a high difference which was found in the accuracy of the models was not reflected in the results of a simulation run.

## 1.2 Related Work

Santana [23] developed an extensible open source large-scale traffic simulator for smart city scenarios. The *InterSCSimulator* is able to simulate big metropolises with more than 20 million actors. Santana focused on the simulation of traffic and the public transport systems and managed to obtain some useful and valid results. Moreover, the *InterSC-*

*Simulator* was used in many projects in the field of smart cities. This work is a promising example for the effectiveness of simulations in the context of smart cities. However, while focusing on large transportation systems, the investigation of a privacy preserving simulation was not the intention of this work.

Olszewski et al. [17] used Cities:Skylines to model a region in Poland which had several ecological and social problems. The area of interest is home to many poultry and pig farms which are a burden for the residents. An over-fertilised soil and an extremely offensive odour being just two of the problems caused. One of the author's aims was to simulate the impact of a biogas plant on the environment and the community. The residents of the investigated area tend to have a sceptical opinion of the local government, so there was interest in finding a realistic model, which could not only be used to try out various options, but which could also be accessible to the public and so help to increase trust in local authorities. This work provides a real-world simulation which explains, due to the good quality of Cities:Skylines graphics, problems and the emerging solutions to the general public. But, in spite of its advantages, this work does not take privacy preserving methods into consideration.

Duncan et al. [6] researched in the field of city planning in their Bachelor thesis. They developed a proof of concept tool to optimize the layout of a city. A machine learning model which learned to play Cities:Skylines was built. The authors were able to optimize the decisions of the machine learning model with a reward function to improve the layout of the played game. This work combines Cities:Skylines with a machine learning model, but with a different city simulation purpose in mind. The research is more concerned with the design of a city than with the simulation of what-if-scenarios.

These works demonstrate the usefulness of simulations to optimize a city and the effectiveness of Cities:Skylines as a basis. My work is an attempt to combine these promising results with privacy preserving methods and in doing so, help to close this research gap.

## 1.3 Structure of this Thesis

This thesis consists of five main chapters. Chapter two sets the ground for a common understanding. The concept and some ideas of smart cities are explained and presented. Further on, Cities:Skylines, a computer game, is introduced which is the framework for this work. Chapter three outlines the transformation of a simulation based on Cities:Skylines to a privacy preserving simulation. The importance of privacy preserving methods is discussed as well as the challenges that had to be solved for a successful transformation. Chapter four handles these challenges in more detail. The generation of a suitable dataset, the configuration of a machine learning model to improve the simulation and the integration with Cities:Skylines are explained. The next two chapters deal with the experiments that were carried out to check the accuracy of the simulation. The first experiment evaluates the accuracy of the simulation in general and the second experiment goes further into detail while researching the possibilities of a what-if-scenario. Further on, the results are discussed and summarized.

# 2

# Background

As mentioned in the introduction, in order that they can continue to provide liveable surroundings, our cities need to prepare for climate change. One way to optimize and prepare a city can be the concept of smart cities. It combines smart technology with city planning. The concepts and basic ideas are introduced in the next section. Further on, Cities:Skylines, the basis of this work, is explained as well as all the necessary steps to turn a computer game into a serious simulation to predict the electricity consumption of buildings.

## 2.1 Smart Cities

Cities are constantly growing and until now they are home for more than half of the worlds population and this will probably increase to 70% in 2050 [16]. Western societies also face demographic changes towards an older population [16]. Simply because of their reliance on a constant power supply, it is cities, rather than the rural areas, which will be the most susceptible to damage following the forthcoming extreme weather conditions [19], which is why it is so necessary that steps be taken now, so as to be able to maintain a reasonable standard of living for city dwellers in the future.

Smart cities are a loose concept without a precise definition. All developments towards a more sustainable and integrated city in combination with smart technology are part of the smart city concept. One idea to improve cities and their traffic system is to make parking spaces more intelligent. An intelligent car park can tell whether it is used or not. This functionality combined in an app allows you to reserve a car park and navigates you towards the free place on the most efficient way. An intelligent parking system can reduce air and noise emissions because people do not have to search for a free car park any more [5].

Another idea for a smarter city deals with the waste collection system of a city. An intelligent waste bin notices when it is full and alarms the waste collection company. This has several benefits. First, the waste collection company does not have to check the waste bins regularly and secondly they can optimize their routes because it is predictable now how fast the truck will be full. This system has the potential to save time, manpower and petrol [7].

Smart buildings take care of light and temperature in each room. This means a smart building notices when people are in the room and light and temperature will be adjusted automatically [27]. These opportunities have the potential to reduce the energy consumption of a building and therefore are part of the smart city concept.

Improving communication with local authorities can also be part of the smart city idea. For example, everything you previously had to go to the town hall for can now be done online. Also, the participation of citizens in political decision making could be optimized to enhance trust in local authorities [14].

But all these solutions can be a risk for the security and privacy of a person as Sookhak et al. [25] pointed out. Internet of Things (IoT) devices are a key technology for a smart building or a smart city, but they are also an entry point for attackers. Therefore, secure solutions are very important for the implementation of smart city ideas. Inibhunu et al. [13] evaluated a highly scalable and secure smart city framework. Oliveira et al. [16] introduced the concept of human smart cities to improve the acceptance of the population for upcoming changes in a city. They pointed out that the smart city concept was developed by industry without taking the opinion of local authorities and the general public into consideration, and found little acceptance. The authors propose a smart cities solution where the emphasis lies in creating a healthier and happier environment for citizens, so as to help counteract their reticence in accepting necessary change.

In contrast to all these ideas and solutions, this thesis deals with a different aspect of smart cities. I am investigating how accurate a simulation of the electricity consumption of buildings in a city can be without using personal data. This also fits into the context of smart cities, because the results could help city planners or the city administration to improve and optimize a city. This work focuses on a tool that could help in optimizing cities in general, rather than on a specific project or issue such as smart bins.

## 2.2 Cities:Skylines

Cities:Skylines is a city building simulation computer game developed by Colossal Order and published by Paradox Interactive in 2015 [3]. The player is in the position of the first city planner or the major. The game is intended to be as close to reality as possible. Not only does it have vivid animations and 3D effects, but it also puts an emphasis on details such as water and electricity infrastructure like pipes and cables. The game provides information about consumption data which is important for this work.

The implemented simulation logic of the game is responsible for the behaviour of everything in the city and thus also for the behaviour of the citizens. Citizens go to work, they go shopping and they enjoy their free time in a nearby park. This indicates a quite realistic simulation and as already mentioned, Cities:Skylines produces beautiful animations and images, as shown in Figure 2.1 for instance, but this is not decisive for a simulation which preserves privacy. The ability to produce results without a huge development effort is however much more important. Cities:Skylines is useful to demonstrate the possibilities of a simulation in the context of smart cities and how it could benefit the privacy of citizens. But besides that, the usage of Cities:Skylines has some limitations.

The reason lies in the purpose of this game itself. It is a computer game, developed for entertainment and not for scientific work. It is also closed source and that is why it has to be treated as a “black box”. Sometimes it is not possible to understand or to prove why the game behaves as it does.

Figure 2.1: Screenshot of the city of Lübeck in Cities:Skylines.

### Modifying Cities:Skylines

It is possible to interact with the game Cities:Skylines and to modify it’s behaviour. Modifications or in short mods are compiled C# libraries which hook into the game code [4]. One way of writing a mod is to use the official modding API that offers programmers a way to overwrite and to extend game features. The other way uses C# reflections to overwrite game functions. The modifications rely on C# because this is the language the game is written in. Writing a mod can be a challenging task. One reason for that is the lack of a good documentation and the other reason is the closed source code of Cities:Skylines. Because of these difficulties, the development of mods is best described as a try and error process. Zeiseweis [28] did some outstanding research on how to write a mod, how to extract data and how to inject and change data in Cities:Skylines.

The overall authenticity and the ability to modify the game logic make Cities:Skylines a reasonable choice as a framework for the simulation of a city. The next section goes further into detail and highlights the importance of modifications.

## 2.3 Cities:Skylines as a Simulation

In order to use Cities:Skyline as a simulation, several steps were necessary. The steps to transform a computer game into a serious simulation are explained in the following.

Richter [22] examined in his Bachelor thesis what kind of framework best suits the requirements of a smart city simulation. He concluded that Cities:Skylines is a reasonable choice because of the overall authenticity and the ability to change the behaviour with modifications. He imported the city of Lübeck with the GeoSkylines Mod [11] into the game and developed a game runtime in which no interaction is needed. This is an important step towards a functional simulation because Cities:Skylines purpose as a computer game is to entertain the player and so different challenges like fires and floodings are unwanted in a simulation. A list of modifications which are used to break the game logic can be found in his thesis. The actual simulation is a save game and when I talk about the simulation I always refer to this save game.

Another important modification towards a serious simulation is the RealTimeMod [21]. The RealTimeMod is used to make the in game simulation more realistic. For example, citizens go to work or to school now in the morning and to bed at night. Without the RealTimeMod there is no logic to this behaviour. So the citizens could go to work or to school in the middle of the night. The RealTimeMod is also responsible for things like the typical rush hour effect and other more realistic behaviour of citizens. For instance, citizens go out for lunch now, so they need some commercial buildings around. They also go shopping in their free time just for fun and they try to protect themselves from rain and storm. All these aspects can have an impact on the electricity consumption, therefore the usage of this mod is obligatory for a realistic and serious simulation.

Zeiseweis [28] did some research about extracting and injecting data out of and into the simulation. In his Bachelor thesis, he developed two modifications, one to extract and the other one to inject data into Cities:Skylines. This is an important contribution for a running and practical simulation. With these modifications it is possible to record the results and to improve some parameters like the electricity consumption of buildings. Zeiseweis also examined whether a correction of parameters in the simulation actually works and how good it can be.

In a following project Zeiseweis and myself did some research on the correction of in game parameters [29]. We improved the correction process through the interaction with a machine learning model. We developed a machine learning model and a modification which connects the simulation with the model. The purpose of the machine learning model was to predict a more realistic electricity consumption of a building based on parameters out of the simulation like time, date, number of residents, size of the building and district. We had to use a local server for the communication between the machine learning model and Cities:Skylines. Although this influenced the overall performance of the simulation, it also gives space for improvements due to the usage of a standardized protocol. It is very easy now to change the machine learning model. The results were mixed due to an error that caused a large difference in scaling.

These works showed that it is possible to use Cities:Skylines as a simulation. The project by Zeiseweis and myself in particular demonstrate that it is possible to use this framework for a serious use case. In this work I focus on the electricity consumption as an example to test the accuracy of a simulation. However, the idea is that the simulation could also be used for other parameters such as water consumption or traffic density, depending on the subject of investigation. My goal is to explore the accuracy of a simulation that does not use personal data.

# 3

# Problem Statement

The last chapter gave some background information about the computer game Cities:Skylines and how it can be used as a simulation in the context of smart cities. In the following chapter it is explained why privacy preserving methods are important and further on I will point out the challenges of this work.

## 3.1 Why Privacy Preserving Methods Are Important

Online services like social media, search engines, mail etc. are free to use for everyone because personal data like interests, age, gender, movement profiles or habits count as a payment method nowadays. The business model of big tech companies (Meta, Google) works in a way where the user pays with his*her personal information to use a free service and the companies countermove is to offer the possibility of personal advertisement to other companies [9]. This business model is probably more annoying than dangerous for citizens of a constitutional state, but political systems in general are not known for their continuity, so it could happen that a government uses this bundled personal data against their citizens. At least in the European Union, there are some rules on the protection of personal data that restrict their use [8]. This is good, as it supports the need for safe tools and privacy preserving methods.

Data, and in particular personal data, is not only used to improve the potential of advertisement, it is also very important for optimisations of all kinds. When there is more information about citizens' behaviour, it is easier to optimise the environment and reduce energy consumption. Suitable ideas like the smart building or the smart parking system were discussed in section 2.1. This could lead to a conflict between the protection of personal data and the need for optimisations due to the climate crisis for example. A simulation of a city can be a useful tool for planners and energy providers. The simulation helps to plan and to design a new district or to react to a new reality where every citizen has an electric car. A typical way to build a city simulation would be to train a neural network which requires a huge amount of personal data. Therefore, sensors must be installed all over the city and a surveillance of the citizens cannot be ruled out. Furthermore, a reconstruction of personal data out of a trained machine learning model is possible [1].

Due to the challenges we are facing at the moment, we cannot neglect the possibilities of optimisations. We need optimisations, but we should protect our personal data because we do not know how a future political system would handle this information. Moreover, we have to ask ourselves whether and to what extent a company is allowed to process our personal data. This is why we have to find alternatives to data intensive methods.

## 3.2 Challenges of this Work

This work evaluates the possibilities of a privacy preserving simulation in the context of smart cities. Further, I am investigating the accuracy of a privacy preserving simulation and its overall flexibility. The challenge is to achieve a certain level of detail to investigate and to answer so called what-if questions which are a possible use case for a simulation.

As outlined in the last chapter it is possible to use Cities:Skylines as a simulation, but in order to use it as a privacy preserving simulation there are some challenges to overcome. In this work, privacy preserving or privacy friendly is defined in such way that no conclusions can be drawn about individuals. And to make it even easier, if no personal information is used, no conclusions will be possible, and this therefore leads to a simulation which preserves privacy. This of course is a very basic understanding of privacy and of a privacy preserving simulation, but it suites the needs. With these definitions taken into account, there are four challenges to overcome in order to transform the simulation based on Cities:Skylines into a privacy preserving simulation.

1. A privacy friendly dataset is needed because it is the basis for the experiments. Each building in the simulation is listed in the dataset with a more realistic electricity consumption value compared to the initial simulation. The dataset is also used to train the machine learning model and to serve as a ground truth when evaluating the experiments. The ground truth is to be understood as the defined reality for an experiment.
2. The machine learning model is trained on the privacy friendly dataset and is used to adjust the simulation towards more realistic results compared to the plain simulation. In this work, the machine learning model will predict the electricity consumption value of the buildings in the simulation.
3. The machine learning model has to be integrated into Cities:Skylines to adjust the simulation and to fulfil its purpose. This integration is important because it opens up the opportunity to define different realities for the simulation, and also to simulate different scenarios.
4. What-if questions define a specific scenario. These scenarios are the main motivation for a privacy preserving simulation of a city. The answers to these questions could help us to improve our cities towards a more sustainable future. Therefore, the design of such question has to be done with great care. There are two approaches to create a what-if-scenario. First, the simulation is changed during a simulation run, e.g. all residents of a district now have an electric car. Secondly, the underly-

ing reality and more specifically the ground truth dataset is changed. Changing the simulation during a simulation run is not possible within the scope of this work, so the second option was chosen, and this is why the dataset and the machine learning model needed to be updated for the defined scenario.

The first three challenges are explained and discussed in chapter 4. The fourth challenge is discussed in the second experiment in chapter 6 while the first experiment in chapter 5 evaluates the overall accuracy of the improved simulation. My aim in this work is to demonstrate that there does not have to be any conflict between innovation and privacy and that a simulation can help to build a resilient and sustainable city.

# 4

# Approach

In order to deploy a privacy preserving smart city simulation in a real-world scenario, four challenges must first be overcome. After having clarified these challenges in the last chapter, this chapter explains how the first three challenges were handled. First, the expectations on a privacy friendly dataset and its generation are discussed; secondly, the creation of a strong and adaptable machine learning model is explained; and thirdly, the integration of the machine learning model with Cities:Skylines is addressed.

## 4.1
## Challenge 1: Generating a Privacy Friendly Dataset

A dataset fulfils privacy requirements when it is not possible to make conclusions about an individual person. This is of course a very simplified and basic understanding of privacy, but it will do as a requirement for the dataset in this work. The dataset is used for the training of the machine learning model and as a basis for the evaluation of the simulations accuracy. Due to the lack of publicly existing data about Lübeck regarding the number of buildings and the electricity consumption of a building, I had to generate a synthetic dataset with hypothetical consumption values. I formulated some assumptions that the dataset should meet:

- The dataset has to suite the simulation. The number of buildings and the number of citizens should be the same.
- The buildings in the dataset should have a realistic electricity consumption. That means, there should be differences between night and day and between workday and weekend.
- An electricity value is provided for every hour.
- Every building is interpreted as a residential building.
- If the building is not occupied because it is a commercial or an industrial building in the simulation, it will have the same consumption as a building with one person living in. This shall symbolize the basic load a building normally has.

A developed mod (CountBuildingMod) extracts statistical data out of the simulation which serves as a basis for the dataset. The dataset now contains information about every building in the simulation like the size of the building and the number of residents

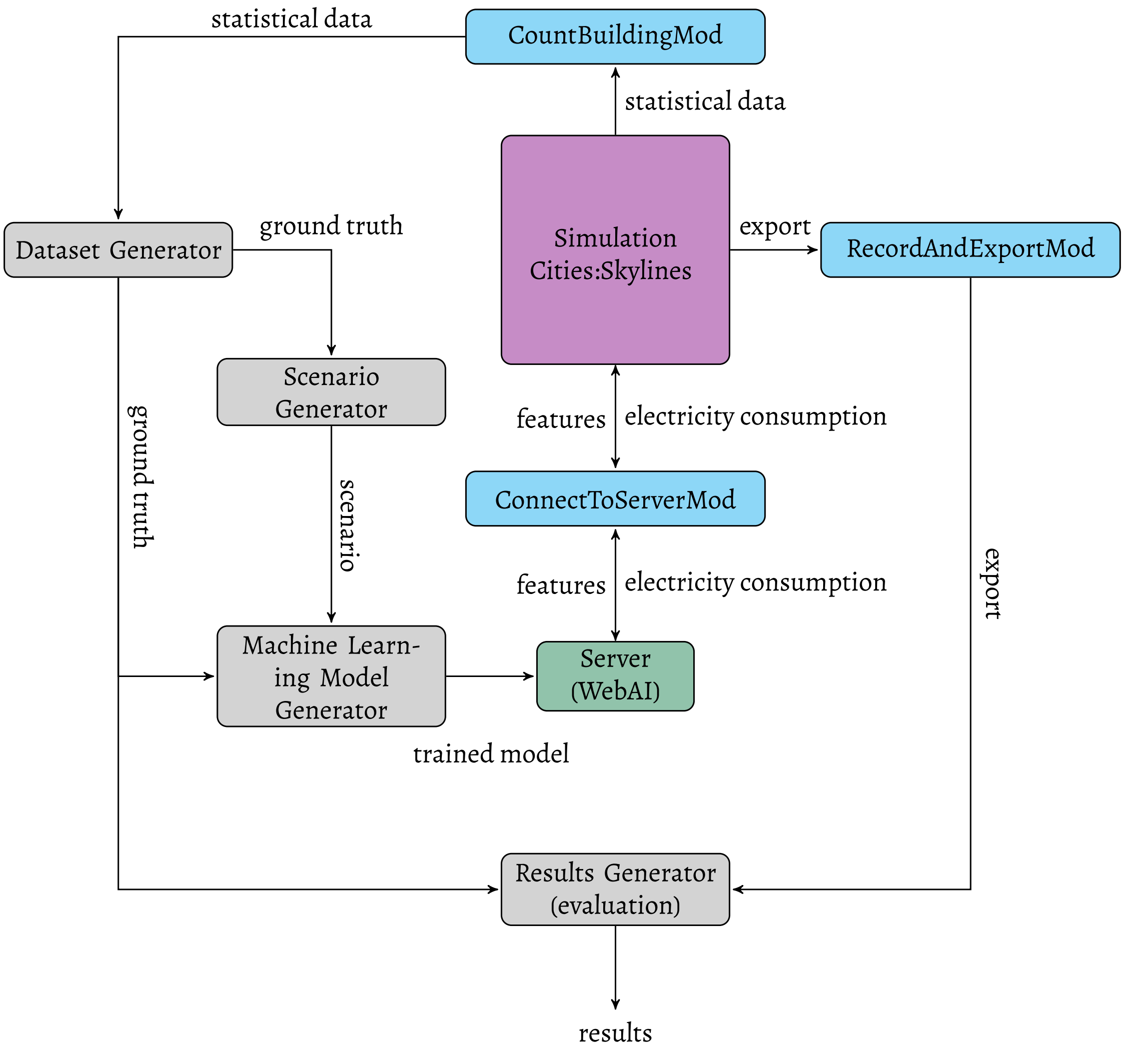


Figure 4.1: System setup. This figure shows how all tools and units work together to serve as a privacy preserving simulation. The CountBuildingMod, RecordAndExportMod and ConnectToServerMod are developed C# modifications for the computer game Cities:Skylines. The Generators are developed Python scripts and the Server is a Python web server that provides the machine learning model.

in general, as well as the number of residents in different age groups. The simulation is designed to simulate a week in September and therefore every building needs 168 entries in the dataset because one week has 168 hours. Table 4.2 gives an example of the important parameters for the calculation of the electricity consumption which was done with a *Standard Lastprofil* of the Bund deutscher Energie Wirtschaft (bdew) [2]. A Standard Lastprofil is a representative method to calculate the electricity consumption of a certain user group if some older data is missing. A Lastprofil provides normalized electricity values for every 15 minutes (the sum of all the 15 minutes consumption values of a year is bounded to 1000 $kwh/a$). The data is separated between workday and weekend and also between winter, summer and the rest of the year. The Lastprofil for the user group Haushalt (household) was suitable for the calculation of the electricity consumption. These profiles are the base for only one person so the calculated electricity values are multiplied for this dataset by the number of residents. Every electricity value was multiplied by 1.5, to denormalize the values from the Lastprofil and to achieve an electricity consumption of an average person [26].

Table 4.2: This table shows the necessary parameters and some example data for the calculation of the electricity consumption with a Standard Lastprofil.

| Time | Weekday | DayOfYear | PersonInBuilding |
|---|---|---|---|
| 15:00 | Saturday | 271 | 11 |
| 17:00 | Sunday | 272 | 4 |

Compared to reality, the dataset and the assumptions I formulated are quite inaccurate, but I only want to investigate how well I can fit the simulation to a given reality. This means, the overall accuracy of the dataset compared to reality is not so important. It is more important, how well it maps the simulation. In other words, it is not important if the consumption value is high or low or if it is interpreted as kilowatt or megawatt. However, kilowatt or kilowatt hour are the units used for the electricity consumption in this work.

Figure 4.3 shows an exemplary electricity consumption curve of a building with 25 residents in the district of St. Jürgen which was generated with the Standard Lastprofil. The assumptions, such as a visible difference between day and night or working day and weekend, were met. Figure 4.4 informs about the distribution of the electricity consumption values in the dataset. The values are not evenly distributed, since 1 527 932 values are equal or less than 5 000 $kw$ in a range from 53 $kw$ to 28 459 $kw$.

Table 4.5 lists all the available parameters in the dataset. These parameters were chosen and extracted from the simulation with challenge four, the what-if-scenarios, in mind. The coordinates for instance are important for the second experiment to determine whether a house is in a specific area or not. The demographic data can be used in a future what-if-scenario. For now, the electricity consumption values in the ground truth only depends on the parameters shown in table 4.2. As an idea to make the dataset even more economical in terms of data usage, some average distribution of residents and buildings in a district could be applied as long as the number of citizens and buildings is still the same, but this investigation will have to be the subject of a future work.

To conclude, this dataset is the base for the machine learning model, so it has to be

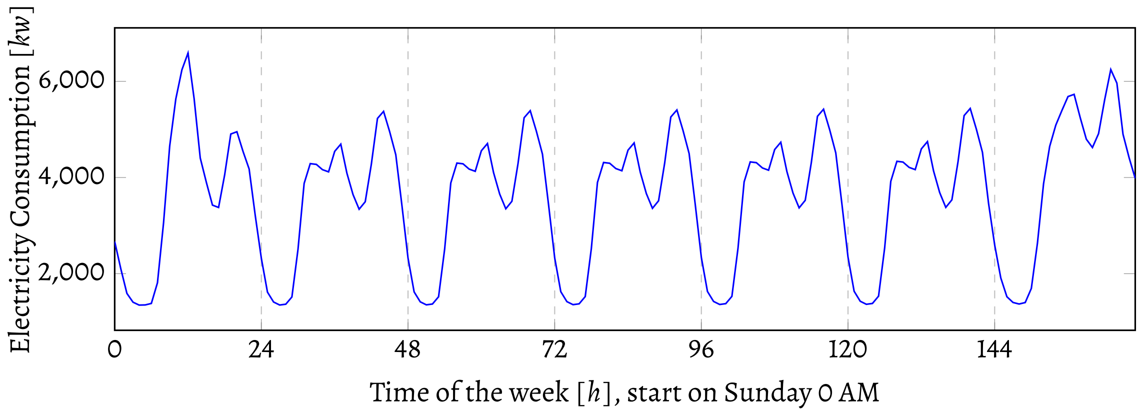


Figure 4.3: Example of a hypothetical electricity consumption of the synthetic dataset for a house with 25 residents in the district of St. Jürgen. The differences between day and night and between weekday and weekend are clearly visible.

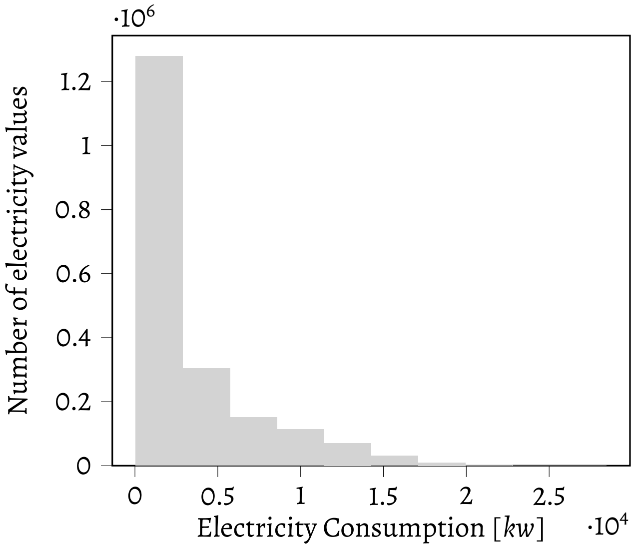


Figure 4.4: This histogram shows the distribution of the electricity consumption values in the dataset. In total there are 1 960 728 values. The data is not evenly distributed since the electricity consumption of 1 527 932 buildings is less or equal than 5 000 *kw*. The maximum consumption is 28 459 *kw* which means that 22 % of the values are distributed in the range from 5000 *kw* to 28 459 *kw*.

created carefully. With regard to the number of buildings, the number of citizens etc., the dataset must be comparable with the simulation. The usage of a Standard Lastprofil for a realistic electricity consumption is not so important for the experiments it can rather be seen as a step towards a more realistic solution.

Table 4.5: All parameters which are used for the training of the machine learning model.

| Parameter | Explanation |
| --- | --- |
| buildingId | Id of the building. |
| time | Time of the day. |
| buildingSize | Size of the building. |
| x | x coordinate of the building. |
| y | y coordinate of the building. |
| z | z coordinate of the building. |
| personInBuilding | Number of persons living in the building. |
| child | Number of children living in the building. |
| teen | Number of teens living in the building. |
| young | Number of youngs living in the building. |
| adult | Number of adults living in the building. |
| senior | Number of seniors living in the building. |
| weekday | Either of Saturday, Sunday or workday. |
| district | District of the building. |
| electricity | Electricity consumption of the building. |

## 4.2 Challenge 2: Using a Machine Learning Model to Improve the Simulation

Cities:Skylines is not able to generate a realistic electricity consumption curve and the logic behind the given values is unclear. Cities:Skylines therefor needs to be adjusted, so as to be able to produce useful results for a what-if-scenario. A machine learning model is trained with the ground truth dataset and then integrated into Cities:Skylines in order to predict the electricity consumption of buildings. After a short summary of the process of choosing the right machine learning model, the requirements for the machine learning model are explained and its accuracy is evaluated.

### Choosing a Machine Learning Model

The machine learning model has to predict the electricity consumption of buildings to improve the simulation. The electricity consumption depends on several parameters like time, day and number of residents in a building. This leads to a regression problem which is an easy to solve problem for a machine learning model.

First, I evaluated and investigated the capabilities of a neural network. Neural networks are a powerful technology, and they are able to solve a wide range of problems. With regard to future what-if-scenarios, which make the dataset more complex as the buildings' electricity consumption then depends on more than just the day, time, and number of residents, I decided to use a neural network in order to handle these more complex scenarios. During the development process, I changed the machine learning model to a *Gradient Boosted Decision Tree*. This decision tree produced better results than the neural network and the development effort was radically minimized. Still, I will use the developed neural network to compare the results and the performance of the the two machine learning models.

### Requirements of a Decision Tree for the Prediction of the Electricity Consumption of a Building

I decided to use the *scikit-learn* library [24] for the Python programming language to develop a decision tree which solves a regression problem. *scikit-learn* is an open source library and widely used in the data science community. The library is very straight forward to use and does not require a huge implementation effort. The decision tree has to investigate and to understand the relationship between all the features listed in table 4.5 and the corresponding electricity consumption value. The decision tree already produced very accurate results without any optimisation. However, two parameters were used to further optimise the decision tree. Both parameters, the number of trees and the depth of a tree, have an influence on the accuracy.

After some training rounds with different assignments of the parameters `max_depth` and `n_estimators`, I used a configuration of 150 trees with a depth of 7. All other parameters were left in the default configuration. RepeatedKFold cross validation was used for the evaluation of the decision tree. KFold means that the dataset is divided into $k$ folds. Then $k - 1$ folds are used for training and one fold is used for validation. This is done $k$ times. RepeatedKFold demands that the KFold process is repeated $n$ times.

### Accuracy of the Gradient Boosted Decision Tree

The gradient boosted decision tree is designed to solve regression problems which is why less optimisation has to be done compared to a neural network in order to develop a decision tree with a high accuracy.

Statistical measurements like the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) were used for the evaluation of the accuracy of the decision tree. Let $a$ be the suitable label, $y$ the predicted output and $n$ the number of samples. According to Fomby [10], the measurements are defined as follows:

$$\text{MSE} = \frac{1}{n}\sum_{j=1}^{n}(a_j - y_j)^2$$
$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(a_j - y_j)^2}$$
$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}|(a_j - y_j)|$$

These are the scale-dependent measures. The MAE is the average difference between the predicted value $y$ and the label $a$ over all $n$ samples. Because of the squared difference, the MSE takes care about the outliers which are very bad predictions. This means, high deviations between simulated values and ground truth have a higher impact on the average error value. The RMSE, which is the square root version of the MSE, shows, compared to the MAE and due to its relationship to the MSE, a higher sensitivity towards the outliers. A low value in this context states a higher accuracy of the predictions and this means that the decision tree managed to understand the connections between the features and the labels quite well. But for a better classification of the results, the scale-independent versions of the MAE, the Mean Average Percentage Error (MAPE) [10], and the Root Mean Square Percentage Error (RMSPE) [10] are very useful.

$$\text{MAPE} = \frac{1}{n}\sum_{j=1}^{n}\left|\frac{(a_j - y_j)}{a_j}\right|$$
$$\text{RMSPE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\frac{a_j - y_j}{a_j}\right)^2}$$

Both values give a percentage when multiplied with 100 and for a high accuracy, the percentage has to be small. High numbers in the problem space can lead to high scale-dependent measures and this makes it difficult to understand and to rank the accuracy of the tree. This is why the scale-independent measures are so useful.

Figure 4.6 shows the results of the cross validation. The dispersion is low and the MAPE of the decision tree is $\approx$ 0.8 %. The decision tree was finally trained and tested on a divided train and test dataset which resulted in a MAE of 15 and a MAPE of 0.8 %. This machine learning model was used for the integration with Cities:Skylines and the simulation.

### Comparison of the Gradient Boosted Decision Tree and the Neural Network

Before I chose a decision tree as the machine learning model for the setup of this work, I evaluated the possibilities of neural networks. Python and the *Pytorch* library [18] were

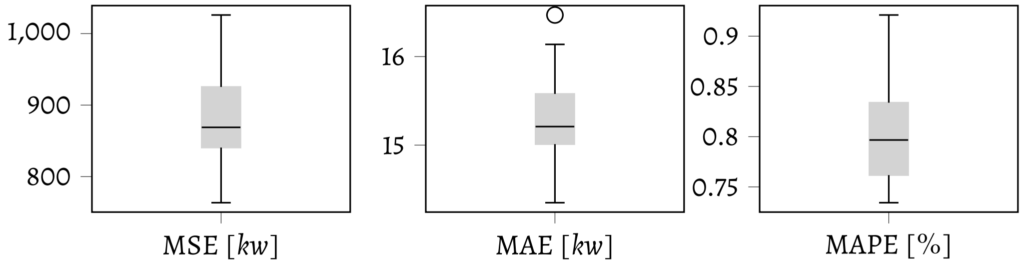


Figure 4.6: These boxplots show the results of the RepeatedKfold cross validation of the decision tree. $k$ was set to 5 and $n$ to 10 which resulted in 50 trained models. The Mean Absolute Percentage Error (MAPE) gives a percentage and ranks the Mean Absolute Error (MAE). The Mean Squared Error (MSE) is, under consideration of figure 4.4 which shows a distribution of possible electricity consumption values of the ground truth, rated as low.

used for the development of a neural network. *Pytorch* is an open source library developed by Meta and widely used in the scientific community. It is known for its flexibility and its optimisation for graphic cards.

A fully connected multi-layer network (Multilayer Perceptron MLP) with four linear layers was used to solve the regression problem. The accuracy of a neural network depends on parameters like learning-rate, batch-size, layer size and number of epochs. These parameters are called hyperparameters. Choosing a working combination requires experience and time. This is why I used a hyperparameter search algorithm to get a suitable combination. Further on, I managed to optimize the results of the hyperparameter search for a higher accuracy of the neural network. Finally, the batch size was equal to 64, the learning-rate was set to 0.001, and the layers were of size 128, 128 and 64. RepeatedKFold cross validation was used for validation purposes.

Figure 4.7 compares the MAPE of the cross validation of the neural network and the decision tree. It is clearly visible that the decision tree achieves higher accuracy and that the results of the different runs are not as widely distributed as the neural network's results.

There are two explanations for the difference in the accuracy of these two machine learning models. First, a decision tree is perhaps more suitable for regression problems than a neural network and secondly, my knowledge about optimizing neural networks is expandable.

## 4.3 Challenge 3: Integration with Cities:Skylines

The machine learning model has to be connected with Cities:Skylines to enhance the electricity consumption of buildings. The integration into the simulation environment was a challenging task that was originally solved in the project by Kilian Zeiseweis and myself [29]. We developed a mod (ConnectToServerMod) for Cities:Skylines that communicated

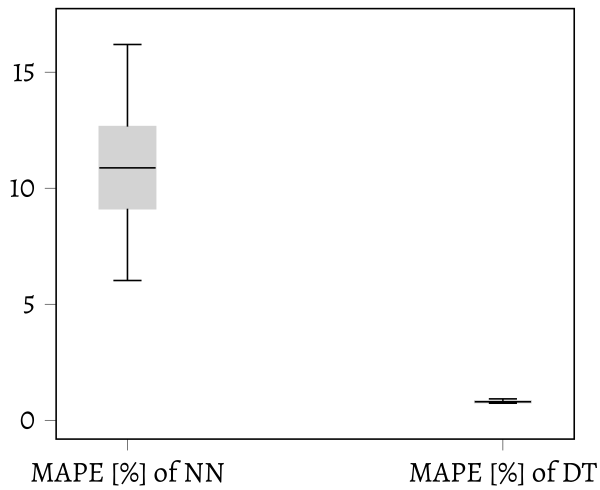


Figure 4.7: Comparison of the distribution of the MAPE score of the cross validation results of the neural network (NN) and the decision tree (DT).

with a server which provided the machine learning model we used in that project.

Only one specific method provides the possibility to change the electricity consumption of a building in Cities:Skylines. This method is the `HandleCommonConsumption` method which is invoked by the `BuildingAI`. The `BuildingAi` handles all the logic about buildings and is called in every simulation step of the in game simulation. Cities:Skylines uses an internal simulation to handle all important tasks. This internal simulation represents the game logic.

To change the electricity consumption of a building, the *Harmony* library [12] is needed. This library provides functionality to inject code into a desired method. We used the postfix functionality of the Harmony library to overwrite the electricity consumption value of the `HandleCommonConsumption` method. Our developed mod detects whether the `HandleCommonConsumption` method is invoked and then calls the postfix which makes a request to the WebAI and returns the predicted electricity consumption value. The postfix extracts all important data from the simulation and requests the machine learning model via a server. The server responds with the predicted value and this value is finally used to overwrite the electricity consumption parameter of the `HandleCommonConsumption` method. Figure 4.8 shows the communication of the different actors during a in game simulation step regarding the `HandleCommonConsumption` method.

This setup is flexible because it does not depend on a specific machine learning framework. In the project, we used the *ML.Net* Framework [15] for instance and in this work I used scikit-learn, Pytorch and Python to build a decision tree and a neural network. The downside of this approach is the speed. The usage of the http protocol is a bottleneck.

The integration of the machine learning model with Cities:Skylines was the third challenge and the final transformation step towards a serious environment for privacy preserving simulations. This chapter dealt with the process of creating a synthetic dataset for training a machine learning model and its integration into Cities:Skylines. Further on, the performance of an adjusted simulation is tested and evaluated.

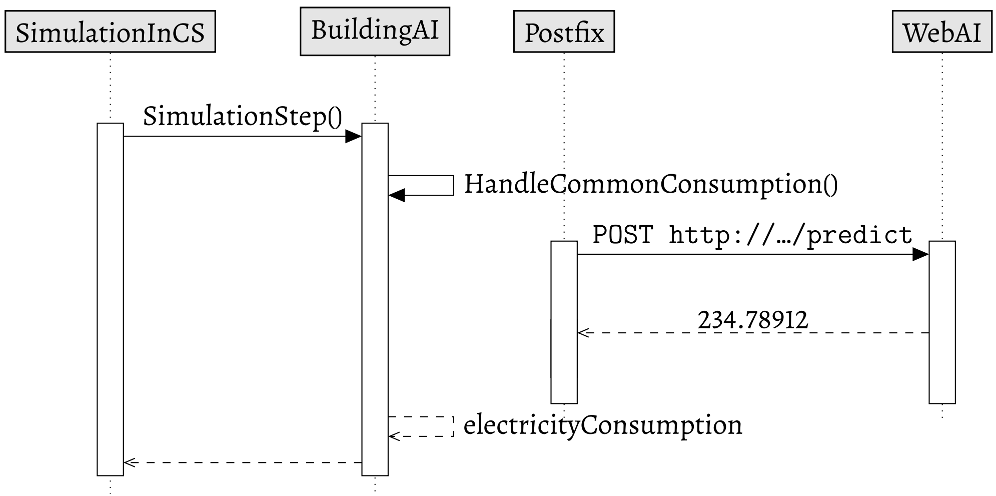


Figure 4.8: This sequence diagram demonstrates how the electricity consumption of a building is set. In every simulation step, the internal simulation calls the `BuildingAI` which is responsible to calculate the buildings electricity consumption. The postfix hooks into the call of the `HandleCommonConsumption` method of the `BuildingAI` and overwrites the electricity consumption value.

# 5

# Experiment I: Accuracy of the Simulation

Having discussed the importance of privacy-preserving methods and the challenges that had to be overcome to turn Cities:Skylines into a privacy preserving simulation, it is now time to evaluate the entire system. This chapter focuses on the accuracy of the simulation in general. The next chapter deals with a what-if-scenario and therefore takes a closer look at the performance of the simulation and its adjustment.

## 5.1 Test Setup

The purpose of the first experiment is to investigate the accuracy of Cities:Skylines as a privacy preserving simulation. This means, the electricity consumption of every building is recorded and compared to the ground truth dataset. To adjust the simulation based on Cities:Skylines to a privacy preserving simulation, four challenges had to be overcome (section 3.2). As described in section 4.1, a ground truth dataset was generated to train a machine learning model to further improve the simulation. See section 4.2 for more information. This dataset is also used to evaluate the simulation. To finish off, a Python web server was built around the machine learning model to provide its functionalities to the simulation.

The simulation run refers to an in game week in September. Over this week, the electricity consumption of every building is adjusted by the decision tree and recorded for a later evaluation. The web server is also logging its in and out going values to determine the impact of the simulation on the electricity consumption. I developed a mod (RecordAndExportMod) to record a simulation run and to export the data as `csv` file. Important parameters such as time, electricity consumption, district and number of residents are recorded. Parameters like coordinates and demographic data are also recorded with future what-if-scenarios in mind.

## 5.2 Results

During a simulation run, the electricity consumption of a building is recorded every hour as well as the number of citizens living in that building, the district and the position of the building to identify and to compare it with the values in the ground truth dataset. The recorded dataset now consists of 168 entries for every building. Due to some unknown behaviour of Cities:Skylines, not every hourly value is present in the recorded export, so the evaluation is based on the intersection of the ground truth dataset and the export.

### A Plain Simulation Run

First, a simulation run without the adjustments of the machine learning model was recorded to evaluate the accuracy of the plain simulation. The electricity consumption in all buildings in a district at a certain time was summed up to give a total consumption value for each hour per district. This approach is justified because it gives a clearer presentation of the results and a hypothetical comparison to realistic values of a local energy provider. This approach allows a comparison of eight districts instead of a comparison of 12 000 buildings.

Table 5.1 lists the comparison values of the evaluation. First, a district named Hamilton Square is listed. This district does not exist in Lübeck. It is not quite clear why Cities:Skylines added this district, presumably this district contains all the buildings that cannot be found in any other district. Secondly, the Mean Absolute Percentage Error (MAPE) is calculated, which is the scale independent version of the Mean Absolute Error. When multiplied with 100, it symbolises a percentage. This is useful to understand and to classify the MAE measurement because the values in table 5.1 seem high. In summary, a high MAE value can be well explained and estimated by the MAPE value. A high MAPE indicates a high MAE and a low accuracy, a low MAPE indicates a low MAE and a high accuracy. If the MAPE is high, the MAE is bad, if the MAPE is low, the MAE is good and the simulation has a high accuracy.

Table 5.1: Evaluation of each district and the whole of Lübeck in the plain simulation. The records of a plain simulation run were compared against the ground truth.

| District | MSE [$kw/h$] | RMSE [$kw/h$] | RMSPE [%] | MAE [$kw/h$] | MAPE [%] |
|---|---|---|---|---|---|
| Buntekuh | $1.221 \cdot 10^{12}$ | $1.105 \cdot 10^{6}$ | 99.45 | $1.03 \cdot 10^{6}$ | 99.45 |
| Hamilton Square | $1.656 \cdot 10^{13}$ | $4.069 \cdot 10^{6}$ | 99.66 | $3.749 \cdot 10^{6}$ | 99.66 |
| Innenstadt | $2.95 \cdot 10^{11}$ | $5.431 \cdot 10^{5}$ | 99.07 | $5.061 \cdot 10^{5}$ | 99.07 |
| Moisling | $5.385 \cdot 10^{11}$ | $7.338 \cdot 10^{5}$ | 99.69 | $6.843 \cdot 10^{5}$ | 99.69 |
| St. Getrud | $1.106 \cdot 10^{14}$ | $1.052 \cdot 10^{7}$ | 99.72 | $9.807 \cdot 10^{6}$ | 99.72 |
| St. Jürgen | $7.005 \cdot 10^{13}$ | $8.37 \cdot 10^{6}$ | 99.56 | $7.804 \cdot 10^{6}$ | 99.56 |
| St. Lorenz Nord | $7.185 \cdot 10^{13}$ | $8.447 \cdot 10^{6}$ | 99.64 | $7.904 \cdot 10^{6}$ | 99.64 |
| St. Lorenz Süd | $4.804 \cdot 10^{12}$ | $2.192 \cdot 10^{6}$ | 99.67 | $2.044 \cdot 10^{6}$ | 99.67 |
| Lübeck | $1.296 \cdot 10^{15}$ | $3.601 \cdot 10^{7}$ | 99.64 | $3.357 \cdot 10^{7}$ | 99.64 |

The same logic also applies to the RMSPE and the RMSE values. The RMSE is more sensible to outliers than the MAE measurement. If there were some outliers, than this would also be noticeable in the comparison of the RMSPE and the MAPE.

Figure 5.2 shows the electricity consumption of Lübeck in the plain simulation. A difference in the electricity consumption between night and day is visible, but the night seems too short and the transition between daytimes is not very smooth. A difference between workday and weekends, as I required for the ground truth dataset in section 4.1, cannot be identified. In summary, the plain simulation produces an electricity consumption graph which is sufficiently accurate for the purpose of the game, but does not fulfil my requirements.

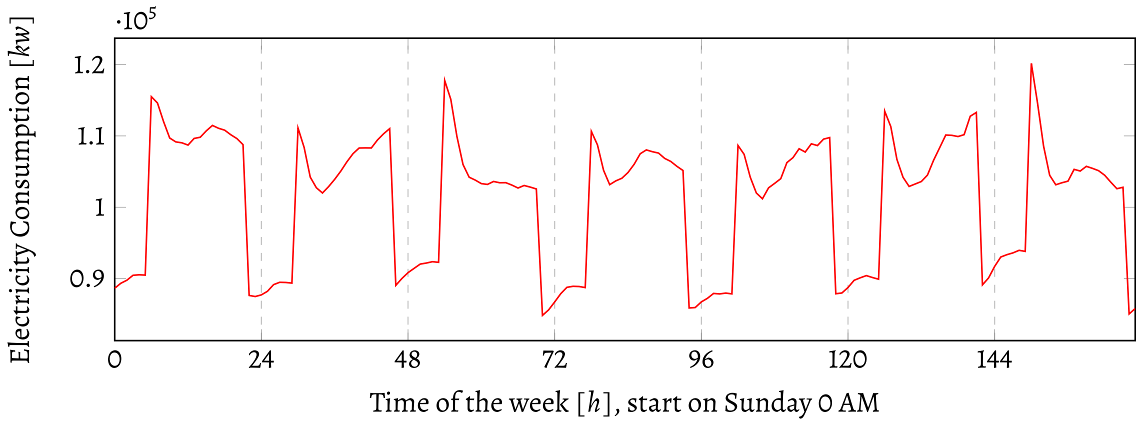


Figure 5.2: Electricity consumption of complete Lübeck, recorded during a simulation run without any adjustments (plain simulation).

## An Adjusted Simulation Run

For the second simulation run, the whole setup with the web server and the machine learning model was used, see figure 4.1. This meant, every electricity consumption value of a building was adjusted by the decision tree. The evaluation process was identical to the first simulation run. The hourly values of every building in a district were summed up for a comparison with the ground truth dataset. Table 5.3 shows the results of the adjusted simulation run.

Figure 5.4 compares the simulated values with the ground truth. The blue graph shows the values of the ground truth dataset and complies with the requirements I have formulated in section 4.1. The red graph shows the electricity consumption of the adjusted simulation. The red graph is very similar to the blue one, so every requirement, like the visible difference between night and day and between workday and weekend, is also fulfilled in the red graph. A graph like this for each district would be very similar which is indicated by a maximum difference of approximately 4 % between the MAPE values in table 5.3.

Table 5.3: Evaluation of each district and the whole of Lübeck in the adjusted simulation. The records of an adjusted simulation run were compared against the ground truth.

| District | MSE [$kw/h$] | RMSE [$kw/h$] | RMSPE [%] | MAE [$kw/h$] | MAPE [%] |
|---|---|---|---|---|---|
| Buntekuh | $8.519 \cdot 10^{9}$ | $9.23 \cdot 10^{4}$ | 8.197 | $7.436 \cdot 10^{4}$ | 7.109 |
| Hamilton Square | $5.491 \cdot 10^{10}$ | $2.343 \cdot 10^{5}$ | 5.34 | $1.402 \cdot 10^{5}$ | 3.505 |
| Innenstadt | $1.254 \cdot 10^{9}$ | $3.541 \cdot 10^{4}$ | 5.985 | $2.245 \cdot 10^{4}$ | 4.192 |
| Moisling | $2.267 \cdot 10^{9}$ | $4.761 \cdot 10^{4}$ | 6.116 | $3.142 \cdot 10^{4}$ | 4.416 |
| St. Getrud | $3.56 \cdot 10^{11}$ | $5.967 \cdot 10^{5}$ | 5.227 | $3.765 \cdot 10^{5}$ | 3.65 |
| St. Jürgen | $2.714 \cdot 10^{11}$ | $5.21 \cdot 10^{5}$ | 5.919 | $3.835 \cdot 10^{5}$ | 4.757 |
| St. Lorenz Nord | $2.521 \cdot 10^{11}$ | $5.021 \cdot 10^{5}$ | 5.527 | $3.31 \cdot 10^{5}$ | 4.035 |
| St. Lorenz Süd | $3.155 \cdot 10^{10}$ | $1.776 \cdot 10^{5}$ | 7.766 | $1.342 \cdot 10^{5}$ | 6.356 |
| Lübeck | $4.685 \cdot 10^{12}$ | $2.164 \cdot 10^{6}$ | 5.634 | $1.482 \cdot 10^{6}$ | 4.242 |

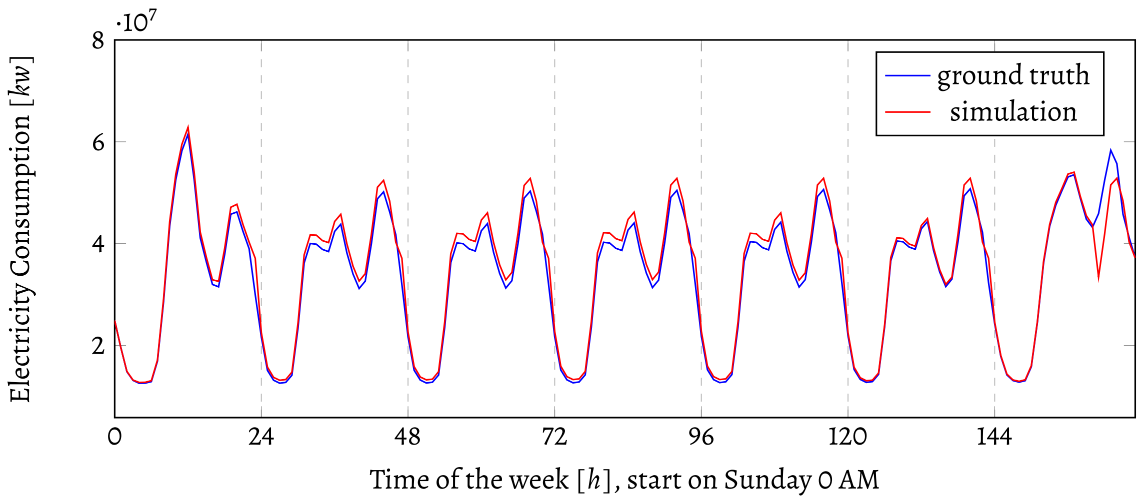


Figure 5.4: Electricity consumption of the whole of Lübeck. Recorded during a simulation run with an active machine learning model for an improvement of the electricity consumption of buildings. The blue graph shows the ground truth and the red graph the results of this simulation run.

### Impact of Cities:Skylines on the Results

To further investigate the simulation's results, it is necessary to examine whether the calculated electricity value of the machine learning model is changed by Cities:Skylines during a simulation run. Therefore, the log of the server was analysed and compared to the recording of the simulation (export). The hourly consumption values for each building for one day (Sunday) were compared. The results of the comparison are listed in table 5.5.

Table 5.5: Comparison of the server log and the export of the simulation. The hourly electricity consumption value of each building on one day (Sunday) was compared. The goal was to investigate whether the simulation had an influence on the predictions of the decision tree.

| MSE [$kw$] | MAE [$kw$] | MAPE [%] |
|---|---|---|
| 0.5992 | 0.506 | 0.23 |

### Sensitivity of the Decision Tree

Furthermore, it is important to know whether the machine learning model is capable to react to the given inputs of the simulation in a proper way. Different inputs have to produce different results. Since the number of residents has a large impact on the electricity consumption of a building, it is expected that this will also be reflected in the predictions of the decision tree. Figure 5.6 shows the electricity consumption of two houses close to each other with the same size. The orange curve represents the results of a house with 62 residents and the other curve represents a house with no residents. The other parameters like district, teen, senior etc. do not have any influence. A house in the district of St. Jürgen will have the same electricity consumption as a house in Moisling as long as the number of residents is the same.

### Results of the Neural Network

The accuracy test of the simulation was also performed with the developed neural network as a machine learning model. The results were compared with the results of the decision tree. The setup was described in section 5.1. A neural network with a MAPE of $\approx$ 11 % was used, as it represented a model with an average accuracy (figure 4.7). Table 5.7 lists the results of a simulation run with the neural network and the results of the decision tree from table 5.3. The average MAPE of the neural network is 5.363 % and the average MAPE of the decision tree is 4.696 %.

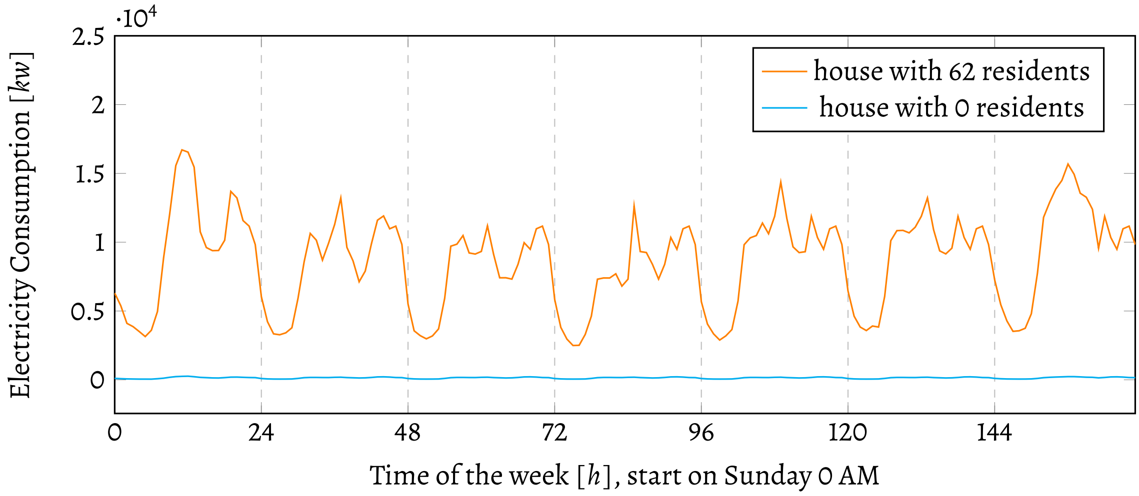


Figure 5.6: Comparison of the electricity consumption of two houses. The orange curve represents a house with 62 residents on average, the cyan curve represents a house with no residents. The number of inhabitants changes every hour, which causes a irregular electricity consumption in the inhabited house. But still, this figure shows the capabilities of the machine learning model to react on different inputs.

Table 5.7: The results of simulation runs with different machine learning models for an adjustment are compared. A decision tree (DT) and a neural network (NN) are used as a machine learning model.

| District | MAPE [%] of NN | MAPE [%] of DT |
|---|---|---|
| Buntekuh | 7.914 | 7.109 |
| Hamilton Square | 4.374 | 3.505 |
| Innenstadt | 4.574 | 4.192 |
| Moisling | 5.797 | 4.416 |
| St. Getrud | 4.465 | 3.65 |
| St. Jürgen | 5.101 | 4.757 |
| St. Lorenz Nord | 4.37 | 4.035 |
| St. Lorenz Süd | 6.893 | 6.356 |
| Lübeck | 4.779 | 4.242 |

## 5.3 Interpretation

The last section showed the results of two simulation runs (plain vs. adjusted), a comparison of the server log and the exported values and a sensitivity check of the setup. The first simulation run was without any adjustments by the decision tree, just to get an impression of the accuracy of the plain simulation. As already stated in the previous section, the graph in figure 5.2 did not meet my requirements for the ground truth dataset. To evaluate the performance of the plain simulation, its results were compared to the ground truth dataset, see table 5.1. The average error MAE between the results of the plain simulation and the ground truth is approximately 33 million for the whole city of Lübeck. The MAPE of this comparison is 99 % which confirms that the plain simulation is unsuitable for producing realistic electricity consumption values.

The adjusted simulation run could achieve better results, meaning the results were closer to the ground truth. The MAPE for the whole city of Lübeck was only $\approx$ 4 % (decision tree) which indicates that the simulation can produce results that are sufficiently close to the ground truth (table 5.3). This is also reflected in figure 5.4 which indicates that the machine learning model can greatly improve the simulation and achieve a higher accuracy compared to the plain simulation. The differences between the adjusted simulation run and the ground truth which is visible in figure 5.4 may have been caused by the imperfect machine learning model (decision tree). Another reason could be the change of the population during a simulation run. Probably, both explanations have an influence, but the deviations in the results are small, so that the overall outcome of this experiment is not affected.

Furthermore, the results showed that the simulation has no impact on the predicted electricity value (table 5.5). The differences in the consumption values between the ones predicted by the decision tree and the export of the simulation are explained by a `float` to `int` conversion which happens in the simulation because the electricity consumption value is of type `int` in Cities:Skylines. This leads to mixed results overall. The fact that the simulation has no impact on the electricity value reveals that the machine learning model takes over the task of predicting the electricity consumption, which is contrary to the original idea that a machine learning model should only optimise the simulation capabilities. Still, as figure 5.6 shows, the machine learning model adopts quite well to different inputs, but this is restricted to the number of residents in the building. This can be attributed to the generation of the dataset. The electricity value in the ground truth dataset only depends on the number of residents, the time and the date due to the underlying Lastprofil [2]. The Lastprofil is very general, as it only uses date and time for the generation of electricity consumption values. I have not extended the generation of electricity consumption further to get a realistic consumption, but I suppose a more customised generation is possible in the way that more parameters are included, e.g. that all houses in a particular district have a high consumption, and that a machine learning model would respond to this. This assumption is based on figure 5.6 which shows that the machine learning model can react on one different input parameters.

Interesting are the results of the comparison between the two machine learning models, which can be seen in table 5.7. Although the accuracy of the decision tree itself is ten times higher than the accuracy of the neural network, this is not reflected in the results of a simulation run. This has probably something to do with a changing number of residents of a building during a simulation run. New citizens move to the city, some die and some move away, these facts can change the number of residents of a building and this causes supposedly an error which has a higher influence on the results than the difference in the accuracy of the machine learning models.

# 6

# Experiment II: Accuracy of the Simulation in a What-If-Scenario

While a general evaluation of the simulation was discussed in the last chapter, this chapter focuses on a more specialized case. The goal is to evaluate the practicability of the simulation for the usage in a what-if-scenario. The question arises, whether an adjusted simulation is accurate enough to simulate such level of detail. Challenge four is addressed in this experiment. See section 3.2 for a description of challenges.

## 6.1
## Test Setup

A simulation must be able to simulate a certain level of detail so as to be useful in answering relevant questions. What-if-scenarios, the main motivation for a privacy preserving simulation of a city, represent a high level of detail. The last experiment showed that the electricity consumption value predicted by the machine learning model only depended on the number of residents in a building, but for a what-if-scenario, it has to depend on more parameters, such as the position of buildings. It follows, that the machine learning model has to be more powerful in order to calculate an electricity consumption value which depends on more parameters than just the number of residents.

The setup of this experiment differs slightly to the setup of the last experiment. First, a what-if-scenario has to be declared. This experiment examines the accuracy of the adjusted simulation for the following scenario: What happens to the power grid if all households in a certain area are using heat pumps? The second experiment does not aim to answer that question, but the goal is to evaluate whether the setup of this work is capable to simulate such a scenario and to investigate if the prediction of the electricity consumption can depend on more parameters than just the date and number of residents of a building. To achieve this, the ground truth dataset has to be adjusted to the scenario. Therefore, a higher electricity consumption value is assigned to all buildings in a certain area in the district of St. Jürgen to simulate the use of heat pumps in those buildings.

Figure 6.1 shows the coordinates of all buildings in the simulation. The purple points symbolise the buildings in the area where heat pumps are being used. To imitate the usage of a heat pump in those buildings, the electricity consumption value was increased by 50 %. Figure 6.2 shows the distribution of the electricity consumption values of the adjusted dataset for the scenario. This distribution is important for a better estimation of the accuracy of the machine learning model. This assumption of the influence of a heat pump is strongly simplified and not based on realistic data regarding the energy consumption of heat pumps. The idea here is, that if the setup produces accurate results with this assumption, it will also produce accurate results with a more realistic assumption.

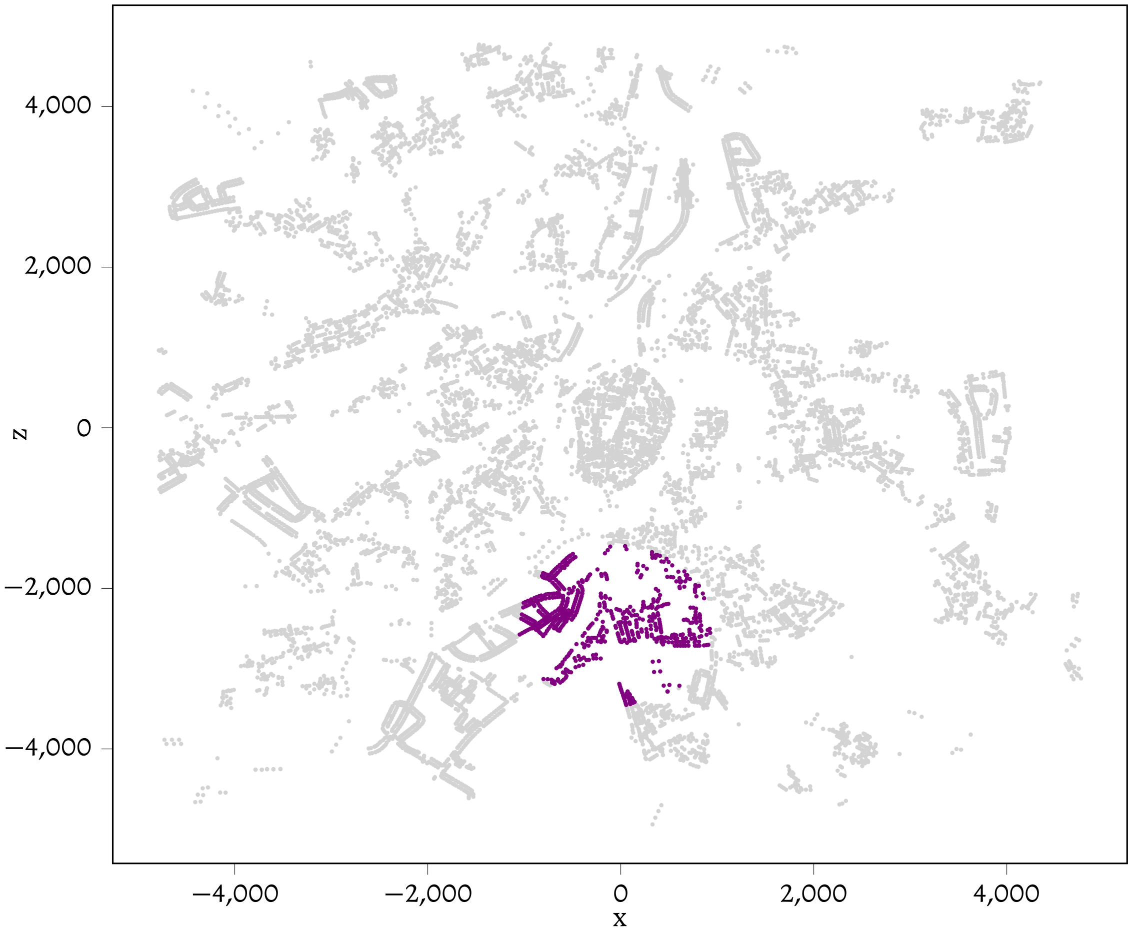


Figure 6.1: Map of all buildings in the simulation. Each building has coordinates and the $x$ and the $z$ coordinate are plotted. The purple points symbolise the buildings with a heat pump in this what-if-scenario. All the houses with a heat pump are in the district of St. Jürgen.

### Training of the Decision Tree

In the next step, the machine learning model is trained again with the adjusted ground truth. The new scenario is more complex, the electricity consumption value now depends on more features, that is why the decision tree has to be more powerful. Therefore the two hyperparameters `n_estimators` and `max_depth` were updated. RepeatedKFold cross

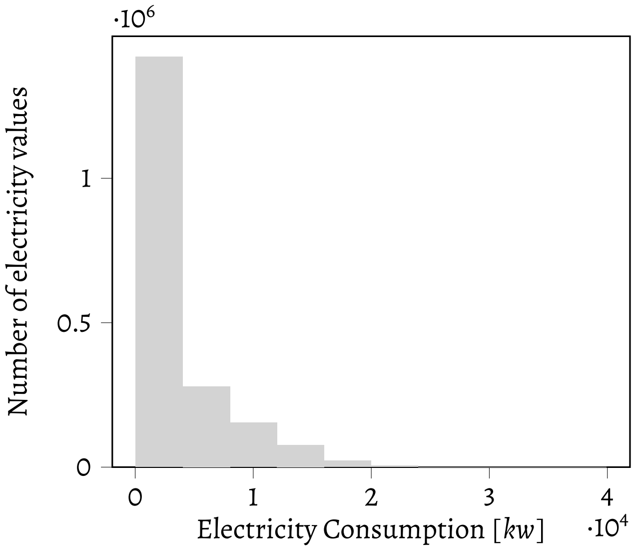


Figure 6.2: This histogram shows the distribution of the electricity consumption values in the dataset for the what-if-scenario. In total, there are 1 960 728 values. The data is not evenly distributed since the electricity consumption of 1 517 394 values is less or equal than 5 000 *kw*. The maximum consumption is 39 921 *kw* which means that 22 % of the values are distribute in the range from 5 000 *kw* to 39 921 *kw*.

validation was used to evaluate the updated machine learning model which had now 175 trees with a depth of 11. Figure 6.3 shows the distribution of the accuracy of 50 training runs. The results of this cross validation are very similar to the results of the cross validation for the decision tree used in the first experiment (figure 4.6). The decision tree used for this experiment was finally trained and tested on a divided train an test dataset which resulted in a MAE of 14 and a MAPE of 0.54 %. It follows that the accuracy of the used decision trees is comparable to each other. From this point on, the setup and the simulation are identical to experiment one.

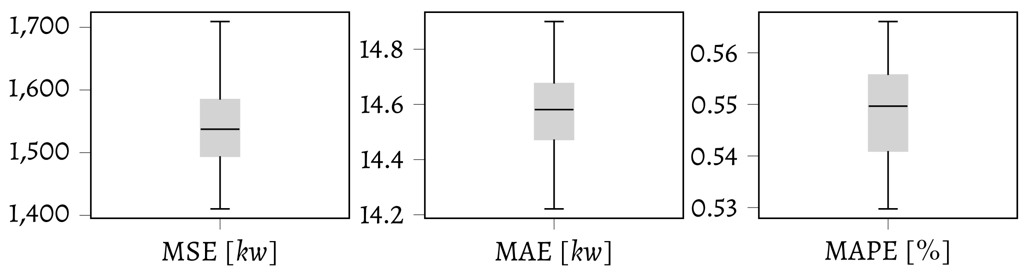


Figure 6.3: Results of 50 cross validation rounds of the decision tree for the scenario. The MAPE gives a percentage and classifies the MAE. The MSE is, under consideration of figure 6.2 which shows a distribution of possible electricity consumption values of the ground truth, rated as low.

### Training of the Neural Network

A comparison of the performance of the decision tree with the neural network is part of the second experiment as it was part of the first experiment. The neural network was trained with the adjusted ground truth for 30 epochs to give the neural network more time to learn due to a more complex dataset. The distribution of the MAPE score of 50 rounds cross validation is listed in contrast to the decision tree's distribution of the MAPE in figure 6.4. An average neural network with a MAPE of 14.5 % was used as a machine learning model for the simulation.

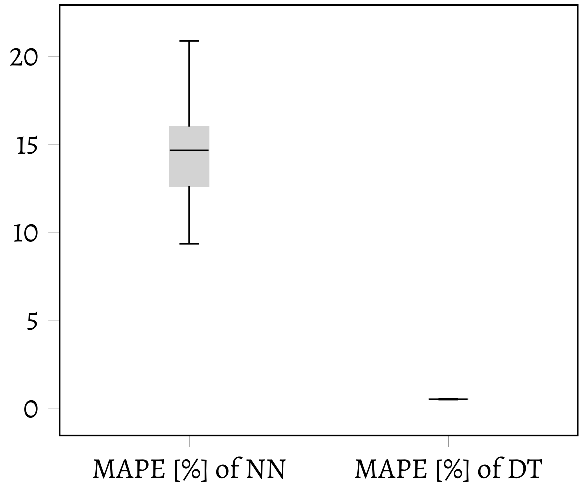


Figure 6.4: Comparison of the distribution of the MAPE score of the cross validation results of the neural network (NN) and the decision tree (DT).

## 6.2 Results

The evaluation of the simulation export follows the same logic as in experiment one in chapter 5. The export of this experiment is compared against the ground truth of the first experiment because the ground truth of the first experiment represents a scenario in which no heat pumps are used in the respective area. This means that the electricity consumption sum of all the buildings in the defined area from export two is compared against the electricity consumption sum of the same buildings based on the ground truth. The expectation of this experiment is that the electricity consumption of all the buildings in the area is 50 % higher compared to the consumption of the same buildings with no heat pumps. This experiment is carried out with both machine learning models.

Table 6.5 list the results of this experiment. Since the electricity consumption of every building with a heat pump was multiplied by 1.5, we can multiply the electricity consumption sum of the buildings with no heat pump with 1.5 to get the target value. At first, the evaluation of the accuracy of the decision tree:

Table 6.5: Comparison of the electricity consumption of the buildings with a heat pump against the same buildings with no heat pump.

| Machine Learning Model | no heat pump $[kw/h]$ | heat pump $[kw/h]$ |
|---|---|---|
| Decision Tree | 350 930 190 | 524 351 868 |
| Neural Network | 350 949 573 | 535 175 692 |

$$\begin{aligned} 350\,930\,190 \cdot 1.5 &= 526\,395\,285 && \mid \text{target value} \\ 526\,395\,285 - 350\,930\,190 &= 175\,465\,095 \\ |526\,395\,285 - 524\,351\,868| &= 2\,043\,417 && \mid \text{difference} \\ 2\,043\,417 \div 175\,465\,095 \approx 0.01 &= 1\,\% && \mid \text{percentage difference} \end{aligned}$$

It follows the evaluation of the neural network's accuracy:

$$\begin{aligned} 350\,949\,573 \cdot 1.5 &= 526\,424\,359.5 && \mid \text{target value} \\ 526\,424\,359.5 - 350\,930\,190 &= 175\,474\,786.5 \\ |526\,424\,359.5 - 535\,175\,692| &= 8\,751\,332.5 && \mid \text{difference} \\ 8\,751\,332.5 \div 175\,474\,786.5 \approx 0.05 &= 5\,\% && \mid \text{percentage difference} \end{aligned}$$

The decision tree achieves a percentage difference of 1 % which indicates a very high accuracy. It follows, the electricity consumption of the buildings with a heat pump is 49 % higher than the electricity consumption of the same buildings with no heat pump. Compared to the decision tree, the neural network is not as accurate. The percentage error is 5 % and the electricity consumption of the buildings with a heat pump is 52 % higher compared to the buildings of the ground truth.

$$\begin{aligned} 524\,351\,868 \div 350\,930\,190 - 1 \approx 0.49 &= 49\,\% \\ 535\,175\,692 \div 350\,949\,573 - 1 \approx 0.52 &= 52\,\% \end{aligned}$$

Figure 6.6 visualizes the results of the decision tree. A difference between the hourly electricity consumption values of all the buildings in the area with a heat pump and with no heat pump is clearly visible. The curve of the neural network would look almost exactly the same. The small difference in the accuracy of these machine learning models would not be visible in this scale.

The percentage difference of 1 % of the decision tree indicates a high accuracy, the outcome of this experiment is also visible when plotting the electricity consumption of the complete district of St. Jürgen as figure 6.7 indicates. This figure shows again the results of the decision tree. The area of interest is placed completely in St. Jürgen.

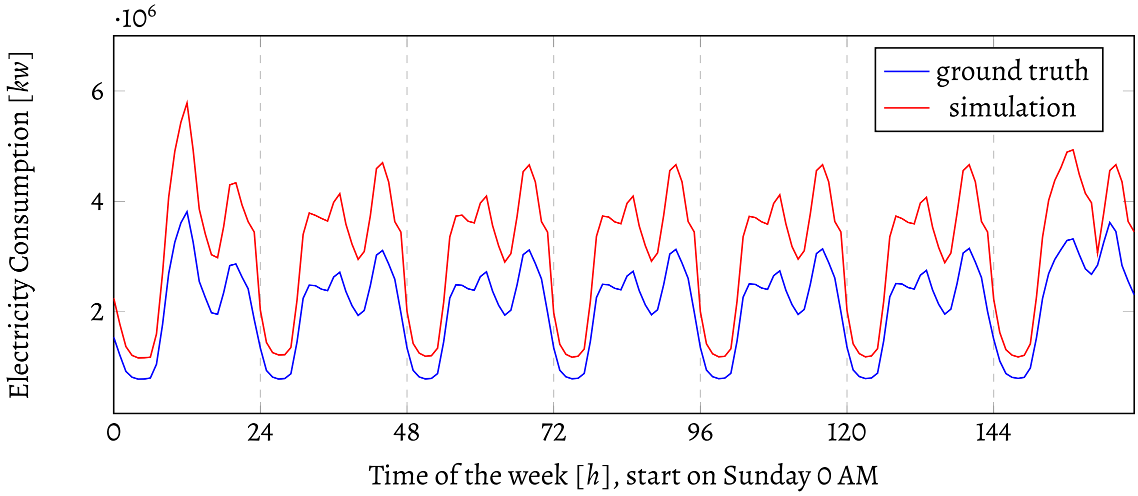


Figure 6.6: Comparison of the electricity consumption of the buildings in the defined area in which either no heat pumps are used in the buildings (blue curve) or heat pumps are used in the buildings (red curve). The decision tree was used as the machine learning model.

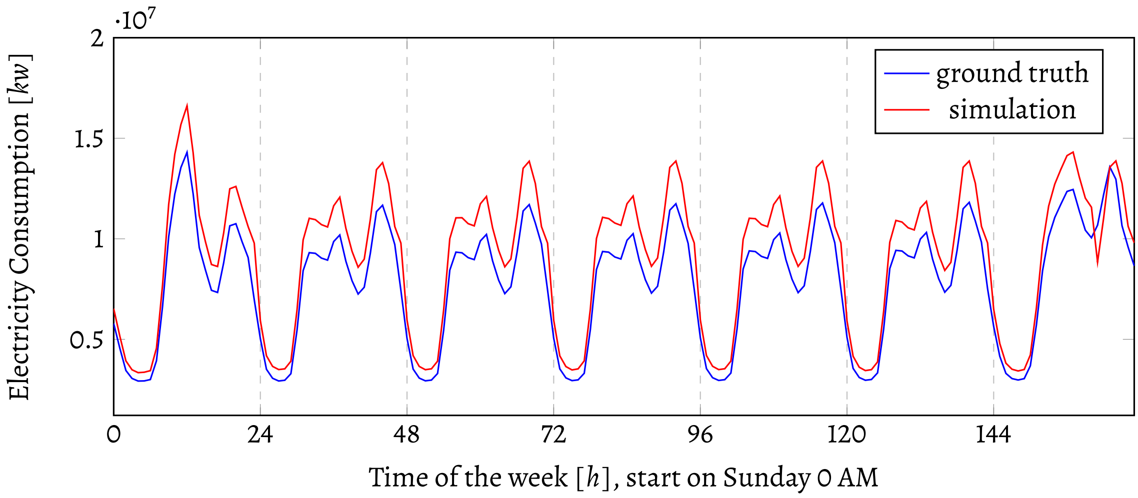


Figure 6.7: Comparison of the electricity consumption of the buildings in the district of St. Jürgen. The blue graph shows the consumption of the buildings with no heat pump and the red graph shows the consumption of the buildings with a heat pump. This results were produces by the decision tree.

## 6.3 Interpretation

After a general accuracy test in the first experiment in chapter 5 of this privacy preserving simulation based on Cities:Skylines, this second experiment focused on a more specific scenario. The goal was to evaluate the simulation capabilities in more detail. A what-if-scenario was defined where some buildings had a higher electricity consumption due to the usage of a heat pump.

The simulated consumption sum of buildings with a heat pump is 49 % higher than the consumption of buildings with no heat pump when using a decision tree as the machine learning model. Expected was a difference of 50 %. This is a good result, the difference of the consumption between buildings with a heat pump and with no heat pump is clearly visible as figure 6.6 demonstrates. The decision tree responds to the defined scenario with a high accuracy. The differences in the results of the cross validation of the two machine learning models (figure 6.4) are even higher than the differences between the two models for experiment one. However, these differences are not reflected in the results of the simulation run. In the simulation there is a percentage difference between the target consumption and the simulated consumption of 1 % for the decision tree and of 5 % for the neural network. This indicates again the existence of an error in the simulation due to a change in the number of citizens which has a bigger influence on the results than the difference in the accuracy of the machine learning models.

The result of the second experiment supports my assumption that a dependence on several parameters of the decision tree for the prediction of electricity consumption values is possible. A difference in the consumption values becomes visible when the generation process of the electricity values in the dataset is more individual and complex.

# 7

# Discussion

The goal of the experiments was to evaluate the accuracy of the simulation in general and in a specific scenario. Even though the results (see section 5.2 and section 6.2) are promising, there are some limitations and conditions that restrict the general validity. As it was investigated in the first experiment, the simulation does not affect the results at all. Although Cities:Skylines offers a method to determine the electricity consumption of a district, it remains unclear how this value is calculated and therefore this value is not used for a comparison with the ground truth. Zeiseweis and myself used this value for a comparison with the ground truth in our Bachelorprojekt [29] and that resulted in a high error with a MAPE of 1 149 %. This means that all results are based on the comparison of the ground truth to the simulations export which is identical to the log of the machine learning model. And this means that the machine learning model alone is completely responsible for the electricity consumption values which undermines the usefulness of Cities:Skylines as a simulation.

This fact clearly reduces the value of the results and a main reason for this is Cities:Skylines itself. Not only this, but also a few more disadvantages question the suitability of Cities:Skylines as a basis for such advanced investigations. The possibilities for modifying Cities:Skylines are limited due to its closed source codebase and its purpose to entertain as a computer game. These disadvantages lead to an unbalanced cost-use ratio where a lot of time was spend to set up Cities:Skylines correctly and to start the simulation. In fact, the start of the simulation must be done in a certain order, otherwise some required extensions will not load and this will lead to a non-functioning simulation. The practicability of Cities:Skylines as a simulation tool for serious investigations is low.

The comparison of the two machine learning models shows that it makes hardly any difference which framework is used. Also, a difference in model accuracy does not affect the simulation as much as it might seem when looking at the accuracy of the model itself. A neural network in general does not have to be more imprecise than a decision tree. This is indeed the case in this work, but it would have been possible to optimise the neural network further. The comparison of the simulation results of the two machine learning models shows that a probable change in the number of citizens has a stronger impact on the overall accuracy than the accuracy of the machine learning model.

The question is now how to assess the results. A privacy preserving setup exists which produces results. In fact, Cities:Skylines as a simulation is extensible with a machine learning model to improve in game parameters for more realistic results. Accurate results were achieved. Also the machine learning model in combination with a running simulation reacts as sensitively as expected to different inputs and thus able to simulate specific scenarios. Moreover, the machine learning model alone is responsible for the electricity consumption values of buildings and this is contrary to the idea of a simulation.

To summarize, Cities:Skylines is not a useful basis for a privacy preserving simulation, but the results of this work can show that the idea of using an existing city simulation to simulate the electricity consumption of buildings in a privacy preserving way has potential.

# 8

# Conclusion and Outlook

## 8.1
## Conclusion

This work evaluated the accuracy of a privacy preserving simulation based on Cities:Skylines. An already existing simulation developed in former projects was improved and extended to serve as a privacy preserving simulation. To achieve this transformation, four challenges had be solved: A privacy friendly dataset was used as a ground truth for the evaluation of the experiments and for the training of the machine learning model (first challenge). A Gradient Boosted Decision Tree and a neural network were the chosen machine learning models to improve in game parameters such as the electricity consumption of buildings (second challenge). Further on, the decision tree was integrated with Cities:Skylines. This was done via a web server which has a negative impact on speed, but opens up several opportunities, such as the changeability of the machine learning model (third challenge). Finally, the definition of a what-if-scenario which required an updated ground truth and a retrained decision tree was also a part of this work (fourth challenge).

The first experiment in chapter 5 evaluated the general accuracy of the used privacy preserving simulation setup. The accuracy was high, a MAPE of approximately 4 % confirms that. The second experiment in chapter 6 tested the capabilities of this setup in a more complex scenario. The question to investigate was, to what extend the simulation is capable of simulating a detailed scenario like the following where all the buildings in a specific area are assumed to use a heat pump. It follows, that the experiment meets its expectations. It is possible to simulate such level of detail, a percentage difference of 1 % between the target consumption and the simulated consumption underlines this fact. Even though there existed a high difference in the accuracy of the machine learning models (the accuracy of the decision tree was ten times higher than the accuracy of the neural network) this was not reflected in the results of a simulation run. This suggests that a change in the number of citizens during a simulation run is likely to have a greater impact than the accuracy of the machine learning model.

The whole setup is privacy friendly because no personal information was used to gain results. One result of this work is that the simulation adapts to the predictions of the decision tree as long as the dataset mirrors the simulations reality. This is a valuable result in terms of privacy because simulation and dataset can both be changed and therefore an average distribution of people and buildings can be assumed as a ground truth for simulation and dataset. This could not only result in a privacy friendly, but also in a realistic simulation. A possible use case of such a simulation was described in the second experiment. Considering the effect of an installation of heat pumps can be an important question for local power providers. Utilities need to know the instabilities in their electricity grid to be able to react to upcoming changes or a sudden increase in consumption due to more heat pumps in an area.

The results also revealed that the machine learning model alone is responsible for the prediction of the electricity consumption values in the simulation. This is contrary to the original idea where the simulation, understood as a black box, should do the work. The simulation has no influence on the electricity consumption values of the buildings. A reason for this is the opaqueness of Cities:Skylines. The machine learning model reacts on different inputs as the second experiment showed, which supports the usefulness of this setup for what-if-scenarios. The results support the research on privacy preserving simulations, but they also show that Cities:Skylines is a limiting factor.

## 8.2 Outlook

This work evaluated the accuracy of Cities:Skylines used as a privacy preserving simulation and researched its potential for simulating what-if-scenarios. What-if-scenarios were modelled through a change of the ground truth in this work. A what-if-scenario based on a change of the simulation, which is the other possible solution can be a further field of investigation. The Realistic Population modification [20] for Cities:Skylines can be helpful for such investigations. It produces a realistic distribution of residents to buildings. However, an important outcome of this work is the ineligibility of Cities:Skylines as a framework. It would not be justifiable to continue using Cities:Skylines, as the time required would be disproportionate to the results. Therefore, a new simulation basis must be found for further research in the field of privacy preserving city simulations. There is still potential in the optimisation of the machine learning model.

As this work indicates, optimisations of a city and privacy do not have to be contrary claims. There are possibilities to combine the concept of smart cities with privacy preserving simulations to achieve a more sustainable city, but more research is needed for a more effective approach.

# Bibliography


[1] Al-Rubaie, M. and Chang, J. Privacy-Preserving Machine Learning: Threats and Solutions. In: *IEEE Security and Privacy* 17(02):49–58, 2019. ISSN: 1558-4046. DOI: 10.1109/MSEC.2018.2888775.

[2] *Bundesverband der Energie- und Wasserwirtschaft (bdew)*. https://www.bdew.de/. Accessed: 2022-10-12.

[3] *Cities:Sklyines*. https://www.paradoxinteractive.com/games/cities-skylines/about. Accessed: 2022-10-12.

[4] *Cities:Skylines Wiki for modding*. https://skylines.paradoxwikis.com/Modding. Accessed: 2022-10-12.

[5] Doulamis, N., Protopapadakis, E., and Lambrinos, L. Improving Service Quality for Parking Lot Users Using Intelligent Parking Reservation Policies. In: *2013 27th International Conference on Advanced Information Networking and Applications Workshops*. Los Alamitos, CA, USA: IEEE Computer Society, 2013, pp. 1392–1397. DOI: 10.1109/WAINA.2013.219. URL: https://doi.ieeecomputersociety.org/10.1109/WAINA.2013.219.

[6] Duncan, C., Cunningham, J., Wang, A., and Kennedy, A. Urban Planning Optimization via "Cities: Skylines". 2021.

[7] Díaz-Díaz, R., Muñoz, L., and Pérez-González, D. Business model analysis of public services operating in the smart city ecosystem: The case of SmartSantander. In: *Future Generation Computer Systems* 76:198–214, 2017. ISSN: 0167-739X. DOI: https://doi.org/10.1016/j.future.2017.01.032. URL: https://www.sciencedirect.com/science/article/pii/S0167739X17301553.

[8] *EUR-Lex*. https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A32016R0679. Accessed: 2022-11-10.

[9] Fainmesser, I. P., Galeotti, A., and Momot, R. Digital privacy. In: *Management Science*, 2022.

[10] Fomby, T. Scoring measures for prediction problems. In: *Department of Economics, Southern Methodist University, Dallas, TX*, 2008.

[11] *GeoSkylines*. https://github.com/gonzikcz/GeoSkylines/. Accessed: 2022-10-12.

[12] *Harmony*. https://github.com/pardeike/Harmony. Accessed: 2022-10-30.

[13] Inibhunu, C. and McGregor, A. C. A Privacy Preserving Framework for Smart Cities utilising IoT, Smart Buildings and Big Data. In: *2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. Los Alamitos, CA, USA: IEEE Computer Society, 2020, pp. 1096–1103. DOI: 10.1109/HPCC-SmartCity-DSS50907.2020.00197.

URL: https://doi.ieeecomputersociety.org/10.1109/HPCC-SmartCity-DSS50907.2020.00197.

[14] Kim, C. and Kim, K.-a. The Institutional Change from E-Government toward Smarter City; Comparative Analysis between Royal Borough of Greenwich, UK, and Seongdong-gu, South Korea. In: *Journal of Open Innovation: Technology, Market, and Complexity* 7(1), 2021. ISSN: 2199-8531. DOI: 10.3390/joitmc7010042. URL: https://www.mdpi.com/2199-8531/7/1/42.

[15] *ML.NET*. https://dotnet.microsoft.com/en-us/apps/machinelearning-ai/ml-dotnet. Accessed: 2022-11-20.

[16] Oliveira, A. and Campolargo, M. From Smart Cities to Human Smart Cities. In: *2015 48th Hawaii International Conference on System Sciences (HICSS)*. Los Alamitos, CA, USA: IEEE Computer Society, 2015, pp. 2336–2344. DOI: 10.1109/HICSS.2015.281. URL: https://doi.ieeecomputersociety.org/10.1109/HICSS.2015.281.

[17] Olszewski, R., Cegiełka, M., Szczepankowska, U., and Wesołowski, J. Developing a Serious Game That Supports the Resolution of Social and Ecological Problems in the Toolset Environment of Cities: Skylines. In: *ISPRS International Journal of Geo-Information* 9(2), 2020. ISSN: 2220-9964. URL: https://www.mdpi.com/2220-9964/9/2/118.

[18] *Pytorch*. https://pytorch.org/. Accessed: 2022-11-20.

[19] Pörtner, H.-O., Roberts, D., Poloczanska, E., Mintenbeck, K., Tignor, M., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., et al. Summary for Policymakers. In: *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, USA: Cambridge University Press, 2022, pp. 3–33. DOI: 10.1017/9781009325844.001.

[20] *Realistic Population*. https://steamcommunity.com/sharedfiles/filedetails/?id=2025147082. Accessed: 2022-11-09.

[21] *RealTime*. https://github.com/dymanoid/RealTime. Accessed: 2022-10-12.

[22] Richter, K. Precision of city simulations using publicly available city data. 2021.

[23] Santana, E. F. Z. InterSCSimulator: a scalable, open source, smart city simulator. PhD thesis. PhD thesis, Universidade de Sao Paulo, 2019.

[24] *scikit-learn*. https://scikit-learn.org/stable/. Accessed: 2022-11-20.

[25] Sookhak, M., Tang, H., and Yu, F. Security and Privacy of Smart Cities: Issues and Challenge. In: *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. Los Alamitos, CA, USA: IEEE Computer Society, 2018, pp. 1350–1357. DOI: 10.1109/HPCC/SmartCity/DSS.2018.00224. URL: https://doi.ieeecomputersociety.org/10.1109/HPCC/SmartCity/DSS.2018.00224.

[26] *Stromspiegel für Deutschland 2021/22*. https://www.stromspiegel.de/fileadmin/ssi/stromspiegel/Tabellen/stromspiegel-2021-tabelle.jpg. Accessed: 2022-11-04.

[27] Taktak, E., Abdennadher, I., and Rodriguez, I. An Adaptation Approach for Smart Buildings. In: *2016 IEEE 18th International Conference on High-Performance Computing*

*and Communications, IEEE 14th International Conference on Smart City, and IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. Los Alamitos, CA, USA: IEEE Computer Society, 2016, pp. 1107–1114. DOI: 10.1109/HPCC-SmartCity-DSS.2016.0156. URL: https://doi.ieeecomputersociety.org/10.1109/HPCC-SmartCity-DSS.2016.0156.

[28] Zeiseweis, K. Evaluation of Injected Data from an Adapted Smart City Simulation. 2021.

[29] Zeiseweis, K. and Hodgkison, P. Untersuchung einer Privatsphäre erhaltenden Smart-City-Simulation auf der Basis von Cities:Skylines und Machine-Learning-Techniken. In: 2022.