LIFT TRAVEL MODELLING USING AI: ASSESSING DIFFERENT LIFT MODERNISATION SOLUTIONS ON ENERGY CONSUMPTION

Lappeenranta–Lahti University of Technology LUT

Master's Programme in Electrical Engineering, Master's Thesis

2025

Nasrin Nouri

Examiner(s): Professor Juha Pyrhönen

Dr. Ilya Petrov, Post-doctoral researcher (Electrical Engineering)

ABSTRACT

Lappeenranta–Lahti University of Technology LUT
LUT School of Energy Systems
Electrical Engineering

Nasrin Nouri

Lift Travel Modelling Using AI: Assessing Different Lift Modernisation Solutions on Energy Consumption

Master's Thesis
2025
87 pages, 27 figures, 27 tables, and 0 appendices
Examiner(s): Professor Juha Pyrhönen and Dr. Ilya Petrov, Post-doctoral researcher, (Electrical Engineering)
Keywords: lift energy consumption, lift modernisation, machine learning, ISO 25745-2, gradient boosting

This thesis explores the use of artificial intelligence (AI) to advance lift traffic modelling, specifically to increase the accuracy of daily trip estimations as well as their impact on estimated annual energy consumption. Based on a dataset of more than 100,000 lifts supplied by one of the world's leading lift manufacturers, the study compares certain assumptions utilised by other articles based on the standard ISO 25745-2 to data-driven models. The paper applies machine learning methods such as XGBoost, LightGBM, and CatBoost to predict average daily trip numbers given building and lift characteristics. The findings reveal that in certain situations the ISO-based approach overestimates trip quantities, especially for tall buildings with greater trip numbers in a day, hence contributing to inaccuracy of energy consumption reporting. The machine learning models exhibit better predictive performance as well as a higher degree of conformity to real-world data fluctuations. This study also analyses the effect of improving trip estimation accuracy in five lift modernisation scenarios, indicating substantial saving by modernisation as well as the relative changes in running energy estimation with improved trip estimation. The implications of this study demonstrate the potential of AI-driven modelling in enabling more accurate environmental assessment as well as sustainable decision-making in the lift sector.

ACKNOWLEDGEMENTS

This thesis was conducted in collaboration with the Finnish lift manufacturer KONE Oyj, with generous financial support from Teknologiateollisuuden 100-vuotissäätiö, whose grant made this research possible.

I would like to express my sincere gratitude to my company supervisor, Toni Tukia, for providing me with the opportunity to work on this topic. His guidance, encouragement, and trust were instrumental in the successful completion of this project.

Special thanks go to my colleagues and manager at KONE, Shahram Heidari, Pekka Perunka, and Semen Uimonen, for their continuous support, collaboration, and encouragement during this journey.

I am deeply thankful to my university supervisors, Professor Juha Pyrhönen and Dr. Ilya Petrov, for their valuable feedback and insightful suggestions throughout the research process.

Finally, my heartfelt gratitude goes to my family, whose unwavering support and belief in me helped me persevere through challenges. Their encouragement, respect for my decisions, and presence in every circumstance have been a constant source of strength.

Espoo, April 2025

ABBREVIATIONS

EPD	Environmental Product Declarations
ISO	International Organization for Standardization
PMSM	Permanent Magnet Synchronous Motor
MRL	Machine Room-Less
DCS	Destination Control Systems
OCL	Operation of Car Light
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
CatBoost	Categorical Boosting
GOSS	Gradient-based One-Side Sampling
TPE	Tree-structured Parzen Estimator
SQL	Structured Query Language
OHE	One-Hot Encoding
TE	Target Encoding

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Key Concepts and Objective

As global awareness of sustainability intensifies, the building industry faces increasing pressure to reduce energy consumption and carbon emissions. Among various building systems, lifts may appear to contribute a modest share of total energy use, yet their impact becomes substantial in mid and high-rise structures. Accurate estimation of lift energy consumption is therefore a key component in sustainability assessments, environmental product declarations (EPDs), and CO_2 emission calculations (Dong et al., 2023).

Governments are also enacting building energy disclosure laws; for instance, certain city ordinances require large buildings to report annual energy consumption, implicitly pushing owners to track lift energy use as part of their total building performance. In Europe, the European Green Deal and related policies aim to transform the building stock by 2030 and 2050 to drastically reduce emissions. This includes improving building energy efficiency through renovations, a context in which lift modernisations are actively encouraged for both safety and environmental reasons (Haavisto, 2022)

In practice, industry guidelines support a hybrid method of energy reporting, favouring direct measurement wherever possible, supplemented by standardized calculations. Both the International Organization for Standardization (ISO) and Germany's VDI have published standards to facilitate consistency in lift energy reporting. Notably, ISO 25745-2:2015 (and its updated series) and VDI 4707 define testing procedures and assign energy-efficiency classes (ranging from A to G) based on measured energy consumption in standby and travel modes, normalized to a standard usage profile (ISO 25745-2, 2015) (Tukia, 2019). These classifications, now commonly cited by manufacturers, allow for meaningful product comparisons.

The ISO 25745 methodology estimates annual energy consumption using test-based power measurements combined with a formula incorporating total lift trips recorded annually. However, one of the primary variables in calculating lift energy use is the daily and annual trip count. Traditional estimation methods, such as those outlined in ISO 25745-2, rely on predefined usage categories based on building type and height, assigning a representative number of trips to each category (ISO 25745-2, 2015) While useful, these assumptions sometimes Can overlook the dynamic and unpredictable nature of lift usage in practice, which can lead to inaccurate energy and emissions reporting in certain cases.

To address this issue, the main objective of this study is to enhance the precision of annual energy consumption estimation in lifts by improving the prediction of average daily trip counts. By leveraging historical data collected from connected lift systems and applying machine learning models, this research aims to uncover more accurate and scalable ways of predicting lift usage patterns.

1.2 Scope, and Research Questions

The dataset used in this study, is provided by KONE company which is one of the global leaders in lift industry. This dataset comprises operational and dimensional data from over 100,000 lifts, enabling the development of robust models for estimating the trips taken on a daily average basis and, subsequently, per year.

This refined trip count estimation is then applied within KONE's internal energy calculation tool to assess its impact on the predicted annual energy consumption — both for existing lifts and those undergoing modernisation. The focus is strictly limited to the effect of trip number estimation on annual energy consumption, excluding other influencing factors such as average load per trip or displacement distance.

To guide the research, the following questions are investigated:

1. Can historical data from connected lift devices be effectively used in machine learning models to predict average trip numbers?
2. Is there a measurable relationship between the dimensional characteristics of a lift, building characteristics, and its daily trip frequency?
3. Can machine learning models outperform traditional estimation approaches, such as ISO-based methods or static floor-based lookups?
4. How does the refinement of trip number estimation influence the output of KONE's internal energy consumption tool?
5. In the context of modernisation, how does more accurate trip estimation affect the comparison of energy consumption before and after upgrades?

By addressing these questions, this thesis aids in the progress of smarter, data-driven tools for environmental assessment in the lift industry, promoting greater transparency and more effective sustainability practices. Artificial intelligence was used in this work for the language maintenance.

2 Literature Review

Lifts typically constitute a modest share of a building's overall energy use on an annual basis. For instance, one study found that a typical range on the order of 2–10% of building energy attributable to lift systems (Al-Kodmany., 2015). This share, however, is not fixed and varies with building characteristics, most notably with building height. In low-rise buildings the lift impact remains at the lower end of this range, whereas in high-rise buildings the proportion grows substantially (Al-Kodmany., 2015). Studies showed that in in tall buildings, lifts can account for around 17-25% of total building energy demand (Makar et al., 2022). That means taller buildings depend more on lifts, so making those systems more energy-efficient can have a bigger impact. Moreover, the effect of efficiency improvement has an effect on energy consumption so that Tukia et al. (2018) further argued that, with efficiency improvements, lift systems may account for just 1–8% of total electricity consumption. This section first reviews the most widely used methodologies for calculating lift energy consumption of builings, then explores the impact of modernisation strategies, and finally provides a theoretical background for the machine learning models applied later in the study.

2.1 Lift Energy Consumption

Energy efficiency in lifts is an important topic in the building and construction industry. Various methodologies exist to estimate a lift energy consumption, each offering a unique perspective on how energy is used. Energy labelling also is one of the most important concepts in this industry since it can determine how efficient a lift is based on its annual energy consumption.

Several international models provide a framework for calculating energy consumption in lifts. Among them, the ISO 25745-2:2015 standard, a global benchmark, and the VDI 4707-1 (2009) guideline offer structured approaches and are the most common frameworks being used in the industry and by research to estimate lift energy consumption characteristics. This section explains these two frameworks, their advantages and drawbacks.

ISO 25745-2:2015 offers a clear framework for estimating lift energy consumption by dividing it into two key areas: running energy (used when the lift is moving) and non-running energy (power consumed while idle or in standby mode). Besides the simulation method, which is accepted by the standard, short-term physical measurements are also used—

particularly in cases where certification of a lift's energy efficiency class by an accredited third party is required. Instead of relying on long-term tracking, the standard allows for short-term measurements taken from either an actual lift in use or a controlled test environment to efficiently estimate overall energy consumption (Barney & Lorente, 2013). The standard specifically calls for measuring running energy by tracking a set reference cycle, which usually means observing how much power the lift uses when it travels empty from the bottom floor to the top and back down again, including door operations at both landings. Additionally, a short cycle needs to be conducted when certain conditions, set in the standard, apply. To comply with the ISO 25745-2 standard, measure of standing idle power which is the energy the lift consumes when it is not moving but power is on. Using these inputs, ISO 25745-2 outlines formulas to project daily and annual consumption. The standard also defines usage categories (based on trips per day) and load factors to account for typical passenger loads, which adjust the reference-cycle energy to more realistic "average trip" energy . These two concepts are discussed in more detail in chapter 2.2.

Furthermore, VDI 4707-1 (2009), an earlier German standard, set the stage for ISO 25745 by using short-term measurements to evaluate how efficiently lifts use energy. It classified lifts into efficiency ratings from A to G, based on two key factors: travel energy, which is the power used while the lift moves with a set load, and standby energy, which is the electricity it consumes while waiting between trips. It categorized lifts based on how often they are used whether low, medium, or high frequency, to better reflect energy consumption across different types of buildings. (Barney & Lorente, 2013).

To compare these two methodologies, it is good to mention that ISO 25745-2 expanded and improved on this approach, closely following the measurement principles of VDI 4707. In fact, ISO 25745-1 and 2 officially standardized the measurement and analysis methods that VDI 4707 originally introduced. However, ISO took this a step further by introducing more standardized duty profiles and widening the range of usage categories, even adding an extra-high usage category for buildings with exceptionally high lift demand. (Chan et al, 2016) One key improvement in ISO 25745-2 is its focus on longer idle measurements such as checking power consumption after 30 minutes of inactivity to account for energy savings in deep standby modes (ISO 25745-2, 2015). In addition, VDI 4707 did not put much focus on longer idle periods, instead looking at shorter inactivity times. ISO 25745-2, however, takes into account how lifts save energy by automatically turning off lights and ventilation after being idle for a while. This means it gives a more accurate picture of energy-saving features like LED lighting and sleep modes when calculating non-running energy. (Barney & Lorente, 2013) While both standards estimate annual energy use using short-term tests and simulations, ISO 25745-2 provides a more thorough and globally recognized method.

Numerous studies have evaluated how well ISO 25745-2 calculation methods perform against real-world data and other approaches. For example, Tukia et al. (2016) collected actual energy use data from a mid-rise office building (via installed energy meters) and compared the results to the estimates produced by the ISO 25745-2 standard and the older VDI 4707-1 scheme. They found that both the ISO and VDI methods provide a useful baseline but can diverge from actual usage if the input parameters like number of trips or idle times are misinterpreted. Their analysis highlighted key points to improve the accuracy of these schemes: for instance, correctly identifying the lift's usage category and total number of starts (trips) is crucial for a good prediction.

2.2 Running Energy Consumption Calculation

The VDI 4707-1:2009 guideline and ISO 25745-2:2015 standard provide two widely utilised methodologies for estimating energy consumption and labelling energy efficiency in lifts. While the VDI guideline takes a focused approach tailored to specific lift technologies, the ISO standard offers a more comprehensive framework applicable across various contexts. Both methodologies emphasize the importance of load modelling, power quality considerations, and the incorporation of energy management strategies to optimize lift performance. In both methodologies the daily energy consumption consists of daily running energy and daily non-running energy. And the annual consumption is obtained by multiplying the daily energy consumption with the number of days the lift is in operation (Tukia, 2019).

Given the primary focus of this thesis on daily energy consumption and the significance of daily trip counts, only the aspects of these methodologies relevant to daily running energy consumption calculation will be discussed.

- VDI 4707-1:2009 daily energy consumption

This approach involves calculating E_{spc} which called the travel-specific energy demand (Wh/kgm)

$$E_{\mathrm{spc}} = \frac{E_{\mathrm{rc}}\, k}{2QS_{\mathrm{rc}}} \tag{1}$$

where Q is the rated load (kg), S_{rc} is cycle length (m), and k is a load factor which depends on the counterbalance ratio of lift, , usually ranging from 0.4 to 0.5 of the rated load.

According to VDI, E_{rd}, daily running consumption (Wh) is calculated by:

$$E_{\mathrm{rd}} = E_{\mathrm{spc}} Q s_{\mathrm{nom}} \tag{2}$$

where s_{nom} is the total distance travelled during a day. This distance is calculated assuming that the lift moves at the nominal speed during the operational hours, which vary depending on the usage category of the lift system (Tukia, 2019).

The total daily energy consumption then is calculated by adding daily non-running energy consumption resulting from standby demand. And finally, the total annual energy consumption is obtained by multiplying the daily energy consumption with the number of operating days which is supposed to be 365 in VDI (Tukia, 2019).

- ISO 25745-2:2015 daily energy consumption

According to (ISO 25745-2:2015), it provides a widely referenced method for predicting lift usage and energy consumption. The methodology is being used in this standard is partially similar to VDI method but with some changes.

The standard defines two methodologies for estimating energy consumption. The first approach determines energy consumption per meter of travel under unloaded conditions. It calculates the energy used per cycle as:

$$E_{\mathrm{rav}} = E_{\mathrm{rc}} \cdot \frac{s_{\mathrm{av}}}{s_{\mathrm{rc}}} \tag{3}$$

Where E_{rav} is the average unloaded running cycle energy and E_{rc} is the total energy for a reference cycle. S_{av} and S_{rc} are the mean travel distance and reference travel distance respectively.

The second approach which is more detailed focuses on acceleration, deceleration and door operation energy use and introducing new method for calculating the cycle energy.

Both methodologies in ISO standard use the same equation to estimate daily running energy According to the ISO 25745-2 standard, the formula for calculating the daily running energy consumption of a lift is provided:

$$\mathrm{E_{rd}} = \left(\frac{\mathrm{K_l n_d E_{rav}}}{2}\right) \tag{4}$$

Where E_{rv} is the running energy consumption of an average cycle (Wh), n_{d} is the number of trips per day and K_{l} is the load factor.

The daily trips number in this standard is estimated based on predefined usage patterns that elaborated in Table 1. And the load factor is calculated as:

$$K_{\mathrm{l}} = 1 - QC_k \tag{5}$$

Where Q is average per-unit load and C_{k} is a coefficient based on counterweight ratio. the different value of C_{k} is provided in the standard based on different values of counterbalance ratio

Both Q and C_{k} values are provided in the standard, with Q determined based on the usage category and rated load, and C_{k} corresponding to specific counterweight settings (ISO 25745-2., 2015).

In this standard, the usage of an individual lift is categorized according to Table 1 and the prediction of usage category is referred to the observation or a trip counter and for cases that data is not available, usage category and accordingly the number of trips is suggested in this table based on categorizing the lifts that have some characteristics. Based on this table, the lifts are separated into six categories based on use. The variables for assessing the use categories are number of daily trips, average range of lifts, and average rated speed. The use categories go from very low use intensity (1) up through extremely high use intensity (6). The use categories are employed in order to provide comparable results based on energy efficiency instead of total energy use (ISO 25745-2., 2015).

Table 1- Number of trips per day and operating days per year.; published by ISO 25745-2, 2015

Usage category	1	2	3	4	5	6
Usage intensity	Very low	Low	Medium	High	Very high	Extremely high
Number of trips per day (Typical range)	50 (<75)	125 (75 to <200)	300 (200 to <500)	750 (500 to <1000	1500 (1000 to <2000	2500 (≥2 000)
Typical buildings and usage (Operating days per year)	Residential building up to 6 dwellings (360d) Residential Care Home Small office or administrative building with few operations	Residential building up to 20 dwellings (360 d) Small office or administrative building with 2 to 5 floors (260 d) Small hotels (360 d) Office car parks (260 d) General car parks (360 d) Main line railway stations (360 d) Library (312 d) Entertainment centres (360 d) Stadia (intermittent)	Residential building with up to 50 dwellings (360 d) Medium-sized Office or administrative building with up to 10 floors (260 d) Medium-sized hotel (360 d) Airports (360 d) University (260 d) Small hospital (360 d) Shopping centre (360 d)	Residential building with more than 50 dwellings (360 d) Large office or administrative building with more than 10 floors (260 d) Large hotel (360 d)	Very large office or administrative building over 100 m height (260 d)	Very large office or administrative building over 100 m height (260 d)

The ISO 25745-2 standard has been widely used as a direct or baseline reference for estimating the number of daily trips in many research studies. For instance, a new estimation method based on this standard is presented in Table 2, where the estimated number of daily lift trips categorized by building type and number of floors, following the classification principles outlined in ISO 25745-2. Each usage category is associated with a normally distributed number of trips, where μ represents the average (mean) and σ the standard deviation, reflecting variations in lift demand across different building profiles.

Table 2-Determination of the number of trips per day per lift estimated by ISO 25745-2 standard. (Tukia et al., 2019)

Number of floors (n_{fl}), {building type}	Usage Category	Number of trips, Normally distributed, average (μ) and standard deviation (σ)
$n_{\mathrm{fl}} \leq 2$	1	$\mu = 50$, $\sigma = 12.5$
$2 < n_{\mathrm{fl}} \leq 5$ OR {entertainment}	2	$\mu = 125$, $\sigma = 37.5$
$5 < n_{\mathrm{fl}} \leq 10$ OR airport {5% of public transport} OR {hospital} ($n_{\mathrm{fl}} < 4$) OR {retail}	3	$\mu = 300$, $\sigma = 100$
$10 < n_{\mathrm{fl}} \leq 25$ OR {hospital} ($n_{\mathrm{fl}} \geq 4$)	4	$\mu = 750$, $\sigma = 125$
$n_{\mathrm{fl}} > 25$	5	$\mu = 1500$, $\sigma = 250$
$n_{\mathrm{fl}} > 25$ (second category)	6	$\mu = 2500$, $\sigma = 250$

Based on the discussion in this section, two primary methodologies are used to estimate running energy consumption. By combining daily running energy consumption with the total number of operating days, the total annual energy consumption can be calculated.

Although these standards provide a structured framework for estimating the number of daily trips, its practical application is sometimes limited. In real-world scenarios, the required detailed data may not be available, and the method's reliance on representative values for each category can restrict the precision of the estimations.

2.3 Lift Energy Efficiency and Modernisation Impact

In recent years, various modernisation solutions have been developed to improve energy efficiency and to reduce energy consumption in lift systems. These solutions can be categorized into hardware and software improvements.

Replacing older geared systems with gearless traction technologies, particularly those using Permanent Magnet Synchronous Motors (PMSMs), has demonstrated significant energy-saving potential (Wang et al., 2017). The adoption of Machine Room-Less (MRL) designs and regenerative drives further enhance performance by reducing space requirements and recovering braking energy (Anh & Duc, 2022). Incorporating energy-efficient lighting, such as LEDs, also contributes to lower energy use and heat generation in confined spaces (Al-Kodmany, 2015).

Software innovations have likewise played a key role in improving efficiency. Advanced control systems, including Destination Control Systems (DCS) and intelligent dispatching algorithms, optimize lift operations by reducing unnecessary stops and minimizing idle times (Zheng et al., 2013; Wang et al., 2012; Latif et al., 2016). Additionally, standby energy-saving solutions, such as motion-sensor lighting and adaptive ventilation can significantly reduce energy consumption during non-operational periods (Sachs et al., 2015). These approaches align with global standards such as ISO 25745 and VDI 4707.

This section explores these modernisation technologies, their functionality, and their impact on energy efficiency in lift systems. Later in the result part, the six case studies energy consumption of six common lift configuration is studied and the effect of modernisation in energy consumption for these cases is analysed.

Lift Technology

Lifts are generally categorized into two main types based on their operating mechanism: hydraulic and traction lifts. The hydraulic lifts are lifts operated by a piston that moves within a cylinder. A motor runs hydraulically supplied oil into the cylinder in order to push the piston. The piston raises the lift cab. Despite the regeneration possibility usually the potential energy of hydraulic lifts is lost and electrical valves control the release of the oil for a gentle descent (De Almeida et al., 2012). Typically, Hydraulic lifts operate at very low speeds and are therefore suited to low-rise situations. Hydraulic lifts are capable of high loads and are often used as a goods lift to carry heavy items. These lifts are less energy efficient as they require more power and generate excessive heat, contributing to higher operating expenditures (KONE, 2020). Conversely, traction lifts use steel or carbon-fibre cables or ropes to hang the car, which is powered through an electrically powered sheave (pulley). Ropes are attached to a counterweight, which counteracts the weight and saves energy required for transportation. Due to their improved energy efficiency, higher speeds, and more comfortable ride, this system makes the use of traction lifts the most popular in the case of mid-rise buildings as well as tall buildings. (De Almeida et al., 2012). In 2012, The research identified that geared traction lifts dominate the market of installed base in Europe with over two-thirds of the market share, being affordable and popular in buildings. Gearless traction lifts with 8% market share have the preference in offices as well as shopping malls because they use less energy as well as perform well. Hydraulic lifts cover the remaining market with their use mainly being in low-rise buildings (De Almeida et al., 2012). However, the market of new installation for lifts is dominated by gearless lifts.

The traction lifts have two different types: geared and gearless. Geared traction uses a gearbox attached to an electric motor to turn the hoist sheave and move the rope. They operate on either Direct Current (DC) or Alternating Current (AC) and can usually reach speeds of up to 2.5 m/s, which is slower than gearless traction lifts but faster than many other types. These lifts can travel generally up to 30 m and lift loads up to 2500 kg with 1:1 roping suspension (KONE, 2020). While a gearless lift operates without a gearbox and use a direct drive system where the motor is directly connected to the hoist sheave. These lifts have higher energy efficiency, greater speed capabilities (up to 12 m/s) and travel distance over 150 meters (KONE, 2020). A study showed that gearless traction uses 25 percent less energy than geared ones (Al-Kodmany., 2015). Also, another study showed that efficient motor selection can have impact on energy saving. For example, Permanent Magnet Synchronous Motors (PMSMs), used in gearless lifts, offer up to 30% energy savings compared to traditional motors. It exhibits a high-power density and a favourable torque-to-inertia ratio, which contributes to their ability to operate efficiently across various load conditions (Wang et al., 2017). This efficiency is particularly evident during startup and acceleration phases, where gearless systems can provide smoother and more energy-efficient operation (Wang et al., 2017). In contrast, geared systems often experience energy losses due to the mechanical friction and inertia associated with the gear train, which can lead to increased energy consumption during operation (Ang et al., 2022).

Regenerative vs Non-Regenerative Drive

Regenerative drives recover energy during specific operating conditions by capturing braking energy and converting the resulting harnesses mechanical energy and converts it into electricity for reuse within the building's electrical system or used for subsequent trips. This process can significantly reduce the overall energy consumption of the lift system. Studies indicate that regenerative lifts can save between 20% to 43% of energy compared to their non-regenerative counterparts, depending on usage patterns and building characteristics (Anh & Duc, 2022). Figure 1 shows the consumption and generation in lifts with electrical drive and in Table 3 all scenarios that drive system cause generating energy or saving energy. Depending on the operational pattern in lift, saving and generating energy are different. For instance, in high-rise buildings where lifts frequently ascend and descend, the potential for energy recovery is maximized, making regenerative systems particularly advantageous (Thebuwena et al., 2024).

Table 3- Scenarios for generating energy through regenerative drive

Scenario	Conventional System	Regenerative Drive
Lightly loaded lift going up	Energy lost as heat; counterweight does most of the work	Harnesses braking energy & converts mechanical power into electricity
Heavily loaded lift going down	Energy lost as heat; gravity does most of the work	Harnesses braking energy & converts mechanical power into electricity
Lift slowing down	Energy lost through braking	Harnesses braking energy & converts mechanical power into electricity
Cooling system efficiency	Excess heat requires cooling system	Eliminates excess heat, reducing cooling needs
Compact design efficiency	Larger, less efficient drives	Smaller, compact, and more efficient

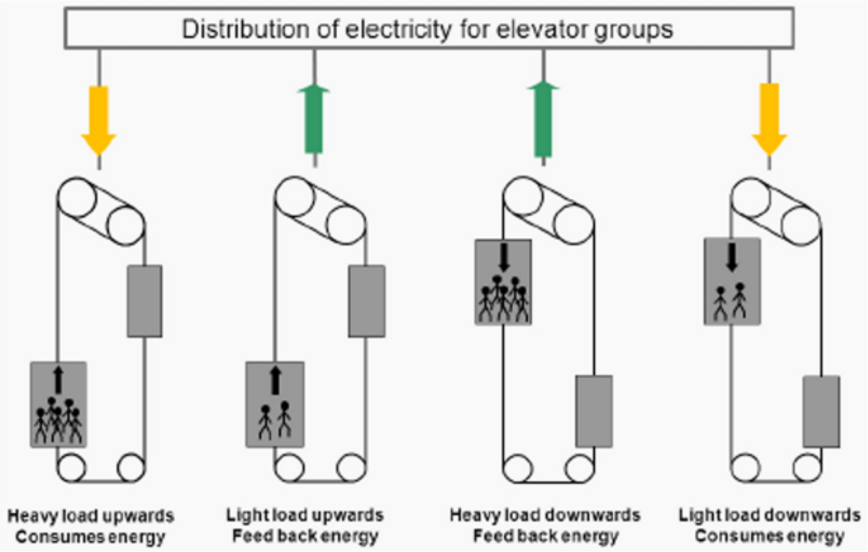


Figure 1- Regenerative drive system functionality (Thebuwena et al., 2024)

Control Systems

Lift control systems are the chief system accountable for lift movement, safety, and operating efficiency. These systems incorporate both hardware and software to ensure optimal operation, smooth trips, and traffic regulation. Lift control technology has evolved significantly in recent years, transitioning from relay-based systems to microprocessor-driven and IoT-enabled controllers. The functionality and evolution of lift control systems have undergone significant advancements to enhance efficiency and improve user experience.

For example, in grouped lifts, the primary function of a lift control system is to manage the dispatching of lifts in response to passenger requests. Traditional systems operated on a simple call-response mechanism, where lifts were dispatched to the nearest call button. However, this method often resulted in inefficiencies, particularly in high-rise buildings where multiple lifts operate simultaneously (Zheng et al., 2013).

The introduction of Destination Control System (DCS) has significantly improved lift efficiency, this system which is widely used by manufacturers, such as KONE, improves lift efficiency by grouping passengers traveling to the same or nearby floors, reducing the number of stops per trip. Unlike traditional dispatching, where passengers enter a lift and select their floor inside the cab, DCS requires users to input their destination before boarding, allowing the system to assign them to the most efficient lift. This method significantly reduces waiting and travel times, as it prevents unnecessary stops and evenly distributes passengers among available lifts. DCS also incorporates AI, travel forecasting, and optimization algorithms, which help predict peak traffic hours and adjust lift operations accordingly. By minimizing stop-and-go movement, DCS cuts down on acceleration and deceleration cycles, leading to lower energy consumption. Additionally, since lifts operate in a more structured and predictable manner, they spend less time idling, allowing them to enter standby mode more frequently, further enhancing energy efficiency and sustainability in high-rise buildings (Latif et al, 2016).

Moreover, the development of advanced algorithms, such as genetic algorithms and fuzzy logic, has further enhanced the efficiency of lift control systems. These algorithms optimize dispatching strategies based on various factors, including passenger demand, travel distance, and energy consumption (Wang et al., 2012). The use of such algorithms has been shown to reduce energy consumption by minimizing the number of trips and optimizing lift routes.

Furthermore, the implementation of advanced control algorithms is crucial for optimizing the use of regenerative energy. These algorithms manage the flow of energy between the lift's motor and the energy storage systems, ensuring that energy is utilised effectively. For instance, Kermani et al. (2021) highlights the importance of indirect field-oriented control strategies, which provide flexibility in managing the regenerative energy received from the lift's motor. This allows for better energy management, particularly during peak demand periods when energy consumption is high.

Although group dispatching can significantly influence lift energy consumption, its impact is considered outside the scope of this project. This is due to the use of ISO 25745-2 as the selected methodology, which does not incorporate group dispatching effects within its energy calculation framework.

Lighting

The role of car lighting in lifts is a critical factor in energy consumption and efficiency, particularly in the context of modern building designs that prioritize sustainability. The implementation of Operation of Car Light (OCL) systems, which automatically switch off lights when the lift is vacant for a programmable time limit, can lead to significant energy savings. (KONE internal,2020). In addition, LED spotlights represent the most energy-efficient lighting option among incandescent, halogen, and fluorescent lamps, consuming up to 80% less energy than incandescent bulbs and lasting significantly longer. The integration of LED technology in lift ceilings not only reduces energy consumption but also minimizes heat generation, which is particularly beneficial in confined spaces. LEDs can operate effectively in various ceiling designs, including those with reflective surfaces, enhancing their efficiency by maximizing light distribution and reducing the need for additional fixtures. Furthermore, recent advancements in LED technology have led to the development of highly efficient white organic light-emitting diodes (WOLEDs), which have shown potential for applications in solid-state lighting, offering luminous efficacy comparable to fluorescent lamps (Al-Kodmany., 2015).

Standby Option

Lifts spend a vast majority of their operational time on standby or in an idle state, with energy utilised in cabin lights, ventilation, and display screens. The installation of standby solutions that limit energy expenditure in these states as a matter of course is seen as a cost-effective measure in improving lift energy efficiency. Sachs, Misuriello, and Kwatra (2015) report that standby energy usage can be a high portion of a complete lift energy expenditure, particularly in buildings with minimal usage, and its minimization can achieve energy savings up to 80%. Such energy-saving technologies are in terms of motion-sensor-based lights, power-saving ventilation shut-offs, and adaptive power management solutions that dynamically reduce standby power utilization.

Standby energy minimization is a central requirement in global energy efficiency requirements, i.e., ISO 25745, VDI 4707, and ASHRAE 90.1, which specify regulatory requirements on power minimization in standby operation (Sachs et al., 2015). Incentive strategies as well as energy code requirements in buildings fuel additional application in sustainable vertical travel. Due to its effectiveness in greatly reducing overall lift energy usage, standby power management in new lift installations as well as lift upgrading is a vital component in making lift operation more ecologically sound.

2.4 Factors Affecting the Number of Lift Trips

This section studies how building characteristics, population, and lift design parameters such as speed and load capacity affect the number of daily lift trips. Both direct and indirect influences on lift demand and trip frequency are examined to provide a comprehensive understanding of lift usage patterns.

Building Population

The total number of people in a building is a primary driver of lift demand. Design guidelines often assume that a certain percentage of occupants will use the lifts during peak periods

(Al-Sharif & Seeley, 2010). A typical case is the morning up-peak in office buildings, which is commonly used to dimension lift systems. According to CIBSE Guide D, in a multi-tenant office, approximately 11–15% of the building population may call lifts within a 5-minute up-peak period (Al-Sharif & Seeley, 2010).

Travel Height and Number of Served Floors

Taller buildings with more served floors typically see greater lift usage, as occupants rely on lifts rather than stairs. More floors also mean longer travel distances per trip, which can reduce the number of trips each lift can make per hour. One study notes that increasing building height and density *"escalate the need for vertical transportation, expanding lift usage"* (Ang et al, 2022). In design calculations, a greater number of floors increases the lift round-trip time, meaning each lift completes fewer trips in a given time, so more or faster lifts are needed to handle the same demand. The number of served floors directly correlates with the potential for stops; more floors typically result in a higher likelihood of passengers requesting service at various levels, thereby increasing the number of trips required to accommodate demand (Al-Sharif et al., 2016). This relationship is particularly pronounced in buildings with mixed-use spaces, where different floors may serve distinct functions, leading to varied traffic patterns throughout the day.

Type of Building and Traffic Patterns

The type of building—whether residential, commercial, or mixed-use—also affects lift traffic patterns and the number of trips. For instance, in office buildings, lifts handle heavy traffic surges during the morning rush, lunchtime, and evening departure, leading to a higher number of trips each day. In contrast, residential buildings have a more evenly spread-out usage throughout the day. Studies have shown that lifts in commercial buildings tend to operate with heavier loads, at higher speeds, and make more trips compared to those in residential settings (Ang et al, 2022). but the presence of amenities such as gyms or pools can create additional demand for lift services. Moreover, the design of the building, including the layout of floors and the distribution of lift shafts, can influence how effectively lifts serve different areas, thereby impacting the number of trips made (Al-Kodmany, 2015).

Group Size and Passenger Demand

The number of lifts in a group has a big impact on how often each lift needs to run. When there are more lifts working together, the overall passenger demand is shared, meaning each individual lift makes fewer trips. Research and simulations have shown that increasing the number of lifts reduces waiting times and minimizes the number of stops per ride, since passengers are spread across multiple cars. For instance, if a building increases its lift group size from 2 to 4, the workload is better distributed, allowing each lift to make fewer round trips in the same amount of time. A study found that adding more lifts improved service, as passengers could ride in separate cars, reducing the number of stops and making trips faster. On the other hand, smaller lift groups mean each lift has to work harder, making more trips to accommodate all calls. (Fernández & Cortés, 2015).

Speed and Load Considerations

The capacity (rated load) and speed of a lift play an important but indirect role in how well it meets passenger demand and how many trips it makes each day. In simple terms, larger and faster lifts can carry more people per trip and complete more trips in the same amount of time, helping to serve more passengers efficiently. However, there are practical trade-offs to consider:

Rated load: A larger lift can move more passenger per trip, meaning that it can potentially reduce the number of trips needed to move a population demand, For instance, doubling capacity could allow half as many trips if usage is at full load. On the other hand, if the car is large, it can cause longer boarding times since each stop takes longer as people shuffle in and out (Al.Sharif et al, 2016). This diminishing return means designers do not simply maximize car size; an oversized car might make fewer trips because each trip has more dwell time at floors. Empirical design rules limit single lift capacity (e.g. rarely above ~2000 kg/26 persons for passenger lifts) to avoid such inefficiency (Al.Sharif et al, 2016). Therefore, capacity must be balanced a suitably large car helps reduce crowding and efficiently handles peak traffic, ensuring that passengers do not have to wait for the next ride. However, if a lift is too large compared to actual demand it might lead to more trips.

Rated speed: The rated speed plays an important role in how quickly it completes a roundtrip. In taller buildings, higher speeds can significantly reduce travel time, allowing the lift to complete more trips per hour, as long as there is enough demand. However, this relationship is not always straightforward. Since lifts need time to accelerate and decelerate, a higher speed does not always make a difference for short trips, where they might not reach their maximum speed. Therefore, for short distance it might have no impact on trip time and trip number. For example, a study showed that increasing lift speed from 2 m/s to 3 m/s had no impact for a short 4-meter rise, because lift cannot exceed the speed of 1.4 m/s (Christy, 2014). In essence, speed is a tool to match lift handling capacity to passenger demand – a higher speed lift can meet a high demand with fewer trips by cycling faster, whereas a slower lift might be overwhelmed unless supplemented by additional units.

Therefore, rated load and speed do not create demand, and they do not have direct effect of trip number, but they determine how well the lift system can accommodate that demand. An inadequate lift will force passenger to wait longer time or even defer trips some people may avoid waiting long time and take stairs instead. These two factors indirectly influence the daily trip count: a high speed, high-capacity lift might make more trips per day if demand calls for it (Al.Sharif et al, 2016).

Thus, in analysing lift trips per day, one must consider both demand side (building type, population) and supply side (lift performance parameters).

2.5 Theoretical Background

In the context of predicting the number of daily trips in lifts using historical data, a comprehensive understanding of the machine learning models employed is essential. This chapter will delve into three Boosted Trees, followed by an overview of performance measures including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 score. Each section will synthesize relevant literature to provide a robust theoretical background.

2.5.1 Gradient Boosting Method

Gradient boosting, as proposed by Friedman in 2001, is an iterative ensemble technique that builds models sequentially, where each new model is trained to correct the errors made by the previous models by combining many weak models, each one focusing on the remaining errors of the ensemble. It generalizes the boosting idea (known from AdaBoost for classification) to arbitrary differentiable loss functions, making it applicable to both regression and classification with different loss choices. Friedman's original work demonstrated how this gradient-based boosting paradigm can be applied to optimize common loss functions (e.g. least-squares for regression) and produced robust, state-of-the-art results in practice (Friedman in 2001). For a regression problem, a common choice of loss is the root mean squared error (RMSE), effectively the square root of the mean squared error. Using RMSE (or equivalently MSE) as the loss, the gradient boosting algorithm can be described in the following iterative steps (Friedman in 2001).

Initialize a constant model ($F_0(x)$) that minimizes the loss on the training data. For RMSE loss, $F_0(x)$ is typically chosen as the mean of the target values (since the mean minimizes the sum of squared errors) for this case, $F_0(x)$ is calculated as follows:

$$F_0(x) = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{6}$$

In this equation y_i is the target value and N, is the total number of training examples.

For each iteration $t = 1, 2, \ldots, T^1$, compute the pseudo-residuals $r_{i,t}$ for each training example i. These are given by the negative gradient of the loss with respect to the current model's prediction:

$$r_{i,t} = -\left.\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right|_{F(x_i)=F_{t-1}(x_i)}, \text{ for } i = 1,\ldots,N \tag{7}$$

Intuitively, $r_{i,t}$ represents the direction and magnitude of the error for each data point under the current model F_{t-1}.Where (L) is the loss function which is considered as MSE (Mean Squared Error) which simplifies the $r_{i,t}$ to:

$$r_{i,t} = y_i - F_{t-1}(x_i) \tag{8}$$

Since $L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$ has derivative $\frac{\partial L}{\partial \hat{y}} = -(y - \hat{y})$. In other words, with MSE loss the pseudo-residual is just the difference between the actual value and the current predicted value (Friedman in 2001).

Next step, fit a weak learner $(h_t(x))$ which is often a decision tree on the training data x_i using the pseudo-residuals $\{r_{i,t}\}$ (Friedman in 2001).

And finally, update the model:

$$F_t(x) = F_{t-1}(x) + \eta\, h_t(x) \tag{9}$$

Where (η) is the learning rate, which is used to scale the addition, controlling how much the new learner corrects the model (Friedman in 2001).

---

[1] Although the iteration index t is defined to range from 1 to a maximum value T, in practice, the actual number of iterations may be determined dynamically. If the addition of new models no longer leads to meaningful improvements in reducing the residuals or overall loss, the boosting process can be stopped early. In such cases, the current value of t is considered the effective stopping point and serves as the final value of T.

Over time, several specialized implementations of gradient boosting have been developed, such as XGBoost, CatBoost, and LightGBM. These variations introduce optimizations and features, like handling of categorical variables or speed enhancements, that will be discussed.

- LightGBM

  LightGBM, introduced by Ke et al. (2017), is an optimized gradient boosting framework designed to enhance efficiency and scalability, particularly for large datasets. It employs a histogram-based algorithm that discretizes continuous features into bins, significantly reducing computational cost and memory usage during training. Unlike traditional level-wise tree growth, LightGBM utilises a leaf-wise growth strategy, which prioritizes splitting leaves that yield the highest loss reduction, leading to faster convergence and improved accuracy. Additionally, LightGBM incorporates Gradient-based One-Side Sampling (GOSS), which retains instances with large gradients while sampling from those with smaller gradients, ensuring more efficient learning. It also features Exclusive Feature Bundling (EFB), which groups mutually exclusive sparse features to reduce dimensionality, further enhancing training speed. Moreover, LightGBM natively supports categorical feature handling, allowing efficient splits without the need for extensive preprocessing (Ke et al, 2017). These optimizations enable LightGBM to outperform traditional gradient boosting implementations in both training speed and predictive accuracy, making it a popular choice for large-scale machine learning tasks.

- CatBoost

  CatBoost, developed by Yandex (Prokhorenkova et al., 2018), is a gradient boosting algorithm designed to handle categorical features efficiently while reducing biases found in traditional boosting methods. One of the key challenges it addresses is prediction shift, a form of bias caused by the way standard boosting models use training targets in iterative updates (Prokhorenkova et al., 2019). In conventional boosting, the model at iteration m is trained using all available target values up to $m - 1$. This means that a sample's prediction is influenced by its own target in earlier iterations, leading to target leakage. Over many iterations, this effect accumulates, causing the model to fit patterns that include partial information about the true outputs. As a result, the predictions for training data become different from those for unseen test data, introducing a bias that reduces the model's ability to generalize (Prokhorenkova et al., 2019). To solve this problem, CatBoost introduces ordered boosting, a technique that ensures target values used for training do not introduce bias into the model's predictions. This method helps prevent target leakage, leading to more reliable and accurate models, especially when dealing with categorical data (Prokhorenkova et al., 2019).

- XGBoost

XGBoost (Extreme Gradient Boosting), introduced by Chen & Guestrin (2016), enhances the traditional gradient boosting framework by incorporating regularization and scalability into its design. Unlike standard boosting methods, XGBoost explicitly includes a regularization term in its objective function, which helps control model complexity and prevents over fitting. If we define $f_k(x)$ as the *k*-th tree, where each tree has leaf weights w and T leaves, the overall objective function in XGBoost can be written as:

$$L = \sum_{i=1}^{n} L(y_{\mathrm{i}}, \hat{y}_{\mathrm{i}}) + \sum_{k=1}^{\Omega} \Omega(f_k) \tag{10}$$

where the first term represents the standard loss function (such as squared error for regression or logistic loss for classification) which in this study Root Mean Squared Error (RMSE) is chosen as loss function, and the second term $\Omega(f)$ is a regularization penalty:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{11}$$

Here, T is the number of leaves in a single decision tree and γ penalizes the total number of leaves in a tree, while λ applies an L_2 regularization on the leaf weights w_j to smooth their values, reducing the risk of extreme predictions (Chen & Guestrin, 2016). Each leaf node in a tree represents a prediction region where a subset of training samples falls. The value assigned to each leaf is called w_j the leaf weight, which determines the predicted value for samples reaching that leaf. Mathematically, the optimal w_j is derived using a Newton step formula:

$$w_j^* = \frac{-(\sum_{\mathrm{i} \in I_{\mathrm{j}}} g_{\mathrm{i}})}{(\sum_{\mathrm{i} \in I_{\mathrm{j}}} h_{\mathrm{i}} + \lambda)} \tag{12}$$

g_i is the first-order gradient of the loss function with respect to the predicted value, h_i is the second-order derivative (Hessian), which adjusts the weight update based on the curvature of the loss function. I_j represents the set of training samples assigned to leaf j. This formulation ensures that each leaf weight is optimized to reduce the overall loss while avoiding extreme values due to the presence of the λ term, which acts as a regularize (Chen & Guestrin, 2016).

In the special case where $\lambda=\gamma=0$, this objective function simplifies to standard gradient boosting, but the inclusion of regularization ensures that XGBoost uses simpler, more generalizable models. (Chen & Guestrin, 2016)

2.5.2 Hyperparameters in Gradient Boosting

Gradient boosting algorithms have several hyperparameters that can significantly influence their performance. Some of the most critical hyperparameters include:

1. Learning Rate (η): This parameter controls the contribution of each tree to the final prediction. A smaller learning rate typically leads to better performance but requires more trees to achieve convergence, increasing computational cost (XGBoost, 2025).
2. Number of Estimators (n_{est}): This hyperparameter specifies the number of boosting stages to be run. More estimators can lead to better performance but may also increase the risk of overfitting (XGBoost, 2025).
3. Minimum Child Weight: In XGBoost, this parameter sets the minimum sum of instance weights (Hessians) needed in a child (leaf) (XGBoost, 2025). This effectively means a node will not be split if the resulting leaf would have too few samples (or too little total Hessian weight). For example, if min_child_weight = 5, any split that results in a leaf with a total Hessian (which for binary logistic roughly corresponds to number of samples times a factor) less than 5 is discarded (XGBoost, 2025). In regression tasks (where Hessian can be constant 1), this is like requiring at least 5 instances in each leaf (XGBoost, 2025).
4. Subsample: This is the fraction of training instances to use for fitting each tree (stochastic boosting). For example, subsample = 0.5 means each tree is trained on a random 50% of the data. (XGBoost, 2025).
5. Gamma (γ): This is the parameter discussed earlier that specifies the minimum loss reduction required to make a further partition on a leaf node. It acts as a regularization term, with higher values leading to more conservative models (XGBoost, 2025).
6. Column Subsampling (*colsample_bytree,,colsample_bylevel, colsample_bynode*): XGBoost allows sampling of features (columns) in addition to rows.

colsample_bytree is the fraction of features to randomly select for each tree; *colsample_bylevel* applies at each tree level; *colsample_bynode* at each split node (XGBoost, 2025). These can be used separately or in combination. For example, *colsample_bytree* = 0.8 means each tree is built using 80% of the features which is randomly chosen (XGBoost, 2025).

7. Lambda (λ): This parameter represents the L2 regularization weight. It helps to control overfitting by penalizing large coefficients in the model. A higher value of (λ) increases the regularization effect, which can lead to simpler models that generalize better to unseen data. Regularization is crucial in preventing the model from fitting noise in the training data (XGBoost, 2025).
8. Scale Positive Class Weight: This hyperparameter is particularly useful for imbalanced datasets. It adjusts the balance of positive and negative weights, allowing the model to pay more attention to the minority class during training. Setting this parameter appropriately can significantly improve the model's performance on imbalanced datasets by reducing bias towards the majority class (XGBoost, 2025).
9. Column Subsampling by Level: This parameter specifies the fraction of features to be sampled at each tree level. By controlling the number of features used for each tree, it can help in reducing overfitting and improving model robustness. A lower value can lead to more diverse trees, which may enhance the ensemble's overall performance (XGBoost, 2025).
10. Column Subsampling by Node: this parameter controls the fraction of features to be sampled at each node. This hyperparameter allows for more granular control over feature selection, potentially leading to better model performance by ensuring that each tree node is trained on a diverse subset of features (XGBoost, 2025).

Certain hyperparameters are essential for inclusion in the model, while others may be considered or omitted depending on the characteristics of the dataset. An increase in the number of hyperparameters results in a more complex system, which in turn demands greater computational resources for implementation. Therefore, selecting the most appropriate hyperparameters is crucial to maintaining both model quality and computational efficiency. In Python, predefined functions assign default values to specific hyperparameters; however, not all potential hyperparameters are incorporated by default. Based on the dataset properties and specific modelling objectives, additional hyperparameters may be introduced or excluded to optimize predictive accuracy and overall model performance.

2.5.3 Importance of Hyperparameter Tuning

The performance of gradient boosting methods is highly sensitive to the choice of hyperparameters. Therefore, careful tuning is essential to achieve optimal results. Various methods can be employed for hyperparameter tuning, including:

- Grid Search: Grid search is a widely used technique for hyperparameter optimization. It involves defining a fixed set of possible values for each hyperparameter and exhaustively evaluating all possible combinations. This method ensures that the optimal combination within the predefined grid is tested. However, it can become computationally expensive, especially as the number of hyperparameters increases, due to the exponential growth in evaluations (Bergstra & Bengio, 2012). Despite its computational inefficiencies, grid search remains popular in scenarios with a manageable search space or sufficient computational resources. It is often coupled with cross-validation to ensure that the selected hyperparameters generalize well to unseen data (Bergstra & Bengio, 2012).
- Bayesian Optimization: This method improves upon random and grid search by using a probabilistic model, often a Gaussian process, to guide the search for optimal hyperparameters. Instead of randomly selecting hyperparameters, it balances exploration and exploitation: (Snoek, Larochelle & Adams, 2012).
    1. Exploration: It searches areas of the hyperparameter space where performance uncertainty is high.
    2. Exploitation: It focuses on regions that have previously shown promising results.

  A key acquisition function in Bayesian optimization is Expected Improvement (EI), which identifies the next hyperparameter set that maximizes the probability of performance improvement. This method is effective for optimizing expensive-to-evaluate models, such as deep learning architecture, by minimizing the number of evaluations required to find the best solution. Compared to random search, Bayesian optimization needs fewer evaluations to find optimal hyperparameters but has higher computational overhead due to maintaining and updating the surrogate model (Snoek, Larochelle & Adams, 2012).

- Random Search: This is a substitute for grid search where hyperparameters are sampled randomly from a defined distribution instead of searching through all possible combinations. Bergstra and Bengio (2012) showed that random search tends to be superior to grid search because random search is a more efficient exploration of hyperparameter space. One advantage of random search is that it mitigates the "curse of dimensionality" better than grid search. In many machine learning problems, only a subset of hyperparameters significantly affects model performance (Bergstra & Bengio, 2012), By randomly sampling, random search can often find good configurations without needing to evaluate every possible combination. This is particularly useful when computational resources are limited. However, random search lacks the ability to incorporate prior knowledge from previous evaluations. Each set of sampled hyperparameters is chosen independently, which means the method may still require a large number of evaluations to find optimal settings. (Bergstra & Bengio, 2012)
- Optuna: is an advanced framework for hyperparameter optimization, developed by Akiba et al. (2019), that improves upon traditional methods by dynamically adjusting the search space during trials. Unlike grid and random search, which require setting the hyperparameter range in advance, Optuna allows for flexible adjustments based on ongoing trial results. Optuna utilises:
    - Tree-structured Parzen Estimator (TPE): A Bayesian optimization technique that effectively models promising hyperparameter regions.
    - Multi-armed bandit strategies: These strategies dynamically allocate resources to promising hyperparameter values.
    - Trial pruning: Optuna can halt underperforming trials early, conserving computational resources if a hyperparameter configuration shows poor performance at the outset.

This adaptive approach enhances efficiency, making Optuna particularly effective for large-scale hyperparameter tuning tasks in deep learning and gradient boosting models (Akiba et al, 2019).

3 Methodology

This chapter summarizes problem identification, sources of data, and analysis strategies applied for improved estimation of lift energy consumption beyond normal ISO-based conditions. This chapter presents the research motivation, a comprehensive description of given data from KONE, and explanations for the two considered modelling strategies—Polynomial Fit and Machine Learning. The chapter further provides descriptions for using the CRISP-DM methodology as the guidance framework, making sure that technical modelling and business targets go hand in hand through the analysis procedure.

3.1 Problem Identification

As it was discussed in literature review, many companies and researchers rely on ISO25745-2 standards to estimate energy consumption in lifts. KONE, as an important member in the lift industry, provides Environmental Product Declarations (EPD). This report evaluates the environmental impact of a product throughout its lifecycle. This includes aspects such as material sourcing, manufacturing, transportation, energy consumption, and end-of-life disposal. KONE's EPD reports energy consumption and CO_2 emissions of its lift systems, helping stakeholders understand their environmental footprint. The methodology used in these reports follows the framework established by ISO 25745-2, which provides guidelines for calculating lift energy consumption. For instance, one of KONE's modernisation solutions, MonoSpace Upgrade DX, showcases how CO_2 emissions are distributed across different lifecycle stages. Figure 2 illustrates the carbon footprint distribution for this modernisation, breaking down emissions from manufacturing, delivery, installation, energy use, maintenance, and end-of-life processes. This figure emphasizes that operational energy use is the largest contributor to overall CO_2 footprint of a lift, reinforcing the importance of improving energy efficiency (KONE EPD, 2023)

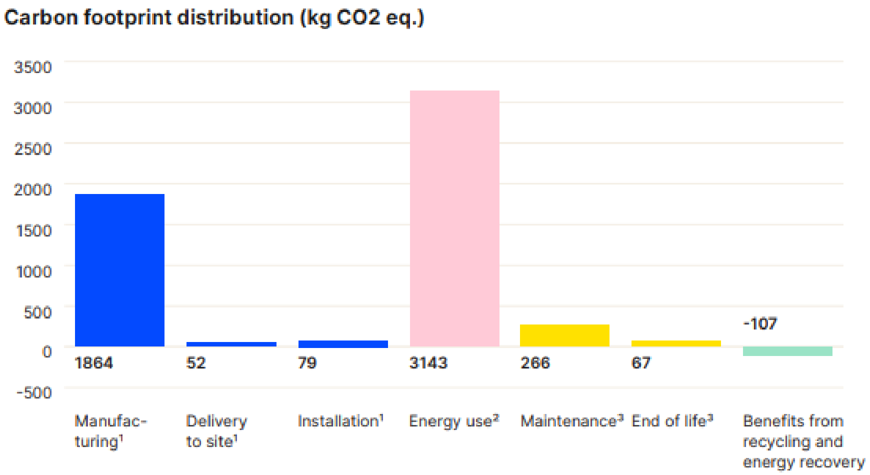


Figure 2- Distribution of carbon footprint for KONE Monospace Upgrade DX (KONE EPD,2023)

In the EPD report, operational energy consumption is calculated using the ISO 25745-2 framework, which defines a structured methodology for assessing energy efficiency. For example, In Table 3, the impact of modernisation on lift energy consumption and CO_2 emissions is illustrated through the KONE MonoSpace Upgrade DX case. For the calculation, KONE, as reported in the Environmental Product Declaration (EPD), follows the ISO 25745-2 methodology. The calculation considers the Belgium national grid mix (0.26 $kgCO_2e/kWh$) and assumes Usage Category 3 (300 trips/day) for an existing gearless traction lift with a non-regenerative V3F PMSM drive, with the following characteristics: rated load of 630 kg, rated speed of 1 m/s, and travel height of 12 meter (KONE EPD, 2023)

Table 3- Annual energy consumption before and after modernisation for the MonoSpace Upgrade DX (KONE EPD,2023)

	Annual energy consumption [kWh]	Operational emissions over 15 years [kgCO2e]
Lift before the modernisation with MonoSpace Upgrade DX	1528	5710
Lift after modernisation with MonoSpace Upgrade DX	841	3143
Reduction before and after modernisation	667	2567

Given the importance of operational energy consumption and its direct impact on CO_2 emissions in environmental reports, a more precise estimation of energy consumption can lead to a better prediction of the environmental impact of lifts. Although ISO 25745-2 provides a structured methodology, it relies on predefined usage categories and representative numbers for estimating the number of trips per day.

Instead of relying solely on simplified ISO assumptions, this study aims to use historical operational data to find patterns between trip numbers, building characteristics, and lift dimensional characteristics. By doing so, the study analyses the effect of trip numbers on annual energy consumption across five modernisation cases. Tukia (2019), in his doctoral dissertation, introduced a simplified version of ISO's trip estimation model. He mapped floor numbers to expected usage categories, making it easier to estimate lift trips in real-world scenarios. Table 4 represents his extracted model, which simplifies the ISO framework.

Table 4- Segmenting the usage category to the number of floors (Tukia, 2019).

Number of served floors	Usage Category	Number of Trips Interval	Represented Trip Number
2	1	10-75	50
3 - 6	2	75-199	125
7 - 11	3	200-499	300
12 - 26	4	500-999	750
>27	5	1000-2000	1500

This study examines historical lift data and introduces two alternative methods for estimating lift trips and energy consumption:

1. Polynomial Fit Model (Baseline model): The baseline model takes a more granular approach, focusing on the number of floors and building types to establish a continuous, empirically derived function for estimating the daily number of lift trips. In contrast to floor-based representation method, this method utilises the historical data that was used in this study to determine a fitted function for only one part of dataset that was introduced as train set that predicts the median number of lift trips per day based on the number of served floors. This approach is applied separately to five different building types: residential, office, hotel, medical, and other sectors, ensuring a sector-specific adaptation. Two functional forms, polynomial and logarithmic, are tested for each sector. The fitted functions are saved and later applied to the test dataset for prediction, after which error metrics are computed to evaluate model effectiveness.

2. Machine Learning: A more advanced technique that integrates multiple lift and building characteristics, enabling the capture of additional variability. This is assumed that at the time of installation all the parameters have been taken into account for designing the lifts, so by looking at historical data and by having trips information of these lifts after installation, machine learning approach can help to model different scenarios not only in terms of floor number but also other factors to introduce more precise estimation than baseline and what have been used so far through ISO standard.

The results from these two approaches will be compared with the simplified ISO-derived table to evaluate their accuracy and effectiveness in estimating energy consumption. To enable comparison with this approach, referred to as the representative number method (first assumption), the test dataset was segmented into floor-based groups based on the classification defined in Table 4. Based on this approach a new column is created that stores the representative trip number based on the number of served floors. And later result of prediction by these two methods will be compare with this method.

3.2 Data Collection

In the pre-processing phase of this work, information extraction was performed using SQL (Structured Query Language) for information extraction from a variety of disparate sources. Information extracted was processed and consolidated for uniformity and completeness. Once pre-processing, the finished dataset was exported in JSON format for easier analysis. Python was utilised for analysis, leveraging a range of specific packages for processing and statistical analysis of information. Anaconda was included in the computational environment, with VS Code and Jupyter Notebook being utilised for development, offering efficient environment.

The dataset for use in this analysis was delivered by KONE, a renowned global lift and escalator manufacturer. It is a rich, full-fledged one with information about 100,000 lifts, providing a picture of their operational and technical information. The dataset was captured via two key sources in the infrastructure of the organisation:

- Internal Database: There is a rich in-house database with a wealth of information regarding the technical and physical specifications of the lifts, including rated speed, rated load, group size, and geographical location. All such information was derived and compiled from rich in-house databases of the corporation.

- Telemetry Data: Some of the information is in the form of telemetry information received through telemetry systems installed in the lifts. These use sensors to track real-time operational data, such as lift stops and usage patterns. Telemetry information mirrors actual lift performance through documentation of operational statistics such as trips and days operated.

Table 5 shows the available features and columns in the aforementioned dataset.

Table 5- Overview of dataset columns and their value ranges

Feature	Description	Value Range
Rated Speed	The maximum speed of the lift, measured in meters per second (m/s)	0.5-8 m/s
Rated Load	The maximum load capacity of the lift car, measured in kilograms (kg).	200-4000 kg
Group Size	The number of lifts managed under the same group control unit	1-8 units
Building Type	The type of building that lift has been installed labelled by the company.	Residential, Office, Hotel, Airport, etc.
Location	The geographical location of the lift installations, including the city and country.	Different countries in Europe, and Asia
Number of Served Floors	The number of floors the lift serves within the associated building.	2-50 floors
Travel Height	The distance between the lowest and highest landings that the lift travels, measured in meters	2-300 m
Average Trips per Day	Extracted by dividing the total number of annual trips to the number of operating days, considering some thresholds to apply the effect of weekly variability in number of trips for some types of buildings such as offices.	10-2000 trips

3.3 Research Strategy

There are various methodologies for conducting machine learning applications in different domains. Since for many machine learning applications, the knowledge of data scientists alone is often insufficient, the involvement of domain experts is essential. One of the most widely adopted data mining methodology frameworks is the Cross-Industry Standard Process for Data Mining. (Martínez-Plumed et al., 2021)

The CRISP-DM framework has emerged as a widely adopted methodology for conducting data science research and projects. This iterative process model provides a structured approach to navigating the complexities of the data analysis lifecycle, from understanding the business context to deploying and maintaining the final solution. (Martínez-Plumed et al., 2021)

The CRISP-DM process consists of six main phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. (Martínez et al., 2021) Each phase requires thoughtful consideration of the inherent variability and dependencies within the data analysis workflow. This iterative, cyclic nature of the CRISP-DM process allows for flexibility and continuous improvement throughout the data science project lifecycle. (Martínez et al., 2021)

Business Understanding phase lays the foundation for the entire project, where the researcher or data science team delves into the specific business objectives, identifies the critical success factors, and develops a project plan that aligns with the organization's strategic goals. This initial phase ensures that the subsequent phases remain focused and relevant to the problem at hand (Jaggia et al., 2020).

Data Understanding phase involves exploring and analysing the available data to gain insights into its quality, relationships, and potential issues. This phase is crucial in identifying the variability inherent in the data and understanding how it might impact the modelling and deployment stages (Ayele, 2020).

Data preparation phase encompasses the tasks of cleaning, transforming, and integrating the data to ensure it is suitable for the modelling process (Singh et al., 2022).

Modelling phase involves the selection and application of appropriate machine learning or statistical techniques to the prepared data, with a focus on capturing and addressing the identified variability (Singh et al., 2022).

Evaluation phase assesses the performance of the models, considering the original business objectives and the identified variability. And finally, the Deployment phase involves the integration of the developed models into the organization's operational systems, ensuring their long-term maintenance and continuous improvement (Singh et al., 2022).

This study utilised a machine learning model to investigate energy usage within the lift industry. Cultivating a thorough comprehension of the industry and augmenting domain expertise prior to data analysis is indispensable. Furthermore, as the CRISP-DM methodology not only addresses data analysis and understanding but also ensures alignment between data and business objectives, it was the preferred approach adopted for this research endeavour. Additionally, since this study necessitated multiple iterative cycles to analyse the data and ensure the reliability of data selection and pre-processing, the CRISP-DM framework was employed. Figure 3 shows the iterative nature of the CRISP-DM process.

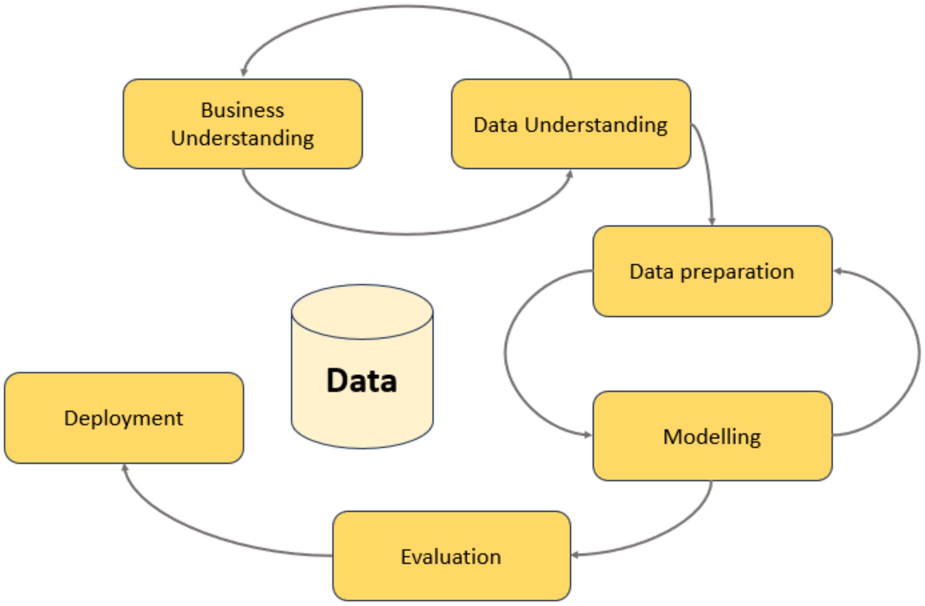


Figure 3-CRISP-DM framework, research methodology

4 Data

4.1 Data Understanding

The primary libraries utilised for analysis in such a case include Pandas, NumPy, Matplotlib, and Seaborn, with each tool having a specific function in both visualization and exploration. Pandas was utilised for data manipulation, including loading, cleaning, and summation of the data. NumPy supported efficient numerical computations, particularly in working with arrays and in calculation of statistics. Matplotlib and Seaborn supported visualization of distributions, trends, and correlations in the data, with a deeper analysis of the dataset being conducted before developing and deploying a machine algorithm.

The first step for understanding data, is finding the distribution of different features and their correlation with target which is the prediction of average daily trips, also since the aim of thesis is finding better model and more precise way to calculate the trip number, historical data also is compared with the first assumption mentioned in Table 4.

4.1.1 Analysing of Historical Data against Representative Number Method (First Assumption)

Since the first assumption is what Tukia (2019), introduced in his dissertation and we call it as a first assumption, the historical data examined against this assumption. For that, data has been categorised in floor bins mapping to the Table 4, for different building types, to see how much percentage of lifts work in the same introduced range in this assumption. Also, since in ISO 25745-2:2015, have a same pattern for labelling the usage category for offices also same based on floors, by analysing the historical data trips number and comparing with what introduced in this assumption and standard, we will have better understanding of real data against this assumption. dataset was subsequently divided into five groups based on the number of building floors. The distribution of usage categories within each of these floor-based groups was then analysed for each of the three building segments: residential, office, and mixed-use. Figure 4 represents the proportion of each usage category for the office sectors as a sample within each floor bin for these respective segments.

As Figure 4 shows, for buildings with 3 to 6 floors, approximately 35% of the lifts were categorized as having "Very Low" usage (Category 1). In buildings with 7 to 11 floors, usage Categories 2 ("Low") and 3 ("Medium") were observed with 31% and 44% respectively. each accounting for 42% of the lifts. This analysis suggests that the anticipated usage category for most floor ranges might be overestimated. lifts in taller buildings (those with more floors) tend to exhibit a lower usage intensity than the standard categorization would predict.

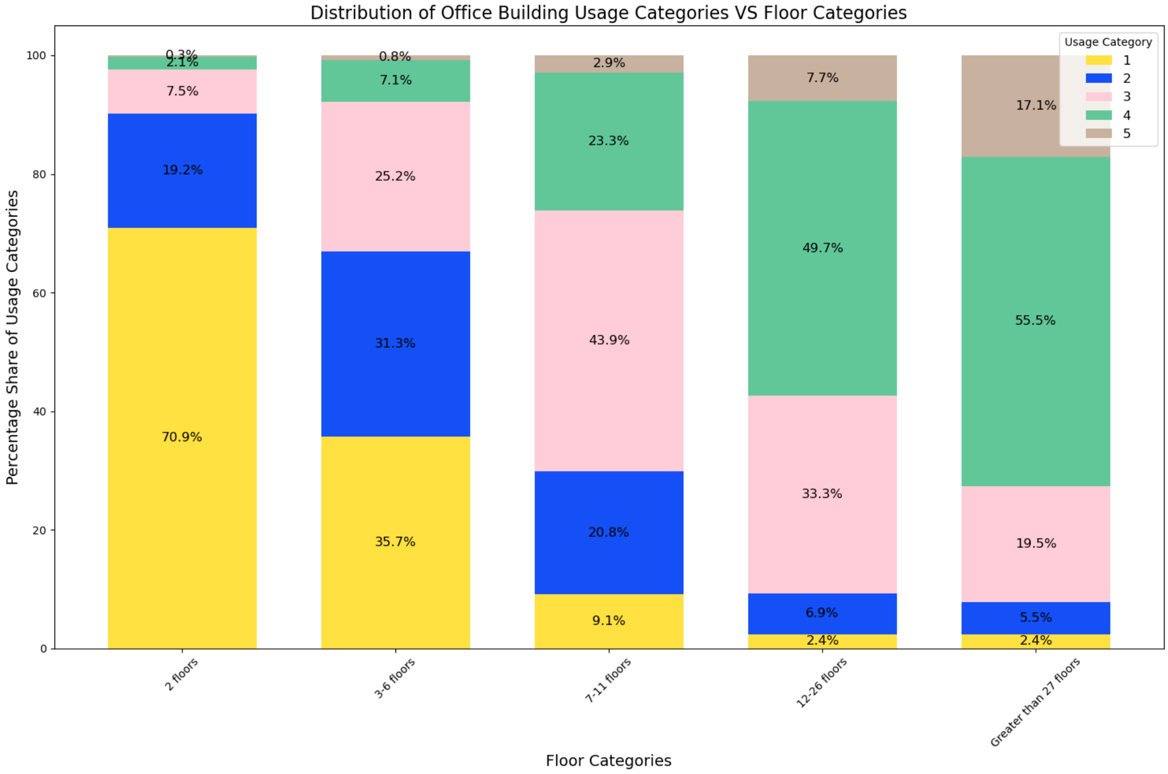


Figure 4- Distribution of trips per day for each floor category in office building. Categories: 1, very low usage; 2, low usage; 3, medium usage; 4, high usage; 5, very high usage

Figure 5 illustrates the 5th and 95th percentiles of daily lift trip counts across different floor levels, using a sample of office buildings for Malaysia. This subset was intentionally selected to explore extreme usage conditions and to evaluate how well the predefined ISO-based assumptions, referred to here as the representative number method, align with actual data. In the figure, the shaded rectangles represent ISO usage categories, while the black dashed lines indicate the representative trip numbers suggested for each floor range. The comparison highlights several inconsistencies. For buildings with more than 27 floors (UCG5), the ISO's recommended trip number (1,500 trips/day) is clearly above even the 95th percentile of observed data, suggesting a significant overestimation. In contrast, for the 3 to 6 floor range (UCG2), the ISO's minimum suggested trip count (75 trips/day) is higher than the 5th percentile of actual usage, again pointing to overestimation at the lower bound. Although some ranges, such as the maximum value for UCG2, do align with observed upper percentiles, others deviate notably from reality. This analysis of extreme cases reveals that ISO-based assumptions do not consistently reflect real-world lift behavior across all building types. Overestimations are most pronounced in taller buildings, while some lower- and mid-rise categories show partial mismatches. These discrepancies can result in substantial miscalculations, exceeding 1,000 trips per day in some cases, which, in turn, significantly impact energy consumption estimates.

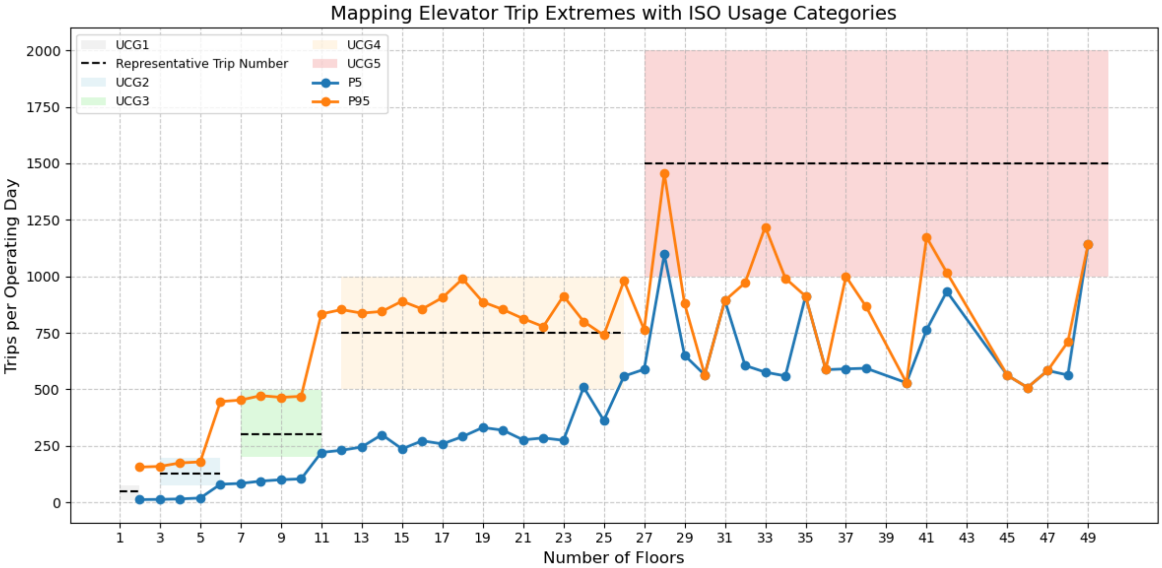


Figure 5-Analysis of extreme trip numbers by served floors

4.1.2 Relationship Between Floor Count and Daily Trip Frequency

Building upon established domain knowledge and insights from prior literature, the number of served floors in a building is a key factor affecting the frequency of lift trips per day. Taller buildings tend to support higher passenger volumes and longer travel distances, which collectively lead to increased lift usage. To investigate this pattern, the dataset was segmented based on the number of floors in each building, grouping together buildings with the same number of floors, and the resulting distribution of daily lift trips was visualized as a boxplot, shown in Figure 6.

In this figure, each box represents the interquartile range (IQR), capturing the middle 50% of trip data, with the lower and upper bounds indicating the 25th and 75th percentiles, respectively. The horizontal line within each box corresponds to the median number of trips per day for that floor group. Whiskers extend to the most extreme data points within 1.5 times the IQR, while points beyond these whiskers are considered outliers. The y-axis shows the number of trips per day, ranging from approximately 0 to over 2,500.

The figure clearly illustrates that, in general, buildings with more floors experience a higher number of trips per day. This is evident from the upward trend in both the median and upper quartile values as floor count increases. However, lower-rise buildings show greater variability in trip frequency, with wider boxes and more outliers, suggesting that additional building-specific factors such as building function, occupancy levels, or lift configurations may strongly influence lift usage in these cases. This observation highlights the limitations of using only the number of floors as a predictor of trip frequency and underscores the value of incorporating other relevant variables for more accurate modelling.

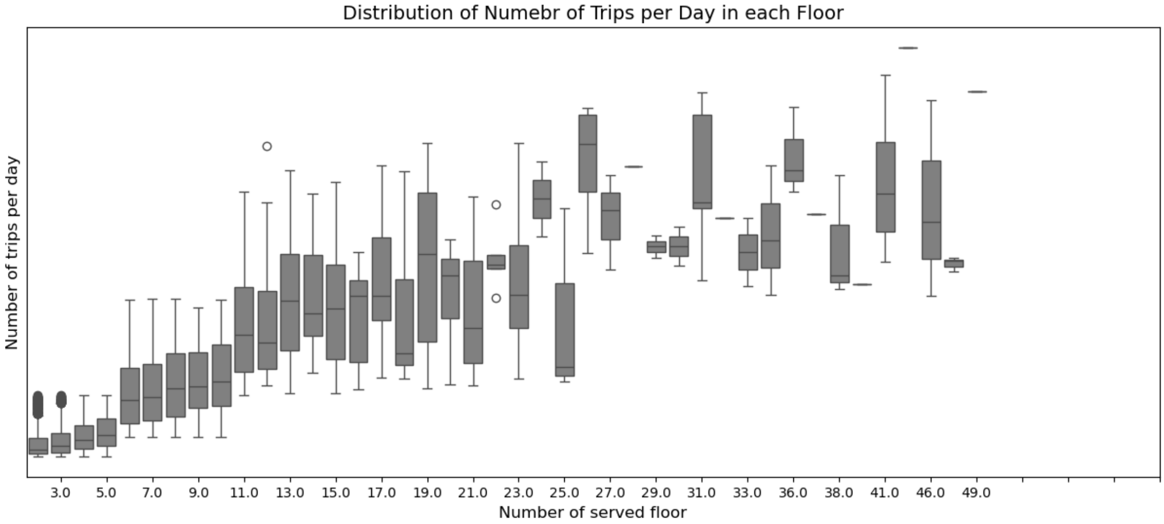


Figure 6-Boxplot of daily lift trip counts across buildings grouped by number of served floors. Each box represents the interquartile range (25th to 75th percentile), with the horizontal line indicating the median value. Whiskers extend to 1.5 times the interquartile range, and points beyond are shown as outliers. Y-axis values are omitted to protect proprietary source information.

4.1.3 Relationship Between Building Type and Daily Trip Frequency

Building type exerts a significant influence on daily trip frequency. This is evident in numerous instances where buildings with lower floor counts exhibit higher trip volumes due to their specific function. To validate this observation, Figure 7 depicts the median daily trip frequency for various building types. As illustrated, the median trip frequency for medical facilities is notably higher compared to that of office or residential buildings.

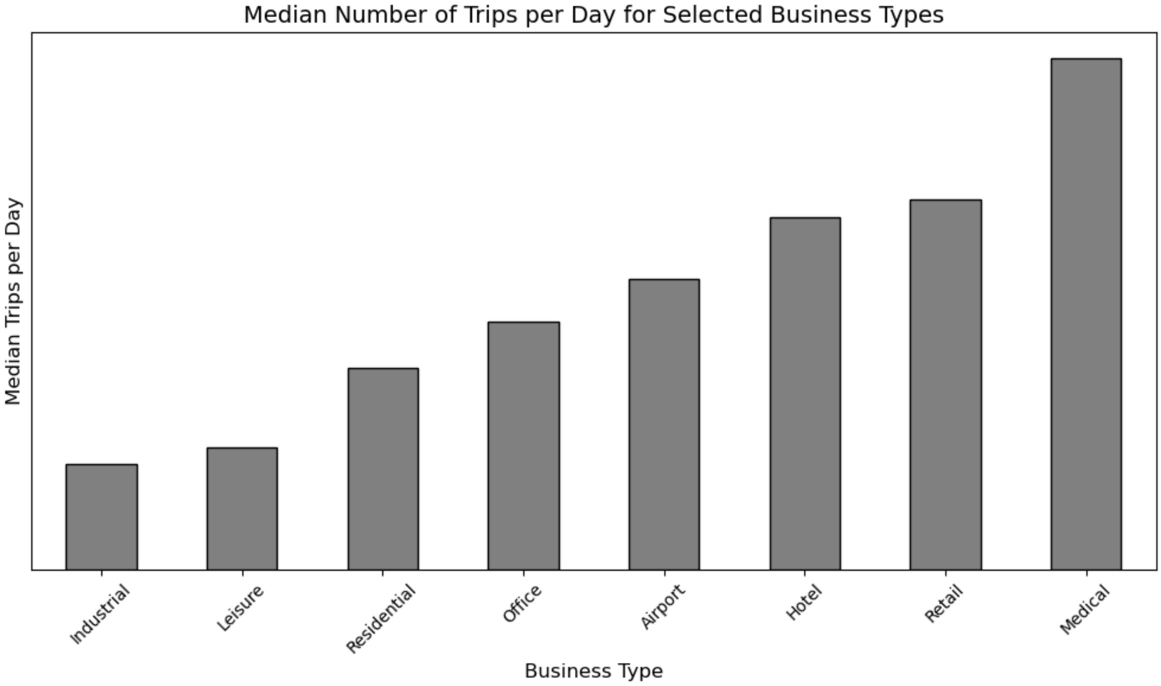


Figure 7- Categorising the median number of daily lift trips across different building types. Y-axis values are omitted to protect proprietary source information; figures are presented to highlight relative differences between categories

4.1.4 Influence of Rated Speed and Load Capacity on Daily Trip Frequency

The boxplots in Figure 8 and 9 show distributions of average daily lift trips for various loads and rated speeds. The initial hypothesis is that larger loads and rated speeds would have more trips because of their higher capacity for situations of higher demand. Generally, the expectation is confirmed by the plots with higher median numbers of trips for higher loads and rated speeds, at least up to about 2-3 m/s for rated speed and to about 1500-2000 kg for loads. There is considerable variation within each class (as gauged by box height) and many outliers indicating that these variables alone cannot be used to make accurate enough predictions. This suggests that other factors like building type, floors served by lifts, and pattern of use take their role in influencing lift demand.

While a general upward trend is evident, correlation of rated speed/load to number of trips appears to be non-linear. Trip frequency increase tends to level off at higher levels of speed/load, suggesting that other limiting factors become dominant. This would make linear model assumptions questionable in properly characterizing the relationship. Models that possess the capability to model non-linearity as well as feature interactions (e.g., Random Forests, Gradient Boosting, or neural networks) would thus be good to consider. Outliers in conditions of variability further make it necessary to include more features and even build interaction terms to make the model more powerfully predictive.

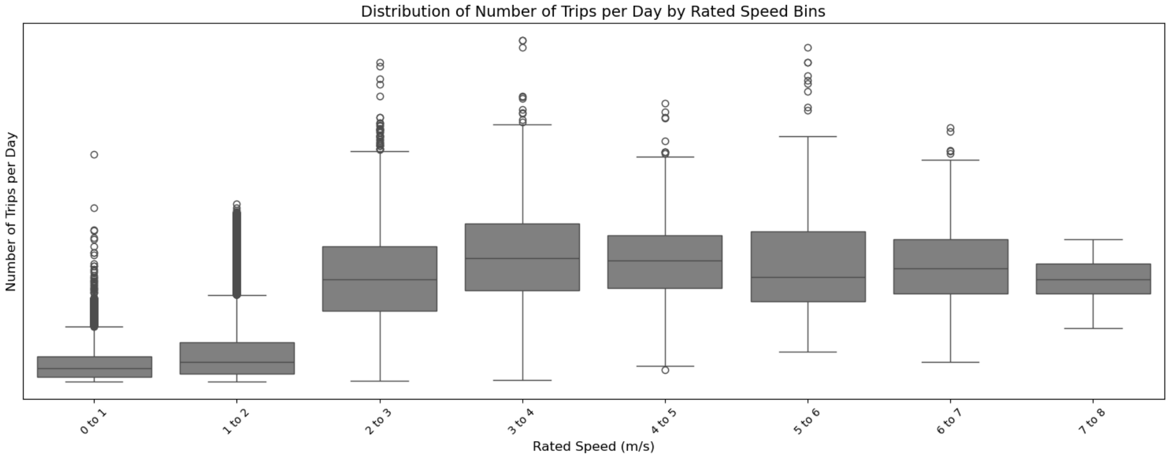


Figure 8- Boxplot of daily trips by lift rated speed. Boxes show interquartile range with median line; outliers are shown as points. Y-axis values are omitted to protect proprietary data and highlight relative differences.

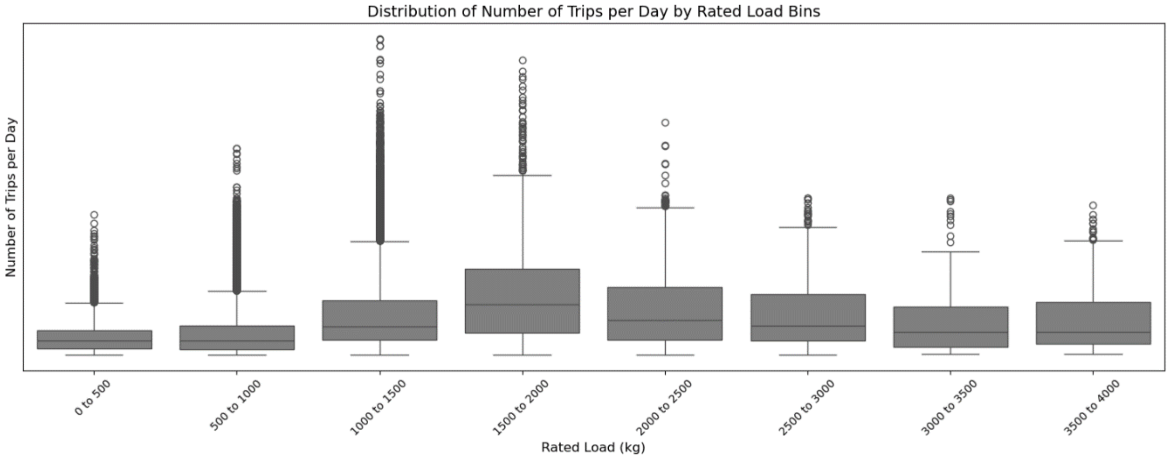


Figure 9- Boxplot of daily trips by lift rated load. Boxes show interquartile range with median line; outliers are shown as points. Y-axis values are omitted to protect proprietary data and highlight relative differences.

4.2 Data Preparation

Data preparation is a critical phase in the machine learning workflow, as the quality and structure of the dataset directly influence the performance of the trained models. A well-prepared dataset ensures that the model can learn effectively from the data, leading to more accurate predictions and generalizations. The importance of data preparation can be underscored by several key factors that need to be considered during the cleaning and preparation of datasets. the process of data cleaning is essential to eliminate inconsistencies and errors that can skew model training. As noted by Rezig et al., (2021) data conditioning often involves addressing discrepancies within the dataset, which can include handling missing values, outliers, and irrelevant features. This step is crucial as it directly affects the model's ability to learn from the data.

In this work, train model data is expansive and dispersed. Additionally, the data consists of measured characteristics of apparatus that have been stored on a human interactive platform. As counts that have been measured daily by linked sensors may contain errors within themselves, proper cleansing of data and in-depth scrutiny must be done before application. Cleansing of data and preparation shall be elaborated in this work part.

4.2.1 Missing Value and Outlier Handling

To ensure data quality before AI model training, initial pre-processing steps were conducted, including missing value handling and outlier detection.

Missing Value Handling

The dataset was first analysed for missing values by calculating the number and percentage of null values in each column. Based on this analysis, some missing values were removed, while others were imputed using appropriate statistical methods such as mean, median, or interpolation. This step was necessary to prevent gaps in the dataset that could negatively impact model performance.

However, a special case arose with the travel height variable, which represents the vertical distance (in meters) that a lift travels. Many values in the column were missing or had a value that was clearly too small, indicating that perhaps the unit of measure was not cohesive. While regression-based imputation using the relationship between floor number and travel height was considered, it was ultimately avoided due to the high correlation between these two features, which posed a risk of overfitting. Instead, the decision was made not to use interpolation, and this column was retained only as an auxiliary feature where valid values were available.

Outlier Handling

The strategy that has been applied for identifying outlier was finding extreme usage of lifts based on the framework Tukia (2019) outlined and in this thesis pointed out in Table 4. This analysis revealed a discrepancy between the predicted and observed usage categories. Specifically, a majority of lifts were observed to operate at a usage level one category lower than the predicted category. While a portion aligned with the predicted category, a smaller subset exhibited usage patterns falling into other categories.

To mitigate the impact of these discrepancies and enhance data quality, 13% of outliers, defined as those falling outside the range of one category above or below the predicted usage level, were removed from the dataset. This outlier removal strategy aimed to normalize the distribution of usage categories and improve the reliability of subsequent analyses.

4.2.2 Feature selection and Engineering

During the data understanding phase, initial exploration revealed complex relationships between the available features and the target variable, average daily trips. While certain features may exhibit a general linear relationship, individually, the rated speed and rated load showed a high variability and no clear pattern with the average daily trips. This suggests that their influence is intertwined and/or moderated by other factors. Furthermore, data limitations restrict the use of potentially crucial features like the number of dwellings in residential buildings or accurate travel height measurements, as these exhibit substantial data incompleteness. This scarcity of reliable data for key variables could limit model performance if those variables were included directly.

To address these challenges and enhance the model's predictive capabilities, feature engineering will be employed. The initial exploration resulted in an engineered the "load in speed" feature (rated load multiplied by rated speed) which is shown in Figure 10 demonstrates a stronger positive correlation with average daily trips. This variable effectively captures the combined impact of a lift's capacity and speed, revealing a more pronounced pattern than either feature alone.

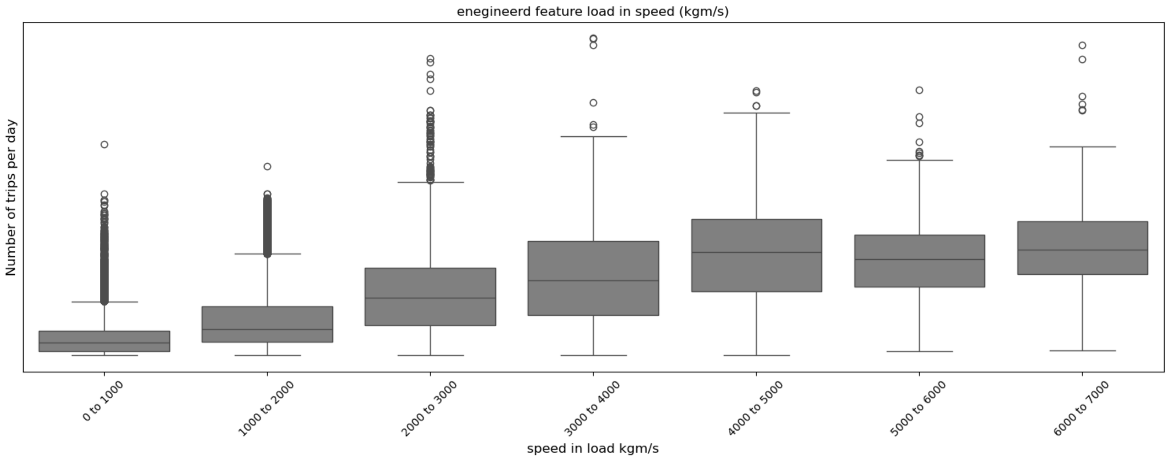


Figure 10-Boxplot of daily lift trips by the product of rated load and rated speed (kg·m/s), used here as an engineered feature representing lift throughput capacity. Boxes show interquartile range with median lines; outliers are shown as points. Y-axis values are omitted to protect proprietary data and emphasize relative patterns.

Also since the combination of floor number and building type can determine the range of trip number, one column is added to the data frame which have 2-story buildings, low-rise, mid-rise, and high-rise buildings which label to the lift by considering their building floor number then a new feature is introduced which taking into account this category and also the type of the building, since the model of this machine learning is tree-based, having better categorical features can lead to better model performance. Figure 11 shows the process of creating and labelling new features.

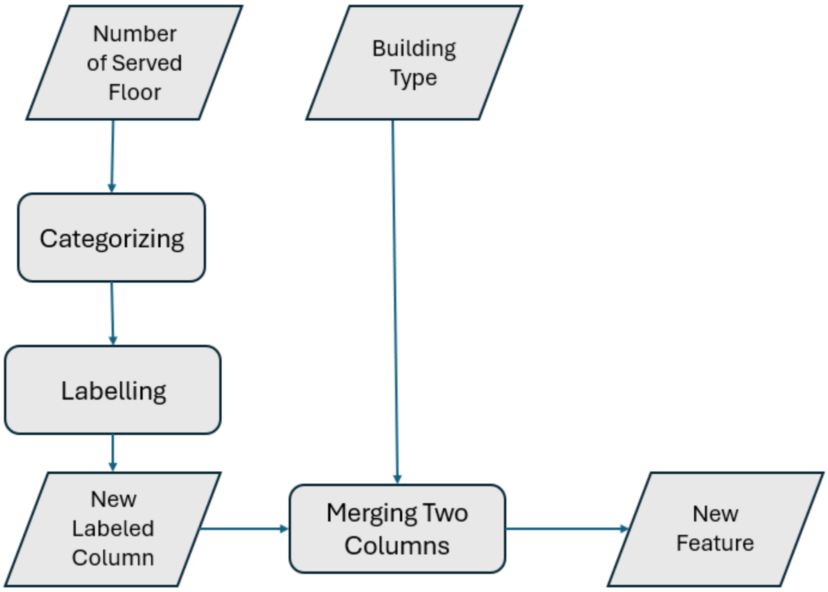


Figure 11- Process of creating a new engineered feature

By creating more features and testing them we will make selection based on feature importances that will bring to the model. Ultimately, the aim is to find and develop an exhaustive list of features that capture well the drivers behind lift trip frequency, which lead to a more accurate and dependable prediction model. The first list of features that were integrated into the model includes the Table 6 below.

Table 6-Initial feature selection

Feature Name	Description	Type
Rated Speed	Maximum operating speed of the lift (m/s).	Numerical
Rated Load	Maximum weight capacity of the lift (kg).	Numerical
Speed-Load (Engineered)	Combination of speed and load to capture their joint impact.	Numerical
Installation Country	Country where the lift is installed, influencing regulations and usage.	Categorical
Number of Served Floors	Total number of floors the lift serves.	Numerical
Group Size	Number of lifts operating as part of a coordinated group.	Numerical
Building Type	Classification of the building (e.g., residential, commercial, mixed-use).	Categorical
Building Height Classification (Engineered)	Merged building type information to define (low-rise, mid-rise, or high-rise)	Categorical

4.2.3 Handling Categorical Features and Numerical Features

Before using the model's functionalities, preparation of the features for usage in the model has to be performed. The Category Encoders library for Python provides a stable collection of encoding methods for categorical values, for use in machine learning. The library provides varying methods of encoding, including notably One-Hot Encoding and Target Encoding, with varying purposes and use.

One-Hot Encoding is a widely used technique that transforms categorical variables into a format that can be provided to machine learning algorithms to improve predictions. It creates binary columns for each category in the variable, where a '1' indicates the presence of a category and '0' indicates its absence. This method is particularly effective for nominal categorical variables with a limited number of categories, as it prevents the model from assuming any ordinal relationship among categories (McGinnis et al., 2018). However, one-hot encoding can lead to high dimensionality when applied to features with many categories, which can negatively impact model performance due to increased computational complexity and potential overfitting (McGinnis et al., 2018).

In contrast, Target Encoding is a technique that replaces a categorical variable with the mean of the target variable for each category. This method is particularly useful in scenarios where the categorical variable has a high cardinality, as it reduces dimensionality while preserving the relationship between the categorical feature and the target variable (Rodríguez et al., 2018). However, care must be taken to avoid target leakage, which can occur if the target information is inadvertently included in the training data (Rodríguez et al., 2018).

Figure 12 illustrates the process of categorical feature encoding on the data before training the AI model. The code example demonstrates a two-step process where low-cardinality categorical values are encoded with One-Hot Encoding (OHE) and high-cardinality categorical features are converted with Target Encoding (TE). The training data are fed into the fit_transform() function first, where the encoders learn from the training data and apply the transformation in the process. OHE expands categorical values into a number of binary columns, and TE replaces categorical values with the target mean for each category. To prevent data leakage, the transform() function is then invoked on the test data, where the learned transforms on the training data are uniformly applied without refitting.

```
ohe = OneHotEncoder(
    cols=cat_low,
    use_cat_names=True,
    handle_unknown='impute'
)

te = TargetEncoder(
    cols=cat_high
)

X_train_enc = ohe.fit_transform(X_train.copy())
X_train_enc = te.fit_transform(X_train_enc, y_train)

X_test_enc = ohe.transform(X_test.copy())
X_test_enc = te.transform(X_test_enc)
```

Figure 12- Categorical feature encoding process

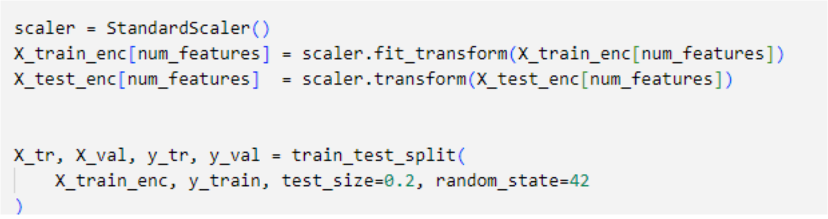
4.3 Modelling

In this subsection, the approach for developing and tuning the predictive models is presented. Three commonly used gradient-boosted frameworks were used, i.e., LightGBM, XGBoost, and CatBoost, with each of them being known for working with structured data and for having ample parameters for solving data imbalance and overfitting. In this subsection, the data splitting strategy is presented. The default parameters of each model, the new parameters, and the process of tuning parameters with the utilization of Optuna are introduced. The approach for maintaining a stable validation process and the factors considered for solving computational cost are also presented.

To facilitate the process, the following Python libraries from the environment for Scikit-learn and some machine learning frameworks such as LightGBM, XGBoost, CatBoost, and Optuna were utilised.

4.3.1 Train-Test Splitting and Data Partitioning:

The encoded dataset was divided into training (80%) and validation (20%) subset. This ensures that the model is trained on one subset of the data while evaluated on unseen data, preventing overfitting. The random seed also was set to ensure reproducibility across different runs. The Figure 13 represents the encoding a well as the train and test process.



```
scaler = StandardScaler()
X_train_enc[num_features] = scaler.fit_transform(X_train_enc[num_features])
X_test_enc[num_features]  = scaler.transform(X_test_enc[num_features])


X_tr, X_val, y_tr, y_val = train_test_split(
    X_train_enc, y_train, test_size=0.2, random_state=42
)
```

Figure 13-Preparing the train and validation set for finding the best hyperparameter

4.3.2 Model Selection and Hyperparameters Tuning

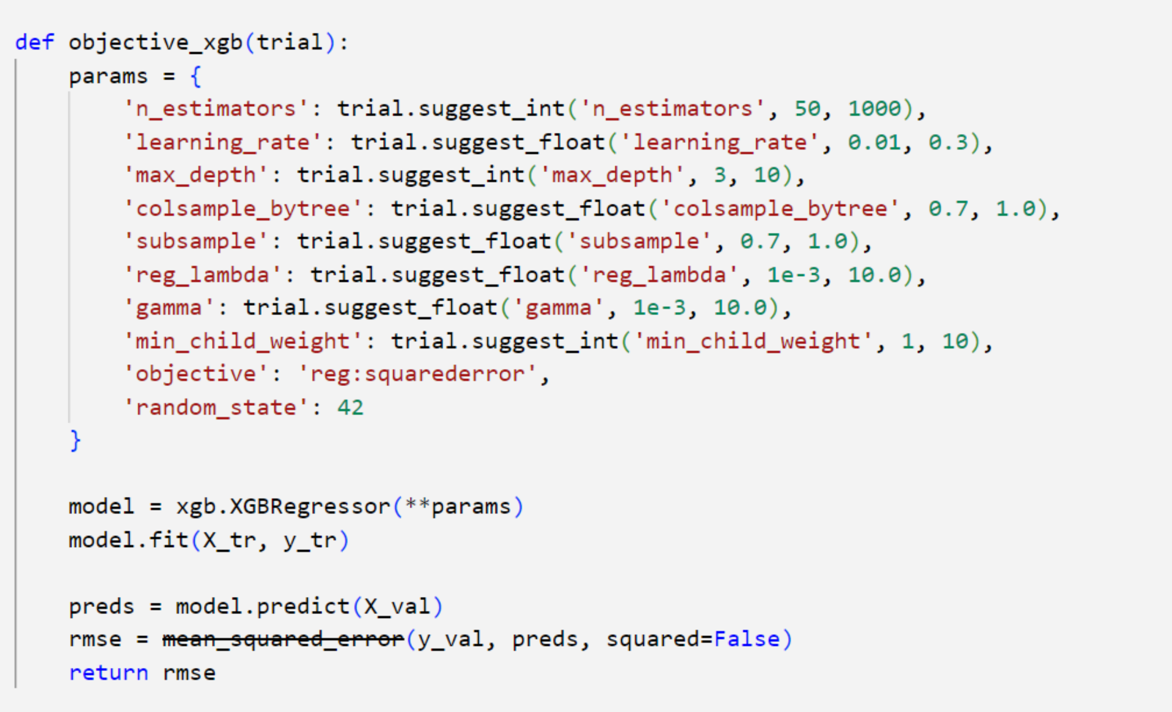
Gradient Boosting Machines (GBMs) were selected for this study due to their high predictive accuracy, resilience to overfitting, and capacity to handle mixed data. Given the three categorical and five numerical features, some of which are highly correlated, GBMs such as LightGBM and CatBoost offer native categorical support, reducing the need for extensive pre-processing. Moreover, GBMs emphasize misclassified points in each boosting cycle, making them effective for imbalanced or complex datasets. Consequently, three different GBM models were implemented using their respective Python libraries. Each library has a set of default hyperparameters which is visible in table 7.

Table 7- Default value of hyperparameters for LightGBM, XGBoost, and CatBoost library

Hyperparameters	LightGBM	XGBoost	CatBoost
Learning Rate	0.1	0.3	0.03
Number of Leaves	31	-	-
Max Depth	No limit	6	6
Subsample	1	1	-
Colsample_bytree	1	1	-
Iteration	-	-	1000
L2 Leaf Regularization	-	-	3

In the experiments, I observed the dataset has some imbalance nature (some target values can potentially be overrepresented compared to others). To counter the potential majority range bias, we added and tuned known hyperparameters for overfitting and handling including: Lambda to control the strength of L2 regularization, Gamma to make the algorithm conservative (fewer splits) and avoid overfitting and finally Min_child_weight to make the model conservative by not allowing small leaves to happen likely.

Figure 14 is the concise illustration of the tuning process for XGBoost.



```
def objective_xgb(trial):
    params = {
        'n_estimators': trial.suggest_int('n_estimators', 50, 1000),
        'learning_rate': trial.suggest_float('learning_rate', 0.01, 0.3),
        'max_depth': trial.suggest_int('max_depth', 3, 10),
        'colsample_bytree': trial.suggest_float('colsample_bytree', 0.7, 1.0),
        'subsample': trial.suggest_float('subsample', 0.7, 1.0),
        'reg_lambda': trial.suggest_float('reg_lambda', 1e-3, 10.0),
        'gamma': trial.suggest_float('gamma', 1e-3, 10.0),
        'min_child_weight': trial.suggest_int('min_child_weight', 1, 10),
        'objective': 'reg:squarederror',
        'random_state': 42
    }

    model = xgb.XGBRegressor(**params)
    model.fit(X_tr, y_tr)

    preds = model.predict(X_val)
    rmse = ~~mean_squared_error~~(y_val, preds, squared=False)
    return rmse
```

Figure 14- Optuna optimization and finding the best hyperparameters

After identifying the best hyperparameters for all models, I re-trained them using the whole train set without the addition of another early-stopping method. This ensures the resulting models make the maximum possible use of all the train data available.

The Figure 15 illustrates the architecture of the predictive model developed in this thesis. The XGBoost model is used, incorporating hyperparameters, tuning, and various building and lift attributes. Inputs such as building type, number of floors, and lift trip distribution percentiles are processed by the trained algorithm to estimate the average number of lift trips.

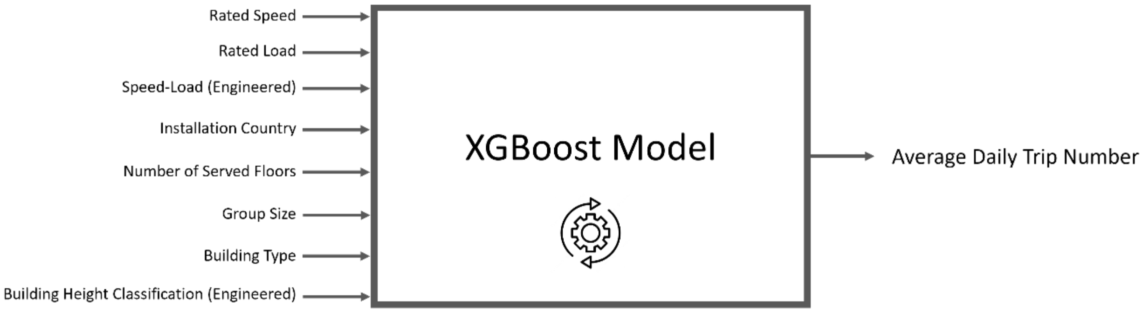


Figure 15- Final model structure

5 Results

In this chapter, I present the results obtained from four distinct approaches to predicting the average number of trips per day for lifts. I begin by introducing each method and then assess their accuracy and reliability using a range of error metrics. The methods under investigation include:

- Representative Number Method (First Assumption):

  This approach employs a representative value based on floor categorization. It is inspired by previous studies and aligns with the usage categories defined in ISO 25745-2.

- Polynomial Fit Model (Baseline):

  Utilizing historical data, this model incorporates building type and the number of served floors to fit a simple polynomial function.

- Advanced Machine Learning Models:

  Also based on historical data, this approach leverages three different gradient boosting models, including Category Boosting, XGBoost, and LightGBM. The models are tested in two cases: first, without applying the travel distance of the lift as an auxiliary feature, and second, after incorporating this feature. The results of both cases are compared to assess the impact of this additional parameter.

Furthermore, we explore five different use cases to evaluate the impact of modernisation under two distinct scenarios: one that does not emphasize energy efficiency and another that focuses on energy efficiency. This analysis investigates how variations in the predicted number of trips influence energy calculations by comparing both the actual values and the results from each predictive method.

5.1 Representative Number Method (First Assumption)

The performance of the Floor-Based Representation Method on the test dataset is summarized in Table 8 Since this method assigns a representative trip number based on predefined categories; the training set does not influence the error metrics. Therefore, only the evaluation metrics for the test set are reported. These results will later be compared with those of other predictive methods to assess relative accuracy and reliability.

Table 8- Performance metrics representative number method

Error Parameter	Value for Test Set
Coefficient of determination (R^2 Score)	0.266
Mean Absolute Error (MAE)	117.42
Root Mean Squared Error (RMSE)	165

The results indicate that this method struggles to provide accurate predictions, as evidenced by the high RMSE (165) and MAE (117.42) values, along with a low R^2 score (0.266), which signifies weak correlation with actual data. The scatter plot in Figure 16 further illustrates the discrepancy between actual and predicted values, where predictions cluster around specific levels, demonstrating the lack of flexibility in this approach. The primary limitation of this method is its rigid categorization and introducing one representative number for many different cases. As a result, the method fails to adapt to variations in real-world data, making it an unreliable approach for precise trip prediction.

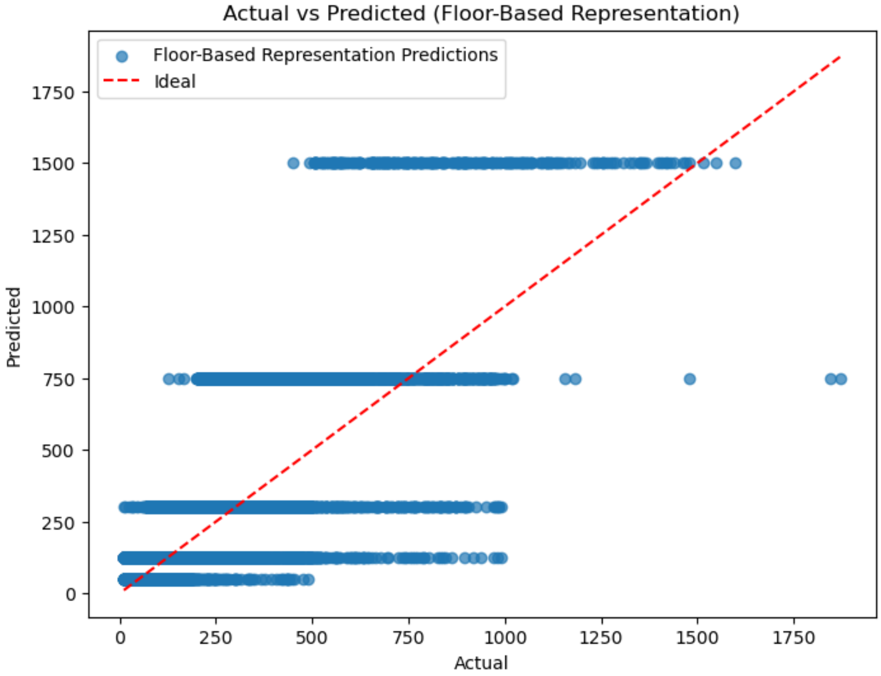


Figure 16- Comparison of the test set with the representative trip number method. The red line indicates the ideal scenario where all predicted and actual values perfectly align

5.2 Polynomial and Logarithmic Fit Model (Baseline Model):

The second predictive method utilises the training dataset to establish a fitted function that predicts the median number of lift trips per day based on the number of served floors. This approach accounts for variations across different building types, ensuring that predictions are tailored to specific sectoral characteristics. The training data was first categorized into five sectors: residential, office, hotel, medical, and others, and within each sector, the median number of trips per operating day was calculated for different floor counts. Two mathematical functions, polynomial and logarithmic, were fitted to these sector-specific datasets to model the relationship between the number of floors and lift usage patterns.

After fitting both functions to the training dataset, their performance was evaluated using R^2 (coefficient of determination) and RMSE (root mean squared error). The polynomial model consistently outperformed the logarithmic model, exhibiting higher R^2 values and lower RMSE values in all sectors. Consequently, the polynomial function was selected as the final model for predicting trips per day in the test dataset. The fitted models were then applied to the test data, where predictions were generated, and error metrics were computed to evaluate the model's accuracy. Figures 17 to 21 illustrate the fitted models for different building types. In these figures, the blue dots represent actual median trip values from the training dataset, while the red and green curves represent the polynomial and logarithmic fits, respectively. It is evident from these visualizations that the polynomial function provides a better approximation of the data points across most sectors. The polynomial function demonstrated superior predictive capability across all building sectors compared to the logarithmic function. In the residential sector (Figure 21), the polynomial model achieved an *R*² of 0.90, whereas the logarithmic model had an *R*² of 0.84, confirming the polynomial function's better fit. Similarly, for office buildings (Figure 20), the polynomial function attained an R² of 0.75, slightly outperforming the logarithmic model. The hotel (Figure 17) and medical (Figure 18) sectors exhibited lower overall *R*² values, 0.75 and 0.45, respectively, indicating that the relationship between floors and trips is less predictable in these building types. In the other building category (Figure 17), the polynomial model achieved *R*² = 0.79, further validating its general applicability.

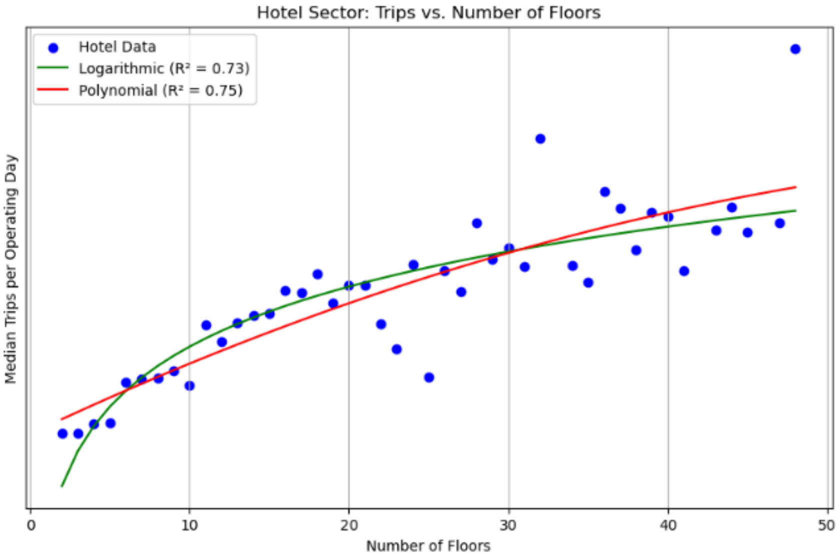


Figure 17- Polynomial and logarithmic trend fitting based on the median of historical data in the hotel sector

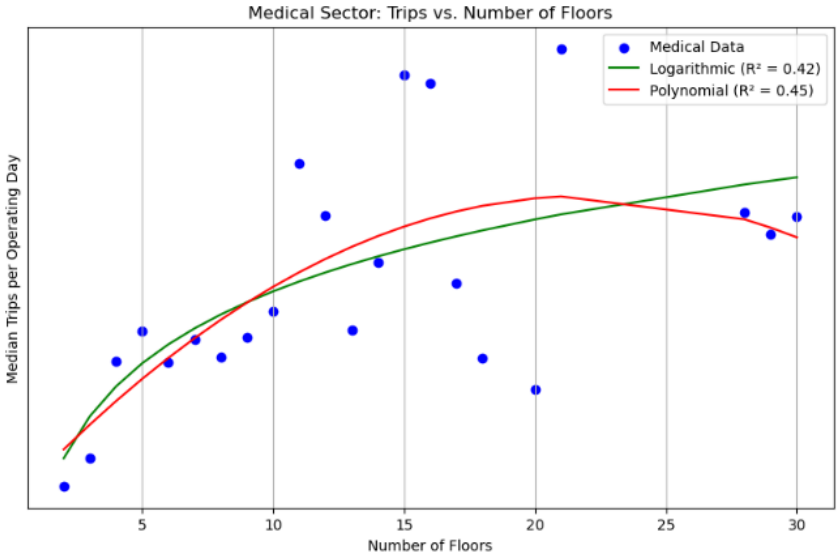


Figure 18- Polynomial and logarithmic trend fitting based on the median of historical data in the medical sector

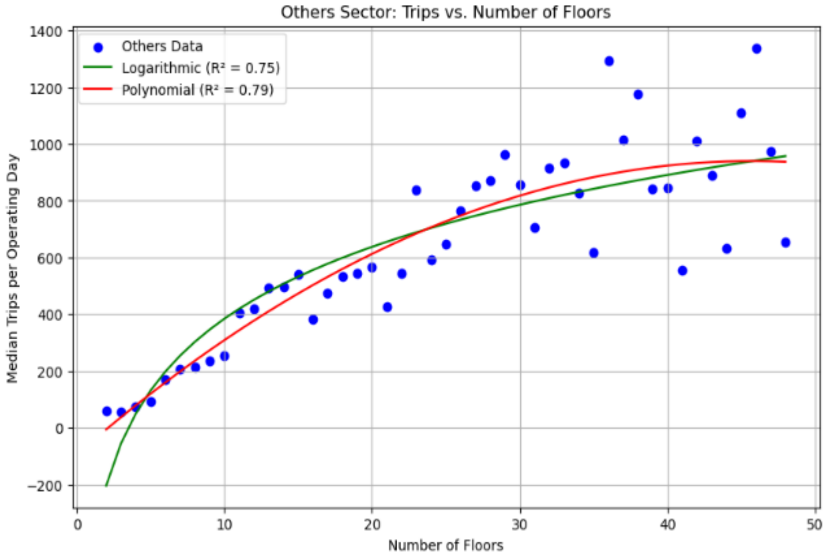


Figure 19- Polynomial and logarithmic trend fitting based on the median of historical data in the other sector

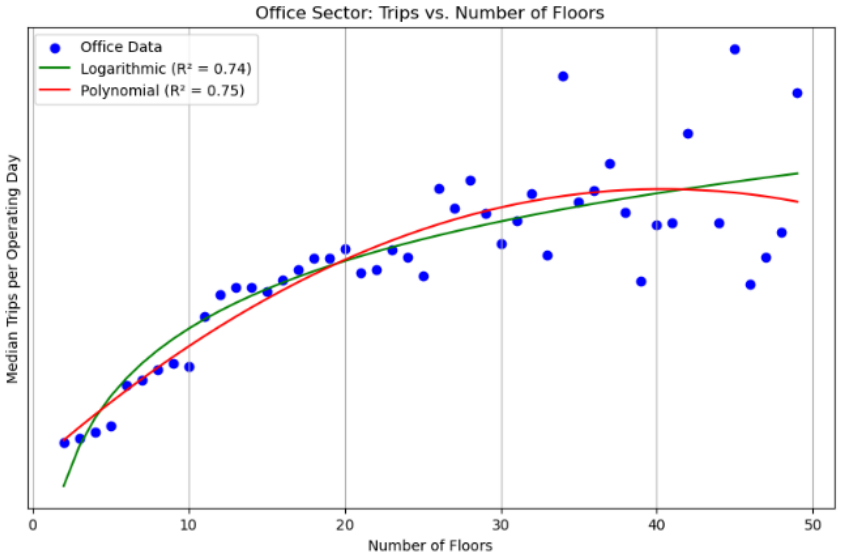


Figure 20- Polynomial and logarithmic trend fitting based on the median of historical data in the office sector

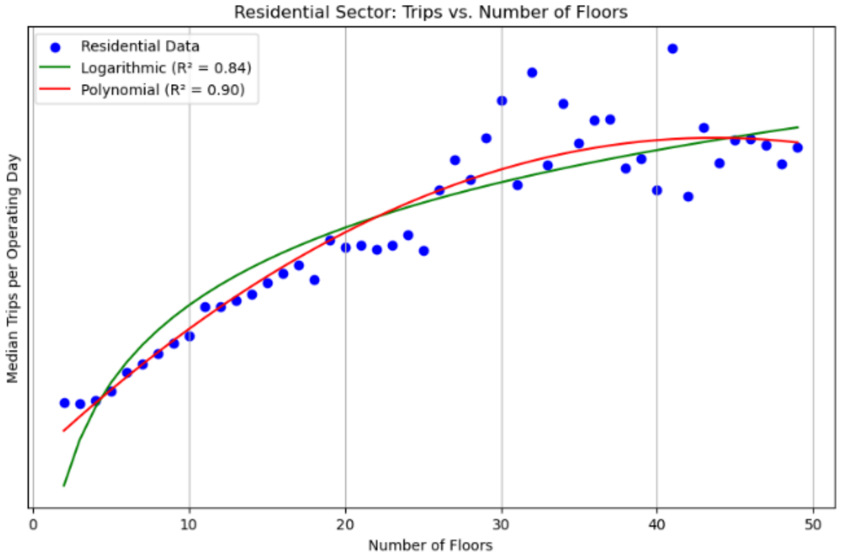


Figure 21- Polynomial and logarithmic trend fitting based on the median of historical data in the residential sector

the overall test set evaluation for this method yielded an RMSE of 120.59, an MAE of 82.80, and an R^2 score of 0.654, showing a moderate correlation between predicted and actual trip values. Compared to the floor-based representation method, which had a significantly lower R^2 score of 0.266, this approach represents a substantial improvement in prediction accuracy. The scatter plot in Figure 22 further highlights this enhancement, where the baseline polynomial predictions (blue dots) are more closely aligned with the red ideal prediction line compared to the floor-based method's results.

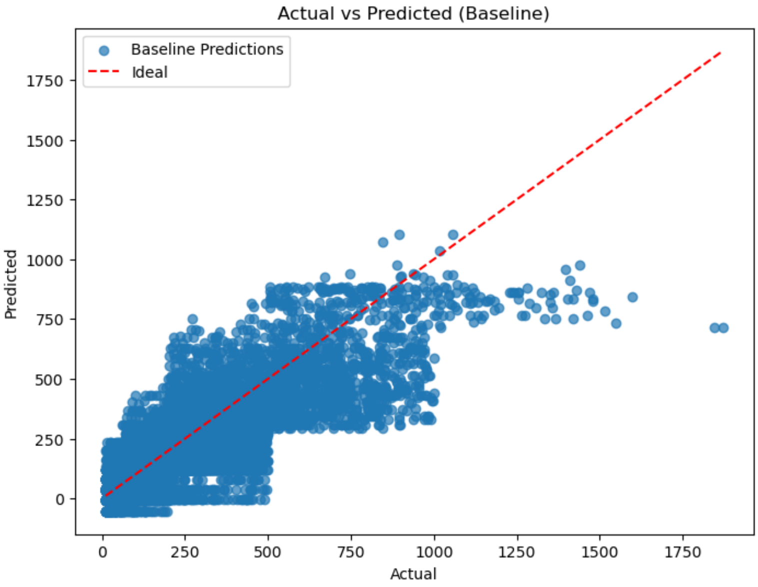


Figure 22- Comparison of the test set with the baseline predicted trip number method. The red line indicates the ideal scenario where all predicted and actual values perfectly align

Furthermore, Table 9 presents the results of the baseline model evaluation, including the calculated error parameters.

Table 9- Performance metrics for the baseline model

Error Paremeter	Value for Train Set	Value for Test Set
Coefficient of determination (R^2 Score)	0.66	0.65
Mean Absolute Error (MAE)	81.98	82.80
Root Mean Squared Error (RMSE)	120.81	120.59

5.3 Advanced Machine Learning Models without Auxiliary Feature (XGBoost)

The first phase of the machine learning-based prediction involved training three gradient boosting models—XGBoost, LightGBM, and CatBoost—on a pre-processed dataset. The pre-processing included encoding categorical variables and scaling numerical features to ensure consistency across different inputs. The training data consisted of features such as the number of floors, business type, geographic location, rated load-speed interaction, and group size, but without incorporating the auxiliary feature (lift travel height). Each model underwent hyperparameter tuning using Optuna to optimize predictive performance. After training, the models were evaluated on the test dataset, and the XGBoost model emerged as the best performer based on error metrics, leading to its selection for further analysis. Figure 23 presents the scatter plot of actual vs. predicted values using this model, where the distribution of points around the red dashed ideal prediction line indicates model performance.

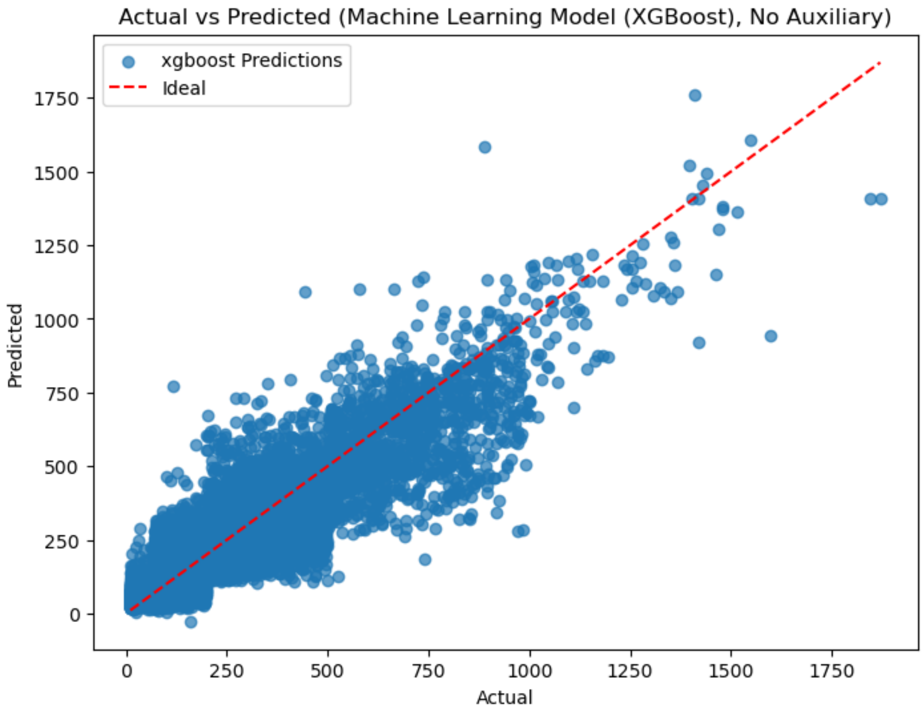


Figure 23- Comparison of the test set with the machine learning (XGBoost) model without an auxiliary feature. The red line indicates the ideal scenario where all predicted and actual values perfectly align

Table 10 shows the error parameters result for this model, as it is shown, The XGBoost model without auxiliary correction achieved an R^2 score of 0.798 on the test set, marking a significant improvement over previous methods such as polynomial regression ($R^2 = 0.654$) and floor-based representation ($R^2 = 0.266$). The RMSE was 92.17, while the MAE stood at 63.62, signifying that the model captures a strong correlation between lift trips and input features

Table 10- Performance metrics for XGBoost model without an auxiliary feature

Error Parameter	Value for Train Set	Value for Test Set
Coefficient of determination (R^2 Score)	0.87	0.80
Mean Absolute Error (MAE)	52.51	62.2
Root Mean Squared Error (RMSE)	73.61	89.62
MAPE	45%	48%
75th Percentile of Error	36%	40%

The close agreement between the model's performance on the training and test sets, as evidenced by comparable R^2 Score and acceptable differences in RMSE and MAE, suggests a lack of overfitting. Furthermore, the high R^2 Score value indicates that the model effectively captures a significant portion of the variability within the data.

Feature importance analysis was then performed, and the results, summarized in Table 11, reveal number of served floor as the most influential factor in predicting daily trip frequency. The table also presents the relative importance of other contributing features after analysing the feature importance, two features that did not have effect on model and only causes overfitting due to high correlation with each other and with rated load-speed engineered feature, were removed and this removal reduced the difference between train and test result. Therefore, although removing these two features did not have effect on test result but made model more reliable.

Table 11- The percentage of important features in model prediction

Important Features	Percentage of Mean Permutation Importance Scores
Number of Floors	65%
Business Type	11%
Rated Load-Speed	11%
Group Size	4%
Installation Country	5%
Building Height Classification (Engineered)	4%

5.4 Advanced Machine Learning Models with Auxiliary Feature (XGBoost)

Domain knowledge suggests that travel height is a potentially significant factor in predicting daily trip frequency. However, the dataset exhibited limitations in terms of travel height data availability and accuracy. Approximately 25% of data points lacked travel height values, while an estimated 50% of the available values were deemed inaccurate, often indicating implausibly low values (less than 3 meters).

In the second phase of machine learning-based predictions, the same XGBoost model was retrained with an additional auxiliary feature: travel height. This feature represents the maximum of vertical distance the lift travels from the first landing door to the last one, which is a crucial factor influencing trip counts. Only test samples with valid travel height values greater than 3 meters were included in this phase to ensure data quality. The model's base predictions were first generated, after which a secondary auxiliary correction model (XGBoost) was trained to predict the residual errors based on the travel height. The final predictions were obtained by adding these corrections to the original model's outputs. Figure 24 illustrates the results of this enhanced model, showcasing the impact of travel height on improving alignment with actual values.

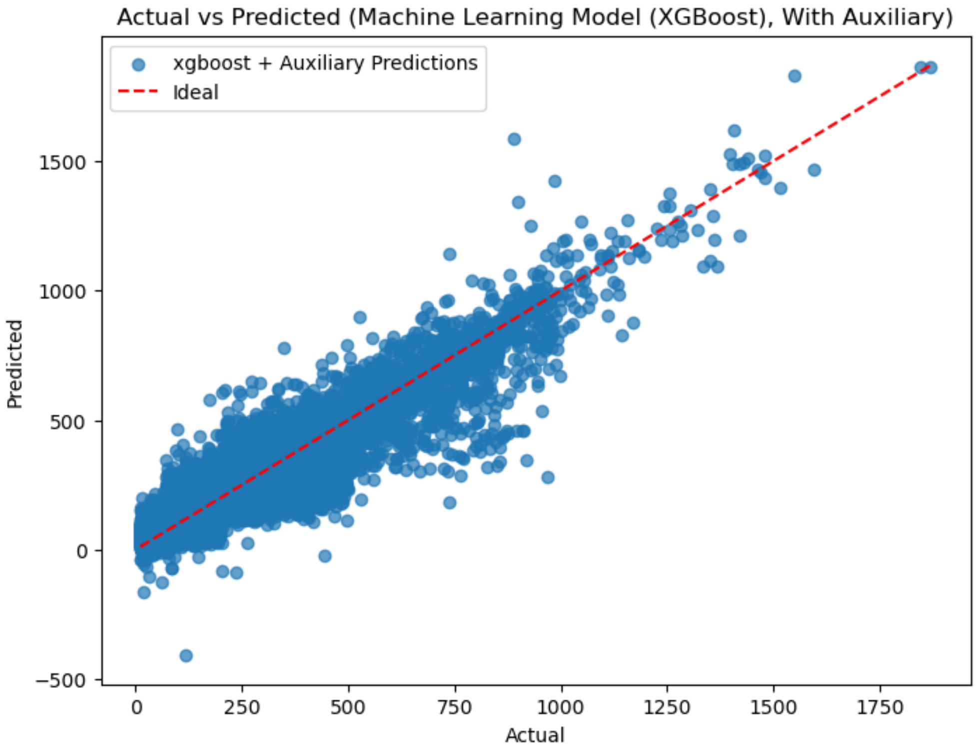


Figure 24- Comparison of the test set with the machine learning (XGBoost) model with an auxiliary feature. The red line indicates the ideal scenario where all predicted and actual values perfectly align

Feature importance analysis within this refined model confirmed the significance of travel height as a predictive factor. Also, as it is shown in Table 12 The introduction of the auxiliary feature significantly enhanced model performance, increasing the R^2 score to 0.850 on the test set. The RMSE was reduced to 75.43, and the MAE dropped to 54.92, demonstrating a clear improvement in predictive accuracy. Moreover, the percentage of predictions within a 30% error margin rose to 73%, confirming the auxiliary feature's role in refining the model's estimations.

By comparing the residuals of the main model (without travel height) to those of the model incorporating travel height, an improvement in model performance was observed.

Specifically, a 33% reduction in RMSE and a 24% reduction in MAE were achieved compared to the baseline model. This improvement underscores the potential of travel height as a valuable predictor when accurate data is available. Furthermore, it suggests that the inclusion of complete and accurate travel height data for all observations could lead to a substantial enhancement in model accuracy and predictive power.

Table 12-- Performance metrics for XGBoost model with auxiliary feature

Error Parameter	Value for Train Set	Value for Test Set
Coefficient of determination (R^2 Score)	0.88	0.85
Mean Absolute Error(MAE)	52.92	54.92
Root Mean Squared Error (RMSE)	70.19	75.43
MAPE	40%	45%
75th Percentile of Error	34%	38%

5.5 Models Comparison and Discussion

A comparative analysis was conducted among four models: the representative number method, the baseline model, the AI model (XGBoost), and the AI model (XGBoost) incorporating travel height as an auxiliary feature. For the representative method the R^2 Score showed that the generalization of this method is not acceptable while this method in some cases can even perform better than AI model, but it cannot be a reliable approach when it comes to different pattern especially for the lifts which work in usage category four. The second approach meaning the baseline model achieved an R^2 value of 0.65 on the test set, indicating although this method has better ability to capture the complex relationships within the data, still some improvements can be done to make generalization better. By using the AI model which consider more features than baseline and uses advance regression methods than simple polynomial model, the performance improved significantly with R-squared (0.80), with better generalization and predictive power.

By adding the auxiliary feature of travel height, performance enhanced even further, as it reached an R-squared of 0.85. It illustrates how valuable the feature of travel height is as a predictor. Though data limitations are present with respect to this feature, this experiment clearly displays its ability to offer improved performance. In summary, R-squared value showed improvement from representative method, baseline model, towards the machine learning model with auxiliary feature by 219%, 30% respectively.

In terms of error metrics, the machine learning model registered significant drops in MAE (47%, 24.39%) and MAPE (51%, 27.18%) relative to the typical method and the baseline respectively. These were improved upon even further with the combined auxiliary machine learning model, witnessing even higher drops in MAE (53%, 34%) and MAPE (54%, 31.8%) for the above models. These findings emphasize the critical role of feature engineering and the effectiveness of advanced AI methods in improving predictive accuracy. The incorporation of complete and accurate travel height data for all observations has the potential to further enhance the performance of these models.

Figures 20 and 21 and 22 visually compared the predictive performance of the representative method, baseline, machine learning model, and machine learning model with auxiliary feature. The plot demonstrates that both AI-based models exhibit superior performance compared to the baseline model. The predicted values from the AI models are more closely clustered around the ideal scenario, represented by the red dashed line, where predicted values perfectly align with actual value

5.6 Analysing the Effect of Trip Numbers on Energy Consumption for Five Cases

In this section, the aim is to analyse how differences in expected traffic quantities affect the projected benefits of modernisation in some of the most common lift modernisation cases today. For this purpose, five different cases have been selected. These cases are defined based on their technical characteristics as well as the components and technologies they incorporate. Two different modernisation approaches have been considered:

1. A typical modernisation solution
2. An energy-efficiency-focused modernisation approach

For each case, energy consumption has been calculated under three different scenarios and compared using four different trip estimation methods:

1. Actual recorded values from connected device
2. Machine learning output with auxiliary features
3. Baseline model predictions
4. Representative model based on ISO assumption

The results will indicate the impact of improved trip estimates on the accuracy of the calculations of the energy consumption in three main categories: running energy consumption, non-running energy consumption, and lighting energy consumption. It is intended that analysis will prove how improved methods of estimating trips can improve the accuracy of modernisation benefit predictions

5.6.1 Case 1

As shown in Table 13, Lift Case 1 is a residential building lift with a gearless, non-regenerative traction system. In a typical modernisation, one of the KONE modernisation products introduces a new gearless drive, an updated controller (KCE), and LED signalization, while keeping most components unchanged. However, in the energy-saving modernisation approach, additional upgrades focus on efficiency, including a regenerative drive, LED ceiling lighting, automatic car light shutoff, and a door upgrade to further optimize energy consumption and sustainability.

Table 13-Characteristics of lift. Case 1: dimensional, building, and technical specifications

Dimensional Characteristics	Value Of the Dimensional Characteristics	Technical Characteristics	Existing Lift	Modernisation typical changes	Modernisation with Energy Saving emphasis changes
Building Type	Residential	Technology	Traction	Traction	Traction
Number of Served Floor	6 floors	Geared vs Gearless	Gearless	Gearless (new drive)	Gearless (new drive)
Rated Speed	1 m/s	Generative/non-Generative	Non-Regenerative	No change	Regenerative
		Ceiling Light Type	Fluorescent ceiling	No change	LED ceiling
Rated Load	1000 kg	Operating of Car Light	No	No change	Yes
		Signalization	Incandescent	LED	LED
Group Size	1	Machine Room Place	Machine Room Less	No change	No change
		Controller Type	25-year-old microprocessor	Modern microprocessor	Modern microprocessor
Travel Height	15 m	Roping Suspension Type	2:1	No change	No change
		Other Upgrade Option	No	No change	Door Upgrade

Table 14 shows the trips number for actual scenario, the output of machine learning prediction, the prediction of baseline model, and representative number reported from ISO 25745-2.

Table 14- Case1- comparing the result of trips number with different methods

Estimation method	Actual value	Machine learning prediction	Simple polynomial fit (Baseline)	ISO-based representative Number
Daily Trip Number	98	104 trips	118 trips	125 trips
Annual Trips	35574	37752	42834	44649

Based on reported KONE internal energy calculation tool the annual energy consumption was calculated and the effect of changing the annual trip number is reported in Table 15. Based on this table, the results indicate that changes in lighting and non-running energy consumption are negligible across different trip estimation methods, as these values remain nearly constant. However, running energy consumption shows variation, highlighting the impact of more precise trip estimations. Compared to actual values, the estimated method results in only a 5.8% difference in running energy, whereas the baseline model prediction leads to a 20% deviation, and the ISO-based floor representative model exhibits the largest error at 25%.

Table 15- Case 1- result of annual energy consumption (kwh/year) calculation in different scenarios of trips number estimation

Daily Trips Estimation Method	Lighting Before Modernisation (kWh/a)	Lighting After Modernisation (kWh/a)	Lighting After Energy-Saving Mode (kWh/a)	Non-Running Before Modernisation (kWh/a)	Non-Running After Modernisation (kWh/a)	Non-Running After Energy-Saving Mode (kWh/a)	Running Before Modernisation (kWh/a)	Running After Modernisation (kWh/a)	Running After Energy-Saving Mode (kWh/a)
Actual (98 trips)	557	557	77	839	601	596	358	311	204
Machine Learning (104 trips)	557	557	77	838	600	596	380	330	216
Baseline model	556	556	78	835	598	594	431	375	245

(118 trips)									
ISO-based representative number (125 trips)	556	556	78	834	597	593	450	391	255

By modernizing the lift, as shown in Figure 25, energy consumption significantly decreased, demonstrating the impact of modernisation solutions. In a typical modernisation scenario, total annual energy consumption (sum of lighting, non-running, and running energy) decreased from 1754 kWh/year to 1469 kWh/year in the actual trip estimation case, representing a 16% improvement in energy efficiency. Similarly, for the estimated trip case (104 trips), total energy consumption dropped from 1775 kWh to 1487 kWh, yielding a 16.2% efficiency improvement. However, when modernisation was implemented with an energy-saving focus, the reduction was even more significant, especially in lighting energy consumption. In this scenario, total energy consumption decreased to 877 kWh in the actual trip case, marking a 50% reduction. The trend continues across other estimation methods, with energy-saving-focused modernisation reducing total energy from 1775 kWh to 889 kWh (estimated case), from 1822 kWh to 917 kWh (baseline case), and from 1840 kWh to 926 kWh (ISO-based representative trip number case).

These results emphasize that while typical modernisation provides notable energy savings, an energy-saving-focused approach leads to even greater reductions, particularly by optimizing lighting energy usage, contributing significantly to overall sustainability and operational efficiency.

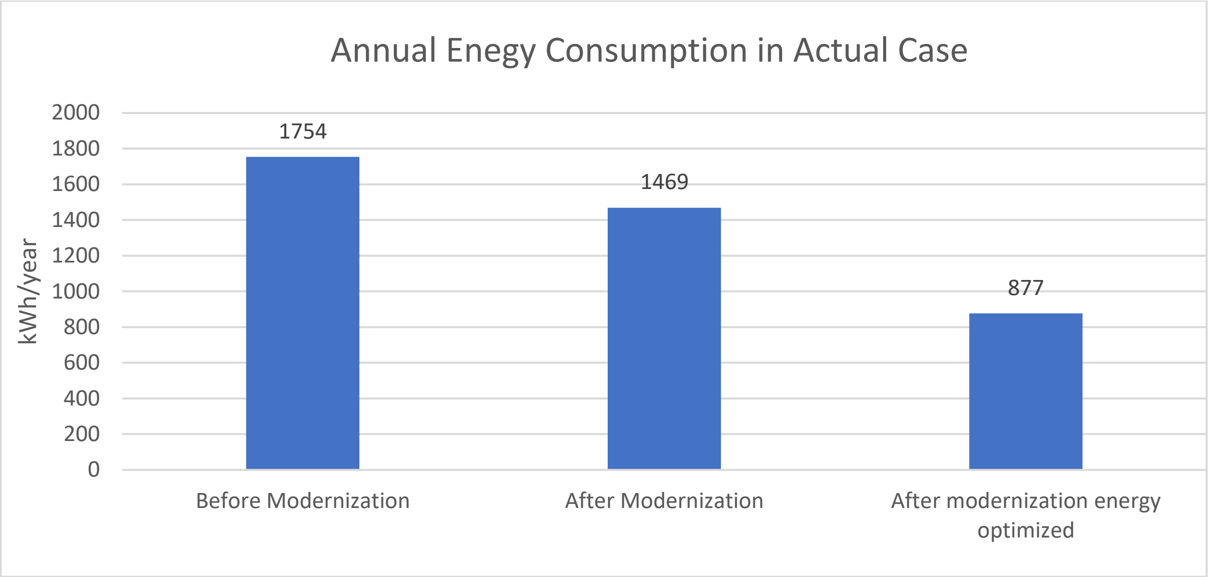


Figure 25-Case1- Annual energy consumption (kWh/year) based on actual trip numbers, comparing three scenarios: before modernisation, after standard modernisation, and after modernisation energy optimized

5.6.2 Case 2

As shown in Table 16, Lift Case 2 also is in a residential building and the other characteristics which is visible in this table. In a typical modernisation, another KONE upgrade solution is implemented, which includes a new geared drive system and an updated KCE controller, while other components remain unchanged. However, in the energy-saving modernisation approach, additional enhancements include LED ceiling lighting, automatic car light shutoff, and improved signalization, optimizing energy efficiency and sustainability.

Table 16- Characteristics of lift. Case 2: dimensional, building, and technical specifications

Dimensional Characteristics	Value Of the Dimensional Characteristics	Technical Characteristics	Existing Lift	Modernisation typical changes	Modernisation with Energy Saving emphasis changes
Building Type	Residential	Technology	Traction	Traction	Traction
Number of Served Floor	7 floors	Geared vs Gearless	Geared	Geared (new drive)	Geared (new drive)
Rated Speed	1 m/s	Generative/non-Generative	Non-Regenerative	No change	No change
		Ceiling Light Type	Fluorescent ceiling	No change	LED ceiling
Rated Load	630 kg	Operating of Car Light	No	No change	Yes

		Signalization	Incandescent	LED	LED
Group Size	1	Machine Room Place	Machine Room above	No change	No change
		Controller Type	Microprocessor	Modern microprocessor	Modern microprocessor
Travel Height	18 m	Roping Suspension Type	1:1	No change	No change
		Other Upgrade Option	No	No change	No change

Table 17 compares the estimated number of trips per year using different calculation methods. The actual recorded trips were 208 per day, while the machine learning model predicted 199 trips. The baseline model estimated 181 trips, whereas the ISO-based representative number projected 300 trips.

Table 17- Case2- Comparing the result of trips number with different methods

Estimation Method	Actual Value	Machine Learning Prediction	Baseline Model	ISO-Based Representative
Trips per Day	208	199	181	300
Annual Trips	75,920	72,635	66,065	109,500

Again, using the KONE internal energy calculation tool, annual energy consumption (kwh/year) was calculated, demonstrating the effect of varying trip estimations. Table 18 shows that, as with case 1, lighting and non-running energy consumption remain relatively stable across different trip estimates. However, running energy consumption varies significantly, reflecting the importance of precise trip estimation. Compared to actual values, the machine learning model introduces 5% difference in running energy consumption, whereas the baseline model deviates by 13%, and the ISO-based floor representative model shows the largest discrepancy at 44%.

Table 18- Case 2- result of annual energy consumption (kWh/year) calculation in different scenarios of trips number estimation

Trips Estimation Method	Lighting Before Modernisation (kWh/a)	Lighting After Modernisation (kWh/a)	Lighting After Energy-Saving Mode (kWh/a)	Non-Running Before Modernisation (kWh/year)	Non-Running After Modernisation (kWh/a)	Non-Running After Energy-Saving Mode (kWh/a)	Running Before Modernisation (kWh/a)	Running After Modernisation (kWh/a)	Running After Energy-Saving Mode (kWh/a)
Actual (208 trips)	2350	209	81	1090	549	549	1047	786	786
Machine Learning (199 trips)	2351	209	55	1077	550	550	1002	752	752
Baseline model (181 trips)	2352	209	54	1082	553	553	911	684	684
ISO-based representative number (300 trips)	2344	208	84	1064	536	536	1510	1134	1134

Furthermore, in this case by modernizing the lift, as shown in Figure 26, total energy consumption significantly decreased. In a typical modernisation scenario, total annual energy consumption reduced from 4487 kWh to 1544 kWh in the actual trip estimation case, representing a 65.6% improvement in energy efficiency.

However, when modernisation was performed with an energy-saving focus, the reductions were even more considerable. In this scenario, total energy consumption dropped to 1416 kWh in the actual trip case, marking a 68.4% reduction. The trend persists across other trip estimation methods, with modernisation energy optimized, decreasing total energy from 4430 kWh to 1357 kWh (estimated case), from 4345 kWh to 1291 kWh (baseline case), and from 4918 kWh to 1754 kWh (ISO-based representative trip number case).

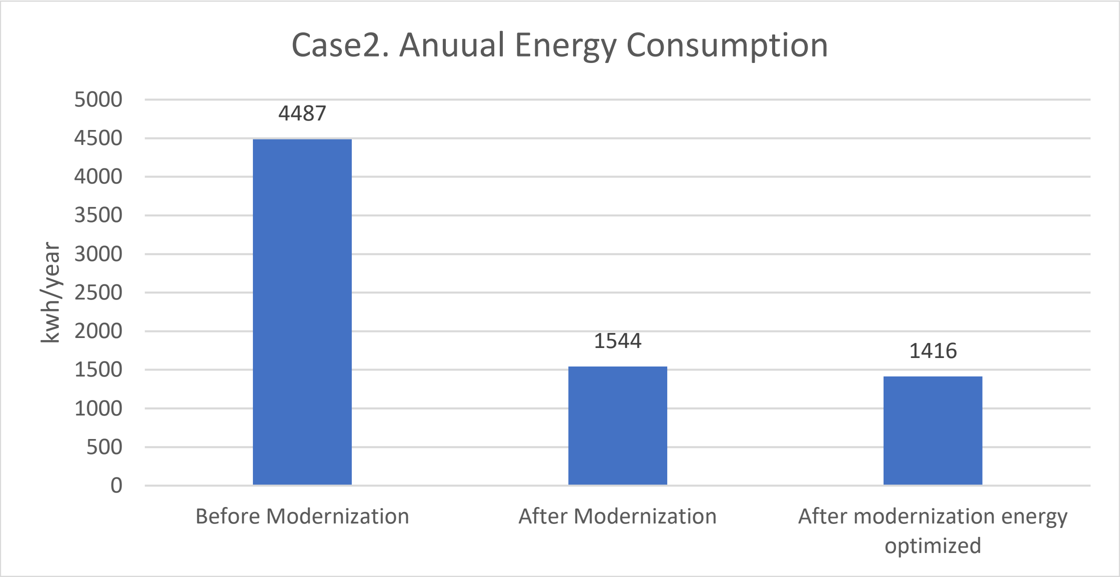


Figure 26- Case2-Annual energy consumption (kWh/year) based on actual trip numbers, comparing three scenarios: before modernisation, after standard modernisation, and after modernisation energy optimized

5.6.3 Case 3

The characteristics of Lift Case 3, including its building type, technical specifications, and modernisation details, are summarized in Table 19. This case is similar to case 1 with only difference in building type that this change has impact on operating days and also trips pattern.

Table 19- Characteristics of lift. Case 3: Dimensional, building, and technical specifications

Dimensional Characteristics	Value Of the Dimensional Characteristics	Technical Characteristics	Existing Lift	Modernisation typical changes	Modernisation with Energy Saving emphasis changes
Building Type	Office	Technology	Traction	Traction	Traction
Number of Served Floor	5 floors	Geared vs Gearless	Gearless	Gearless (new drive)	Gearless (new drive)
Rated Speed	1 m/s	Generative/non-Generative	Non-Regenerative	No change	Regenerative
		Ceiling Light Type	Fluorescent ceiling	No change	LED ceiling
Rated Load	1000 kg	Operating of Car Light	No	No change	Yes
		Signalization	Incandescent	LED	LED

Group Size	2	Machine Room Place	Machine Room Less	No change	No change
		Controller Type	25-year-old microprocessor	Modern microprocessor	Modern microprocessor
Travel Height	15 m	Roping Suspension Type	2:1	No change	No change
		Other Upgrade Option	No	No change	Door Upgrade

The lift operates 260 days per year, and trip estimations were derived using different methods. The actual recorded trips, the machine learning model predicted, the baseline model, and the ISO-based representative number projected can be found in Table 20.

Table 20- Case3- Comparing the result of trips number with different methods

Estimation Method	Actual Value	Machine Learning Prediction	Baseline Model	ISO-Based Representative
Trips per Day	168	160	150	125
Annual Trips	42000	40000	37500	31250

In this case, the machine learning model had a 5% difference in running energy prediction, whereas the baseline model deviated by 10%, and the ISO-based model showed the largest discrepancy at 24%. The detail reported consumption for different parts (running, non-running, and lighting) is shown in Table 21.

Table 21- Case 3- result of annual energy consumption (kWh/year) calculation in different scenarios of trips number estimation

Trips Estimation Method	Lighting Before Modernisation (kWh/a)	Lighting After Modernisation (kWh/a	Lighting After Energy-Saving Mode (kWh/a)	Non-Running Before Modernisation (kWh/a)	Non-Running After Modernisation (kWh/a)	Non-Running After Energy-Saving Mode (kWh/a)	Running Before Modernisation (kWh/a)	Running After Modernisation (kWh/a)	Running After Energy-Saving Mode (kWh/a)
Actual (168 trips)	398	398	57	565	415	412	422	367	240
Machine	383	383	55	544	399	396	402	350	229

Learning (160 trips)									
Baseline Model (150 trips	383	383	55	545	400	397	377	328	214
ISO-Based Representative (125 trips	399	399	56	571	419	416	314	273	179

Also from the modernisation effect perspective, this case is the same as case 1, Modernisation significantly reduced total energy consumption, as shown in Figure 27. In the actual trip case, total energy decreased from 1385 kWh to 1180 kWh (15% improvement) with standard modernisation. The modernisation energy optimized approach further reduced it to 709 kWh (49% improvement). Similar trends were observed in other estimation methods, with energy-saving.

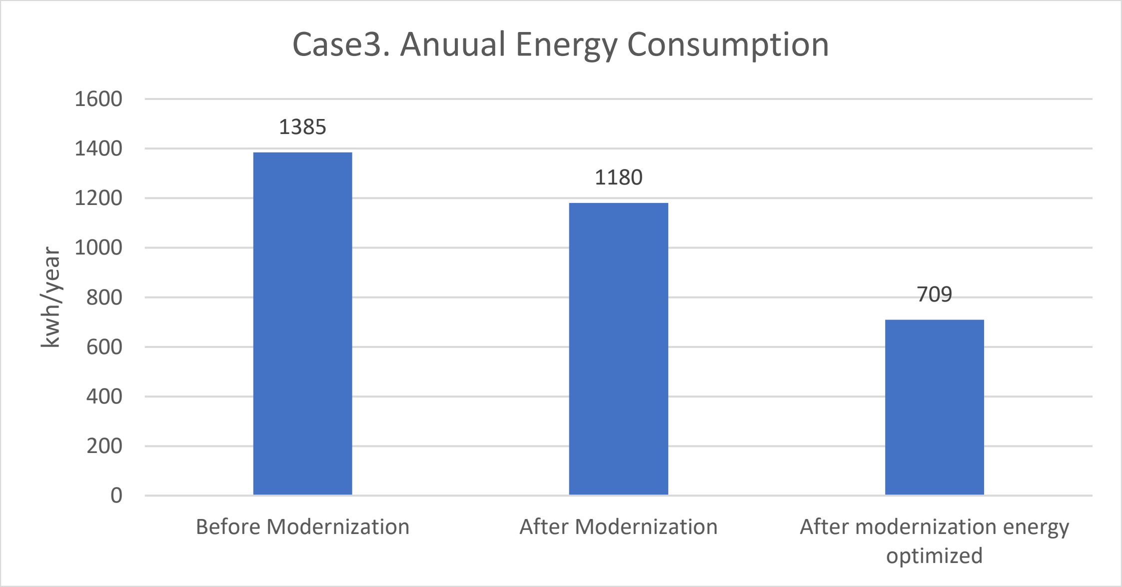


Figure 27- Case3-Annual energy consumption (kwh/year) based on actual trip numbers, comparing three scenarios: before modernisation, after standard modernisation, and after modernisation energy optimized

5.6.4 Case 4

In lift Case 4, modernisation is performed using KONE modernisation solution, replacing the gearless drive system, upgrading the controller to modern microprocessor, and enhancing signalization with LED ceiling lighting. An energy-saving-focused modernisation further introduces automatic car light shutoff and a door upgrade. The details of the modernisation solution can be found in Table 22.

Table 22- Characteristics of lift. Case 4: dimensional, building, and technical specifications

Dimensional Characteristics	Value Of the Dimensional Characteristics	Technical Characteristics	Existing Lift	Modernisation typical changes	Modernisation with Energy Saving emphasis changes
Building Type	Office	Technology	Traction	Traction	Traction
Number of Served Floor	11 floors	Geared vs Gearless	Gearless	Gearless (new drive)	Gearless (new drive)
Rated Speed	2.5 m/s	Generative/non-Generative	Regenerative	No change	No change
		Ceiling Light Type	Fluorescent ceiling	No change	LED ceiling
Rated Load	1600 kg	Operating of Car Light	No	No change	Yes
		Signalization	Incandescent	LED	LED
Group Size	1	Machine Room Place	Machine Room above	No change	No change
		Controller Type	Microprocessor	Modern microprocessor	Modern microprocessor
Travel Height	45 m	Roping Suspension Type	1:1	No change	No change
		Other Upgrade Option	No	No change	Door Upgrade

This lift operates 305 days per year, and trip estimations were derived using different methods. The trips number in different methods can be found in Table 23.

Table 23- Case4- Comparing the result of trips number with different methods

Estimation Method	Actual Value	Machine Learning Prediction	Baseline Model	ISO-Based Representative
Trips per Day	653	584	381	300
Annual Trips	199,165	178,120	116,205	91,500

In this case also the effect of trips deviation is on running energy consumption mostly and as it is shown in Table 24, compared to the actual running energy consumption, The machine learning model had a 10% error in prediction of running energy consumption, whereas the baseline model deviated by 36%, and the ISO-based model showed the highest discrepancy at 50%, emphasizing the importance of precise trip estimations.

Table 24- Case 4- result of annual energy consumption (kwh/year) calculation in different scenarios of trips number estimation

Trips Estimation Method	Lighting Before Modernisation (kWh/a)	Lighting After Modernisation (kWh/a)	Lighting After Energy-Saving Mode (kWh/a)	Non-Running Before Modernisation (kWh/a)	Non-Running After Modernisation (kWh/a)	Non-Running After Energy-Saving Mode (kWh/a)	Running Before Modernisation (kWh/a)	Running After Modernisation (kWh/a)	Running After Energy-Saving Mode (kWh/a)
Actual (653trips)	2378	2378	561	7619	6448	6387	26379	13413	13390
Machine Learning (584trips)	2380	2380	552	7780	6584	6522	23592	11996	11975
Baseline Model (381trips)	2389	2389	442	7495	6232	6166	16696	8565	8551
ISO-based Representative (300 trips)	2392	2392	429	7675	6381	6314	13146	6744	6733

Also, Modernisation significantly reduced total energy consumption. In the actual trip case, total energy decreased from 36,376 kWh to 22,239 kWh (39% improvement) with standard modernisation. The energy-saving-focused approach further reduced it to 20,338 kWh (44% improvement). Similar reductions were observed across other methods.

5.6.5 Case 5

This case is same as case one and only the type of building is hotel which will affect the pattern and trips number. The details of the modernisation solution can be found in Table 25.

Table 25- Characteristics of lift. Case 5: dimensional, building, and technical specifications

Dimensional Characteristics	Value Of the Dimensional Characteristics	Technical Characteristics	Existing Lift	Modernisation typical changes	Modernisation with Energy Saving emphasis changes
Building Type	Hotel	Technology	Traction	Traction	Traction
Number of Served Floor	6 floors	Geared vs Gearless	Gearless	Gearless (new drive)	Gearless (new drive)
Rated Speed	1 m/s	Generative/non-Generative	Non- Regenerative	No change	Regenerative
		Ceiling Light Type	Fluorescent ceiling	No change	LED ceiling
Rated Load	630 kg	Operating of Car Light	No	No change	Yes
		Signalization	Incandescent	LED	LED
Group Size	1	Machine Room Place	Machine Room Less	No change	No change
		Controller Type	25-year-old microprocessor	Modern microprocessor	Modern microprocessor
Travel Height	16 m	Roping Suspension Type	2:1	No change	No change
		Other Upgrade Option	No	No change	Door Upgrade

This lift operates 365 days per year, and trip estimations were derived using different methods. The actual recorded trips were 398 per day, while the machine learning model predicted 291 trips. The baseline model estimated 265 trips, and the ISO-based representative number projected 125 trips which is summarised in table below:

Table 26- Case5- Comparing the result of trips number with different methods

Estimation Method	Actual Value	Machine Learning Prediction	Baseline Model	ISO-Based Representative
Trips per Day	398	291	265	125
Annual Trips	145,270	102,565	96,725	45,625

Table 27 also shows the energy consumption in different parts which shows the estimation of running energy consumption in comparison with actual consumption differs by changes in trips number where the machine learning model had a 10% error with actual value, whereas the baseline model deviated by 18%, and the ISO-based model had the highest deviation at 53%.

Table 27- Case 5- result of annual energy consumption (kwh/year) calculation in different scenarios of trips number estimation

Trips Estimation Method	Lighting Before Modernisation (kWh/a)	Lighting After Modernisation (kWh/a)	Lighting After Energy-Saving Mode (kWh/a)	Non-Running Before Modernisation (kWh/a)	Non-Running After Modernisation (kWh/a)	Non-Running After Energy-Saving Mode (kWh/a)	Running Before Modernisation (kWh/a)	Running After Modernisation (kWh/a)	Running After Energy-Saving Mode (kWh/a)
Actual (398 trips)	416	416	86	688	465	461	982	859	600
Machine Learning (291 trips)	417	417	83	709	479	475	693	607	423
Baseline Model (265= 96725 trips)	417	417	82	711	481	477	654	572	399

ISO-based Representative (125 trips)	419	419	52	721	482	478	308	270	188

In addition, Modernisation decreased overall energy consumption considerably. In the real trip scenario, overall energy fell from 2086 kWh to 1740 kWh for standard modernisation (17% improvement). The focus on the energy-saving approach decreased it eventually to 1147 kWh (45% improvement). Such declines were typical for the other methods: Machine Learning: 1819 kWh down to 1503 kWh (17% reduction), again down to 981 kWh (46%), Baseline: 1782 down to 1470 kWh (17% reduction), again down to 958 kWh (46%), and ISO-Based: 1448 down to 1171 kWh (19% reduction), again down to 718 kWh (50%).

5.7 Comparison of different cases and effect of modernisation and trips number in results

This research proves the dramatic reduction in lift energy consumption due to modernisation solutions. In all cases, both conventional as well as energy-saving-oriented modernisations showed considerable cuts in power consumption. In Case 1, for example, modernisation decreased energy consumption by as high as 50%, Case 2 demonstrated 65% savings, and Case 4 resulted in an astounding reduction of 54%.

1. Trip Estimation Accuracy and Running Energy Deviation

   In terms of trip estimation accuracy, Cases 1, 2, and 3 closely aligned with the ISO-defined usage categories. However, despite this alignment, deviation in running energy estimation remains significant due to variations in representative numbers and actual trip values. In some cases, where the representative number significantly differs from the actual range, running energy estimation errors become more pronounced. for instance, the deviation in running energy estimation was 25% in Case 1, 44% in Case 2, and 24% in Case 3. However, in Cases 4 and 5, where the actual trip numbers did not fit within the ISO-defined representative categories, the discrepancy increased substantially to 50% and 53%, respectively. This indicates the limitation of covering all trip variations within this classification. Also, as it was shown in Section 4.1.1 based on real data analysis, all chosen cases characteristics belonged to the range where the 5th and 95th percentiles of trip numbers fell into the ISO category. However, if these cases would have been chosen from the characteristics that, for example, the floor number was more than 20 floors, the inability to predict trip numbers precisely became more evident.

2. Impact of Trip Estimation Error on Total Energy

Despite these deviations, the impact of trip estimation errors on total energy consumption was relatively minor in cases where running energy contributed a smaller share of overall consumption. For example, in Case 1, while the running energy deviation from actual values was 25%, its share of total energy was only 20%, leading to an overall deviation of just 5%.

6 Discussion and Conclusion

The findings of this study demonstrate that utilizing historical data from connected lift systems in combination with machine learning methods can significantly improve the estimation of average daily trip numbers. This advancement directly contributes to more accurate predictions of annual energy consumption in lifts, which is vital for sustainability assessments and CO_2 emissions reporting.

However, the model's accuracy is inherently influenced by the quality and scope of the input data. While the dataset used in this research was large and representative in many aspects, the inclusion of additional contextual variables could have further enhanced the predictive power of the model. For instance, detailed demographic data on the number of residents in buildings and population density information across different building types could introduce valuable variation, allowing the model to better distinguish usage patterns and improve trip number estimation.

Moreover, this study focused solely on trip number estimation. Yet, the machine learning approach has promising potential for predicting other influential factors in lift energy consumption. These include the load factor per trip and the average travel distance, both of which can also be derived or approximated using connected device data. By extending the current modelling framework to include these variables, even more comprehensive and accurate energy consumption estimates can be achieved.

Another important insight is the opportunity to combine machine learning-based predictions with the structure of existing standardized methods, such as those defined in ISO 25745-2. While ISO provides a useful framework for classification and estimation, it relies on assumptions that may not always reflect real-world usage. By integrating data-driven predictions for key input parameters, like trip count, load, and travel distance, into the ISO calculation structure, the limitations of static assumptions can be mitigated, leading to more reliable and case-specific results.

Ultimately, the convergence of traditional engineering standards with real-time data and machine learning models offers a powerful pathway toward enhanced energy transparency and precision in sustainability reporting for lift systems. This synergy not only improves the internal accuracy of energy calculation tools but also strengthens the credibility of environmental product declarations (EPDs) and life cycle assessments (LCAs) that depend on these estimates.

This thesis contributed towards improvement in the accuracy of the annual lift energy consumption estimates by bridging the gap in the prediction of daily trip counts. Conventional ISO-based methods tend to employ representative trip number based on building height, type, and population in the residential sector. In certain instances, whenever the ISO-proposed representative trip number greatly varied as compared to the actual value, the estimate would not be accurate. Using machine learning methodologies on a large dataset provided by KONE, with the addition of dimensional as well as operational features, this research indicated improvement in the accuracy of trip estimates. Using models like XGBoost, especially with the addition of auxiliary features such as travel height, had higher performance as compared with the baseline method as well as the ISO-based method. R^2 value improved from 0.65 (baseline) up to 0.85 (AI model + auxiliary features), while error metrics like MAE as well as MAPE decreased. In addition, this research evaluated the effect of modernisation on the energy consumption in five typical cases of modernisation as well as the effect of the accuracy in trip estimates on the outcomes before versus after modernisation. The research findings show that in situations where the ISO-representative trip value differed considerably as compared to the actual value, the model could offer an improved prediction, especially in estimating running energy consumption.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M. (2019) Optuna: a next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Al-Kodmany, K. (2015) Tall buildings and lifts: a review of recent technological advances. *Buildings*, 5(3), pp. 1070–1104. https://doi.org/10.3390/buildings5031070

Al-Sharif, L. and Seeley, C. (2010) The effect of the building population and the number of floors on the vertical transportation design of low and medium rise buildings. *Building Services Engineering Research and Technology*, 31(3), pp. 207–220.

Al-Sharif, L., Al Sukkar, G. M., Hakouz, A. and Al-Shamayleh, N. (2016) Towards a systematic methodology for the design of lift traffic systems in high-rise office buildings. In: *Proceedings of the Symposium on Lift and Escalator Technologies*.

Al-Sharif, L., AlOsta, E., Abualhomos, N. and Suhweil, Y. (2016) Derivation and verification of the round-trip time equation and two performance coefficients for double-deck lifts under incoming traffic conditions. *Building Services Engineering Research and Technology*, 38(2), pp. 176–196. https://doi.org/10.1177/0143624416672505

Ang, J. H., Yusup, Y., Zaki, S. A., Salehabadi, A. and Ahmad, M. I. (2022) Comprehensive energy consumption of lift systems based on hybrid approach of measurement and calculation in low- and high-rise buildings of tropical climate towards energy efficiency. *Sustainability*, 14(8), p. 4779. https://doi.org/10.3390/su14084779

Anh, A. T. H. T. and Duc, L. H. (2022) Super-capacitor energy storage system to recuperate regenerative braking energy in lift operation of high buildings. *International Journal of Electrical and Computer Engineering*, 12(2), pp. 1358–1367. https://doi.org/10.11591/ijece.v12i2.pp1358-1367

Ayele, W. Y. (2020) Adapting crisp-dm for idea mining. *International Journal of Advanced Computer Science and Applications*, 11(6). https://doi.org/10.14569/ijacsa.2020.0110603

Barney, G. and Lorente, A. (2013) Simplified energy calculations for lifts based on ISO/DIS 25745-2. In: *3rd Symposium on Lift and Escalator Technologies*. Available at: https://liftescalatorlibrary.org/paper_indexing/papers/00000048.pdf [Accessed 15 March 2025].

Bergstra, J. and Bengio, Y. (2012) Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, pp. 281–285. Available at: https://jmlr.org/papers/v13/bergstra12a.html [Accessed 24 May 2025].

Chen, T. and Guestrin, C. (2016) XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Christy, T. M. (2014) Common misconceptions regarding lift traffic simulations. *Lift World*, April, pp. 61–64

De Almeida, A. et al. (2012) Energy-efficient lifts and escalators in Europe: an analysis of energy efficiency potentials and policy measures. *Energy and Buildings*, 47, pp. 151–158. https://doi.org/10.1016/j.enbuild.2011.11.053

Dong, Y., Liang, C., Guo, L., Cai, X. and Hu, W. (2023) Life cycle carbon dioxide emissions and sensitivity analysis of elevators. *Sustainability*, 15(17), p. 13133. https://doi.org/10.3390/su151713133

Fernández, J. and Cortés, P. (2015) A survey of lift group control systems for vertical transportation: a look at recent literature. *IEEE Control Systems*, 35(4), pp. 38–55. https://doi.org/10.1109/MCS.2015.2427045

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

Haavisto, A. (2022) CO2 emissions of elevator modernization: life cycle assessment. Master's thesis. Tampere University.

International Organization for Standardization (ISO) (2015) *Energy performance of lifts, escalators and moving walks – Part 2: energy calculation and classification for lifts (elevators)*. ISO 25745-2

Jaggia, S., Kelly, A., Lertwachara, K. and Chen, L. (2020) Applying the crisp-dm framework for teaching business analytics. *Decision Sciences Journal of Innovative Education*, 18(4), pp. 612–634. https://doi.org/10.1111/dsji.12222

Ke, G. et al. (2017) LightGBM: a highly efficient gradient boosting decision tree. In: *Proceedings of Advances in Neural Information Processing Systems* 30, pp. 3146–3154. Available at: https://proceedings.neurips.cc/paper_files/paper/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf [Accessed 10 April 2025].

Kermani, M. et al. (2021) Lift regenerative energy applications with ultracapacitor and battery energy storage systems in complex buildings. *Energies*, 14(11), p. 3259. https://doi.org/10.3390/en14113259

KONE (2020.) 'Energy-efficient MRL Elevators', *KONE Americas Blog*. Available at: https://www.kone.us/blog/energy-efficient-mrl-elevators.aspx [Accessed 15 January 2025].

KONE Corporation (2023) Environmental product declaration: lift with KONE MonoSpace® Upgrade DX. EPD International AB. Available at: www.environdec.com [Accessed 2 March 2025].

Latif, M. N., Kheshaim, M. and Kundu, S. (2016) A review of lift dispatching systems. In: *World Congress on Engineering*, Vol. II, pp. 671–673. International Association of Engineers. ISBN 978-988-14048-0-0

Makar, M., Pravica, L. and Kutija, M. (2022) Supercapacitor-based energy storage in lifts to improve energy efficiency of buildings. *Applied Sciences*, 12(14), p. 7184. https://doi.org/10.3390/app12147184

Martínez-Plumed, F. et al. (2021) Crisp-dm twenty years later: from data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), pp. 3048–3061. https://doi.org/10.1109/TKDE.2019.2962680

McGinnis, W. D., Siu, C., Andre, S. and Huang, H. (2018) Category encoders: a scikit-learn-contrib package of transformers for encoding categorical data. *The Journal of Open Source Software*, 3(21), p. 501. https://doi.org/10.21105/joss.00501

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. and Gulin, A. (2019) CatBoost: unbiased boosting with categorical features. Yandex & Moscow Institute of Physics and Technology. Available at: https://github.com/catboost/catboost [Accessed 20 March 2025].

Rezig, E. K., Cafarella, M. and Gadepally, V. (2021) Technical report on data integration and preparation. *arXiv*. https://doi.org/10.48550/arxiv.2103.01986

Sachs, H., Misuriello, H. and Kwatra, S. (2015) Advancing lift energy efficiency. American Council for an Energy-Efficient Economy. Available at: https://www.aceee.org [Accessed 20 March 2025].

ingh, V. K., Singh, A. and Joshi, K. (2022) Fair crisp-dm: embedding fairness in machine learning development life cycle. In: *Proceedings of the Annual Hawaii International Conference on System Sciences*, p. 190. https://doi.org/10.24251/HICSS.2022.190

Snoek, J., Larochelle, H. and Adams, R. P. (2012) Practical Bayesian optimization of machine learning algorithms. In: *Advances in Neural Information Processing Systems*, 25, pp. 2951–2959.

Thebuwena, A. C. H. J., Samarakoon, S. M. S. M. K. and Ratnayake, R. M. C. (2024) Optimization of energy consumption in vertical mobility systems of high-rise office buildings: a case study from a developing economy. *Energy Efficiency*, 17, p. 68. https://doi.org/10.1007/s12053-024-10246-5

Tukia, T. (2019) Evaluating power consumption of lifts: the perspective of power systems and demand response models. Doctoral dissertation. Aalto University. Available at: http://urn.fi/URN:ISBN:978-952-60-8570-8 [Accessed 10 December 2025].

Tukia, T., Siikonen, M.-L., Donghi, C., Uimonen, S., Hakala, H. and Lehtonen, M. (2016) Practices to improve the annual elevator energy consumption estimates and measurements. *International Review on Modelling and Simulations*, 9(3), p. 134. https://doi.org/10.15866/iremos.v9i3.8757

Tukia, T., Uimonen, S., Siikonen, M.-L., Donghi, C. and Lehtonen, M. (2018) High-resolution modeling of lift power consumption. *Journal of Building Engineering*, 18, pp. 210–219. https://doi.org/10.1016/j.jobe.2018.03.008

Wang, G., Wang, B., Li, C. and Xu, D. (2017) Weight-transducerless control strategy based on active disturbance rejection theory for gearless lift drives. *IET Electric Power Applications*, 11(2), pp. 289–299. https://doi.org/10.1049/iet-epa.2016.0474

Wang, J., Mu, X., Xu, J. and Wang, S. (2012) Design and optimization of dispatching rules for lift group control aiming at energy saving. In: *2012 IEEE International Conference on Information Science and Technology*. https://doi.org/10.1109/ICIST.2012.6221620

XGBoost Developers (2025) XGBoost parameters: XGBoost latest documentation. Available at: https://xgboost.readthedocs.io/en/latest/parameter.html [Accessed 10 November 2025].

Zheng, Y., Zhang, Z. T. and Xu, H. (2013) A novel intelligent lift group control algorithm based on corridor passenger detection and tracking. *Applied Mechanics and Materials*, 336–338, pp. 815–819. https://doi.org/10.4028/www.scientific.net/AMM.336-338.815