DTU Civil Engineering
Department of Civil Engineering

Machine Learning Based Diagnosis of Indoor Environment in Offices

Sophia T. L. Wesche

Master Thesis

Machine learning-based diagnosis of indoor environments in offices

Master Thesis
29th of January, 2023

By
Sophia T. L. Wesche

Cover photo: Vibeke Hempler, 2012
Published by: DTU, Department of Civil Engineering, Brovej, Building 118, 2800 Kgs. Lyngby Denmark
www.byg.dtu.dk

Approval

This Master Thesis was carried out as the last project of a Master degree in Architectural Engineering, at the Technical University of Denmark. The project ran from the 29th of August to the 29th of January and corresponds to 30 ECTS points. It was carried out under the supervision of Jørn Toftum.

It is assumed that the reader has a basic knowledge in the area of statistics.

Sophia T. L. Wesche (s173828)

....

Signature

....

Date

Abstract

In this project, a data-driven tool is developed that predicts the well-being and performance of occupants, given the indoor climate and setup of the office they are in, as well as their personal characteristics. This is done through three main supervised machine learning models: One predicting the satisfaction with the overall indoor climate, one predicting occupant perceived performance, and one predicting the experience of office-related symptoms. Through these models, it is possible to diagnose what parts of the office's indoor climate are problematic - and thereby what areas are most beneficial to improve.

In all three models, background noise from speech and high CO_2 concentrations are the indoor environment quality parameters that have the most negative influence. One of the most influential features in the models is the sex of the occupant, as women seem to be far more negatively influenced by the indoor climate than men.

The project was a success and the three main models were shown to have predictive power, which lead to an in-depth analysis of each of the three targets. It is, however, clear that re-training the models as more data is gathered, could improve them even more. You could, thus, say that the more the tool is used, the more the models learn. Due to the limited size and range of the data set used to train the models, the current models do have some limitations. They are not trained to assess buildings during a summer situation or in countries with climates that are different from Denmark and Greenland.

Acknowledgements

I would first and foremost like to thank my supervisor, Jørn Toftum, who has been excellent at following my process and providing me with useful feedback along the way.

The success of the models designed was made possible because of the thoroughly designed surveys and checklists, as well as the cleaning done to make the measurement data useful. I would, therefore, like to thank Emilie Patricia Dam-Krogh, for letting me use the data she has gathered thus far, as part of her PhD. Not only did she produce data of high quality, she was also very helpful in answering any questions I had about data origin or setup. In extension, I would like to thank everyone who helped Emilie Patricia Dam-Krogh produce the data, as well as the offices that were willing to participate in her research.

Contents

1 Introduction

The relationship between the well-being of occupants and the indoor environment of an office is complex. And yet, when assessing the indoor environment quality, IEQ, we like to look at the performance of specific aspects of it such as the light quality, the thermal environment, the air quality and the acoustics. The well-being of an occupant is, however, also influenced by how the different aspects of the IEQ are interconnected with each other, as well as the characteristics of the occupant and the workplace as a whole [1]. The purpose of this thesis is, therefore, to find a way to make a holistic assessment of indoor office environments.

To do this, the performance and well-being of occupants are predicted, based on the characteristics of the office and the occupants. By using the well-being and performance of the occupants as a design parameter, the goal is to make it easier to prioritise different indoor environment domains in a cohesive way.

Traditionally when designing an office, the indoor environment quality is calculated theoretically. Here, I will instead be turning it on its head and use supervised machine learning to predict the well-being of the occupants based on data from a collection of offices of different designs. This should make it quicker and easier to make the right improvements to the indoor environment.

Data on office layouts, indoor environment measurements and occupants' perception is used to train the supervised machine learning models. The models should predict occupant perception of performance, satisfaction with the overall indoor climate, and experience of building-related symptoms such as *dry or irritated eyes*, *headaches*, *tiredness or fatigue*, and *difficulty to concentrate*. Based on these models, the goal is to diagnose what building attributes most influence the occupants' well-being and performance.

Problem
How can machine learning be used to give a holistic diagnosis of the indoor climate in offices? How do the indoor environment, the characteristics of an office and the occupants within it influence the well-being of the occupants? How do the characteristics of the occupant influence their perception of the office?

1.0.1 Learning objectives

Machine learning

- To analyse data and assess its limitations.
- To compare different supervised machine learning models.
- To design supervised machine learning models that can identify how office characteristics influence the well-being of occupants.

Indoor environment quality and well-being

- To narrow down relevant parameters regarding indoor quality and occupant well-being.
- To be able to diagnose the indoor climate and office parameters based on the well-being of the occupants.

- To communicate how the indoor environment as well as the office and occupant characteristics have an influence on the well-being and performance of the occupants.

2 Background

Improving the well-being of occupants does not only improve their comfort and health - it also increases their performance. As illustrated in Figure 2.1, the occupant's well-being is connected to many different factors.

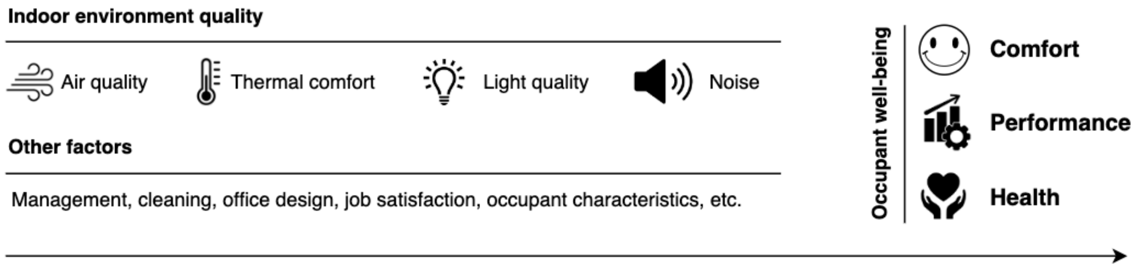


Figure 2.1: Influences on occupant well-being

A study based on 5732 questionnaires in 57 different office buildings showed that perceived comfort is influenced by several personal, social and building-related factors [1]. Therefore, we can't fully rely on the parameters of indoor environment quality, IEQ, when predicting perceived comfort. In order to truly assess the well-being of the occupant, a holistic approach is necessary.

In the following, examples of the influence of indoor environment quality on the well-being of occupants are presented. Afterwards, metrics used to assess and quantify occupant well-being in the form of comfort, performance and building-related health are described. Finally, it is described how machine learning works, and what type of machine learning might be useful when modelling the indoor climate.

2.1 Indoor environment quality and well-being

The indoor environment quality, IEQ, has been shown to influence the well-being and performance of occupants. Examples of Sick Building Syndrome (SBS), Building-Related Illness (BRI) and productivity loss have been seen in relation to office occupants exposed to indoor parameters such as extreme thermal factors, poor lighting, moisture, mould, noise, chemical compounds, etc. [2].

The IEQ is divided into the four areas seen in Figure 2.1: Air quality, thermal comfort, light quality and noise. In an office space, discomfort will often appear due to a combination of these. Table 2.1 shows a selection of studies that details how the IEQ parameters might influence occupants.

Table 2.1: Overview of studies used in the following description of IEQ parameters.

Study	IEQ measure	Result
A metaanalysis of 24 studies, including 6 office studies. [3]	Air quality and temperature	High temperatures and poor air quality, lowered performance by up to 10% on measures like typing speed and unit output
Climate chamber studies, simulating office situations. [4]	VOC and CO_2 concentrations	Lowered concentrations of VOC and CO_2 independently increased cognitive scores.
Controlled experiment simulating office situation. Performed on 12 persons. [5]	Temperature	Lowered performance at 30°C, to thermal neutrality at 22°C.
State-of-the-art study quantifying the health and human impacts of daylighting strategies and views quality from windows on employees' health in offices. [6]	Light levels and views	Workers in offices with poor ratings of light quality and in offices with poorer views used significantly more sick leave hours. Taken together, the two variables explained 6.5% of the variation in sick leave use, which was statistically significant.
Tests on pupils from 8 UK primary schools, with ordinary ventilation rate and increased rates. [7]	CO_2 concentrations	Low ventilation rates in classrooms significantly reduce pupils' attention and vigilance, and negatively affect memory and concentration
A field study assessed subjective reports of distraction from various office sounds among 88 employees at two sites. [8]	Noise	Out of the sample, 99% reported that their concentration was impaired by various components of office noise, especially telephones left ringing at vacant desks and people talking in the background.
Three experiments performed in an open plan office doing office tasks. [9]	Noise	Background noise from speech reduced performance regarding mental calculation tasks and communication tasks.

Research methodology
The studies seen in Table 2.1 are chosen because they all assess the indoor climate based on how it influences the occupant's well-being, performance/productivity or health. It was also a priority to choose studies that assessed office situations, and employees' well-being. The scope of the search was therefore defined by the following:

- Research on office environments.
- Must include indoor environment quality parameters: E.g. air quality, thermal com-

fort, light quality, and acoustics.

- Must include specific metrics: E.g. VOC and CO_2 concentration, temperature, illuminance and sound level.
- Must include conclusions on the influence on occupant well-being, health and performance.

A simplification of the process of selecting the papers is shown in Figure 2.2.

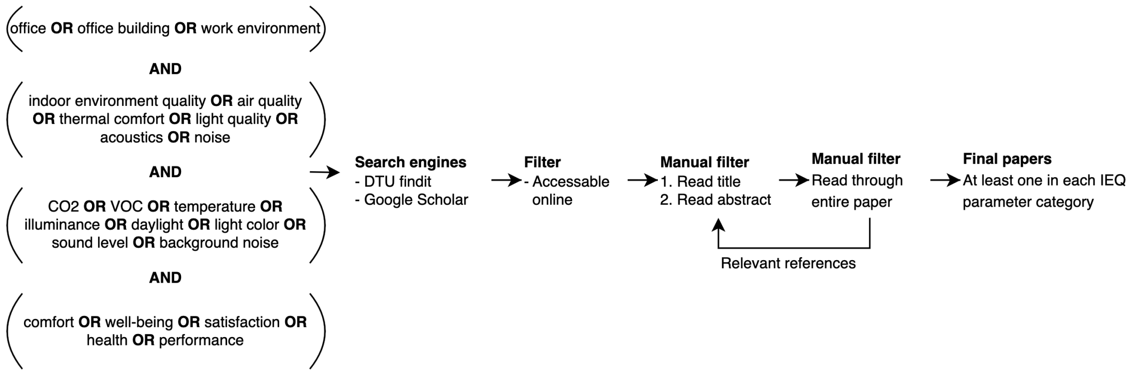


Figure 2.2: Research process for selection of papers.

While this shows the general process, not every key word was used in each search, as the search was adapted to narrow down the results depending on what was needed. All synonyms or alternative key words used in the search are also not seen in the figure. These are words like "productivity", "concentration" and "cognitive performance" to describe performance for instance. If relevant referenced papers were found when reading through a paper, these would be assessed as well.

In Table 2.1, it is seen that one of the studies is based on pupils in classrooms, which is stretching the scope slightly. It was still included as a classroom does have some similarities to an open plan office. Furthermore, it is an example of an experiment that assesses the same work environment before and after the air quality has been improved. It is therefore more representative of a real life situation than some of the controlled climate chamber experiments simulating offices might be.

The final studies seen in Table 2.1 are used to discuss the influence of indoor environment quality in the following.

2.1.1 Air quality

An occupant will be exposed to a range of different pollutants. These are usually pollutants from other people, particles, gaseous emissions from building parts, etc. [10], as illustrated in Figure 2.3.

Higher activity levels and large numbers of people will increase human bio-effluents such as CO_2, heat and smells. When assessing air quality, the CO_2 concentration is, thus, often used as an indicator of human pollutants. In Denmark, it is recommended that the CO_2 concentration is kept below 1000 ppm = 0.1%, while it must be below 2000ppm to comply with the Danish Labor Inspection Authority (at.dk, [12]).

Another parameter often measured is volatile organic compounds, VOCs. VOCs are chemicals, found in products such as paint, cleaning products, building materials, etc. Because they are found in such materials, concentrations are usually much higher inside

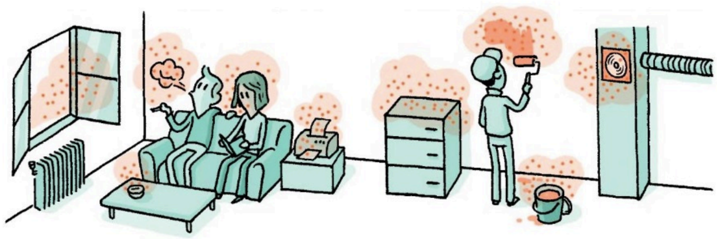


Figure 2.3: Pollutants from people, materials and processes. [11]

than outside [13]. This, however, will depend on the function of the building as well as the ventilation rate.

Consequences
The pollutants described are not problematic as long as their concentrations stay at an acceptable level. The focus on air quality was particularly heightened as a response to building envelopes becoming better due to the energy crisis in the 1970s [14]. The new building envelopes both decreased the heat loss and the air exchange, making ventilation systems extra important.

As seen in Table 2.1, high concentrations of CO_2 and VOC are shown to decrease occupants' cognitive ability [4]. Bar plots of the air quality's influence on the results of different cognitive domains are seen in Figure 2.4. In the study, VOC was high in the conventional situation and low in the Green and the Green+ situation, while CO_2 was high in the conventional situation, lower in the Green situation, and lowest in the green Green+ situation.

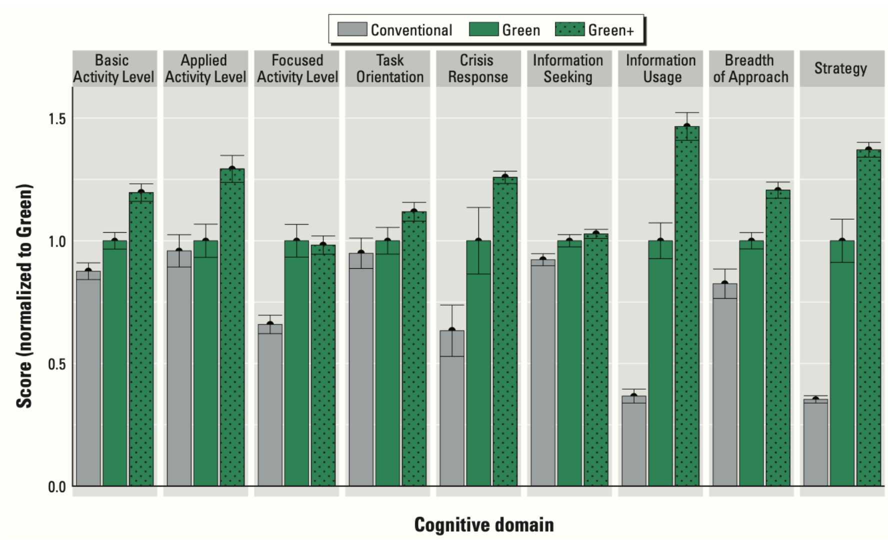


Figure 2.4: Average cognitive function scores and standard error bars from paper. [4]

Another study found that lowering ventilation rates would reduce pupils' attention as well as influence their memory and concentration [7]. While the studies show that the specific concentrations of the pollutants have an influence on the well-being of the occupant, realistically, it will be the overall cocktail of pollutants that has an impact. Fortunately, ventilating for the most pressing pollutant will bring down the concentration of all of them.

2.1.2 Thermal comfort

Thermal comfort is a measure of whether a person feels too cold or too hot, and is influenced by both environmental and personal factors. Environmental factors could be the air temperature, the radiant temperature, the relative humidity, air velocity, etc. Personal factors, on the other hand, relate to the occupant's clothing and activity level, often measured in clothing insulation and metabolic rate, respectively. See Figure 2.5.

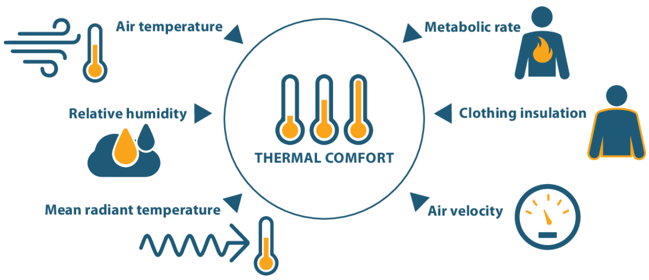


Figure 2.5: Thermal factors [15]

In the late 1960s, Fanger conducted controlled experiments in a climate chamber on university students' assessing the thermal environment whilst wearing standardized clothing and performing controlled activities [16]. This resulted in the Predictive Mean Vote (PMV) model, that rates the thermal environment on a 7-point ASHRAE (The American Society of Heating, Refrigerating and Air-Conditioning Engineers) scale of thermal comfort, seen in Figure 2.6.

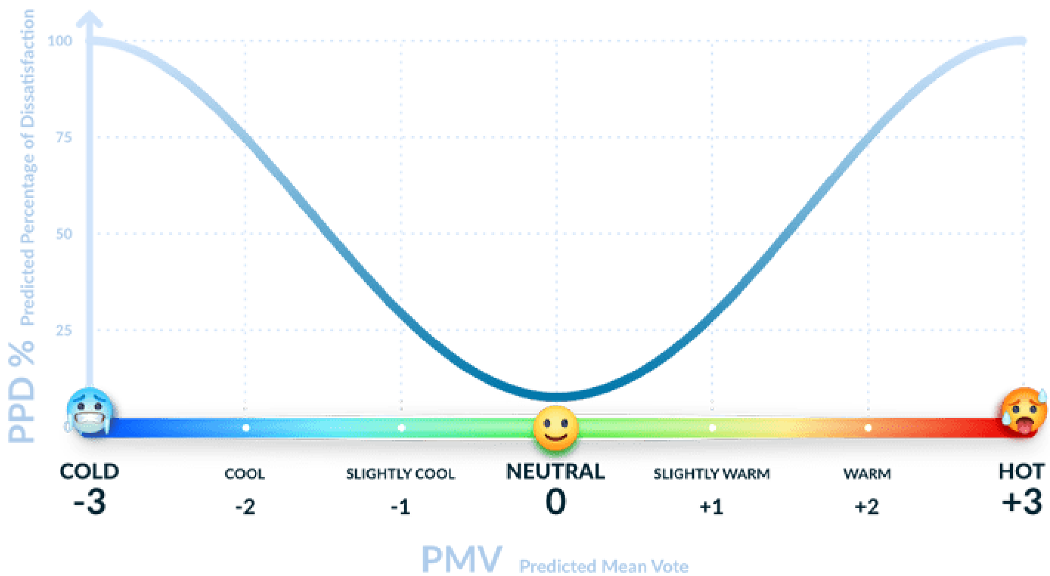


Figure 2.6: PMV in relation to PPD [17]

Here, -3 means that the occupant feels too cold while +3 means that the occupant feels too hot. Thermal balance is found at PMV = 0. In Figure 2.6, PMV is seen in relation to the Predicted Percentage of Dissatisfaction, PPD. According to this model, most people should be satisfied at PMV = 0.

Multiple design standards such as ISO 7730, ASHRAE Standard 55, DS 16798, and CEN CR 1752 are basing recommendations on Fanger's models. Recommended temperatures here, are based on assumed clothing insulation in the summer and the winter, as well as the metabolic rate of 1.2 met, corresponding to a sedentary person doing office work.

Consequences
It can be complicated to evaluate temperature's direct influence on occupants' well-being and performance because changing the temperature might also change other factors related to the occupant's comfort. An example is the study seen in Table 2.1, simulating an office situation, and then increasing the temperature from 22 °C to 30 °C [5]. This lowered the occupants' general well-being: Reduced performance, building-related symptoms, expressions of negative mood, and less willingness to exert effort. Furthermore, the temperature change would also cause the test persons to assess the air quality to be worse.

Relative humidity is the ratio between water vapour in the air and the maximum amount of water vapour that the air can hold at the same temperature and pressure. The thermal comfort won't be influenced too much by the relative humidity, as long as it is held between 40% and 70% [18]. If the relative humidity is much higher, it can, however, prevent sweat from evaporating from the skin, which is a way for the body to cool itself down. If the relative humidity is kept above 70% for long periods of time, it can also cause mould to occur, which can influence the health of the occupant.

2.1.3 Light quality

There are several ways to measure light quality. In this project illuminance and light colour are used.

The illuminance is the total luminous flux per incident on a surface, per unit area. The general requirement for office work is to have a minimum illuminance of 500 lux [19].

The light colour is measured in kelvin, K, and consists of a spectrum of cold, neutral and warm light [20]. Light with a temperature lower than 3300K is warm while light with temperatures above 5000 K is cold. Light between these temperatures is neutral. You get warm light from candlelights, old-school light bulbs and sunsets, while you will see cold light in a blue sky, for instance.

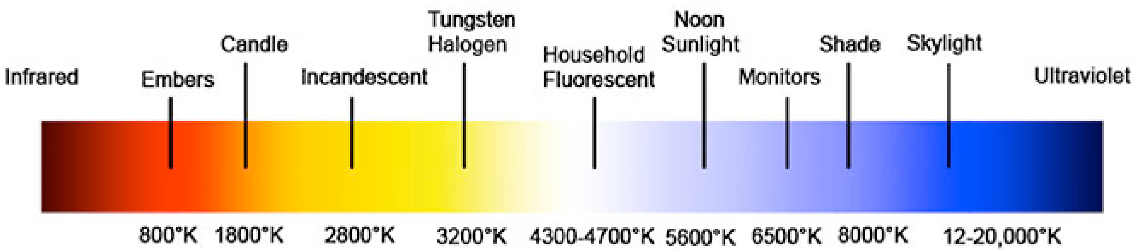


Figure 2.7: Spectrum of light colors in Kelvin [21].

Consequences
Since our ability to see is connected to the quality of light, it seems obvious that it would play a role in our well-being and performance at work. Light, however, also influences our

mood, health and circadian rhythm.

The light-sensitive ganglia cell controls our hormonal and bodily circadian rhythm [22]. The cells are connected to the Suprachiasmatic nucleus, which is the part of our brain that regulates the circadian rhythm, e.g., sleep rhythm, body temperature and hormonal production. This includes the production of the sleep hormone, melatonin. Light will, thus, influence the ganglia cells to inhibit the production of melatonin, supporting a higher activity level.

The colour temperatures in the cold part of the spectrum, are most effective at influencing the ganglia cells to inhibit the production of melatonin. For the ideal circadian rhythm, one should, thus, be exposed to cold light in the active hours of the day, warm light in the evening and complete darkness throughout the night. Reductions in light during daytime can even result in depression, stress and seasonal affective disorder (SAD). This has led to the use of intense light in the morning as a treatment for SAD [22].

Therefore, it would seem obvious to ensure as much natural light as possible. This can, however, pose a risk of overheating the office and causing glare on computer screens. On top of that, it can be difficult to provide sufficient daylight in the winter in northern countries, such as Denmark and Greenland, making the quality of artificial lighting extra important for the well-being of the occupants [22].

2.1.4 Noise

The sound level is used when measuring sound. A-levelled decibels, dB(A), specifically, is used as a unit to quantify sound based on what the human ear can hear. Sound equivalent to 85 dB(A) or above, across eight hours, is more likely to cause hearing loss over time. The sound level is recommended to be kept below 50 dB(A) in offices, according to the Danish Labor Inspection Authority (at.dk, [23]).

Different sounds can influence us in different ways. Physically, noise can cause the release of stress hormones such as cortisol [24]. On the other hand, music can for some people reduce anxiety and stress [25]. It makes sense that we are wired to be alarmed by some sounds, and calmed by others. Cognitively, we only have the bandwidth to follow 1.6 conversations at once [26]. We can, therefore, easily get distracted by other people's conversations.

Consequences

There is no doubt that high sound levels are unhealthy and uncomfortable. Distractions from noise are one of the leading causes of dissatisfaction with the indoor climate [27].

A psychological study looked into the influence of background noise on occupant performance doing office-related tasks [9]. Here it was seen that general background noise only influenced mental arithmetic tasks (being able to calculate in your head). If the speech was part of the background noise, it would, however, both influence the occupant's performance on mental arithmetic tasks and memory for prose tasks (ability to express ideas). Interestingly, the background speech did not need to be relevant, or even in a language the occupant understood, to impair the performance. Similarly, another study showed that concentration was impaired by various office noises - particularly telephones ringing and people talking in the background. [8]

2.2 Occupant-centric building assessment metrics

The studies seen in Table 2.1 are assessing the connection between specific IEQ parameters and well-being. They are applying the metrics that work best for that parameter. Different metrics are, however, necessary to assess the indoor climate as a whole. In the following, metrics used to assess well-being in the form of comfort, performance and health are described.

2.2.1 Comfort

ASHRAE uses a 7-point scale of comfort to rate thermal comfort. This, of course, assumes that comfort is found at thermal neutrality. The problem with this metric is that it does not apply to other IEQ parameters.

When assessing the quality of a wider range of indoor climate parameters, a scale going from dissatisfaction to satisfaction is often used. This scale, however, does not explain what the dissatisfaction is caused by. The advantage is that it allows you to quantify the perception of multiple different types of indoor parameters using the same metric.

In order to reduce comfort into one parameter, the occupant perception of the overall indoor climate can be assessed using the same type of scale. The quality of asking a generalized question like this is that it allows you to see how much each part of the indoor climate matters to the overall perception. This ability to compare IEQ parameters is particularly important if the goal is to diagnose problem areas within the indoor climate. Because the method relies on the perception of the occupant, it does contain some bias. It is therefore important to ask many people.

2.2.2 Occupant performance

In Table 2.1, many of the studies are using performance or concentration as a measure of how the IEQ influences the occupants. To understand how an occupant's performance is influenced by the indoor climate in an office, they are usually tested or asked how they perceive the indoor climate influences their performance.

The tests are mostly cognitive tests. Cognitive tests are tests of how your brain functions work and can involve answering simple questions and/or doing tasks. The advantage of such tests is that they make it possible to quantify performance. This makes it a good tool when you want to compare two different indoor climate situations. The disadvantage is that you always need to have a baseline situation to compare the results to. This makes it complicated if you are working with the influence of multiple parameters at the same time.

The other option is to ask the occupant how they believe their performance, productivity or concentration is influenced by the indoor environment in the office. The answer to this question is usually made quantifiable by letting the occupant answer on a scale. This scale will go from the indoor climate having a poor influence to having a positive influence on their performance, with no influence as a midpoint. The advantage of this method is that it does not need a baseline, and therefore is quicker to perform. While the cognitive testing method might be biased by exterior factors between the two tests, such as lack of sleep, stressful workday, etc., this method assumes that the occupant takes this into account automatically. Thereby it assumes that the occupant knows how they should be performing in the ideal indoor climate. The disadvantage of this method is that it is biased by the person answering the question and that people might perceive the steps on the scale in different ways. When using this method, it is therefore important to gather a large quantity of data, so biased observations don't influence the result too much.

In conclusion, then cognitive testing is best when isolating specific changes in the indoor climate, while the perceived performance is better when you want a more generalized assessment of the indoor climate.

2.2.3 Building related health

It is difficult to measure building-related health. In Table 2.1 it was seen that some studies use sick leave as a measure. While this is a measure that is easy to quantify, it is impossible to differentiate between what is caused by the office indoor climate, and what is caused by lifestyle, stress, etc.

Sick Building Syndrome, SBS, and Building-Related Illness, BRI, are both related to how the indoor climate can affect employees' health. While BRI is connected to diagnosable symptoms caused by a specific agent (eg. legionella), SBS covers a range of unspecific symptoms that are not connected to anything in particular. Because BRI is easily diagnosed and connected to the cause, it is somewhat easy to solve.

According to the World Health Organization, WHO, SBS is identified when people experience a range of common symptoms more frequently than expected [28]. The symptoms increase with time spent in the building and improve or disappear when people are away from the building. SBS is, therefore, today often referred to as Building-Related Symptoms, BRS, as it is the occupants that are experiencing the symptoms and not the building. The diagnosis will, however, be referred to as SBS throughout this report for continuity. The most common symptoms of SBS are:

- Irritated, runny or blocked nose
- Dry or sore throat
- Dryness of skin, or rashes
- Less specific symptoms such as headache, tiredness, irritability and poor concentration.

According to WHO, SBS causes reduced work performance and increased sick days, resulting in a total cost in the range of 0.5-1.0% of the gross national product [28]. Even though the cause-effect relationship is unclear, there are some areas that are more connected to SBS than others. Cleaning and ventilation have both been shown to reduce SBS, by providing fresh air, diluting air pollution and removing pollution from surfaces. The combination of comfort factors such as noise, uncomfortable temperatures, low relative humidity and poor lighting can, however, make SBS more likely. Stress or dissatisfaction with work is also likely to increase the chance of SBS. Solving all problems regarding the indoor climate is, thus, not necessarily going to eliminate the SBS [29].

2.3 Machine learning

Machine learning is a subarea of artificial intelligence, that uses algorithms to solve tasks in a way that imitates how they would be solved by people, but with far more computational power and memory. Machine learning uses past experience (data) to learn how to solve the task, in ways that also apply to future similar tasks. The process of the algorithm figuring out how to solve the tasks is the "learning" part of machine learning. [30]

The three main areas of machine learning; (1) Supervised learning, (2) reinforcement learning, and (3) unsupervised learning, are presented in the following.

2.3.1 Supervised learning

Supervised learning uses two key parts of a data set. The first part are the features and the second part are the labels. The features are the bits of information we have for all the observations we have made; the labels are the bits of information we are trying to predict, based on the features. For example, suppose we take 100 people and record their height, weight and sex. We would then like to train a model that uses people's heights and weights to predict their sex. This would give us 100 observations with the features being height and weight, and the label being the sex.

Here, the information used for the model is split into training data and testing data. The training data is used to teach the model how to predict the label. The testing data, on the other hand, tests how well the model does at predicting the label. For example, we could train the model on 80 people. This would mean we feed the model someone's height and weight and then let the model make a prediction on their sex. If the model is correct we will reward it, so it will keep using the same (or similar) algorithm to predict in the future. If the model is incorrect, we will punish it so it will change its algorithm for prediction. Once we have trained on 80 people, we can then use the other 20 people for testing. Here we will just feed the model their heights and weights, and then see how many have their sex correctly guessed.

One issue of machine learning is overfitting. This occurs when a model is not trained well and does not follow the principles of model training. This can lead to the model only predicting a single output, which may be accurate for the sample data it has been fed, but will not be an accurate depiction of the population it is trying to model. Generally, the goal of machine learning is to automate solving tasks as much as possible, and it is, therefore, preferable to use a model that generalizes the problem enough to give an accurate depiction of a population and not just the sample data.

Going back to the example, then a good model will correctly guess most (or all) of the 20 people's sex. An overfit model may train well but would struggle to predict the sex in the testing set. For example, if the training data featured people from Denmark (generally larger) and we then tested it on people from Indonesia (generally smaller), we may see that the model is overfitted to large people, and would predict most Indonesians as women.

Supervised learning problems are split into two main categories: regression problems and classification problems.

Regression

In regression problems, we predict a continuous value, y, based on the observation, x [30]. This could, for instance, be making a prediction based on fitting a line to your data. Your function,

$$b + ax = y,$$

might not be a very good model, but it would be a supervised regression model. An example would be predicting someone's height, y, based on their weight x.

Classification

In classification problems, the discrete response is predicted. This means that the model is predicting what class, y, an observation, x, belongs to. The target classes we wish to predict could, for instance, be an occupant's level of comfort given the indoor climate of their office. Examples of how supervised classification models have been used to predict thermal comfort can be seen in the Literature review, Chapter 3.2.

Classification example with K-nearest neighbour

An example of a simple supervised learning algorithm, often used for classification, is K-nearest neighbours (KNN). Based on the distance between the observations, the algorithm divides the data into subgroups belonging to each class. This is visualized in Figure 2.8. Based on the training data, the algorithm has determined that the blue observations belong to one class and the green observations belong to another.

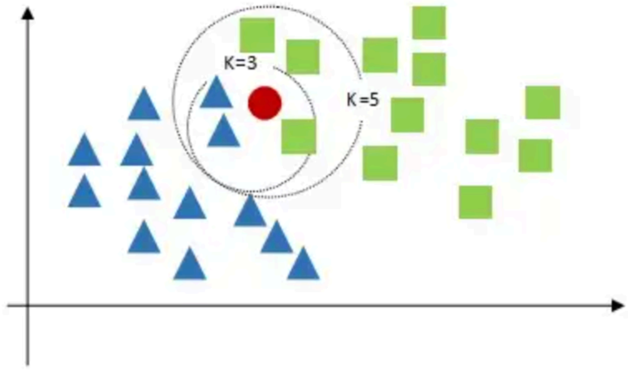


Figure 2.8: Illustration of how KKN classifies [31].

What class the new observation (red circle) belongs to, is determined by what the most common class of the K-neighbours is. If it is decided to look at K = 3 neighbours, the observation is, therefore, classified as blue, while it is classified as green if we look at K = 5 neighbours.

2.3.2 Reinforcement learning

Reinforcement learning is similar to supervised learning in that it makes an initial prediction/action based on feature data. Unlike supervised learning, reinforcement learning will use feedback from its output to reinforce the learning process. This can be compared to a child being told off when misbehaving and adapting its behaviour because of the ”reward signals” it receives from its parents. [30]

Reinforcement learning has, for instance, been used to create adaptive HVAC systems [32]. Here, an initial prediction of the best set-points is made and implemented in the office. The occupants will then give their feedback, which the reinforcement model uses to adapt the indoor climate to the occupants’ liking. The HVAC system will, hence, be controlled by a feedback loop between the occupants’ perception and the adaption of the algorithm. The only problem with this type of machine learning is that it requires a lot of feedback from the employees to tailor the final system to the employees.

2.3.3 Unsupervised learning

Unlike the two other machine learning methods, unsupervised learning isn’t predicting a specific target. It works by identifying underlying patterns in data, and grouping together similarities [30]. This is similar to how a child learns. The child might not know that a tiger is a tiger, but it knows that it looks more like a cat, than it looks like an elephant.

This method can for instance be used for image recognition. Here it would take ages to train a model to learn what every single person looks like if supervised learning was used. By instead using unsupervised learning, the algorithm can simply compare two images and figure out if it is showing the same person or not.

3 Literature review

Throughout this literature review, it is assessed how comfort previously has been modelled, and what we can learn from it. Furthermore, it is seen how different post-occupancy tools are assessing and diagnosing the indoor climate in offices.

3.1 Learning from Fanger's thermal comfort model

The PMV comfort model was briefly introduced in Chapter 2.1.2. As mentioned, PMV is widely used when designing thermal environments and is a good model for capturing the connection between heat balance and occupants' perception of the thermal environment. The model has, however, received some points of criticism, we might be able to learn from. Figure 3.1 illustrates how these areas of criticism are interconnected.

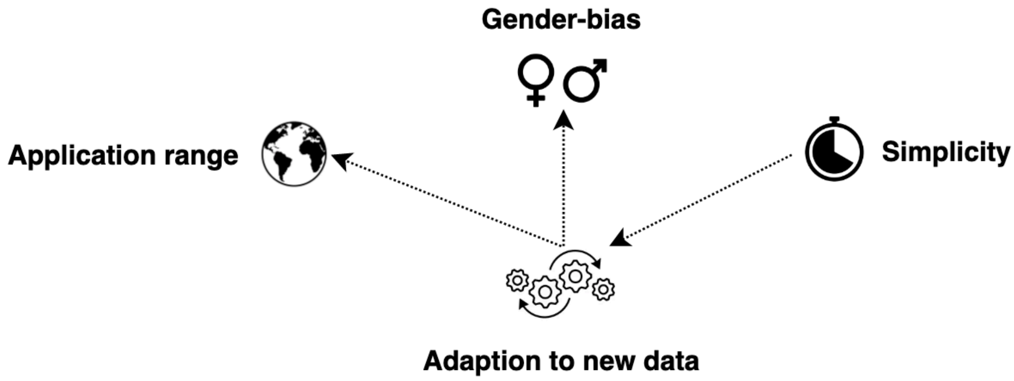


Figure 3.1: Areas of criticism and their connection to each other.

Application range

The model only applies to specific geographical conditions and types of offices. It is, for instance, assuming that the office is mechanically ventilated and has closed windows. Generally, the accuracy of the model is reduced when it is tested in actual offices, which is partly because only a few offices provide a setting that is as controlled as the climate chamber the model is based on [33].

Making a data-driven model, based on a range of actual offices would reduce the model's accuracy, but perhaps limit over-fitting to the controlled climate chamber. This might not be the best solution when working with specific parts of the heat balance. It could, however, be a good solution when working with broader and less specific metrics.

Simplicity

Another point of criticism is the very specific input parameters needed, making the model complex to use [33]. Because of the difficulty of measuring some input parameters, standard values and simplifications will often be used, reducing the overall accuracy.

Basing a model on easily accessible information from actual offices and occupants could make the final model easier to use. Using broader input parameters does, however, also open the door to some uncertainty.

Gender-bias
When design criteria are made based on the comfort model, the metabolic rate is set to $1.2met = 70Wm^{-2}$, corresponding to someone doing sedentary work (eg. DS 16798). The issue with this value is that it actually is the resting metabolic rate of a 70kg, 40-year-old male. For comparison, the average metabolic rate of young adult females that are performing light office work is $48 \pm 2Wm^{-2}$ [34].

A paper from 2019 [34] argues that these gender-biased design requirements ultimately can result in the building being less energy-efficient because women will have to adapt the indoor environment to fit them. They argue that the design requirements should be based on a broader range of age groups and genders. In other words, they want the standard metabolic rate to reflect a broader range of people than the current one does.

Adaption to new data
This brings us to the last issue, which touches on some of the previous ones. The empirical aspect of the model makes it hard to adapt or expand the model as new information is gathered [35]. Every time new factors are considered, the model must be re-calibrating based on new information, which isn't a particularly sustainable solution.

The issue regarding the implementation of new information could be solved through machine learning models, as they can easily be retrained as new information is gathered. Machine learning is, however, only effective and representative when enough quality data is gathered.

3.2 Thermal comfort predicted with machine learning

Table 3.1 shows three studies where supervised machine learning is used to predict thermal comfort response on an ordinal scale, going from too cold to too hot. Some are using the ASHRAE 7-point scale while others are using a simplified version of only three steps. Each study is using different measures of performance. To make them somewhat comparable, only the accuracy is noted in the table.

Table 3.1: Literature review of studies using machine learning to predict thermal comfort.

Study	Models	Highest accuracy	Prediction	Data origin
A comparative study of predictive individual thermal sensation and satisfaction using a wrist-worn temperature sensor, thermal camera and ambient temperature sensor [36]	subspace KNN, Random Forrest, Subspace Discriminant KNN, Quadratic SM	0.73-0.76	ASHRAE 7point scale of thermal comfort and sensation.	Controlled experiment with temperature sensors
Personal thermal comfort models with wearable sensors. [37]	Different models types: lda, rehLogistic, nnet, svmRadical, KNN, Naive Bayes, rpart, J48, Part, C5.0, treebag, gradient boost model, extraTrees, random forest.	0.84-0.88	Cooler, no change, warmer.	Wearables during daily activity
Machine learning algorithms applied to a prediction of personal overall thermal comfort using skin temperatures and occupant's heating behaviour. [38]	Support vector machines; linear, quadratic, gausian.	0.95	Thermal comfort votes and thermal sensation votes.	Controlled experiment with temperature sensors

Data quality > machine learning algorithm
In Table 3.1, it is seen that the studies reviewed all have tested a large range of machine learning algorithms, to find the best model. In each of the three studies, it is seen that multiple different machine learning algorithms result in almost the same accuracy. This

shows that the quality of the data might be more important than the specific algorithm. This of course does not mean that different machine learning algorithms shouldn't be compared and optimized.

Control or range
Like in the original PMV model, we still see that the models are based on controlled and fairly individualized experiments. The data is gathered by testing a few people multiple times or over a longer period of time. Whether this is in a climate chamber or by following them around in their everyday life with sensors, the extensive data collection is limited to very few test persons. When working with machine learning a large amount of data is a priority. Assessing a few people in different situations allows a large quantity of data to be gathered relatively quickly while doing a range of specific measurements that can be compared to the person in other situations.

In a project where the focus is the diagnosis of the overall indoor climate of offices in connection to multiple occupants, this type of experiment might not work as well. When covering a broader group of people and office environments is a necessity, these controlled methods would be way too time-consuming. As mentioned, the lack of control might reduce the accuracy of the model, but will also reduce the chances of over-fitting to one situation, making it more applicable in real-life offices.

3.3 Post-occupancy evaluation, POE

Post-occupancy evaluation, POE, tools are set up in an attempt to close the gap between design and building performance. In the next section, three post-occupancy tools, that are set up in different ways, are reviewed. An overview of the tools can be seen in Table 3.2.

The InnoByg tool was developed as part of InnoByg's development project "Totalværdi og indeklima" in collaboration with the Technical University of Denmark (DTU) and the Technological Institute (TI) [39]. Their goal was to make a simple tool that translated indoor climate improvements into monetary gain. In contrast to the other two tools, this tool is only taking temperature and CO_2 concentration into account.

Another tool is an IEQ survey tool, developed by the Center for the Built Environment at Berkeley University [40]. The tool is built on online reporting and feedback loops, based on surveys on office layout, office furnishings, thermal comfort, indoor air quality, lighting, acoustics, and building cleanliness and maintenance.

Finally, the BOSSA tool (Building Occupancy Survey System Australia) [41] is able to diagnose the indoor environment parameters based on a rating system. Here, the design of the office is rated based on feedback from building end-users and quantitative data from the offices. When developing the tool they first defined nine different uncorrelated IEQ dimensions:

- Spatial comfort
- Indoor air quality
- Personal control
- Noise distraction and privacy
- Connection to the outdoor environment
- Building image and maintenance

- Individual space
- Thermal comfort
- Visual comfort

The nine dimensions were then used to develop separate multiple regression models on four indexes: Work area comfort, building satisfaction, productivity, and health. The tool rates how well a building performs in each of the 9 categories, based on their importance regarding the four indexes.

Table 3.2: POE tools.

	InnoByg Tool	Berkeley IEQ tool	BOSSA
Interface	Excel document	Website with surveys and results.	Website
Input data	Number of employees, average pay, and percentage hours of temperatures and CO_2 concentrations in different categories across a year	Office characteristics and occupant perception of indoor climate.	Physical measurements, office characteristics and work area comfort, building satisfaction, productivity and health
Data gathering	Consultant's perception of temperature and CO_2 concentration, and owner's numbers on employees and pay.	Surveys that ask more questions if the occupant is dissatisfied with parts of the indoor climate.	Snapshot physical measurements in offices, at the time the occupants fill surveys out on their perception of the indoor environment.
Data handling	Calculates percentage productivity improvement regarding temperature and CO_2 concentration based on two studies.	Diagnoses problem areas of the indoor climate by analysing the surveys, and comparing them to baseline buildings, such as a previous office building.	Uses principal component analysis and linear statistical models to create a type of scoring system that rates indoor climate parameters.
Output	Possible percentage productivity increase and money saved if temperature and CO_2 concentration improvements are implemented	Figures of different indoor climate parameters that show the percentage of employees that feel satisfied, neutral or dissatisfied. The website then allows you to see what specific features in the office the occupant believes is making them dissatisfied.	Output is a numerical scoring of nine different IEQ dimensions based on models of work area comfort, building satisfaction, productivity and health.
Advantages	Simple, cheap and quick to use and get results without needing an engineering background. Calculates the economic gain which is useful in a sales pitch.	A cheap and quick tool that allows a large quantity of data. The survey questions digging into problem areas makes the diagnosis more specific.	While the development of the model is more complicated than the other tools, the final rating system is simple to use and intuitive to understand.
Disadvantages	You do not gain much more from the tool than you would from look ing at design standards. It is also very simplified predicting the economic gain based on one study of each IEQ parameter.	The tool does not learn when new data is gathered. It only looks at the data from a specific building and then compares that to the data from another baseline building.	The model is better the more data is gathered, but added data also means a complete change of the model and the rating system.
Plans to improve tools	Plans to include sound and light as IEQ parameters.	It is planned to use the data to develop a data mining tool, that uses all the data from different buildings to find patterns that are useful to improve the diagnosing process.	The tool can be improved as more data is gathered.

Occupant-centric rating metrics
The three models each have their own way of diagnosing the office indoor environment. The InnoByg tool and the BOSSA tool use productivity as at least one of the rating measures, while the Berkely IEQ tool and BOSSA both are using scales, going from dissatisfaction to satisfaction, to rate how the occupant perceives parts of the indoor climate. What all the tools have in common is that the rating or diagnosis of the indoor climate is determined based on perception rather than tests and measurements.

The BOSSA tool uses the occupants' rating of how they believe the office indoor climate influences their productivity and health. They could have chosen to test the occupants' performance through cognitive tests or assessed the occupant's health by looking at employee sick days. By instead asking the occupant, the creators of the tool assume that the occupant has a better idea of what is office related and what might just be a general health or performance issue.

Diagnosing the office environment
The holistic approach used when gathering information about the building performance for BOSSA and the Berkeley tool makes it more difficult to fall into the trap of purely basing the building performance on quantitative measurable IEQ parameters. By adding the extra dimension of office design, it becomes easier to understand why the office is performing as it is, as well as figure out what could be changed.

A simple tool
All three models are trying to find a balance between making a tool that is simple and easy to use, whilst also being predictive and informative. Both the InnoByg tool and the Berkeley tool are simplifying the use of the models by relying on surveys and perception.

In the InnoByg tool, it is the consultant's assessment of how many hours they believe temperature and CO_2 concentration levels are within different categories. This makes sense as it would be time-consuming and expensive to measure or simulate the values across a whole year. Still, it is hard to believe that a person actually is able to give realistic input.

Rather than getting people to guess IEQ parameters, the Berkeley tool, asks the occupants how satisfied they are with different parts of the indoor climate. They are also asking questions about the general office setup, but are never performing any measurements. Basing the tool on surveys is both their strength and weakness. The strength is that they are able to reach a lot of occupants with their survey, thereby being able to gather information from the entire building in an inexpensive non-time-consuming way. The weakness is that the assessment of the entire building will be tied to the people filling out the surveys, making it necessary to have enough respondents to get a useful result and diagnosis.

The BOSSA tool uses a combination of surveys and quantitative measurements. Here, snapshot measurements of the IEQ parameters are performed in the office at the same time as occupants are assessing the indoor climate. Doing measurements of the entire office will be a simplification of what the occupants are experiencing in comparison to the sensor measurements used in the studies from the previous chapter. Still, it is more costly and time-consuming than just relying on surveys. The advantage of the simplification is that it makes it possible to reach many occupants in an efficient way, whilst also getting quantitative data that is comparable to other buildings.

4 Methodology

The data analysis consists of the following stages:

1. Data cleaning
2. Feature selection
3. Model training and evaluation

Although this is the overall structure of the data analysis, features are also removed in both the cleaning- and the training phase, just like the data cleaning is improved based on learnings from the two other stages. It is, thus, not a completely linear process. A simplification of the processes is visualized in Figure 4.1.

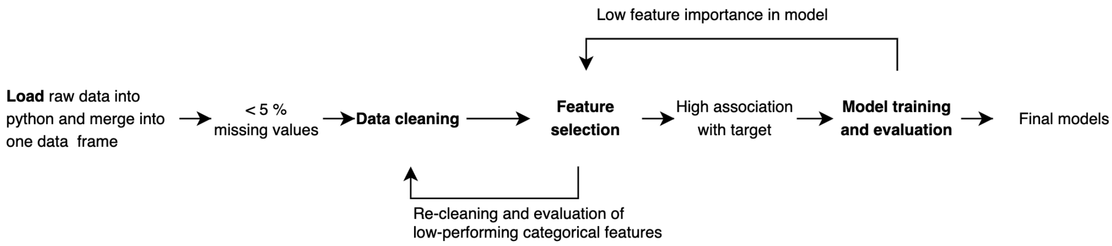


Figure 4.1: Visualization of processes.

4.1 Data

The data is gathered by Emilie Patricia Dam-Krogh, who also designed the surveys and building checklists, as part of her PhD. The data covers 18 offices in Denmark and 22 offices in Greenland.

4.1.1 Data gathering

In each office, a building checklist and personal questionnaires were filled out and physical measurements performed.

The physical measurements include relative humidity, air temperature, CO_2 concentration, VOC concentration, sound level, illuminance, and light colour, measured continuously every minute for about a week. The available features vary between different buildings.

In each building, three different checklists were filled out: A building checklist, a manager checklist and an office checklist. Both the manager and the building checklists would contain general information about the building such as when it was built, the address, last major renovation, typology etc. On top of that, the building checklist contained information about the entire office building and its layout. This would be information about the HVAC system, the building envelope and ventilation of non-office areas. The manager checklist, on the other hand, would contain information about how the office is run. This includes outdoor contaminant sources in the vicinity of the building, past or present damage to the building as a result of fire or water, occupants in the building as well as occupancy rate in general and during covid, cleaning frequency and cleaning product type.

The office checklist contained information about the specific offices that the physical measurements are made in. This contained questions about the physical layout and materials in the office, the number of people in the office, light fixtures, ventilation fixtures, but also the damages to the room. On top of this, the inspector would note what they observed as they enter the room, noise during the monitoring, pollutants and the cleanliness of the office.

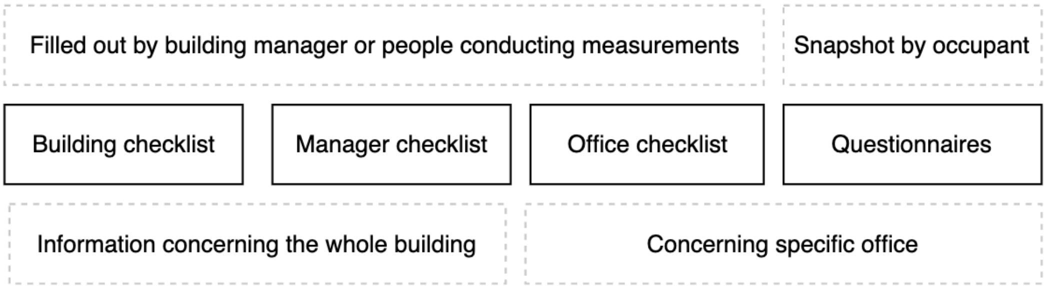


Figure 4.2: Overview of checklists and questionnaires.

The personal questionnaire included questions about the rate at which the occupants were dissatisfied with air movement, temperature, air humidity, smells, light level, light reflections, noise level and cleanliness in the last four weeks. Furthermore, the occupants would answer questions about what they thought of the overall indoor environment and how they felt it influenced their performance. Here, they also answered questions about their sex, age, smoking status, use of eyewear, and job type.

4.1.2 Predictions and predictors

Figure 4.3 shows an overview of how the information gathered is used as predictors/features in the models. This covers the building/office design, how the building is run, quantitative measurements of the indoor environment, occupant perception of the IEQ parameters and occupant characteristics. Supervised learning is then used to predict the occupant's opinion on how the indoor climate performs, how they believe the indoor climate is influencing their performance, and how often they experience at least two building-related symptoms. These are the three performance parameters used to rate or "diagnose" the office's indoor environment.

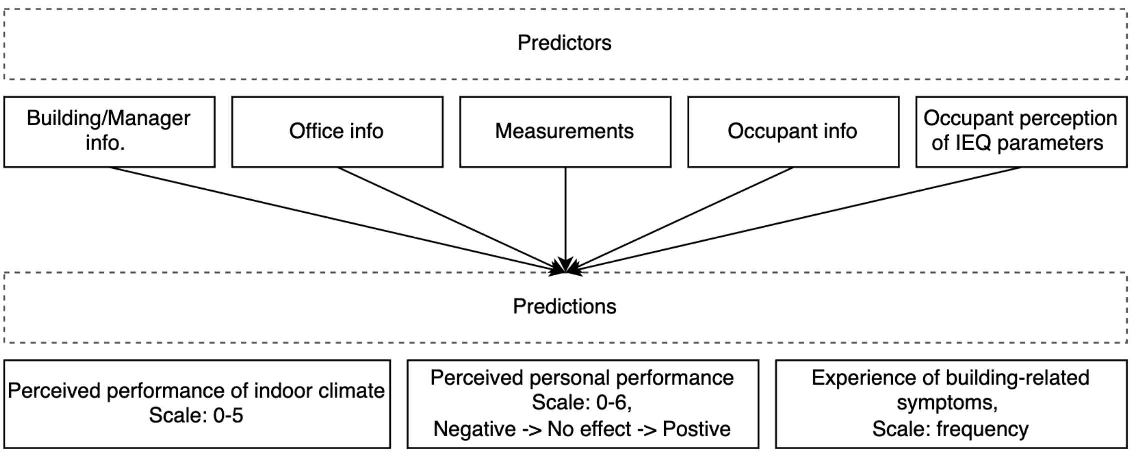


Figure 4.3: Overview of information used as predictors and predictions.

The indoor climate
For the first building performance parameter, the occupants were asked to answer the question: "Considering the overall indoor climate, how do you think the building is performing?" on a 6-point scale from 0-5, where 0 was "Not at all" and 5 was "Very well".

Performance
For the second building performance parameter, the occupants were asked to answer the question: "As a whole, how do you estimate the impact on your performance of the indoor environment, the building and the office?" on a 7-point scale from 0-6, where 0 was "Very negative", 3 was "No effect", and 6 was "Very positive".

Building-related symptoms
For the third building performance parameter, the occupants were asked to answer how often they felt a series of symptoms, such as headaches, tiredness and fatigue, difficulty concentrating, dry/irritated eyes, etc. They would be asked: "During the past four weeks you were at work, how often have you experienced the following symptoms:", and answer (0) "Not in the past 4 weeks", (1) "1-3 days in the past 4 weeks", (2) "1-3 days per week in the past 4 weeks", or (3) "Every workday in the past 4 weeks". To figure out how the experience of the symptoms related to the office, they were also asked "How did you experience the symptom when you were out of the building?". As an answer to this question, they had three different options: (1) "Symptoms got worse", (2) "Symptoms got better", or (3) "Symptoms stayed the same". This information is compressed into a single model of the experience of building-related symptoms.

The point of predicting the experience of building-related symptoms is to make a model that covers building-related health. This is done by approximating the definition of Sick Building Syndrome, SBS, described in Chapter 2.2.3. To make sure the symptom is actually building-related, only observations where the occupant experienced a symptom in the office and felt better after they left the office, are used. In cases of sick building syndrome, multiple symptoms will often be experienced [28]. The model is, thus, predicting when at least two building-related symptoms are experienced by the occupant.

According to the definition of SBS, 20% of the occupants should be experiencing symptoms. This would be impossible to check and base the model on, given the available data from the surveys. In general, the definition of SBS is fairly broad and unspecific, making it difficult to diagnose. While this building-related health model is based on the definition, it is, thus, important to remember that it is not actually able to diagnose SBS.

4.1.3 The problem of few observations

Because much of the data was gathered during a pandemic, not many employees were actually working from the offices that were assessed. This means that the subjective opinion of the few people that were in the office might skew the results. The small number of observations will equally mean that the models won't be as representative as they otherwise could be.

In the final data set, there are only 200-221 observations, depending on the prediction/target. When we wish to predict targets that have between four and seven different classes, very little data will be available to train each class. Regarding perceived performance, for instance, the model would have to predict each of the seven different classes. This means that there are only about 30 observations available to both train and test the model on. This is of course assuming that the classes are completely balanced. In most cases, they are not, which means that there might be 7 observations in one class, 70 in another and 13 in a third. Ultimately, this would result in poor models, that are over-fitted to the small sets of observations. To solve this issue, the classes of each target are merged into two. Since all targets are ordinal, the merge seems somewhat natural. See Figure 4.4.

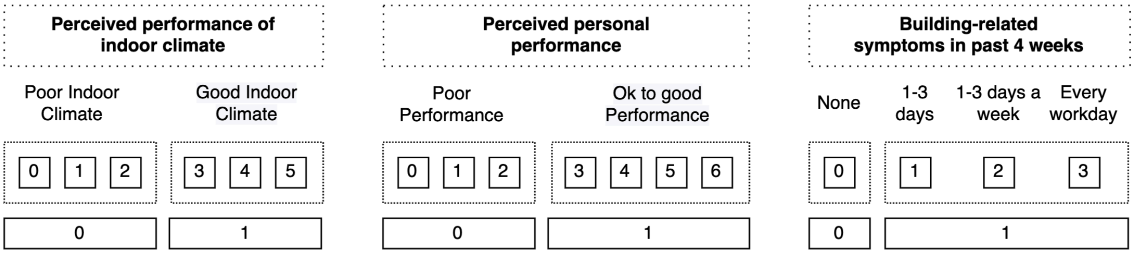


Figure 4.4: Reduction of target classes.

As is seen Figure 4.4, the classes have been merged in a way that prioritizes isolating the critical situation. While this does mean we miss out on some of the complexity of the original scale, it will make it easier to pin-point what features are more likely to cause the occupant's dissatisfaction.

4.2 Data cleaning

The data cleaning is centred around (1) gathering available data in useful data frames, that can be read by the machine learning algorithms, (2) analysing the data quality and (3) finding solutions for missing data. The overall data cleaning process is seen in Figure 4.5.

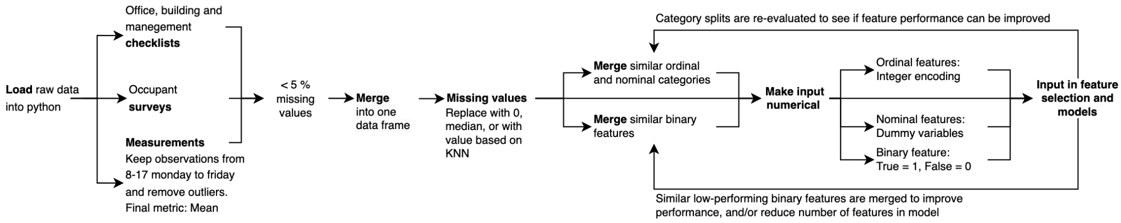


Figure 4.5: Data cleaning process

4.2.1 Creating the initial data frames

The raw data consists of multiple excel and CSV files. They were loaded into python, where all code and calculations throughout the project are executed. The data were merged into a final data frame as it is illustrated in Figure 4.6.

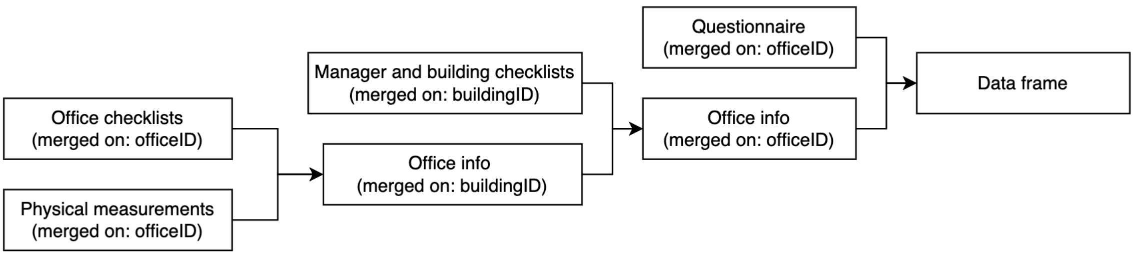


Figure 4.6: Merge of data into final data frame.

As both measurement data and office checklists are set up for each office, these are simply merged based on the office IDs. Based on the building ID, the manager and building checklist data are then exploded to merge with each of the relevant offices. This means that the building and manager data is multiplied as many times as there are different offices in the building. Finally, all the information from the offices is merged with the questionnaire data frame containing the occupant experience. If multiple people answering the questionnaire are sitting in the same office, that means that this row of office information will be multiplied. This also means that the office information is removed if no occupant in the office has filled out the questionnaire.

Physical measurements
The physical evaluation of the indoor environment consisted of measurements taken every minute across a week. Given that the occupants are only at work from 08:00 to 17:00 from Monday to Friday, all measurements outside this time frame were removed. Unrealistic values and outliers in each category were removed, in order to avoid them skewing the result. This was done by removing values outside the following intervals:

- $15°C$ < Temperature < $35°C$
- 0 < Relative Humidity < 100%
- 380 ppm < CO_2 concentration
- $0 \leq$ Illuminance [lux], Light Colour [K], Sound Level [db(A)], VOC [ppb]

To generalize each of the physical features in the office, the mean of the remaining observations was calculated and added to a data frame consisting of the means of the psychical measurements in each of the offices.

When deciding what metric to use to best communicate the physical measurements, the minimum and maximum values, as well as the 25th, 50th, and 75th percentiles were also assessed as contenders to the mean. First, the minimum and maximum values were dismissed because of their sensitivity to potential outliers and inability to represent the general indoor climate. The 25th and the 75th percentiles would be a way to represent both high and low values, in a way that is less sensitive to outliers. It was, however, found that the two percentiles of each physical measurement were highly correlated (0.71-0.93). This means that it would be possible to get rid of one of the features without losing too much information. However, if we were to represent all the measurements through just one feature, the mean or the median would be more representative. The mean is more sensitive to extreme values than the median. But, since it is such values of the IEQ parameters that are likely to influence the overall indoor climate poorly, the mean was chosen to describe the physical measurements in the models. This way, the mean measurement will pick up on abnormalities in the indoor climate if they are either present across many of the observations or very high or low values are seen.

Checklists and questionnaires
Checklists and questionnaires were loaded in separately and merged into data frames, and later merged into the final data frame as illustrated in Figure 4.6. Before the separate data frames were merged, features that were not relevant enough to say anything about the offices were removed. This meant that only information about the building envelope and the HVAC system was kept from the building checklist, while only the cleaning schedules were kept from the manager checklist data frame. Many features, such as history of damages due to fires, water, etc. were removed - not because it wasn't relevant - but because they occurred in less than 5% of the observations, and would, thus, not explain nearly as much about the building as some of the other features, likely making the model more imprecise. This, of course, is also a convenient way to reduce the number of features.

4.2.2 Dealing with categorical features

Only a few machine learning models can handle non-numerical input. It is, thus, crucial to transfer the features into numerical values without losing too much information. How this is done, depends on the feature type. In the following, different relevant categorical feature types from the data set are described as well as how they are transformed into numerical input.

Binary feature

A binary feature is a discrete feature which can only take two values, denoted 0 (false) or 1 (true). Most of the checklists are built as yes/no questions, making it fairly simple to give the features a binary setup. In some cases, binary categories, are so similar to each other, that it was decided to merge them into one feature. This was for instance done when the three binary features: 'noisePhoneTalk','noiseOverhearing' and 'noiseNeighborTalk', were merged into one feature called 'talk'. This new binary feature is 1 if either of the above is true.

Nominal features

A categorical feature is nominal when it is not ordered and only uniqueness matters, such as the profession of the occupant. These features will usually be processed through one-hot encoding, making a binary feature for each of the categories, as illustrated in Fig 4.7.

id	color
1	red
2	blue
3	green
4	blue

One Hot Encoding

id	color_red	color_blue	color_green
1	1	0	0
2	0	1	0
3	0	0	1
4	0	1	0

Figure 4.7: One-hot encoding example [42].

In the data analysed, most categorical features were already presented in the one-hot encoded format. This was in some instances reversed in order to get an overview of the representation of each category and to merge them if necessary. When combining the binary features into one categorical feature, it also automatically removed all categories that were not seen in the data set.

When using nominal features in the model, each category is represented as its own binary feature. This means that they will be inversely correlated with each other, which can be difficult to handle for some machine learning algorithms. In this project, dummy variables are used to deal with this. When creating dummy variables you automatically one-hot-encode the feature and remove one of the one-hot-encoded new binary features as it will already be inversely represented by the other features.

An example is the feature of occupant sex. When one-hot-encoding the feature, it will create two binary features. One called called "sex_male", where 1 represents that the occupant is male and 0 represents that the occupant is female, and another feature called "sex_female", where 1 represents that the occupant is female and 0 represents that the occupant is male. By using dummy variables, the "sex_female" feature is removed, as being female is already inversely represented by the "sex_male" feature.

Ordinal features
If the categorical variables have an order or ranking, it is an ordinal feature. This is for instance the cleaning frequency or the age group of the occupant. To keep the natural relationship of the feature, ordinal encoding (or integer encoding) can be used. This means that each ordinal category will be assigned an integer starting at zero, according to it's rating [43].

This can be exemplified with the frequency of how often the floor is mopped. Here, "never" is encoded as 0, "less frequently " is encoded as 1, "once a month" is encoded as 2, "every second week" is encoded as 3, "once a week" is encoded as 4, "2-4 times a week" is encoded as 5, and "every day" is encoded as 6. Had the feature "mopping frequency" been one-hot-encoded as you do with nominal features, important information would be lost about the ranking of the frequencies.

4.2.3 Missing values

Few machine learning algorithms can run if there are missing data. Missing values are easy to handle with nominal features, because the one-hot encoded sub-feature showing missing values, simply can be removed. It is, however, a bit more complicated in the case of numerical and ordinal features.

Here, the problem can be handled in two ways: By removing rows of missing values completely, or by replacing the missing values with other values.

For classification, the target values cannot be missing as they are needed to train and test the model, and it is, thus, natural to remove all such rows. Because of the low number of observations and the large number of features, we would, however, lose too much data if we delete all rows with missing feature values.

That leaves us with replacing the missing values, in three different ways: (1) Replacing with specific integers (like 0), (2) Inserting a value based on a metric (like the mean or median), or (3) Inserting a value based on an algorithm [44].

Replacing with zero
When answering a survey, it might seem natural to skip questions that do not concern your situation. If you do not have any ventilation in your office, it would, therefore, seem natural to skip a question asking how many supply air devices are present in the office, rather than answering 0.

In situations where zero means that something is not present, it is therefore assumed that missing values can be replaced with zero. This, of course, only counts in situations where zero would be a likely answer. Replacing something like missing floor areas with zero would, for instance, make no sense, and would, therefore, create some extremely misleading results.

Replacing with the median
If you take the median of a sample and add that median value to the sample hundreds of times, the median of the sample will stay the same. Replacing missing values with the median, is, therefore, a way to minimize the influence of the missing value.

Compared to replacing missing values with the mean, the median is less influenced by extreme values, that otherwise could skew the overall result. The median is, thus, used to replace missing values in the features consisting of means of measured CO_2, temperature, relative humidity, sound level, illuminance, light colour, and VOC. Box plots and distributions of these features before and after they are replaced with the median, can be seen in the Appendix, Figure A.6. By replacing the missing values with the median, the missing values should have as little influence on the predictions as possible.

The sound level was only measured in about 50% of the office buildings, but because it was the most quantifiable measure of sound available, it was not removed as a feature. Replacing the missing values with the median here makes the observations with missing sound levels less predictive regarding sound. This is on purpose, as there would be a fairly large risk of creating misleading data patterns when making more extreme assumptions, based on such few observations.

When working with ordinal features, each discrete number will represent a class. It is, therefore, crucial that the missing value is replaced with one of these discrete numbers and not some continuous number, as this essentially would create a new class between the classes. The missing values are, therefore, replaced with the median. By doing so, the missing value only adds volume to the least extreme ordinal class.

Replacing missing values with the median is not a great solution in situations where the median might mean different things in relation to other features. A small office that is missing the window area value, would for instance get an unrealistically big window area if the median is inserted. Ultimately this could do more harm than good in the final model, and it is, thus, crucial to find a better solution to this type of missing value.

Replacing with mean based on KNN

In the case of relative numerical features such as window areas and office volumes, K nearest neighbour, KNN, has been used to replace the missing values. An example of how KNN works is seen in Chapter 2.3.1.

Here, KNN identifies 'K' observations in the dataset that are similar. These 'K' samples are then used to estimate the value of the missing data. The mean value of the 'K'-neighbors found in the data set is then inserted in the spot of the missing value.

In order to do this successfully, the other features in the data used to predict the missing value must have actual relevance to the prediction. In the case of predicting office volumes and window areas, the ceiling height, floor area and the number of occupant workstations were, thus, used to predict the missing values.

This means that only the mean of window areas in offices with a similar amount of workstations and floor areas are used. Going back to the case of a small office, this means that a mean window area that is appropriate to the rest of the office is used to replace the missing value. In the Appendix, Figure A.7, you can see distributions of window areas and office volumes before and after the missing values are replaced. Here it is seen that the observation count changes slightly, while the distribution stays almost completely the same.

4.3 Feature selection

Even after this initial cleaning, the data frame contains a high number of features. It is crucial to perform a deeper feature selection, as it can (1) decrease over-fitting, (2) improve accuracy, and (3) reduce training time. Due to the small number of observations, the training time is not the biggest issue for now. Decreasing over-fitting, however, is essential, as irrelevant features could influence the model in the wrong direction. As we get down to only the most relevant features, the accuracy of the model should be improved.

For supervised learning, two common feature selection methods are wrapper methods and filter methods [45]. The advantages and disadvantages of each method are described in Table 4.1.

Table 4.1: Filter and wrapper selection methods.

	Selection method	Advantages	Disadvantages
Filter selection	Filter methods select subsets of features based on their relationship with the target using statistical methods and feature importance.	Fast, less complex and less prone to overfitting than wrapper methods.	It only analyses the relationship between the individual features and the prediction
Wrapper selection	Searches for well-performing subsets of features based on specific machine learning algorithms.	Takes the features' relationship to each other into account.	Is prone to overfitting and has a high computational time for data sets with many features.

Due to the large number of features in the data set, the filter feature selection methods are used to reduce the number of features. After all redundant features are removed, the remaining features are used in the final model training stage. Here, features that have little influence on the prediction are removed to improve the model performance. You could, thus, say that both methods are used in the project. An overview of the feature selection process is seen in Figure 4.8.

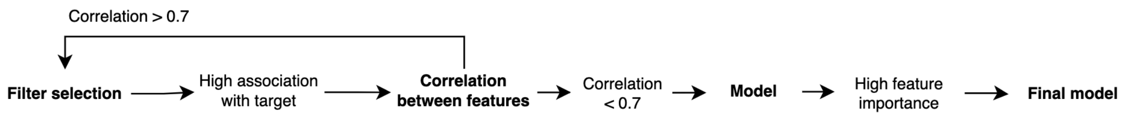


Figure 4.8: Visualization of feature selection.

4.3.1 Filter-based feature selection

For the filter-based feature selection, correlation-type statistics are used to assess the importance of each feature in relation to the target. The necessary statistical measure is dependent on the input and output. The input is the group of features we wish to use for prediction and will consist of many different feature types. Output variables, on the other hand, are the targets the model intends to predict. In our case, we remember that all our targets are binary and, hence, categorical with two possible classes:

- Satisfaction with indoor climate (dissatisfaction, satisfaction)
- Occupant performance (poor, ok-good)
- Office-related symptoms (no, yes)
- Single office-related symptoms (no, yes)

The input, however, is a combination of both numerical and categorical features. Figure 4.9 shows how this knowledge has been used to decide what correlation statistics to use [46].

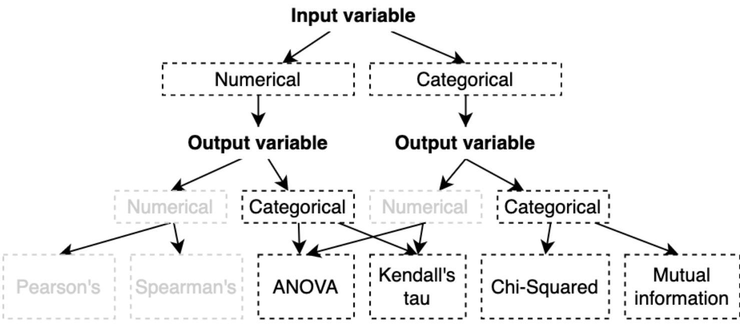


Figure 4.9: Correlation statistics. [46]

The different correlation-type statistics used, have different ways of assessing each feature's correlation with the target, and are described in Table 4.2.

Table 4.2

Method	ANOVA	Kendal's Tau	Chi-squared	Mutual information
Feature type	Numerical	Numerical and ordinal	Categorical	Categorical
Requirements	Feature must be continuous (interval or ratio)	Assumes the feature to be either continuous or ordinal, and that the relationship between the input andoutput feature is monotonic (move in the same direction).	Features must be categorical.	Works best for categorical feautres.
What it does	ANOVA (Analysis of Variance), checks if the means of two or more groups are significantly different from each other.	Measures strength and direction of association between two variables [47].	Measures association between two variables [48] [49].	Based in information theory, and uses entropy to quantify how much information there is in a random variable. The less entropy there is, the more alike the two variables are [50] [51].
Ranking measure	F-score	Tau-B	Chi^2	I(x; y)
Interpretation	The higher the F-value, the higher the variation between sample means is relative to the variation within the samples.	Tau-B ranges from -1 (inversely monotonic) to 1 (monotonic).	The higher the Chi-squared value is, the higher the association between the feature and the target.	Mutual information is given on a score from 0-1. The higher it is, the more information is shared between the two variables.
Passing criteria	H0: Means of all groups are equal. H1: At least one mean of the groups is different. If p value < 0.05, H0 is disproven.	If the p-value is less than or equal to 0.05, the result is statistically significant and we can trust that the difference is not due to chance alone.	H0: There is no relationship between categorical feature and target variable. H1: There is some relationship between categorical feature and the target variable. If p value < 0.05, H0 is disproven.	Only when I(x;y) = 0, the two variables completely independent.
Function documentation	[52]	[53]	[54]	[55]

The zero-hypothesises are described in the table to explain how the significance of the correlation can be proven. It is, however, mainly the ranking of how well each feature is correlated with the target that is used for the filter selection.

In Appendix B.1, the association results for each of the features can be seen, as well as whether it was decided to keep or drop the feature based on this. Here it is seen that there are features not significantly correlated with the target, that are kept in the models. In some situations, the feature might have a data structure, that performs poorly with the specific correlation-type statistic. It could also be the case that the feature does not have a direct correlation with the target, but that it in interaction with other features in the final model has a larger importance. Features that theoretically should have an influence on the target are, therefore, kept to see if they might have importance in the actual model.

There are also features that have high correlations with the target but are dropped because they have low performance in the model. You could, therefore, say that correlation-type statistics are used as an initial rough filter, to make it easier to focus on the more relevant features.

4.3.2 Removing features with high inter-correlation

Having input features that are highly correlated with each other, will bring very little additional information to the model. After the initial filter selection is performed, the correlation between the remaining features is, therefore, checked. If two features have an absolute correlation higher than 0.7, one is dropped. The decision of which one is dropped is decided based on (1) theoretical knowledge of indoor climate, (2) how well-correlated the feature is with the prediction, and (3) which feature has the fewest missing values.

To assess the correlation between each variable, I used the dython library. The dython library is used to identify different types of features, and based on this, use an appropriate correlation statistic or strength-of-association method [56]. Heat maps of the correlation between each of the features in the final models can be found in Appendix B.2.

An example of highly correlated features are the VOC and the CO_2 concentration, as well as the illuminance and the light colour. As described in the chapter, 2 Background, ventilating for one pollutant will reduce all pollutants, and the correlation of the two concentrations was, thus, to be expected. The light colour will usually have a higher temperature at higher illuminance, making this correlation equally predictable. Because the VOC concentration and the light colour were measured in less than 50% of the buildings, they were the obvious choices to drop.

4.4 Model training and evaluation

To create the best models, the performances of different machine learning algorithms are compared. The models are then optimized by tuning the best-performing algorithms. Finally, the features' influence on predicting the target is analysed. In other words, we assess how different indoor climates and design parameters influence the perceived indoor climate, and experience of building-related symptoms.

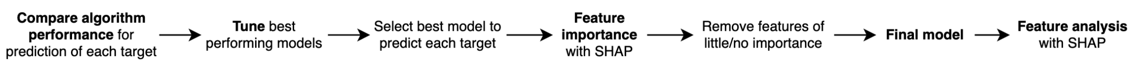


Figure 4.10: Model training and evaluation process.

4.4.1 Training, testing and cross-validation

When training the supervised machine learning models, the data is split into a training set and a testing set. The model is, thus, trained using 80% of the data and tested using the remaining 20% of the data.

However, if you just do a classic 80%-20% split, the result might be dependent on the specific split of data. To avoid this, a 5-fold cross-validation is used when comparing the different models. This means that the data is split in K = 5 folds. Five models will then be built and tested on each fold.

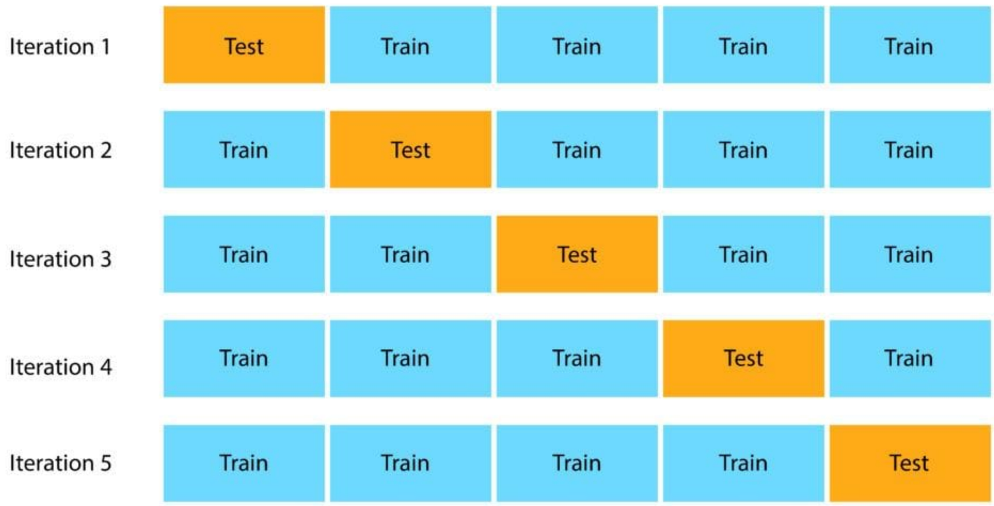


Figure 4.11: Illustration 5 fold cross-validation. [57]

By using cross-validation, we essentially are able to use all the data. This method is, thus, particularly useful when working with small data sets. Furthermore, by training models on each fold, we will get different performances. Through the variation of performance, it is possible to see if the model has a good performance in all folds, or if it is performing very well in some but poorly in others. Preferably, the model should do well in all folds, and box plots of the performance metrics are, thus, used to compare the machine learning algorithms.

4.4.2 Assessing model performance

The performance of a model can be assessed using many different metrics, and different metrics have different strengths and weaknesses. It is, thus, common to asses a model using multiple metrics. In this project, the accuracy, F1 score and AUC are used to assess the model performance. The performance according to each metric is interpreted as seen in Table 4.3.

Table 4.3: Interpretation of performance metrics. [58] [59] [60]

Interpretation	**Accuracy**	**F1 score**	**AUC**
Very good	>0.9	>0.9	>0.8
Good	0.7-0.9	0.8-0.9	0.7-0.8
Acceptable	0.6-0.7	0.5-0.8	0.51-0.7
Poor	<0.6	<0.5	$\leq$0.5

Accuracy
Accuracy is a simple and commonly used metric to assess model performance. It is defined as the number of test cases correctly classified divided by the total number of test cases.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- TP is number of true positives,
- TN is the number of true negatives,
- FN is the number of false negatives,
- FP is the number of false positives

Accuracy works best when assessing balanced data sets, meaning that there are equally many observations in each predicted class. The less balanced the data is, the more likely the accuracy is to make the model appear either better or worse than it is. [61]

F1 score
A way to better focus on the underrepresented class is the F1 score, which is a weighted metric of precision and recall. [62] [61]

$$F1 = 2\frac{Precision \times Recall}{Precision + Recall}$$

Precision identifies the correctness of classification, by finding the percentage of a predicted class that was classified correctly:

$$Precision = \frac{TP}{TP + FP}$$

The greater the fraction, the higher the precision, which means better ability of the model to correctly classify the positive class.

Recall shows the percentage of the of observations actually in the class, that were predicted to be in the class:

$$Recall = \frac{TP}{TP + FN}$$

A model with high recall is successful in predicting the class, but might also falsely predict other observations to be in that class. When combining the two metrics into the F1 score, a model will obtain a high score when both recall and precision are high, but be reduced if at least one is low. It should, thus, make up for the shortcomings of either model.

AUC
Another way to assess the quality of an imbalanced classifier is the AUC (the Area Under the ROC Curve). The ROC (Receiver Operating Characteristics) curve, shows the performance of a model at different thresholds, by plotting the TPR (true positive rate) along the y-axis and the FPR (false positive rate) along the x-axis, as illustrated in Figure 4.12. [63]

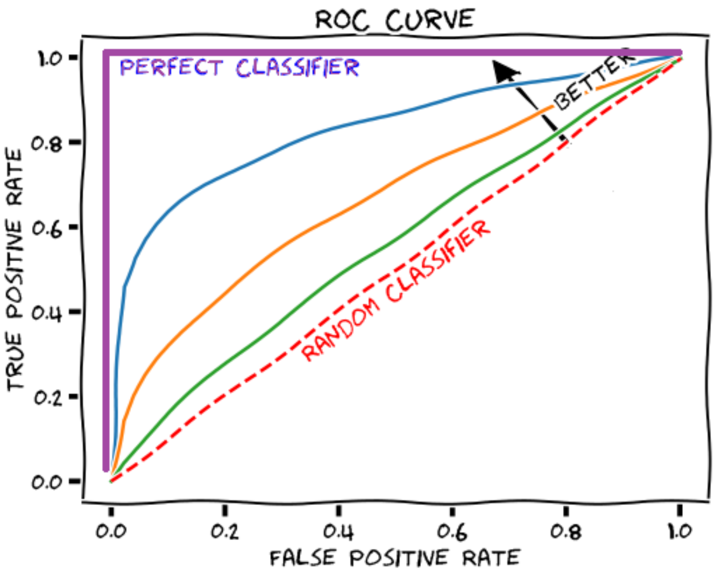


Figure 4.12: Illustration of ROC curve [58]

The true positive rate is the same as the recall for the specific threshold, while the false positive rate is defined:

$$FPR = \frac{FP}{FP + TN}$$

If the AUC is 0.5, the classifier is random, and the closer to 1 it gets, the better it is.

4.4.3 Comparing models

For each prediction, different machine learning algorithms are assessed using 5-fold cross-validation. The algorithms compared are a logistic regression (LR), a K-nearest-neighbour analysis (KNN), decision trees (DT), Gaussian Naïve Bayes models (GNB), support vector machine classifiers (SVM), extreme gradient boosting models (XGB), random forest classifiers (RF), and Artificial Neural networks (ANN). Table 4.4 shows how each algorithm is used as a classifier, as well as what their qualities are. [64] [65]

Table 4.4

ML alg.	Classification method	Advantages	Disadvantages
LR	Logistic regression models predict the probability of an event by fitting data to a logit function.	Tuning of hyperparameters is not needed.	Poor performance on nonlinear data.
KNN	The K-Nearest-Neighbour algorithm estimates how likely a data point is to be a member of one group or another.	Simple and has no assumptions about data structure.	It is slow. Scaling of data is necessary. It does not work well on imbalanced data and with missing values.
GNB	The Gausian Naïve Bayes classifier is based on Bayes' theorem and classifies every value as independent of any other value. It predicts a class using probability.	Quick, and good at multi-dimensional data.	Assumes features to be completely independent and equally relevant.
SVM	Support Vector Machines filter data into categories. The algorithm then works to build a model that assigns new values to each category	Performs well in higher dimensions	Slow and has a poor performance when classes are overlapping
DT	A decision tree uses a branching method to illustrate every possible outcome of a decision. Each node within the tree represents a test on a specific variable.	Normalization or scaling of data is not needed, and missing values or irrelevant features have no considerable impact.	Slow and prone to overfitting.
RF	Random forests is an ensemble learning method based on decision trees, combining multiple algorithms to generate better results.	Good performance on imbalanced datasets and little impact of outliers.	Appears as black box.
XGB	Extreme Gradiant Boosting algorithm is a treebased boosting algorithm.	No need for scaling or normalizing data. Quick and less prone to overfitting than decision trees.	Overfitting is possible if parameters are not tuned properly
ANN	Artificial neural networks are inspired by biological systems, and comprise of 'units' arranged in a series of layers, each of which connects to layers on either side.	Useful for modelling non-linear relationships in high-dimensional data.	Heavy, complicated and prefers normalised data. There should be 10 times as many observations as weights.

As seen in Table 4.4, artificial neural networks need 10 times as many observations as there are weights. This would only allow very few features, or a neural network so simple that only the activation function is used. Therefore, it was decided not to use artificial neural networks, and only test the performance of the other seven algorithms.

When deciding on what algorithms to use, the discussed performance metrics are used as well as our knowledge of the qualities of each algorithm from Table 4.4.

4.4.4 Model tuning

The chosen algorithms must be tuned to best avoid over-fitting and ultimately improve each model. In this project, the tuning is done through grid search, using the python package GridSearchCV, from the sklearn library. GridSearchCV uses 5-fold cross-validation to iterate through different combinations of hyper-parameters. [66]

Some hyper-parameters will make the algorithms less strict on the rules it is based on, while others might make the rules even stricter. Tuning them can, thus, make the model more or less likely to over-fit to the training data. Other hyper-parameters control what rules or functions to use in the algorithm. Tuning such a hyper-parameter might help you fit the structure of the data better. There are also hyper-parameters that help the specific algorithm deal with imbalanced classes, to avoid the the model only focusing on the over-represented class. Because hyper-parameters are specific to each algorithm, it is easier to explain their function in connection with how the algorithm works.

Looking at the performance of the models (See Table 5.1 and 5.2), it is seen that particularly the tree based models are doing well. Specifically the decision tree, the random forest and the extreme gradient boost model. Besides from these, both the Gaussian Naive Bayes algorithm and the support vector machine classifiers are performing well. These algorithms are, thus, described in the following to give an idea of how they have been tuned to improve the models.

Gaussian Naive Bayes models

The Gaussian Naive Bayes algorithm is based on the Bayes Theorem, which is used to calculate conditional probability. For continuous features, the algorithm assumes that the values associated with each class have Gaussian (normal) distributions. Based on this assumption, the likelihood of the feature having the observation, x, given the class of y, is calculated as:

$$L(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp(\frac{(x_i - \mu_y)^2}{2\sigma_y^2})$$

Here, the standard deviation, σ^2, and the mean, μ, are found based on the training observations belonging to the specific class of the target, y. x_i is the observation of the given feature [67]. For each feature, the likelihood of each target class is calculated as seen in Figure 4.14.

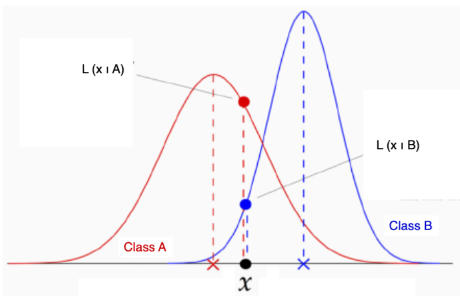


Figure 4.13: Illustration of Gaussian probability of each class based on a single feature. [67]

Here, x is the observation and the red and the blue dots are the probabilities of x given that they belong to each class.

When making a prediction using the Gaussian Naive Bayes classifier, an initial guess based on the probability of belonging to one class is made. This is called the prior and is based on the number of observations belonging to each class. If there is an equal amount of observations belonging to each class, this prior will give the probability of 0.5, while it will be far more skewed if the classes are imbalanced. If the prior is set to ”None” in the model, the model is not taking the prior probability into account.

Given two classes, A and B, for instance, the score of class A is calculated by:

$$Score_A = p(A) \cdot \Pi_{i=1} logL(feature_i = x|A)$$

where $p(A)$ is the prior and $L(feature_i = x|A)$ is the likelihood of the observation given class A. To avoid problems of underflow (when the likelihood is so low that the computer picks it up as zero) the logarithmic function is used.

Because each feature is seen as equally important, a thorough feature selection is important when using the Gaussian Naive Bayes model. If this is the case, the model is very durable, as it can be used on even very small training sets. Due to the simplicity of the model, it is not necessary to tune any hyper-parameters. A hyper-parameter that can be tuned is the variance smoothing. This parameter, adds a value to the variance in order to smooth the curve and can be a useful tool to improve the model. The default model setting of variance smoothing is 1e-9.

Further reading on the library of the Gaussian Naive Bayes algorithm used, can be found in the official documentation [68].

Support Vector Machine models

Support vector classifiers work by separating different target classes with a hyperplane in an n-dimensional space. This means that two-dimensional data, is split by a 1-dimensional support vector classifier line in a 2-dimensional space, 3-dimensional data is split by a 2-dimensional support vector classifier plane in a 3-dimensional space, and in general, data in n dimensions is split by the support vector classifier hyperplane in an n-dimensional space. An illustration of a support vector classifier is seen in Figure 4.14.

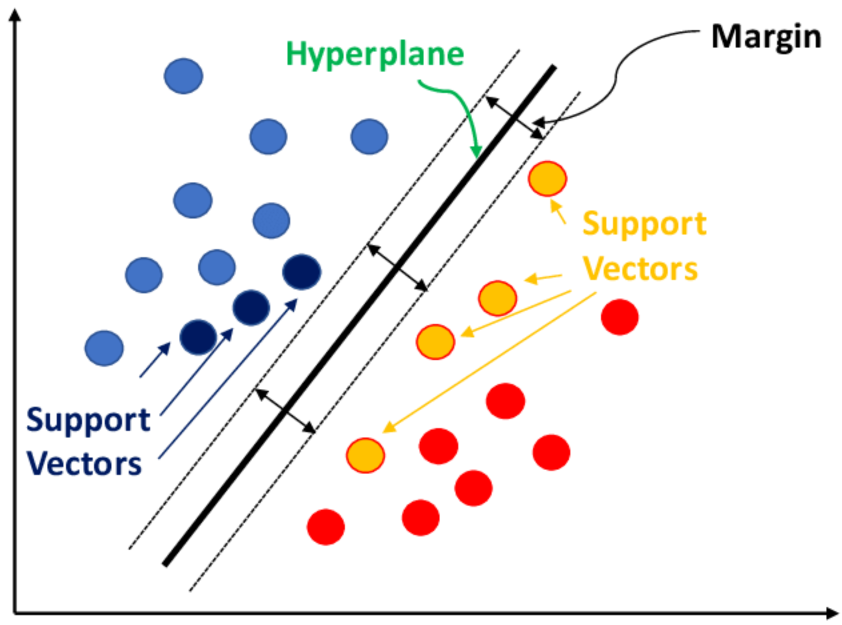


Figure 4.14: Illustration of support vector classifier [69].

By just splitting classes with a linear hyperplane, the chance of miss-classifying non-linear data is fairly high. To better work with non-linear data, a support vector machine is used. The support vector machine will use a kernel to transfer the data into a space that best separates the different classes with the support vector classifier. Here, linear, polynomial, radial basis function (RBF), and sigmoid kernels are the most popular. [70]

All kernels have the hyper-parameter, C, in common. The higher C is, the smaller the hyperplane margin will be. This means that the model will be fitted strictly to the training data, not allowing many observations to be misclassified. A lower C will widen the hyperplane margin and, thus, allow for more miss-classified observations in each group. The default C-value is 0. [71]

In this project, the best-performing kernel has been the sigmoid kernel. The hyperbolic sigmoid kernel uses the function:

$$K(x_i, x_j) = tanh(\gamma \vec{x_i} \cdot \vec{x_j} + r)$$

Where γ, and r are generalization parameters. The r shifts the hyperbolic tangent. High positive or negative values of r will make the impact of the scalar product smaller. The default value is 0. Gamma, γ, controls the influence of the training observation closest to the hyperplane. A high γ will, thus, only consider the observations closest to the hyperplane while a lower gamma will consider observations further away as well. The default setting in the model used:

$$\gamma = \frac{1}{n_{features}\sigma^2}$$

where σ^2 is the variance of the training observations.

Further reading on the library of the support vector classifier algorithm used can be found in the official documentation [72].

Decision tree models
Figure 4.15, illustrates how a decision tree is set up. At the base of every decision tree, is a root node, containing an initial question, forming the initial split. The split can result in branches or leaf nodes. They are branches if they are asking new questions and have arrows going away from them, and leaf nodes if they are not continued.

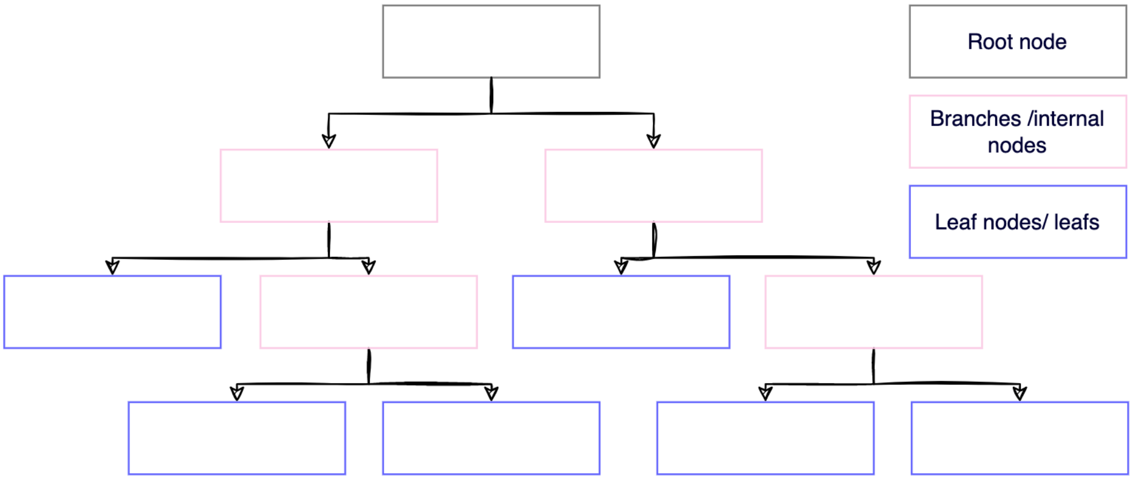


Figure 4.15: Illustration of a decision tree setup.

What question is better than the other, is decided based on impurity as well as the balance of the question. The balance of the question is important, as there will be very little information in a split with only a few observations on one side and all the other observations on the other. The purity is the degree to which the split only contains one target class. The higher the impurity, the more the target classes are mixed. We wish to minimize the impurity of each split in order to most efficiently predict what class the observation belongs to. The quality of a split from either the root or the K'th branch is, thus, found by using an impurity measure. In this project, the Gini impurity measure is used. [30]

$$Gini(v) = I(v) = 1 - \sum_{c=1}^{C} p(c|v)^2$$

Here, C is the number of target classes and v is the branch. The probability, $p(c|v)$, is the relative size of the classes in the given branch. To compare the impurity before and after the split, the purity gain,Δ, is calculated:

$$\Delta = I(r) - \sum_{k=1}^{K} \frac{N(v_k)}{N(r)} I(v_k)$$

Here, r is the root, v_k is the k'th branch, K is the total number of branches, and N is the number of observations in the given node. A high purity gain indicates that the classes have become more pure relative to the root impurity, $I(r)$.

A tree will continue growing branches until purity of the classes is found, and a leaf node can be made. These overly detailed trees, can, however, lead to over-fitting. To avoid this, the model should be tuned. This is often described as pruning when working with tree-based models. In this project, the decision tree is tuned by setting a maximum depth of the tree, setting a minimal amount of samples required to split an internal node and a minimal number of samples required to be at a leaf node. When working with imbalanced target classes, it can be useful to tune the class weight. If you don't use it, all classes have the weight of one. When the "balanced" setting is used, it uses the values of the target to automatically adjust weights inversely proportional to class frequencies in the input data.

Further reading on the library of the decision tree classifier used, can be found in the official documentation [73].

Random Forrest models

Instead of building one decision tree, the random forest model builds and combines multiple decision trees. Each individual tree in the random forest will result in a class prediction and the class with the most votes becomes our model's prediction. This makes the model slower, but also less likely to over-fit than the decision tree model.

The idea is that by building a forest of uncorrelated trees, the errors of one tree will be outweighed by the group. To ensure that the trees in the forest are not correlated the model uses bagging and feature randomness. Bagging means that the model splits the data into random subsets, resulting in unique trees. Feature randomness means that instead of building the decision tree with all the features at once, each tree is built using a random sub-group of features.

Because the model is built very similarly to the decision tree, the tuning hyper-parameters are almost the same. The main difference in the hyper-parameters tuned in this project is the class weight. Instead of just being able to set the class weight to ”balanced”, the random forest model allows it to be set to ”balanced subset”. The ”balanced subset” setting does the same as the ”balanced” setting, except that weights are computed based on the sample for every tree grown, rather than using the same weighting on all trees. Another hyper-parameter seen here is the ”n estimators”, which is the number of trees in the forest. The higher the number is, the slower the model is to run, and it is, thus, desirable to find the lowest number of trees while achieving the improvement in performance that comes from having many trees.

Further reading on the library of the random forest algorithm used, can be found in the official documentation [74].

Extreme Gradient Boost models

Another tree-based model is the extreme gradient boost model. Instead of bagging as is seen in the random forest model, this model uses boosting. This means that it boosts the observations that are hard to classify by selecting them more often. This is done by running multiple iterations of the tree model and updating the weighting of the observations along the way. The simplified process of the extreme gradient boost algorithm is shown in Figure 4.16.

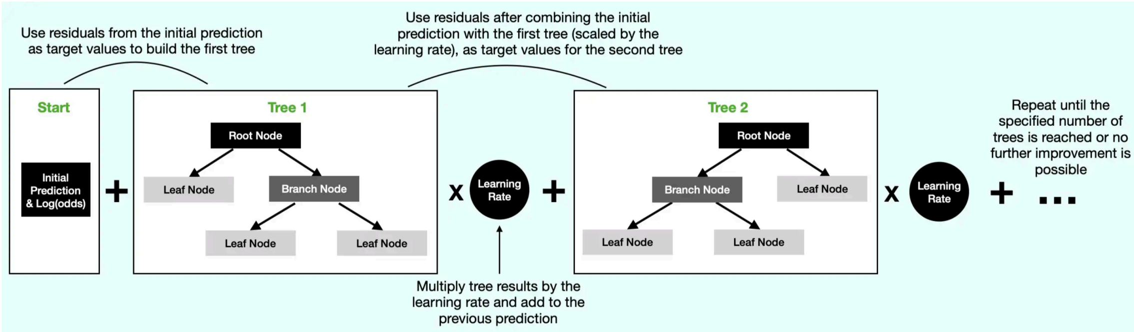


Figure 4.16: Iteration process of extreme gradient boost model. [75]

Unlike normal decision trees, the splits of the extreme gradient boost tree are based on the similarity score, which is based on the residuals from the previous prediction.

$$Similarity = I(v) = \frac{(\sum_{i=1}^{n} Residual_i)^2}{\sum_{i=1}^{n} p_i(1 - p_i + \lambda)}$$

The residual is the predicted values subtracted from the actual value. p is the previous probability, which is the probability of an event calculated at the previous step. The first tree is built using $p = 0.5$. Lambda, λ, is a regularization parameter, helping the tuning of the model. Increasing lambda will reduce the influence of leaves with few observations. The default value of lambda is $\lambda = 1$. Another important hyper-parameter is the gamma, γ, which allows you to prune nodes with minimal purity gain, Δ. The tree will, thus, not create new nodes unless $\Delta - \gamma > 0$.

Between each new tree, the result is scaled by a learning rate. The standard learning rate is 0.3. Higher learning rates will give each tree more influence on the final model, and reducing it can, thus, reduce the risk of over-fitting. The model will continue creating trees until the max number of trees is reached or the residuals are very small.

Besides the described hyper-parameters here, many of the hyper-parameters of classic decision trees can also be applied to tune the extreme gradient boost algorithm. In this project, tree depth, for instance, is used in the extreme gradient boost model. Further reading on the library of the extreme gradient boost algorithm used, can be found in the official documentation [76].

4.4.5 Assessment of features

The goal is to diagnose the indoor climate, and it is, thus, essential to be able to understand each predictor's influence on the output.

In this project SHAP (SHapley Additive exPlanations) is used to interpret the machine learning models. SHAP is an Explainable Machine Learning technique based on cooperative game theory. In SHAP, the "game" is reproducing the outcome of the model for a single observation, and the "players" are the features in the model. SHAP re-trains the model that is already created and tuned, but with different combinations of features. This way, SHAP is essentially reverse-engineering its way back to understanding the influence of each feature in relation to the other features and the target. [77]

SHAP is visualizing the influence feature values have on non-linear models, as if there was a linear relationship. This gives a good overview of how the input values will influence the output of the models, but it should be remembered that some of the models are more complex than that and that each feature value, in reality, will have an influence on the influence of other feature values.

4.4.6 Visualization of feature effect and importance

To present the influence of the features according to each model, SHAP summary plots are used. SHAP summary plots explain both feature importance and effect. See the example of a SHAP summary plot in Figure 4.17.

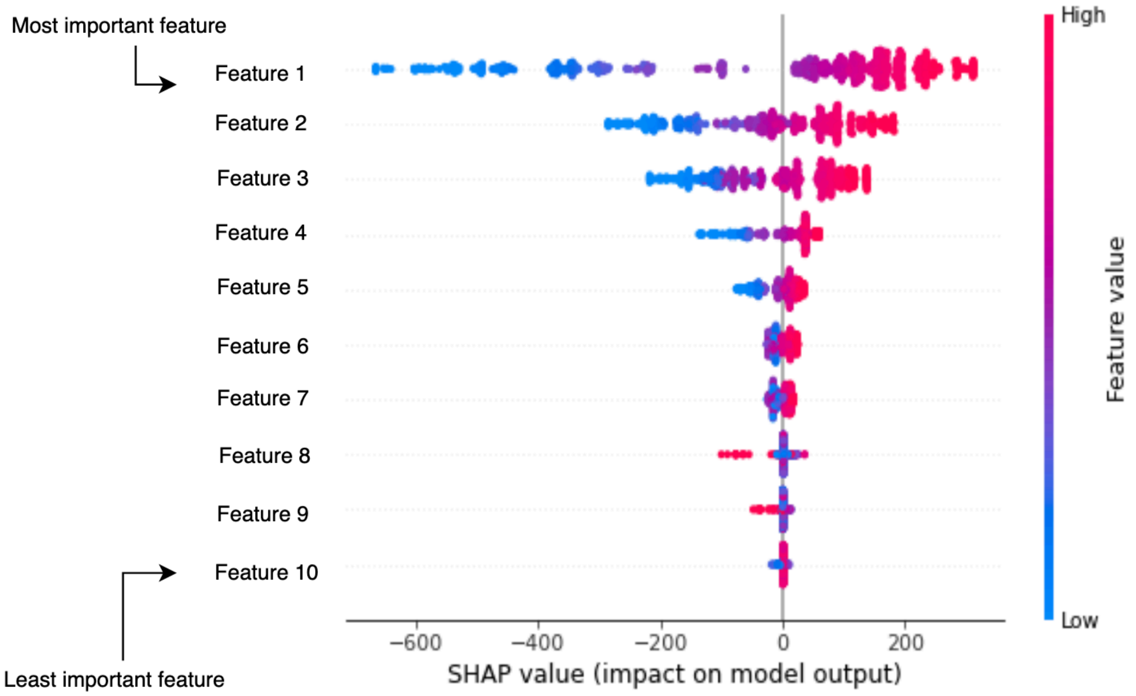


Figure 4.17: Example of SHAP summary plot.

In this plot, each point is a SHAP value for a feature and an observation. The negative SHAP values mean that the feature value influences the model towards class 0, while positive SHAP values are influencing the model towards class 1. The colour bar is used to present the feature values, going from low values to high values.

The features shown in the plot are ordered according to their importance. The feature importance is used to rank the importance of each feature's influence on the model's prediction. Besides better understanding each feature's influence on the result, the feature importance is used to cut out features with little or no importance.

4.4.7 Visualization of influence from single input

When using the models, a SHAP force plot is used to help diagnose the indoor climate parameters based on a single observation. A force plot visualises the influence of each feature value in a single prediction. It can, thus, be used to see what has the most positive influence on the outcome of the model and what has the most negative influence on the outcome of the model. An example of a force plot is seen in Figure 4.18.

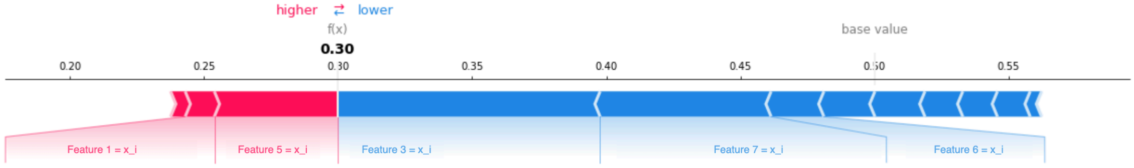


Figure 4.18: Example of SHAP force plot.

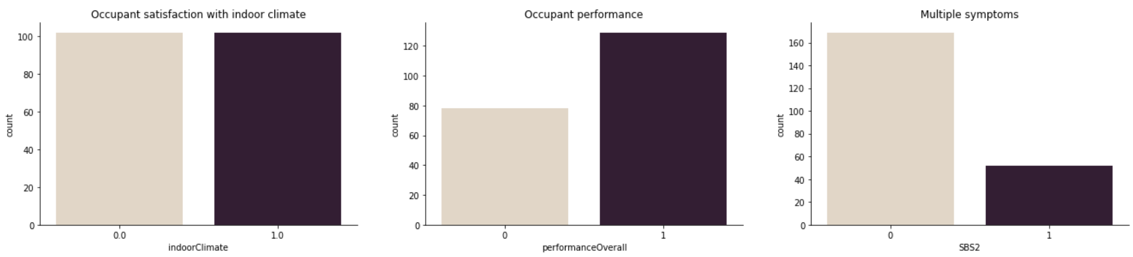
In the force plot, each feature value is a force that either increases or decreases the prediction. The prediction starts from the baseline, which is the average of all predictions. For completely balanced sets the base value will, therefore, be at 0.5, while it will be further towards 0 or 1 depending on what class the imbalance is leaning towards. If the forces push the prediction enough towards 1, one class is predicted, while the model predicts the other class it is pushed further towards 0.

If you combined the influence of each feature, for every row of data, you would end up with the summary plot. You could, therefore, say that the SHAP summary plot visualizes the results of force plots of all observations.

5 Results

5.1 Target class balance

The final models are predicting three main targets: Indoor climate, performance and the experience of office-related symptoms. The class balance of the targets is seen in Figure 5.1. In Figure 5.1a, 0 means occupant dissatisfaction, while 1 means occupant satisfaction with the indoor climate. In Figure 5.1b, 0 means that the occupant believes the office indoor climate has a negative influence on their performance, while 1 means they believe it has a neutral or positive influence on their performance. In Figure 5.1c, 1 means that the occupant has had at least two office-related symptoms in the last four weeks, while 0 means that they haven't.



(a) Indoor Climate (b) Performance (c) Office-related symptoms

Figure 5.1: Model targets.

In Figure 5.1a it is seen that the two classes are almost completely balanced. This means that there are as many people who find that their office indoor climate is poor as people finding that their office indoor climate is good.

In Figure 5.1b and 5.1c, it is seen that the target classes are far more imbalanced. Class 1 is over-represented in Figure 5.1b, while class 0 is over-represented in Figure 5.1c. The imbalance means that the models will be better at, and more likely to correctly predict the larger class, unless something is done to handle the imbalance. The model will be better at predicting the over-represented class because it will have been trained on a larger amount of observations belonging to the class. It will be more likely to correctly predict the class because there will be more testing data belonging to the over-represented class.

The "office-related symptoms" target is a little more complicated in its structure than the two other targets. It consists, not only of all the symptoms experienced but the symptoms experienced along with at least one other symptom. The symptoms included in the final target, are shown in Figure 5.2.

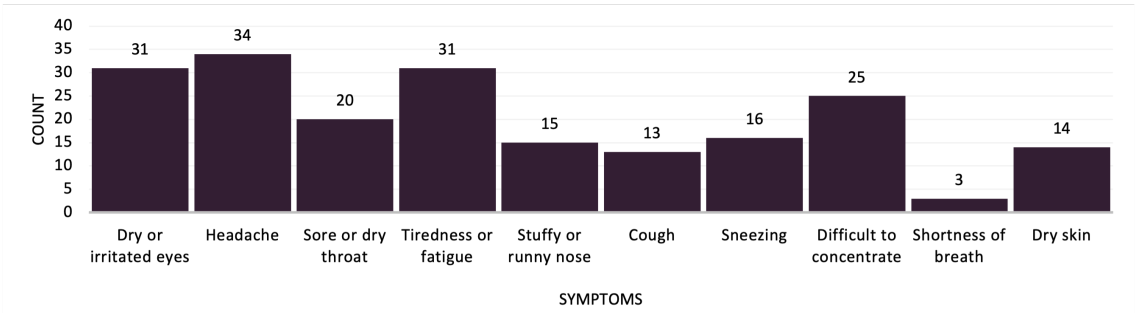


Figure 5.2: Count of each symptom used in the multiple symptoms target

In Figure 5.2 it is seen that the most common symptoms are (1) dry or irritated eyes, (2) headache, (3) tiredness and fatigue, and (4) difficulty concentrating. These symptoms will, thus, have a larger influence on the outcome of the final "multiple symptoms" model.

To better understand what parts of the indoor environment lead to specific symptoms, models are made for each of the most common targets. The target of each of these models will, thus, be the experience of that specific symptom. The target balance of these models can be seen in the Appendix, Figure C.1. As expected, they are fairly imbalanced, with the symptom class being the under-represented class.

5.2 Model selection

Table 5.1 shows the mean and standard deviation of three different performance metrics when models of the main targets are trained on seven different algorithms, using 5-fold cross-validation. The algorithms compared, are a logistic regression (LReg), a K-nearest-neighbour analysis (KNN), a Gaussian Naïve Bayes model (GNB), a support vector classification (SVC), a decision tree (DT), a random forest classifier (rFor), and an extreme gradient boosting model (xgB). The performance metrics are accuracy, f1 score and AUC.

In Table 5.1, the best-performing algorithms are written in bold. The decision of the best algorithm is based on the highest mean and smallest standard deviation (STD) of each of the performance metrics, as well as our knowledge about the tuning opportunities of each algorithm.

Table 5.1: Mean and (STD) of accuracy, f1 score and AUC in each fold.

		LReg	KNN	**GNB**	**SVC**	DT	**rFor**	**xgB**
Indoor Climate	Acc.	0.66 (0.05)	0.62 (0.05)	0.60 (0.07)	**0.62 (0.05)**	0.61 (0.06)	0.62 (0.11)	**0.65 (0.12)**
	f1	0.65 (0.04)	0.58 (0.04)	0.55 (0.01)	**0.63 (0.03)**	0.59 (0.07)	0.62 (0.10)	**0.62 (0.15)**
	AUC	0.70 (0.02)	0.66 (0.05)	0.66(0.02)	**0.69 (0.03)**	0.65 (0.07)	0.72 (0.13)	**0.68 (0.13)**
Performance	Acc.	0.69 (0.11)	0.67 (0.11)	0.63 (0.13)	**0.67 (0.10)**	0.65 (0.13)	**0.69 (0.14)**	0.65 (0.14)
	f1	0.75 (0.09)	0.73 (0.11)	0.68 (0.12)	**0.74 (0.08)**	0.68 (0.10)	**0.73 (0.13)**	0.70 (0.14)
	AUC	0.70 (0.12)	0.69 (0.12)	0.69 (0.11)	**0.69 (0.11)**	0.62 (0.10)	**0.75 (0.12)**	0.73 (0.12)
Office-related symptoms	Acc.	0.78 (0.07)	0.79 (0.03)	**0.72 (0.07)**	0.76 (0.09)	0.66 (0.08)	**0.71 (0.06)**	0.68 (0.06)
	f1	0.38 (0.18)	0.39 (0.10)	**0.30 (0.17)**	0.19 (0.20)	0.27 (0.10)	**0.26 (0.02)**	0.25 (0.10)
	AUC	0.75 (0.08)	0.66 (0.07)	**0.75 (0.10)**	0.73 (0.11)	0.54 (0.09)	**0.69 (0.08)**	0.64 (0.07)

As we know, the "office-related symptoms" target is more imbalanced than the other targets. All f1 scores are, therefore, fairly low here. The model performing best is the Gaussian Naive Bayes. As described, this is a fairly simple model that does not need many observations to train a sufficient model.

The performance of the models trained to predict different symptoms can be seen in Table 5.2.

Table 5.2: Mean and (STD) of accuracy, f1 score and AUC in each fold.

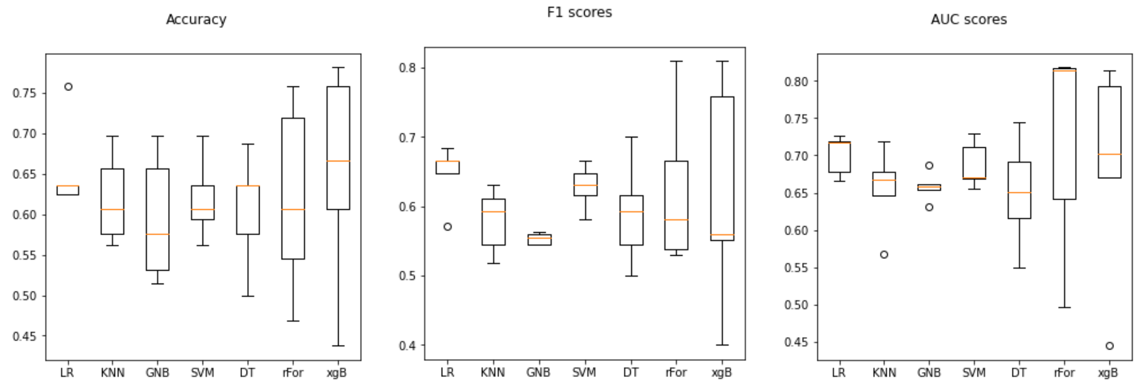
		LReg	KNN	**GNB**	SVM	**DT**	**rFor**	xgB
Symptom:	Acc.	0.85 (0.03)	0.85 (0.04)	0.80 (0.04)	0.85 (0.03)	0.75 (0.05)	**0.80 (0.03)**	0.80 (0.04)
Dry or	f1	- (-)	- (-)	0.16 (0.14)	- (-)	0.12 (0.16)	**0.18 (0.18)**	0.16 (0.15)
irritated eyes	AUC	0.61 (0.14)	0.56 (0.14)	0.60 (0.07)	0.57 (0.10)	0.53 (0.08)	**0.51 (0.15)**	0.54 (0.10)
Symptom:	Acc.	0.85 (0.06)	0.84 (0.06)	**0.73 (0.05)**	0.85 (0.06)	**0.74 (0.05)**	0.80 (0.08)	0.76 (0.08)
Headache	f1	- (-)	- (-)	**0.14 (0.08)**	- (-)	**0.11 (0.15)**	- (-)	- (-)
	AUC	0.57 (0.08)	0.53 (0.04)	**0.59 (0.12)**	0.61 (0.05)	**0.52 (0.09)**	0.58 (0.08)	0.48 (0.10)
Symptom:	Acc.	0.86 (0.05)	0.86 (0.05)	**0.78 (0.03)**	0.86 (0.05)	**0.77 (0.05)**	0.83 (0.05)	0.80 (0.07)
Tiredness	f1	- (-)	- (-)	**0.31 (0.11)**	- (-)	**0.35 (0.09)**	0.21 (0.19)	0.11 (0.16)
or fatigue	AUC	0.59 (0.09)	0.57 (0.09)	**0.68 (0.10)**	0.60 (0.19)	**0.58 (0.09)**	0.65 (0.14)	0.59 (0.13)
Symptom:	Acc.	0.84 (0.04)	0.81 (0.05)	**0.79 (0.047)**	0.84(0.03)	0.74 (0.09)	**0.82 (0.02)**	0.81 (0.03)
Difficult to	f1	0.11 (0.14)	0.29 (0.05)	**0.43 (0.13)**	- (-)	0.19 (0.18)	**0.14 (0.11)**	0.29 (0.09)
concentrate	AUC	0.83 (0.06)	0.75 (0.06)	**0.82 (0.09)**	0.79 (0.05)	0.54 (0.13)	**0.80 (0.07)**	0.79 (0.07)

In Table 5.1, it is again seen that the best-performing models generally are the simple Gaussian Naive Bayes models. Since the tree-based models are doing almost as well and we know that they can be tuned to better handle imbalanced data, these are also selected as the best models.

In the following, the performance of the models predicting different targets is described in more detail.

Indoor Climate

In Table 5.1, it was seen that the support vector machine and the extreme gradient boost algorithm had the best performance when predicting the occupant's dissatisfaction with the indoor climate. To better understand the performance of the models, Figure 5.3 shows box plots of the models' accuracy, f1 score and AUC in each fold of the cross-validation. The median is marked with orange.



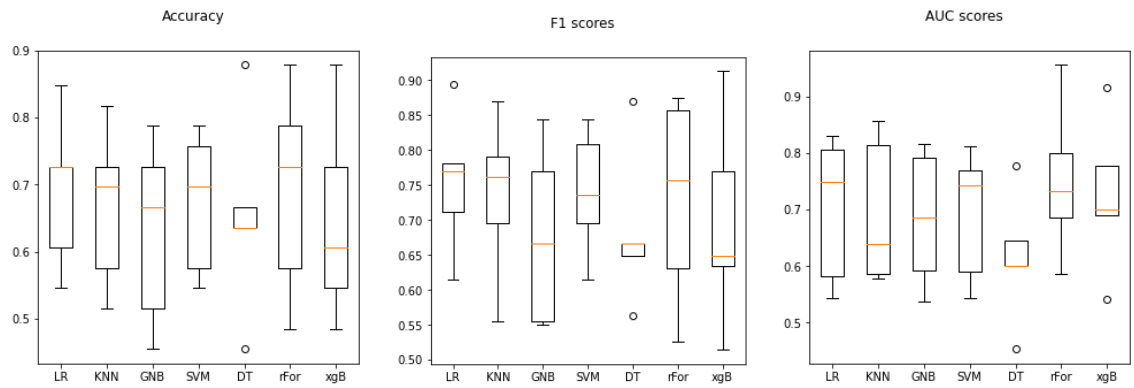
(a) Accuracy (b) f1 score (c) AUC

Figure 5.3: Target: Indoor Climate

Here, it is seen that the support vector machine performs best across all parameters while the extreme gradient boost performs really well regarding AUC. Both models should, thus, be tuned to find the best model.

Performance

In Table 5.1, it was seen that the random forest algorithm and the support vector classifier had the best performance when predicting the occupant's perception of the indoor climate's influence on their performance. To better understand the performance of the models, Figure 5.4 shows box plots of how the models perform in each fold.



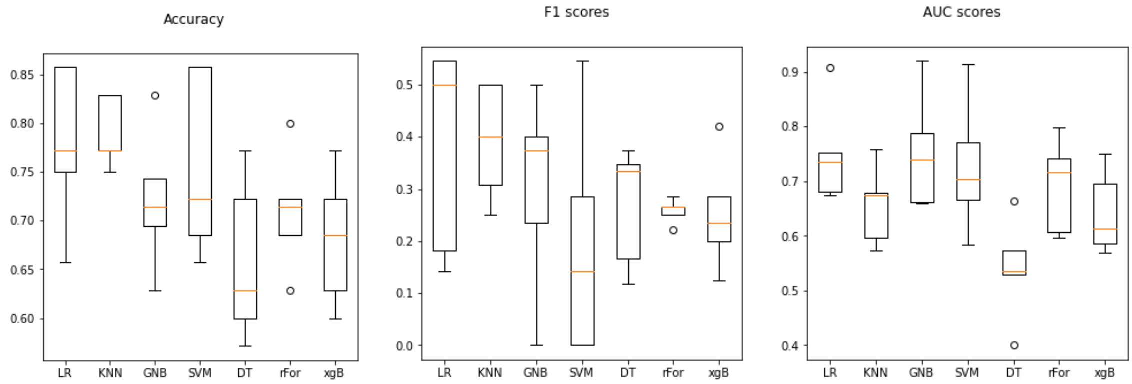
(a) Accuracy (b) f1 score (c) AUC

Figure 5.4: Target: Performance

Here it is seen the support vector does well on all parameters with low variation. The random forest model is, however, still tuned to see if better results can be found here.

Multiple symptoms

In Table 5.2, it was seen that the Gaussian Naive Bayes algorithm had the best performance when predicting the occurrence office-related symptoms. To better understand the performance of the models, Figure 5.5 shows box plots of how the models perform in each fold.



(a) Accuracy (b) f1 score (c) AUC

Figure 5.5: Target: Multiple symptoms

In the box plot in Figure 5.5b, it is seen that while the GNB model has a higher median, one fold has an f1 score of zero, which means that nothing is correctly predicted in some folds. The random forest model will, thus, also be tuned to see if a better result can be achieved.

The performance box plots of the single symptom models can be seen in the Appendix, Figure C.2, C.3, C.4 and C.5. Here it is seen that Gaussian Naive Bayes and tree based models have the best performance.

5.3 Tuning and final model selection

The tuning was done through 5-fold cross-validation in GridSearchCV. See Chapter 4.4.4 for further explanation of how this was done. All models are run using the random state seed 42 to ensure that the "random" result can be repeated, and the result tested.

Indoor Climate
The extreme gradient boost model and support vector machine model predicting the satisfaction with the indoor climate were tuned through the hyper-parameters seen in Table 5.3

Table 5.3: Hyperparameters in models predicting satisfactionwith indoor climate.

	random seed	max depth	gamma	lambda	learning rate	n estimators	kernel	Acc	f1	AUC
SVM	42	-	-	-	-	-	sigmoid	**0.71**	**0.71**	**0.71**
xgBoost	42	3	0.5	1	0.1	100	-	**0.63**	**0.63**	**0.63**

While both models achieve acceptable performance on all metrics, the support vector machine model is the best. No regularization parameters are chosen to tune the support vector model, as the the best result was found using a simple sigmoid kernel with its default hyper-parameters described in Chapter 4.4.4.

Performance
The tuning hyper-parameters and performance metrics for the prediction of occupant performance are seen in Table 5.4.

Table 5.4: Hyperparameters in models predicting occupant performance.

	random seed	max depth	min samples leaf	min samples split	max features	class weight	n estimators	kernel	Acc	f1	AUC
SVM	42	-	-	-	-	-	-	sigmoid	**0.69**	**0.79**	**0.61**
Random forest	42	6	1	3	None	None	100	-	**0.74**	**0.81**	**0.69**

The best random forest model is found when the class weight is set to $None$, despite the classes not being completely balanced. After tuning both models, the random forest has the best performance and is chosen as the final model.

Office-related symptoms
The Gaussian Naive Bayes model and the random forest model predicting the occurrence of office-related symptoms were tuned as is seen in Table 5.5

Table 5.5: Hyperparameters in models predicting experience of multiple symptoms.

	Random seed	max depth	min samples leaf	min samples split	class weight	Max features	n estimators	Prior	Variance smoothing	Acc	f1	AUC
GNB	42	-	-	-	-	-	-	None	1.23e-07	**0.84**	**0.59**	**0.74**
Random forest	42	4	2	5	balanced_subsample	12	100	-	-	**0.84**	**0.59**	**0.74**

Here, it is seen that the two models when rounded to two decimals, have the same performance. The two models are, thus, equally good at making the prediction, and the decision on what model to use must be based on what we want out of the model.

The GNB algorithm does well on small data sets but won't improve much when trained on more data. The random forest model has a good potential to get better if given more training data. The random forest model is, therefore, the more sustainable choice. The GNB model also assumes that all the data is normally distributed, and requires that the input is normalized. The random forest model, however, is non-linear and is really good at picking up on the relationships between the features. On top of that, it can even deal with missing values, which makes it a very robust model. Based on all this, the random forest model is superior and will be the one that is used.

Single symptom models
The tuning of the best-performing models, predicting each of the symptoms, is shown in Table 5.6.

Table 5.6: Hyperparameters in models predicting occupant performance.

Target	Model	random seed	max depth	min samples leaf	min samples split	class weight	n estimators	Prior	variance smoothing	Acc	f1	AUC
Dry or irritated eyes	Random forest	42	4	7	9	balanced	100	-		**0.80**	**0.47**	**0.71**
Headache	GNB	42	-	-	-	-	-	None	0.0008	**0.84**	**0.46**	**0.66**
	Decision tree	42	4	9	2	None	100	-	-	**0.71**	**0.52**	**0.78**
Tiredness or fatigue	GNB	42	-	-	-	-	-	None	5.34e-07	**0.82**	**0.50**	**0.72**
	Decision tree	42	5	10	2	balanced	100	-	-	**0.64**	**0.27**	**0.56**
Difficult to concentrate	GNB	42	-	-	-	-	-	None	5.34e-07	**0.80**	**0.47**	**0.67**
	Random forest	42	2	8	2	balanced_subsample	100	-	-	**0.84**	**0.63**	**0.78**

These models are all based on highly imbalanced target classes, which is reflected in their performance. The best-performing models are either the simple GNB model that only requires little training data to make a decent model or the tree-based models that are able to take the imbalanced targets into account. Here it is seen that tuning the tree-based models to work with the imbalanced data, in most cases makes them better than the GNB models.

5.3.1 Final models

Figure 5.6 shows the performance metrics of the final models used to predict each target. The interpretation of each performance metric was described in Chapter 4.4.2.

Target		Best model	Acc.	f1	AUC
Main models	Indoor climate	SVM	0.71	0.71	0.71
	Performance	Random Forrest	0.74	0.81	0.69
	Office-related symptoms	Random Forest	0.84	0.59	0.74
Symptoms	Dry or irritated eyes	Random forest	0.80	0.47	0.71
	Headache	Decision tree	0.71	0.52	0.78
	Tiredness or fatigue	Gaussian Naive Bayes	0.82	0.50	0.72
	Difficulty concentrating	Random Forest	0.84	0.63	0.78

Interpretation

Poor	Acceptable	Good	Very good

Figure 5.6: Performance of best and final models.

Notice that compared to the original tuned models, the accuracy, f1 score and AUC are now generally closer to each other. By tuning the models it has, thus, been possible to reduce over-fitting to the over-represented class, meaning that the f1 score generally has gone up while the accuracy has decreased.

Generally, it is seen that the main models predicting satisfaction with indoor climate, occupant performance and office-related symptoms are performing well on all parameters, although they are far from perfect. This is to be expected when working with such a small number of observations. It can, however, be assumed that as more data is gathered, the performance will increase.

The issue of few observations becomes more apparent regarding the most imbalanced target classes seen in the single symptom models. For some of these, the f1 score is still very low after the tuning of the models. It should be mentioned that an f1 score of 0.5, does not mean that the model is completely random, as it would be if the AUC was 0.5. Since the AUC is still decent, the low f1 score just means that the model is struggling to predict the under-represented class.

A low f1 score like this would be unacceptable if the model was used to predict credit card fraud or something else where you would rather be too sensitive to the fraud than not. In our case, we are more focused on the features' influence on the prediction than the output of the prediction. A low f1 score will mean that model is less likely to diagnose parts of the office indoor climate as being problematic, and, therefore, only diagnose parts that are very likely to cause the symptom. As long as the low-performing models are used with a grain of salt, they can, therefore, still be used to understand the features' influence on the experience of the building-related symptom.

5.4 Feature analysis

Throughout the feature selection, the number of features in the final models have been reduced. An overview features kept in the models can be seen in Figure 5.7. Here, the feature type - numerical, binary, ordinal or nominal - is represented by different colours. The features have been divided up into the categories: Office design, cleaning, exterior, occupant, HVAC, Light and noise.

Figure 5.7: Features in final models.

The final features used in each model are selected as described in Chapter 4.3. These features are the ones that are most influential in the prediction of the specific target. Looking at Figure 5.7, it is seen that there are fewer features in the imbalanced models. The fewer observations available, the higher the risk is of over-fitting the model when increasing the number of features. It was, therefore, necessary to be even pickier when selecting the most important features.

Notice that there are several different noise-related features. Because the "mean sound level" is a fairly weak feature, due to the high number of missing values, the other more specific noise-related features help explain the influence of noise in general.

5.4.1 SHAP feature interpretation

The influence on model output is interpreted through SHAP summary plots. The features shown in the SHAP summary plots are ordered by importance to the model output. This means that the features at the top of the plot have the highest importance and the features at the bottom of the plot have the lowest importance. The more negative the SHAP values are, the more the feature is pushing the output towards class 0, and the more positive the SHAP values are, the more the feature is pushing the output towards class 1.

The colour bar presents the feature values. For the nominal or binary features, the red dots represent observations with the value 1, while the blue dots represent observations with the value 0. For the ordinal and numerical features, the colour of the dots represents the observation values on a scale from low (blue) to high (red).

To help the interpretation of the colour bar with regard to each feature, the range of data represented in the models is shown in the following figures. In the ordinal, binary and nominal features, different points of the bar represent different categories. These are visualized with black lines, so it is possible to see exactly what colour represents what category. For the continuous features, the minimum and maximum values are marked, to make it easier to get an idea of what high or low values mean with respect to the model.

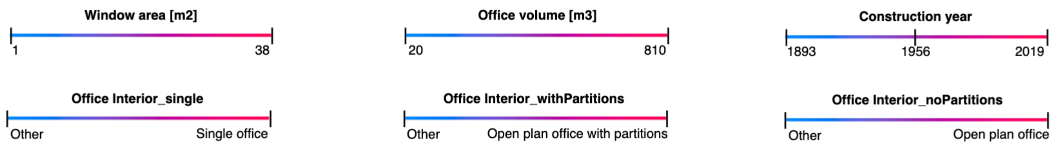


Figure 5.8: Office design features.

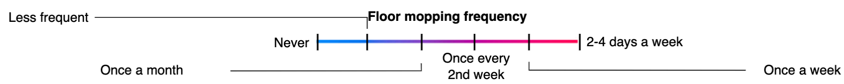


Figure 5.9: Cleaning features.

Notice that there are only six points on the colour bar of floor mopping frequency in Figure 5.9. This is because there were no observations in which the office was cleaned every day. The highest cleaning frequency is, thus, 2-4 days a week in the models.

Figure 5.10: Exterior features.

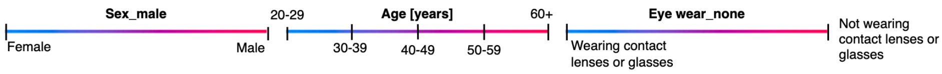


Figure 5.11: Occupant features.

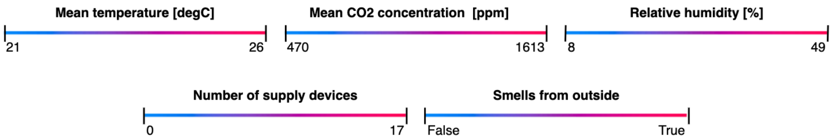


Figure 5.12: HVAC features.

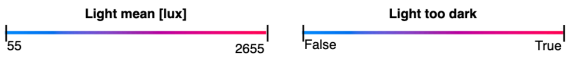


Figure 5.13: Light-related features.

The feature ”Light too dark”, is based on the occupant survey follow-up questions. If the occupant says they are unhappy with the lighting, due to lack of lighting, the feature value will be set to ”True”.

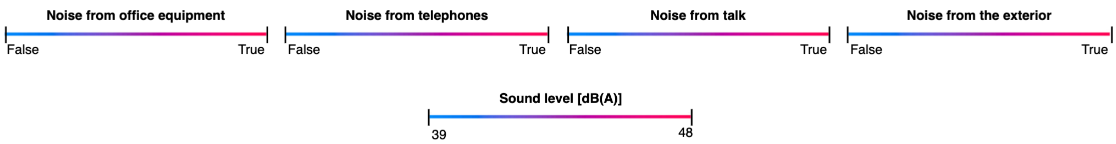


Figure 5.14: Noise-related features.

Indoor Climate
The SHAP summary plot of the model for satisfaction with the indoor climate is seen in Figure 5.15. The more positive the SHAP values are, the more likely the feature is to influence the model to predict that the occupant is satisfied with the indoor climate. The more negative the SHAP values are, the more likely the feature is to influence the model to predict that the occupant is dissatisfied with the indoor climate.

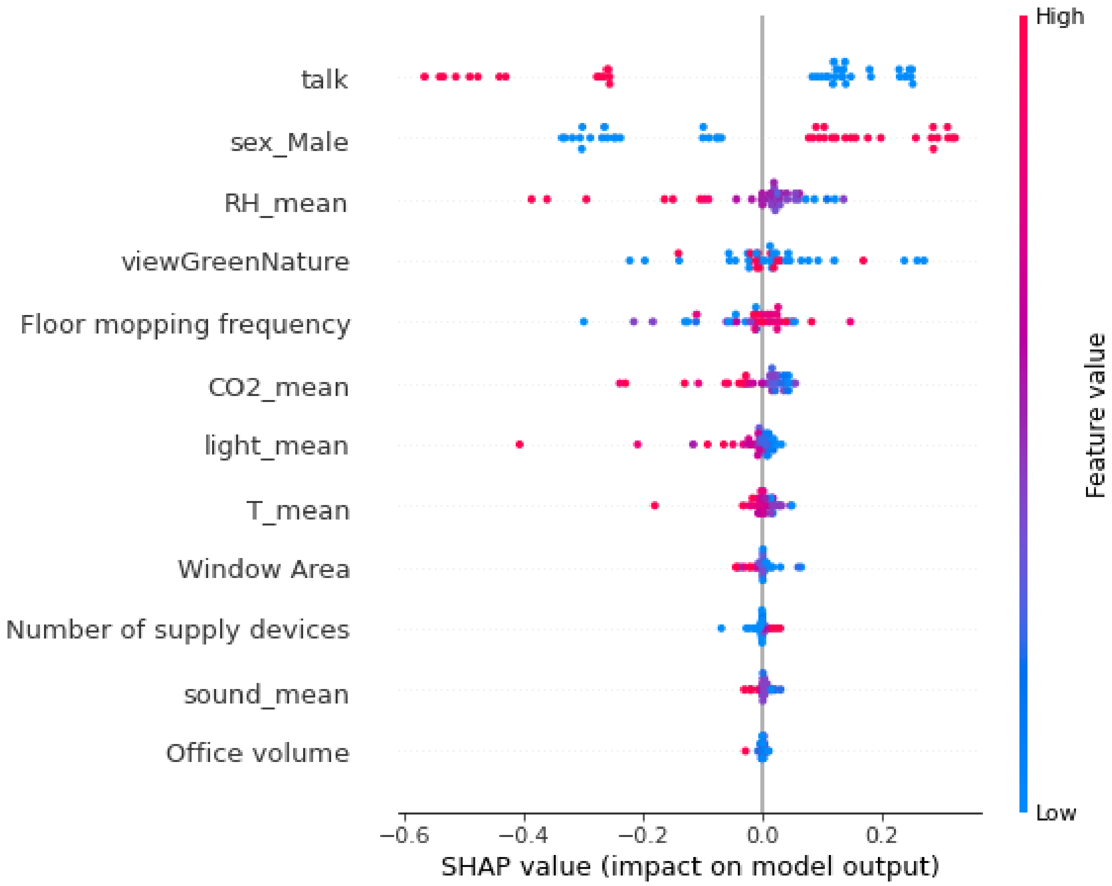


Figure 5.15: SHAP summary plot of indoor climate model.

Talk is the feature that influences the model most to predict dissatisfaction. Equally, a high mean sound (sound_mean) level skews the prediction in the same direction, although it isn't as influential.

Occupant sex is the second most important feature and it is seen that women are more likely to be dissatisfied with the indoor climate than men.

Regarding HVAC, it is seen that high mean relative humidity, high CO_2 concentration, high temperature and a low number of supply devices will cause the model to be more likely to predict dissatisfaction with the indoor climate.

It is seen that the highest mean illuminance and large window areas will cause dissatisfaction.

Regarding the feature representing the frequency of how often the floor is mopped, and thereby the general cleaning and maintenance habits of the office, it is seen that frequent mopping leads to satisfaction with the indoor climate.

Performance
The SHAP diagram of the prediction of the occupant's perceived performance in the office is seen in Figure 5.16.

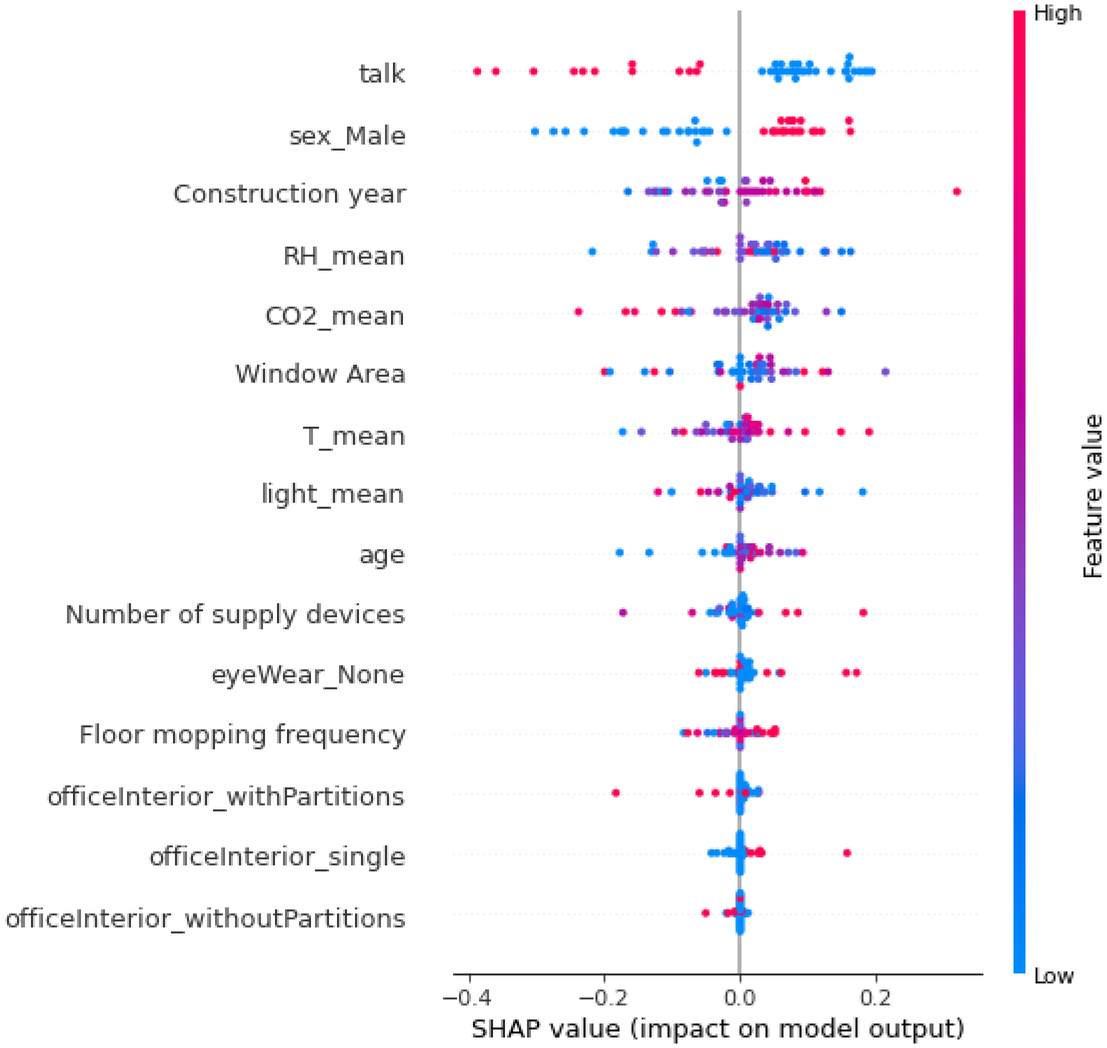


Figure 5.16: SHAP summary pot of performance.

For some features, such as sex, talk, and mean CO_2 concentration, the pattern is the same as was seen in the summary plot of the indoor climate model.

Regarding the prediction of the occupant performance it is, however, also seen that the construction year and the office interior, occupant age, and whether the occupant is wearing glasses or contact lenses has an influence (eyeWear).

Both low and high relative humidity influences performance negatively. Unlike in the indoor climate model, it is here seen that low temperatures have a poor influence on performance. Regarding the number of supply devices, it is seen that a high number of devices has a positive influence, while a medium number of devices actually has a negative influence.

Regarding office design, it is seen that newer office buildings have a more positive influence on performance. It is also seen that occupants in single offices have a much better performance than occupants in open space offices.

Office-related symptoms
The SHAP diagram of the prediction of the occupant's experience of office-related symptoms in the past four weeks is seen in Figure 5.17. Here, the positive SHAP values mean that feature values are influencing the model to predict the occurrence of office-related symptoms.

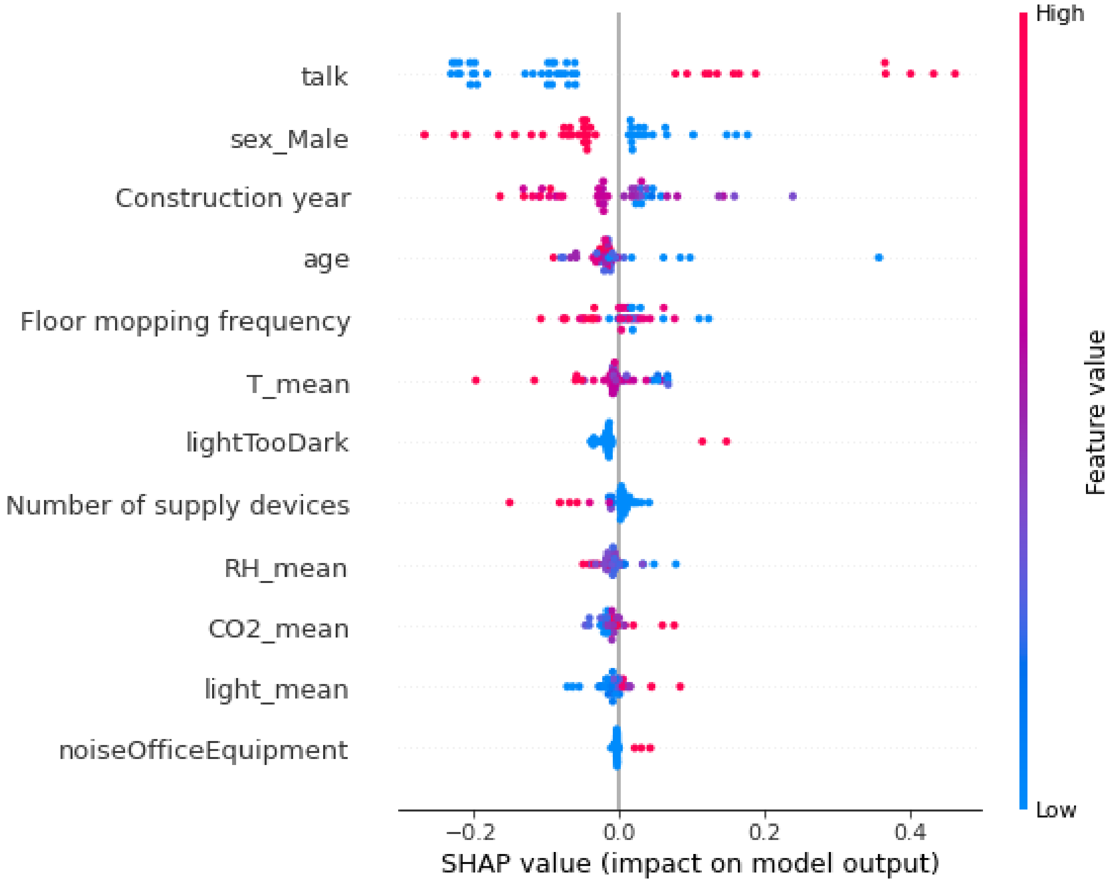


Figure 5.17: SHAP summary plot of the office-related symptoms model.

The model generally shows the same pattern as the two others models. The main difference is that it also takes background noise from office equipment and whether the light level is too dark into account.

Regarding lighting, it is similar to the other models in that high mean illuminance has a negative impact. In this model, it is, however, also seen that lack of lighting has an even larger negative impact.

It is seen that background noise from talk has a big influence on the occurrence of symptoms, while noise from office equipment has a smaller influence.

Compared to the other models, the HVAC-related features generally have less influence. Particularly the CO_2 concentration is seen to have much less influence than it has in the other models.

6 Development of tool to diagnose office environments

The final tool is an occupant-centric tool that diagnoses the indoor climate parameters and office setup, based on three models: One on the perceived overall indoor climate, one on the occupant performance, and one on the office-related health of occupants. Naturally, it would make sense to improve the indoor climate in a way that satisfies all targets. By having three different models, it does, however, make it possible to prioritise one problem area. If problems are seen regarding building-related health, the results of the single-symptom models can give a more detailed diagnosis of the specific symptoms.

6.0.1 When to use the tool

The models are based on offices in movement. The tool is, thus, meant to help make design improvements to existing offices by diagnosing problem areas.

Because the tool is taking occupant characteristics into account, it also allows you to design the best possible indoor environment for your specific employees. For instance, we see that women tend to be far more influenced by a poor indoor climate than men. This could be because women are more sensitive and/or because the current offices are not designed for them. If there are more women in an office space, it would, however, be possible to use the tool to see how changes to the environment might help. This, of course, does not have to be a full renovation. The solution could be something as simple as finding the optimal temperature set point.

While the tool is not designed for it, it could also be used when designing new buildings. They will naturally not have the necessary input to make useful diagnoses. They will, however, be able to use the tool to see how each aspect of the office should play together in order to optimize occupant well-being, performance and health in the final office. The tool could, thus, be used to better connect the design phase to the use phase.

6.0.2 Target audience

The tool is designed to apply to building consultants and company owners. This, however, does not mean that the tool cannot be beneficial to other parts of the building industry.

The company owner can benefit from the tool if they have been receiving complaints about the office environment, and want to incentivize their employees to work from the office rather than working from home. It can also help them to improve employee well-being, performance, and health. The tool doesn't have to be the first step in a larger renovation. It can also be used to point out problem areas in the office that might be easy to solve.

A building consultant could use the tool to make holistic design choices that optimize the benefit to the indoor environment without having to tick off each building standard. By combining the building consultant's knowledge of the expense of a design choice with the tool's calculated possible occupant benefit of an improvement to the office, it will make it far easier to make well-documented decisions. Later they can even use the tool to document the benefits of the specific design choices when pitching the renovation project - making it a strong sales tool as well.

6.0.3 A simple tool to use

Through the literature review, we saw how other tools diagnosed or rated the indoor climate. Many of these had the goal to make the tool as simple as possible. One tool was, therefore, only using occupant perception of the indoor climate, rather than doing physical measurements. Since that method only requires surveys, it makes each observation far cheaper and easier to gather, allowing a larger quantity of data from each building. When using the data from one building to strengthen the overall model, it does, however, make it problematic if there are no physical measurements to tie the occupant experience to.

In this data-driven tool, the necessary parameters are a mix of each. There are measurements, but they are not as specific as the ones we saw in some of the controlled experiments. Only one measurement device is needed for the office (depending on how large it is). This device makes measurements across an entire week without interference, and will thereby not complicate the natural flow and action of the office.

The survey method is also used. For instance, people who are dissatisfied with parts of the indoor climate are asked further follow-up questions to better understand the problem area. This is how we get features such as whether it is too dark or if there are specific noise disturbances in the office. This sort of feature allows us to make a more balanced diagnosis of the indoor climate, that you couldn't get from just looking at a design standard.

One could, thus, say that the tool is combining the successful parts of different tools to make a tool that is simple while also gathering building information that improves the tool in the long term.

6.0.4 How to use the models

To use the tool, the following must be done:

- Building, management, and office checklists should be filled out by the building owner/manager.
- Measurements of CO_2 concentration, temperature, relative humidity, illuminance and sound level must be performed in each office.
- Surveys must be filled out by occupants in each assessed office.

The building owner/manager fills out checklists that are similar to the ones used for the development of the tool. The reason full checklists and surveys must be filled out, despite having reduced the predictive features in the final models, is that this allows us to adapt the model to include more predictive features about the office design as more data is gathered. It is, thus, an essential part of improving the tool in the long term.

It is important to make sure that the measurements are representative of the occupants' experience in the office. The occupant survey should, therefore, be sent out and answered by occupants during the week that the physical measurements are performed. This does mean that the assessment of the office only reflects the indoor climate during that specific week. It can, therefore, be beneficial to do it multiple times across a year. Doing it once in the summer and once in the winter, would reflect the most extreme weather conditions.

This input will allow the models to predict whether there should be problems regarding overall satisfaction with the indoor climate, performance, or office-related symptoms. If there are issues regarding any of the models, the user will be able to see what parts of their office indoor climate are skewing the models in an unwanted direction. Even if the models are giving good results according to each model, it will still be possible to see what parts of the indoor climate are doing better than others.

To display this, the SHAP force plots, described in Chapter 4.4.5 are used in the output. They are used because they make it possible to show what feature values have the most significant influence on the model predictions. The plots are, thus, a visual way of showing what features should be changed to improve the indoor climate. With this diagnosis of the office's indoor climate, it is possible to make design recommendations. This is further visualized and explained in the use case seen in Chapter 6.2.

6.1 General diagnosis

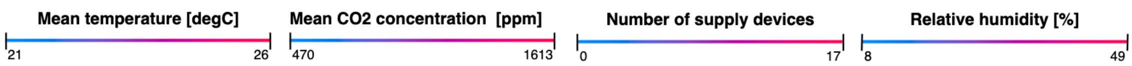
The following shows some of the general indoor climate problems that can be extracted from the patterns of the models, and how they could be resolved through design improvements.

6.1.1 Improvements to HVAC

The HVAC system of the office building will influence the feature's mean temperature, the mean CO_2 concentration, the number of supply devices, and the mean relative humidity. Table 6.1 sums up how different ranges of values of these features influence each model according to what was seen in the SHAP summary plots. Regarding CO_2 concentration, it was, for instance, seen that high concentrations lead to the prediction of (1) dissatisfaction with indoor climate, (2) poor performance, and (3) office-related symptoms, while a medium to low mean concentration did the opposite.

Table 6.1

	Indoor climate		**Performance**		**Office-related symptoms**	
	Dissatisfaction	Satisfaction	Poor	Ok-good	None	Symptoms
CO2 concentration	high	medium - low	high	medium - low	low - medium	high
Temperature	high	medium - low	low, high	medium, high	high	low
RH	high	medium - low	high , low	medium, low	medium - high	low
Supply devices	low	high	medium	high	high	low



A high mean CO_2 concentration causes dissatisfaction according to all models. Regarding relative humidity, it is seen that both high and low values have a poor influence on mainly indoor climate and performance. When looking into the single symptom model of "dry or irritated eyes" (Figure C.6) it is seen that the symptom is influenced by a low relative humidity and a low CO_2 concentration.

In conclusion, it is crucial to ventilate sufficiently according to the occupancy of the office, and thereby keep the CO_2 at a decent concentration. Over-dimensioning the ventilation system, can, however, lead to the occupants experiencing symptoms such as dry or irritated eyes. A design solution could be to split the office into smaller spaces, allowing the ventilation system to be targeted at a smaller group of people.

Low temperatures lead to satisfaction with the indoor climate, but to poor performance and office-related symptoms. Generally, Danish and Greenlandic buildings will have heating systems in place to increase the temperature if needed. This is seen as the feature specifying the heating type, didn't have any predictive importance in any of the models. It is, hence, not the type of heating system that is the limiting factor. Problems regarding low temperatures could have more to do with the cost of heating. This could be helped through an energy renovation of the building envelope. As the energy loss of the building and costs are not included in the models, this is, however, hard to recommend.

Another explanation is that if an office does not have a ventilation system, and the windows have to be opened to ventilate the space, opening such windows in the winter, can decrease the temperature and create problems. Therefore, solving problems regarding ventilation, could in some situations also solve problems regarding temperature.

6.1.2 Noise and office spaces

Due to the relatively weak mean sound feature, the general sound level of the office is not telling us as much as the noise from specific sources. While a high mean sound level does have a slightly negative influence on the perception of the indoor climate, it does not have a large influence on the prediction of the two other targets. It is, however, seen that background noise in the form of talk has a negative influence on all models while noises from office equipment only influence office-related symptoms.

Table 6.2

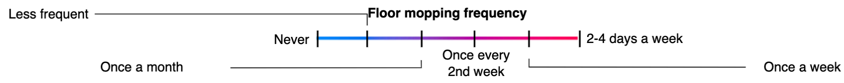
	Indoor climate		Performance		Office-related symptoms	
	Dissatisfaction	Satisfaction	Poor	Ok-good	None	Symptoms
Sound level	high	low	-	-	-	-
Talk	yes	none	yes	none	none	yes
Equipment	-	-	-	-	none	yes
Interior	-	-	Open plan with and without partitions	Single office	-	-

While the overall sound level can be brought down through acoustic solutions, it seems to be specific noise sources - particularly from people talking - that have the most negative influence. Furthermore, it is seen that the occupants have a better performance in single offices than in open space offices. Therefore, it is clear that if distractions from colleagues and office equipment can be reduced, the general office environment would be improved.

Open space offices do, however, have other qualities such as fitting more people and supporting collaborative co-working strategies. A solution could, thus, be to either divide a larger open space office into smaller offices with fewer people or as a minimum implement partition walls. Here, implementing acoustic solutions would further limit the noise between each of the office zones.

6.1.3 Cleaning

To get an idea of the cleaning and maintenance patterns of the office, the feature of floor mopping frequency is used. Here it is seen that high frequencies have a positive influence on the occupants' perception of the indoor climate as well as the chance of not having building-related symptoms, while low and medium frequencies skew the models in the opposite direction. It is also seen that the low cleaning frequencies have a negative influence on occupant performance. In conclusion, low cleaning frequencies are likely to influence all the models negatively.



The mopping frequencies go from ”2-4 days a week”, ”once a week”, ”every other week”, ”once a month”, ”less frequently” to ”never”. To perform well regarding cleaning according to all models, the office should at least be cleaned once a week. The higher the frequency, the further the models are skewed in a positive direction.

6.1.4 Lighting

High mean illuminance has a negative influence on the perception of the indoor climate and the experience of symptoms, while both low and high illuminances have a negative influence on performance. High light levels can cause annoying reflections on computer screens and force the occupant to squint their eyes. Lighting can change a lot throughout a workday, and the mean is, thus, only telling us anything if there is a continuous extreme situation. Therefore, if your office performs poorly because of a high mean illuminance, it would seem obvious to review the solar shading and artificial lighting.

Regarding the experience of building-related symptoms, it is seen that the feature describing the lack of lighting also has a high impact. The lack of daylight during the winter in nordic countries makes it extra important to have sufficient artificial lighting for both health and well-being. If there are complaints regarding the office being too dark, it could, hence, be beneficial to look into better solutions for artificial lighting.

6.2 Use case

A use case is set up to explain how to interpret the model results based on input from an office. The input used here, is an example of some of the office information that the models have been tested on. The input is seen in Figure 6.1.

	Occupant specific features	Occupant 1	Occupant 2	Occupant 3
Occupant characteristics	Sex	Female	Male	Female
	Age	20-29	30-39	30-39
	Eye wear	Contacts or glasses	None	None
Noise	Talk	True	False	True
	Office equipment	False	False	True
Visual	Light too dark	False	False	False
	View green nature	True	True	True

	Office-specific features	
IEQ measurements	Mean CO2 concentration	1613 ppm
	Mean relative humidity	29 %
	Mean temperature	24 degC
	Mean illuminance	841 lux
	Mean sound level	42 dB(A)
Building/office-related paramters	Office Interior	Shared with partitions
	Window area	17.3 m2
	Office volume	242m3
	Construction year	1985
	Number of supply devices	None
Cleaning	Floor mopping frequency	Never

Figure 6.1: Input parameters

That there is only input from three occupants, does not necessarily mean that there were only three people present in the office. All it means is that only three people filled out the survey.

The perception of indoor climate, performance, and experience of symptoms in relation to the office, predicted for each occupant, can be seen in Figure 6.2.

	Performance models	Occupant 1	Occupant 2	Occupant 3
Main models	Indoor climate	Dissatisfied	Satisfied	Dissatisfied
	Performance	Poor	ok-good	Poor
	Office-related symptoms	Yes	No	Yes

Figure 6.2: Indoor climate performance according to model output

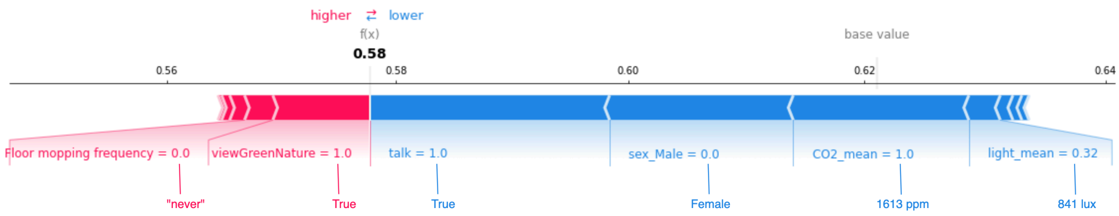
6.2.1 The occupants

In Figure 6.2, it is seen that two out of three occupants are poorly influenced by the indoor climate, according to all models. The differences in model output for each occupant, despite the office characteristics being the same, have to do with the occupant-specific features.

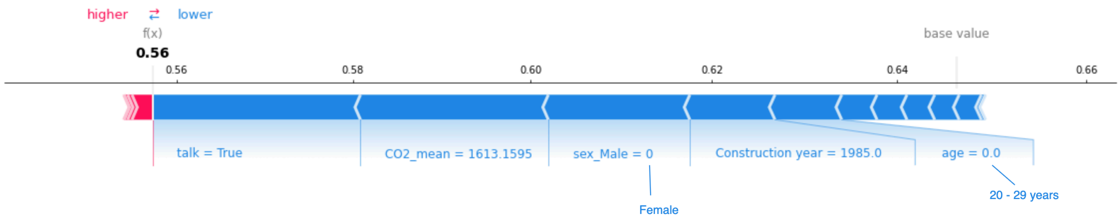
If problematic parts of the indoor climate are improved to suit the person that is most negatively influenced by the office indoor climate, it can be assumed that it will also influence the other occupants positively. Occupant 1 is female and between 20 and 29 years old, which are both occupant characteristics that will make the occupant more likely to be negatively influenced by a poor indoor climate. Therefore, the focus will be on the input from Occupant 1 in the following.

6.2.2 Problematic aspects of the office indoor climate

To simplify the process of diagnosing the indoor climate, we will be focusing on the model of satisfaction with the indoor climate and occupant performance. In Figure 6.3 two force plots based on the input for occupant 1, are seen. In these plots, the most influential feature values are highlighted. The red feature values are influencing the models to predict satisfaction with the indoor climate in Figure 6.3a and to increase performance in the model in Figure 6.3b, while the blue feature values are doing the opposite. In order to change the predicted class, more features need to be on the red side.



(a) Indoor climate model. Prediction: Dissatisfaction with indoor climate.



(b) Performance model. Prediction: Poor performance.

Figure 6.3: Force plots of occupant 1 input.

Notice that the indoor climate model is using normalized values and that a mean CO_2 concentration of 1, therefore, means the highest possible mean CO_2 concentration the model has been trained on, which is the same as the value seen in the force plot for performance. This can be hard to interpret straight from the plot, and I have therefore added annototations to explain what some of the numerical feature values mean.

The input values that the models pick up as being problematic, paint a clear pattern that can be divided into two problem areas:

1. Ventilation
2. Distractions

The high mean CO_2 concentration, as well as the lack of supply devices, suggests that there is no mechanical ventilation system in place. Implementing this and, thus, reducing the CO_2 concentration, would therefore improve the overall performance of the office.

The mean sound level is not particularly high, but two of the occupants are influenced by background noise from speech or office equipment. The office is also a relatively large shared office ($242m^3$). This could be a sign that it is more about the distractions in the office than the actual sound level. Since background noise from talk also is the most problematic feature according to both plots, this needs to be dealt with.

6.2.3 Testing recommendations

With these problem areas in mind, the models are used to iteratively see what changes are needed to achieve good results in all models. The necessary input values can be seen in Figure 6.4. Here, the feature values that had to be changed are blue.

	Occupant specific features	Occupant 1	Occupant 2	Occupant 3
Occupant characteristics	Sex	Female	Male	Female
	Age	20-29	30-39	30-39
	Eye wear	Contacts or glasses	None	None
Noise	Talk	**False**	False	**False**
	Office equipment	False	False	**False**
Visual	Light too dark	False	False	False
	View green nature	True	True	True

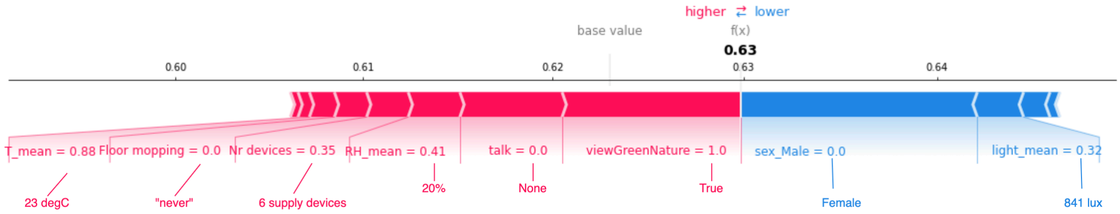
	Office-specific features	
IEQ measurements	Mean CO2 concentration	**750 ppm**
	Mean relative humidity	**20 %**
	Mean temperature	**23 degC**
	Mean illuminance	841 lux
	Mean sound level	42 dB(A)
Building/office-related paramters	Office Interior	Shared with partitions
	Window area	17.3 m2
	Office volume	242 m3
	Construction year	1985
	Number of supply devices	**6**
Cleaning	Floor mopping frequency	Never

Figure 6.4: Input with improvements in bold.

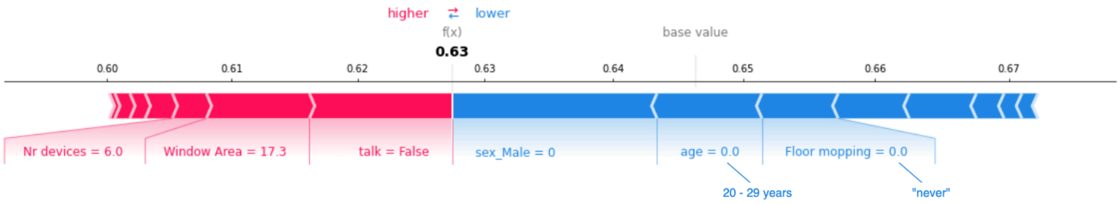
To satisfy all models it was necessary to drastically decrease the CO_2 concentration, slightly lower the temperature, lower the relative humidity, and increase the number of supply devices. Such changes to the IEQ parameters could be achieved through the implementation of mechanical ventilation.

It is also assumed that a solution to avoid background noise from talk and office equipment is found. This doesn't have to be a design idea. It could be something as simple as investing in noise-cancelling headphones for the occupants.

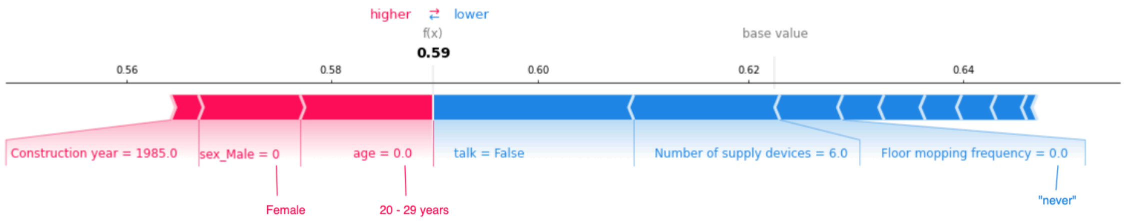
In Figure 6.5, the force plots show the influence of the new input on each model. Notice that in the model for office-related symptoms, we wish the model's prediction to go towards 0, as this makes the model more likely to predict that the occupant is *not* experiencing symptoms.



(a) Indoor climate model. Prediction: satisfaction



(b) Performance model. Prediction: ok-good performance



(c) Office-related symptoms model. Prediction: No symptoms.

Figure 6.5: Force plots based on occupant 1, after improvements are made.

As is seen in the force plots of each model after the improvements are implemented, some of the most problematic features' ”force” have now changed colour, and are, therefore, seen as positive aspects of the office rather than problematic parts. Some of the most beneficial parts are the lack of background noise from speech and the implementation of supply devices.

In the original summary plots, it was seen that, regarding performance, a medium amount of supply devices actually had a more negative effect on the indoor climate than not having any devices, but that a high number of devices had a positive influence. When compared to all the data, that the models are trained on, the input of 6 supply devices is fairly low. A natural conclusion would be that such an implementation would have a negative influence on performance. In reality, the model will assess the number of devices in relation to the input of the other office size-related features such as office volume and window area. Given these feature values, the models conclude that the number of devices is high enough to successfully ventilate that size office. If the number of devices was set to five instead of six, they will, however, stop having a positive effect. Six supply devices are, thus, the lowest number of devices needed to create a beneficial influence given that specific office size. It should also be remembered that this is a tool to guide the user to figure out what to aim for when improving the indoor climate in the office. The tool should not be used to dimension the final ventilation system.

When the improved values are implemented in each model, all occupants are predicted to be satisfied according to each model. The improvements made to satisfy occupant 1, are, thus, also helping the other occupants. Therefore, it has been shown that this tool, based on the developed models, can directly help people/businesses to identify, and correct, areas of weakness within their office environment. This can lead to a direct increase in well-being and performance.

7 Discussion

The goal of this project was to create a data-driven tool that predicts performance and well-being based on the characteristics of the office and the occupants. The supervised machine learning models predicting this should be able to diagnose problematic parts of the indoor climate. Through this, the main problem of the project should be answered:

"How can machine learning be used to give a holistic diagnosis of the indoor climate in offices? How do the indoor environment, the characteristics of an office and the occupants within it influence the well-being of the occupants? How do the characteristics of the occupants influence their perception of the office?"

This chapter will be a discussion of how this has been achieved as well as what could be improved in the future.

7.0.1 The models

This project resulted in a tool consisting of three main models. One predicting the occupant's satisfaction with the indoor climate using a support vector machine classifier. The model has an accuracy of 0.71, an f1 score of 0.71, and an AUC of 0.71. Another model predicts the occupant's satisfaction with how the indoor climate influences their performance, using a random forest algorithm. This model has an accuracy of 0.74, an f1 score of 0.81, and an AUC of 0.69. The last model predicts whether the occupant has experienced office-related symptoms over the previous four weeks, using a random forest algorithm. This model has an accuracy of 0.84, an f1 score of 0.59, and an AUC of 0.74. This means that all models have good accuracies, and acceptable to good f1 scores and AUC values. It has, thus, been possible to create three predictive models based on the data.

It should be mentioned that the models are based on a fairly small data set of 221 observations. A small data set makes the models more likely to overfit to the training data, which reduces the models' performance. Despite this, the models have a good performance. It can, however, be assumed that the performance of the models would improve, with more available training data.

7.0.2 The diagnosis

Based on each model, SHAP summary plots were used to visualize the influence that each feature has on the final prediction. This allows us to diagnose what parts of the indoor climate are problematic.

IEQ parameters

The problem areas influencing all three models negatively were high CO_2 concentrations, high illuminance, and noise from specific sources. Particularly the negative effect of high CO_2 concentrations and background from noise and speech was to be expected. Both low and high temperatures had a negative influence. Temperature as a general feature was, however, not nearly as influential as some of the other features. As expected, it was seen that high-frequency cleaning had a positive influence on all measures of occupant well-being.

Building design

Regarding the general office, it was seen that newer office buildings contributed to a better perceived performance and a lower risk of experiencing office-related symptoms. It did, however, not have a significant influence on the perception of the overall indoor climate.

The focus on ventilation has increased in the last 30 years. Requirements regarding both energy use and indoor climate in newly designed buildings are continuously tightened as we become more aware of the consequences. It is, thus, not surprising that this is a factor.

It was also seen that high numbers of supply devices in the office had a positive influence on all models. This of course shows that the indoor climate is better in mechanically ventilated offices than in offices that do not have a ventilation system, which isn't too surprising. What is interesting regarding the model of performance, is that offices with a medium amount of devices are worse than offices without any devices at all. Offices that are not mechanically ventilated will usually be naturally ventilated through operable windows, for instance. It has previously been shown that people are more likely to be pleased with the indoor climate if it is naturally ventilated and they are able to control it themselves [78]. It would, therefore, make sense that occupants are happier with an office that is naturally ventilated than with an office that is poorly ventilated by an under-dimensioned mechanical ventilation system.

Large window areas had a negative influence on satisfaction with both indoor climate and its influence on performance. Normally this would be assumed to be caused by overheating. Because (1) the temperatures in the study were relatively low, (2) low temperatures generally caused more dissatisfaction than high ones, and (3) temperature wasn't a particularly influential feature in the models, this could, however, not be proven here. Another assumption is that there are larger windows in larger offices, with more occupants. This could have some truth to it as open-plan offices were shown to have a poor influence on performance. Still, the office volume does not seem to have nearly as big an influence, and this can, thus, not be the entire reason. A better conclusion would be that the large window areas cause higher illuminance, causing reflections on computer screens, as very high mean illuminance, is shown to be problematic according to all models.

7.0.3 The occupant's influence on their perception

The well-being of occupants is not purely dependent on the indoor environment quality. To better take some of the occupant-specific parameters into account, it is crucial to include occupant characteristics in the models - especially when working with occupant-centric model targets.

In all three models, sex was one of the most important features. Age was relevant in two of the models, and whether the occupant was wearing glasses or contact lenses only had an influence on the prediction of performance.

Sex

Sex is an influential feature in all models, and in all models, women are more likely to be negatively influenced by the indoor climate. The problematic parts of the office's indoor climate are, thus, disproportionally influencing women.

It is also seen that being female makes the indoor climate far more likely to negatively influence your performance, than being male is to improve your performance (See performance model, Figure 5.16). In other words, there is more to win in also fitting the indoor climate to the female employees than there hypothetically would be by only hiring male employees to fit the current indoor climate.

The disproportion between each sex's experience of the indoor climate could be traced back to some of the physiological differences, and the gender-biased design standards our buildings are based on. While the models developed in this project stay far away from the physiological aspects of the occupant, the models can be used to diagnose the indoor climate with respect to each sex, and find a solution that might optimize the outcome for both.

In response to how Sick Building Syndrome hits each sex disproportionally, WHO also suggests that it could be because women tend to be in worse jobs, as people in low-paid jobs are also more likely to experience SBS [28]. This goes back to occupant well-being not only being influenced by the IEQ but by the experience of going to work as a whole.

Age
While WHO's argument might not explain the entire reason behind the differences in each sex's experience, it does link nicely with the role age plays in the models of occupant performance and experience of office-related symptoms. In the model predicting performance, the youngest occupant's performance is more likely to be negatively influenced by the office while the other age groups are more likely to be either positively influenced or not influenced at all by the indoor climate in the office. Age as a feature isn't as impactful in the prediction of office-related symptoms. It does, however, still show that older occupants are less likely to experience office-related symptoms than young occupants.

You would think that older people are more likely to be poorly influenced by the indoor climate. The young age group of people between 20-29 years should be healthier and more adaptive to the indoor climate than people in the oldest age group of 60+ years olds, for instance. An explanation, could be that people in their twenties will be in the office as student workers or at their first job. This could mean that they are assigned more routine assignments than more senior employees. Interestingly it is specifically low-paying, routine jobs with a lack of personal control that are more likely to cause SBS [29]. Equally, it is not hard to imagine that a bored occupant is easier to distract, and, hence, might be more influenced by a poor indoor climate.

In conclusion, job type might have a larger influence on the well-being of the occupant than the physiological aspects of age. Given that this is the case in these models, you could argue that it seems odd that the job description did not have a statistically significant influence on the target of any of the models. This can, however, be explained by the survey only differentiating between administrative, technical, professional, and managerial jobs - job categories that are not saying much about the actual work conditions.

7.1 Model limitations and discussion of features

In the next section, some of the choices made in the development of the model are discussed, as well as some of the models' limitations.

7.1.1 Occupant characteristics

The well-being of the occupant won't only be influenced by the IEQ. The well-being will in part also be influenced by personal factors such as sex, age, etc. Because such parameters have an influence on how the indoor climate is experienced, they are important to have in the models. When looking at age, for instance, it was seen that the youngest age group's disposition to be poorly influenced by the indoor climate might have more to do with job satisfaction than their general physical well-being. This, however, does not make the feature irrelevant to the model. On the contrary, it allows the model to take some of the more personal factors into account, that otherwise are not considered.

It should also be remembered that the occupant-related parameters are a generalization and a simplification in itself. They show how each sex and age group statistically are perceiving the indoor climate. There will, thus, always be people that do not fit the mould. This can be compared to how young people pay higher car insurance because they statistically drive more recklessly. Of course, not all young people drive recklessly, but enough do to make it a beneficial model for insurance companies.

7.1.2 IEQ measurements

The measurements of IEQ parameters make up some of the most influential parameters in the models. This is a strong argument as to why it was important to include quantitative measurements in the models rather than just using surveys. In the following, the quality of these quantitative features and the decisions made to make them as representative as possible in the models are discussed.

Not a snapshot measurement

Unlike the BOSSA tool, which takes one snapshot measurement of each indoor climate parameter at the time the occupant survey is filled out, the models here are built on the means of measurements taken across a week. The quality of the snapshot method is that you get an idea of the exact indoor climate at the time of the survey. It does, however, require that you stop everything happening in the office at that time to make everyone fill out the survey at once. Given that the data for this project was gathered during the pandemic while many were working from home, there would simply be too few people in the offices at a time, to do this in a somewhat efficient way. By instead measuring the indoor climate parameters across a week in each office and allowing the occupants to fill out the survey in their own time, it made it possible to get more data.

The week-long measurements also make it possible to gain an idea of the overall indoor climate. This makes the models less sensitive to bias. Bias could, for instance, be from someone opening a window right before taking the measurement, although the window might usually stay closed during winter days.

Using the mean of measurements

In this project, it was decided to generalize the week-long measurements by using the mean. The advantage and disadvantage of using the mean is that it only picks up on odd parts of the indoor climate if they are very extreme or continuous. Furthermore, it does not show polarized situations, such as a mix of very high and very low values. Instead of picking up on these as being problematic, it will most likely view the value as being normal.

This issue is countered by the use of occupant surveys. Here, the occupants who are dissatisfied with the indoor climate are asked what the issue is. This way we pick up on patterns that might be lost in the quantitative metrics.

Sound level

The mean sound level feature in the models is fairly weak due to the many missing values. This should not skew the overall influence of the sound level, as missing values are set to the median. It does, however, mean that the overall feature has an artificially low predictive importance in the models, compared to how it might be in the real world. This must be taken into account when using the tool.

7.1.3 Covid-biased data

The data the models are built on were gathered in 2020 and 2021, during the covid-19 pandemic. This means that people were allowed to work from home, and there would generally be fewer people in the offices than under normal circumstances. In an office with a lot of people and poor ventilation, high CO_2 concentrations will usually be an issue. As expected, high CO_2 concentration is indeed an influential feature in the models built. It must, however, be assumed that human bio-effluents will have an even larger impact in offices at full capacity.

7.1.4 Seasonal limitations

As it is seen in Figure 7.1, none of the data used for the models are from either the summer or winter months. Most of the data is gathered in March and April, while a smaller proportion of it is from September, October, and November.

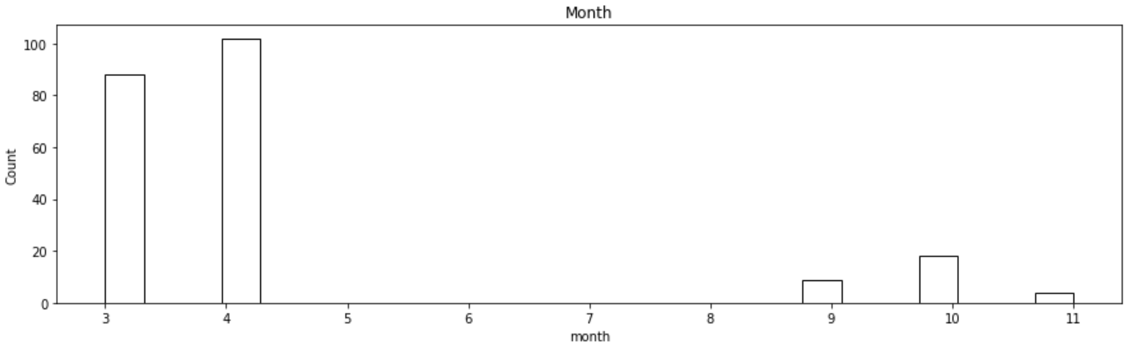


Figure 7.1: Distribution of months the data is gathered in.

The data from months in the heating season can be assumed to represent similar indoor climates to the observations taken in the winter months. There are, however, very few observations that are even close to being outside the heating season. These observations (September), are still not reflecting a situation that is as extreme as it could be in the middle of the summer. In the data, the highest mean temperatures are about $26°C$. You would, however, expect an even higher mean temperature if an office is overheated during peak summer.

The models are ultimately flawed by not being trained on the extreme weather conditions you might see in the summer or the winter. This could particularly be a problem when using the tool in a summer situation. In the models, it is seen that high CO_2 concentration is a very influential IEQ-related feature. This is very indicative of a winter situation, where it is cold outside and people don't open their windows. In the summer, windows in a naturally ventilated office will often be opened, and CO_2 concentrations will be lower. Therefore, high temperatures should be the most influential feature. Since the models are not trained in such situations they are, however, likely to underestimate how problematic overheating actually is.

7.1.5 Geographical limitations

The data the models are built and trained on were gathered in office buildings in Denmark and Greenland. While this does create some range in the climate the model is based on, it definitely does not represent all climates. If applying the models to office buildings in other countries, it is, therefore, important that the countries have a similar climate. Here, a beneficial feature to add could be the average temperature in the given month at the specific location.

7.2 Future work

In the following, different ways of improving the models and the overall tool are discussed.

7.2.1 Models improvements with new data

The more office buildings are assessed, the more data is gathered that can be used to train and improve the models.

Reassessment of models

At the moment some of the best-performing models are fairly simple to avoid over-fitting when working with a small data set. As more data is gathered it would, thus, make sense to optimize the current models to more detailed models that might have a better performance.

Equally, it would make sense to reassess the predictive features and see if more descriptive models can be made. As a rule of thumb, the larger the data set, the more features can be used without reducing the quality of the model. More features would make the diagnosis of the office indoor climate more tailored to the specific office.

A broader range of training data
Gathering data during both the summer and the winter, would not just improve the models. It would be essential to the quality of the models if they are to be used efficiently during these months. With more data from the summer, features such as window shading percentage, window orientation and air condition systems might also become influential in the models, giving the model a broader range of designs to diagnose the indoor climate based on.

This is the same regarding the geographical limitations. To make the model applicable in other climatic situations, it needs data from a wider range of places.

Allowing multi-categorical targets
If the available number of observations moved into the thousands or millions one day, it would make sense to go back to the multi-categorical targets rather than using the current binary targets. The binary targets are not limiting the use of the current models for the purpose they were intended for. Using the ordinal scales as targets would, however, allow a higher level of detail in the assessment of the indoor climate. A model like this would make a distinction between performance being a little influenced by the indoor climate or being very influenced by the indoor climate. Regarding the experience of office-related symptoms, it would be able to differentiate between the frequency the symptoms are experienced at. Ultimately this would give more range to the diagnosis of the indoor climate and make it easier to make prioritized design decisions.

7.2.2 The economic aspects

The tool developed by InnoByg, outputs the monetary gain connected to improving the indoor climate. There is no doubt that the economic aspect combined with the tools diagnosing abilities make it easier to make prioritized decisions on how best to improve the office. If the output of the model, for instance, shows that the indoor climate can be improved by changing one of three things, this could help you decide what design decision might be more economically viable.

Looking even further, the cost of different design choices could also be based on the use of natural resources or the CO_2 footprint. To do this, it would make sense to have each design choice linked to a simplified life cycle assessment (LCA).

8 Conclusion

Throughout this project, a tool that diagnoses the indoor climate based on three supervised machine learning models was developed. The models are built on data about the office building, measurements of IEQ parameters, and occupant surveys. The output of the tool is an assessment of the indoor climate based on satisfaction with the indoor climate, the indoor climate's influence on occupant performance, and the experience of office-related symptoms. Based on this, it is possible to pinpoint the problem areas of the indoor climate. By including occupant characteristics, it is possible to tailor the diagnosis and recommendations better to the employees in the office.

The model predicting the occupant's satisfaction with the indoor climate, is a support vector machine classifier, the model predicting the indoor climate's influence on the occupant's perceived performance is a random forest model, and the last model predicting the experience of office-related symptoms is a random forest model. The models have an accuracy between 0.71 and 0.84, an f1 score between 0.59 and 0.81, and an AUC between 0.69 and 0.74. The models are, thus, proven to have acceptable to good predictive power. It is, however, assumed that retraining the models with more data will improve their performance further.

According to the models, background noise from speech and high CO_2 concentrations are the IEQ parameters that have the most negative influence on the well-being and performance of the occupants. Regarding the management of the office, it is seen that cleaning frequency should be at least once a week in order to have a positive influence on the well-being and performance of the occupants. Regarding office design, it is seen that new buildings and single offices perform better than older buildings and open-plan offices. One of the most influential features in the models is the sex of the occupant, as women seem to be far more negatively influenced by the indoor climate than men.

Due to the limited size and range of the data set, the models, and thereby the overall tool, do have some limitations. They are not trained to assess buildings during a summer situation or in other countries with climates that are different from Denmark and Greenland. The models are, however, easy to re-train as more data is gathered. The tool will, thus, improve the more it is used.

Overall, the project was a success, as it was possible to create a data-driven tool that is able to give a holistic diagnosis of the indoor climate in offices.

Bibliography

[1] P.M. Bluyssen, Myriam Aries, and Paulavan Dommelenc. “Comfort of workers in office buildings: The European HOPE project. (English)”. In: *Building and Environment* 46.1 (2011), pp. 280–288. DOI: https://www-sciencedirect-com.proxy.findit.cvt.dk/science/article/pii/S0360132310002350.

[2] Philomena M. Bluyssen and et. al. “Assessment of wellbeing in an indoor office environment. (English) [On]”. In: *Building and Environment* 46.1 (2011), pp. 2632–2640.

[3] P. Wargorcki, O. Seppänen, and et al. *REHVA Guidebook: Indoor Climate and Productivity In Offices: How to integrate productivity in life–cycle cost analysis of building services*. Undervisningsnotat. REHVA, 2006. ISBN: 2–9600468–5–4.

[4] Joseph G. Allen and et. al. “Associations of Cognitive Function Scores with Carbon Dioxide, Ventilation, and Volatile Organic Compound Exposures in Office Workers: A Controlled Exposure Study of Green and Conventional Office Environments. (English) [On Environmental Health Perspectives]”. In: *Research* 124.6 (2016), pp. 805–812. DOI: https://ehp.niehs.nih.gov/doi/10.1289/ehp.1510037.

[5] L. Lan and et. al. “Effects of thermal discomfort in an office on perceived air quality, SBS symptoms, physiological responses, and human performance. (English)”. In: *Indoor Air 2011* 21 (2011), pp. 376–390. DOI: https://onlinelibrary-wiley-com.proxy.findit.cvt.dk/doi/full/10.1111/j.1600-0668.2011.00714.x.

[6] Ihab M.K. Elzeyadi. “Daylighting-Bias and Biophilia: Quantifying the Impact of Daylighting on Occupants Health. (English)”. In: *Greenbuild 2011 Proceedings* (2011). DOI: https://www.researchgate.net/profile/Ihab-Elzeyadi/publication/344361245_Daylighting-Bias_and_Biophilia_Quantifying_the_Impacts_of_Daylighting_and_Views_on_Occupants_Health/links/5f6c4023458515b7cf497bc6/Daylighting-Bias-and-Biophilia-Quantifying-the-Impacts-of-Daylighting-and-Views-on-Occupants-Health.pdf.

[7] J.F.Nicol and M.A.Humphreys. “Ventilation rates in schools and pupils’ performance. (English)”. In: *Building and Environment* 48 (2016), pp. 215–223. DOI: https://www.sciencedirect.com/science/article/pii/S0360132311002617?via\%3Dihub.

[8] S. P. BANBURY and D. C. BERRY. “Office noise and employee concentration: Identifying causes of disruption and potential improvements. (English)”. In: *Ergonomics* 48.1 (2005,), pp. 25–37. DOI: https://www-tandfonline-com.proxy.findit.cvt.dk/doi/abs/10.1080/00140130412331311390.

[9] Diane Berry and et al. “Disruption of office-related tasks by speech and office noise. (English)”. In: *British Journal of Psychology* 89.3 (1998), pp. 499–517. DOI: https://bpspsychub-onlinelibrary-wiley-com.proxy.findit.cvt.dk/doi/abs/10.1111/j.2044-8295.1998.tb02699.x.

[10] B. Howald Petersen. *Komfortventilation*. Undervisningsnotat. Byg DTU, 2005.

[11] Hana Abdel. *How to Transform a Polluted Indoor Environment into a Healthy Home*. URL: https://www.archdaily.com/938491/how-to-transform-a-polluted-indoor-environment-into-a-healthy-home?ad_medium=gallery. (accessed: 7.1.2023).

[12] Arbejdstilsynet. *Indeklima*. URL: https://at.dk/regler/at-vejledninger/indeklima-a-1-2/. (accessed: 6.1.2023).

[13] EPA. *What are volatile organic compounds (VOCs)?* URL: https://www.epa.gov/indoor-air-quality-iaq/what-are-volatile-organic-compounds-vocs. (accessed: 14.10.2022).

[14] Mariel Wolfson. “Defining Healthy Housing: The Competing Priorities of Energy and Indoor Air Quality in the Early 1980s”. In: *Joint Center for Housing Studies* (2013). DOI: https://www.jchs.harvard.edu/research-areas/working-papers/defining-healthy-housing-competing-priorities-energy-and-indoor-air.

[15] Paraskevi Vivian Dorizas and et al. *The inner value of a building. Linking indoor environmental quality and energy performance in building regulation*. URL: https://www.semanticscholar.org/paper/The-inner-value-of-a-building.-Linking-indoor-and-Dorizas-Groote/9aaaf6f9b4dedcc3c1456b7e2eaca681574968e8/figure/2. (accessed: 7.1.2023).

[16] P.O. Fanger. “Calculation of thermal comfort: introduction of a basic comfort equation. (English)”. In: *ASHRAE Trans.* 73 (1967).

[17] Sebastian Guenther. *What Is PMV? What Is PPD? The Basics of Thermal Comfort*. URL: https://www.simscale.com/blog/what-is-pmv-ppd/. (accessed: 15.1.2023).

[18] HSE. *Thermal comfort*. URL: https://www.hse.gov.uk/temperature/thermal/index.htm. (accessed: 18.10.2022).

[19] Lysviden.dk. *BELYSNINGSSTYRKE*. URL: https://lysviden.dk/grundviden/emne/belysningsstyrke/artikler/introduktion/. (accessed: 19.10.2022).

[20] Lysviden.dk. *FARVETEMPERATUR*. URL: https://lysviden.dk/grundviden/emne/farvetemperatur/artikler/introduktion/. (accessed: 19.10.2022).

[21] The Lighting Design Studio. *Colour temperature*. URL: https://lightingdesignstudio.co.uk/colour-temperature/. (accessed: 7.1.2023).

[22] Lysviden.dk. *LYSETS BIOLOGISKE VIRKNINGER*. URL: http://www.lysviden.dk/lys-og-menneske/emne/lys-og-sundhed/artikler/lysets-biologiske-virkninger/. (accessed: 19.10.2022).

[23] Arbejdstilsynet. *6. Støjgrænser og unødig støjbelastning*. URL: https://at.dk/regler/at-vejledninger/stoej-d-6-1/#:~:text=og%5C%20un%5C%C3%5C%B8dig%5C%20st%5C%C3%5C%B8jbelastning-,6.1.,over%5C%20137%5C%20dB(C).&text=Overskrides%5C%20st%5C%C3%5C%B8jgr%5C%C3%5C%A6nserne%5C%2C%5C%20skal%5C%20der%5C%20derfor,at%5C%20bringe%5C%20st%5C%C3%5C%B8jen%5C%20under%5C%20gr%5C%C3%5C%A6nsev%5C%C3%5C%A6rdierne.. (accessed: 15.1.2023).

[24] M. Spreng. “Possible health effects of noise induced cortisol increase. (English)”. In: *Noise Health* (2000), pp. 59–64. DOI: https://pubmed.ncbi.nlm.nih.gov/12689472/.

[25] Joke Bradt, Cheryl Dileo, and Noah Potvin. “Music for stress and anxiety reduction in coronary heart disease patients. (English)”. In: *Cochrane Database Syst Rev.* (2003). DOI: https://pubmed.ncbi.nlm.nih.gov/24374731/.

[26] Julian Treasure. *Sound Business*. Management Books 2000 Limited, 2006. ISBN: 1852525282.

[27] World Green Building Counsil. *Health, Wellbeing Productivity in Offices*. The next chapter for green building. World Green Building Counsil, 2014.

[28] WHO. “Sick building syndrome. (English)”. In: *World Health Organization Regional Office for Europe* (). DOI: https://www.wondermakers.com/Portals/0/docs/Sick%20building%20syndrome%20by%20WHO.pdf.

[29] Jack Rostron. “Sick building syndrome: A review of causes, consequences and remedies . (English)”. In: *Journal of Retail and Leisure Property* 7 (2008), pp. 291–303. DOI: http://new-courses.justice.eku.edu/OSH/OSH366/Docs/OSH366W6_SickBuildingSyndrome.pdf.

[30] Tue Herlau and et al. *Introduction to Machine Learning and Data Mining*. Lecture notes, Fall 2021, version 1.2. DTU, 2021.

[31] Ben Mirbozorgi. *Theory of K-Nearest Neighbors (KNN)*. URL: https://medium.com/analytics-vidhya/theory-of-k-nearest-neighbors-knn-234654826c50. (accessed: 7.1.2023).

[32] José Joaquín Aguilera and et. al. “Thermal adaptation in occupant-driven HVAC control”. In: *Journal of Building Engineering* (2019). DOI: https://www.sciencedirect.com/science/article/pii/S2352710219303237?casa_token=OPmdv-WTwPMAAAAA:cLy87z-uD0pJP-25UanKjGCIldcene23-rD9h0LvUqVYfnlevlF4Eb6XhSfMgSSIz0ae7PNewvs.

[33] J. van Hoof. “Forty years of FangerÕs model of thermal comfort: comfort for all?. (English)”. In: *INDOOR AIR* 18 (2008), pp. 182–201. DOI: https://onlinelibrary.wiley.com/doi/10.1111/j.1600-0668.2007.00516.x.

[34] Boris Kingma and Wouter van Marken Lichtenbelt. “Energy consumption in buildings and female thermal demand. (English)”. In: *Nature Climate Change* 5 (2015), pp. 1054–1056. DOI: https://www-nature-com.proxy.findit.cvt.dk/articles/nclimate2741.

[35] J.F.Nicol and M.A.Humphreys. “Adaptive thermal comfort and sustainable thermal standards for buildings. (English) [On Energy and Buildings]”. In: *Annalen der Physik* 34.6 (2002), pp. 563–572. DOI: https://www.sciencedirect.com/science/article/pii/S0378778802000063?casa_token=CD_omsR-IZ4AAAAA:30w7i7NhFcAm1f4K-TcxfPEI-ruffzqlsHlb-CA5I8mn4W4so1WJ62OAa2WEFFS1kpEAcIDCiSA.

[36] Ashrant Aryal and Burcin Becerik-Gerber. “A comparative study of predicting individual thermal sensation and satisfaction using wrist-worn temperature sensor, thermal camera and ambient temperature sensor. (English)”. In: *Building and Environment* 160 (2019). DOI: https://www-sciencedirect-com.proxy.findit.cvt.dk/science/article/pii/S0360132319304330.

[37] Shichao Liu and et. al. “Personal thermal comfort models with wearable sensors. (English)”. In: *Building and Environment* 162 (2019). DOI: https://www-sciencedirect-com.proxy.findit.cvt.dk/science/article/pii/S0360132319304913.

[38] Shichao Liu and et. al. “Machine learning algorithms applied to a prediction of personal overall thermal comfort using skin temperatures and occupant’s heating behaviour.. (English)”. In: *Applied Ergonomics* 85 (2020). DOI: https://www-sciencedirect-com.proxy.findit.cvt.dk/science/article/pii/S0003687018305763.

[39] Geo Clausen and et al. *Totalværdi og indeklima*. ET INNOBYG UDVIKLINGSPROJEKT. Totalværdi og indeklima, 2017.

[40] Leah Zagreus. “Listening to the occupants: a Web-based indoor environmental quality survey. (English)”. In: *Indoor Air* (2014), pp. 65–74. DOI: https://www.researchgate.net/publication/8069025_Listening_to_the_occupants_A_Web-based_indoor_environmental_quality_survey.

[41] Christhina Candido et al. “BOSSA: a multidimensional post-occupancy evaluation tool. (English)”. In: *Building Research Informatio, 2016* 44.2 (2016), pp. 214–228. DOI: 10.1080/09613218.2015.1072298.

[42] George Novack. *Building a One Hot Encoding Layer with TensorFlow*. URL: https://towardsdatascience.com/building-a-one-hot-encoding-layer-with-tensorflow-f907d686bf39. (accessed: 18.11.2022).

[43] Jason Brownlee. *Ordinal and One-Hot Encodings for Categorical Data*. URL: https://machinelearningmastery.com/one-hot-encoding-for-categorical-data/. (accessed: 18.11.2022).

[44] Vasile Păpălu□ă. *What’s the best way to handle NaN values?* URL: https://towardsdatascience.com/whats-the-best-way-to-handle-nan-values-62d50f738fc. (accessed: 18.11.2022).

[45] Vikas Verma. *A comprehensive guide to Feature Selection using Wrapper methods in Python*. URL: https://www.analyticsvidhya.com/blog/2020/10/a-comprehensive-guide-to-feature-selection-using-wrapper-methods-in-python/. (accessed: 18.11.2022).

[46] Jason Brownlee. *How to Choose a Feature Selection Method For Machine Learning*. URL: https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/#:~:text=Filter%20feature%20selection%20methods%20use,be%20used%20in%20the%20model. (accessed: 18.11.2022).

[47] StatsTest.com. *Kendall's Tau*. URL: https://www.statstest.com/kendalls-tau/. (accessed: 18.11.2022).

[48] Akhil Anand. *Categorical Feature Selection using Chi- Squared Test*. URL: https://medium.com/analytics-vidhya/categorical-feature-selection-using-chi-squared-test-e4c0d0af6b7e. (accessed: 11.1.2023).

[49] Saptarsi Goswami. *Using the Chi-Squared test for feature selection with implementation*. URL: https://towardsdatascience.com/using-the-chi-squared-test-for-feature-selection-with-implementation-b15a4dad93f1. (accessed: 11.1.2023).

[50] Guhanesvar. *Feature Selection Based on Mutual Information Gain for Classification and Regression*. URL: https://guhanesvar.medium.com/feature-selection-based-on-mutual-information-gain-for-classification-and-regression-d0f86ea5262a. (accessed: 22.11.2022).

[51] Bob Rupak Roy. *Mutual Information Score — Feature Selection*. URL: https://bobrupakroy.medium.com/mutual-information-score-feature-selection-8eb19071664b. (accessed: 22.11.2022).

[52] scikit learn. *sklearn.feature*$_selection.f_classif$. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_classif.html. (accessed: 11.1.2023).

[53] scipy. *scipy.stats.kendalltau*. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kendalltau.html. (accessed: 18.11.2022).

[54] scikit learn. *sklearn.feature*$_selection.chi2$. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html#sklearn.feature_selection.chi2. (accessed: 11.1.2023).

[55] scikit learn. *sklearn.feature*$_selection.mutual_info_classif$. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_classif.html#sklearn.feature_selection.mutual_info_classif. (accessed: 22.11.2022).

[56] dython. *nominal - dython*. URL: http://shakedzy.xyz/dython/modules/nominal/. (accessed: 22.11.2022).

[57] Iniabasi Affiah. *How to Implement K fold Cross-Validation in Scikit-Learn*. URL: https://www.section.io/engineering-education/how-to-implement-k-fold-cross-validation/. (accessed: 7.1.2023).

[58] Rachel Draelos. *Measuring Performance: AUC (AUROC)*. URL: https://glassboxmedicine.com/2019/02/23/measuring-performance-auc-auroc/. (accessed: 22.11.2022).

[59] Stephen Allwright. *What is a good F1 score and how do I interpret it?* URL: https://stephenallwright.com/good-f1-score/. (accessed: 23.11.2022).

[60] Stephen Allwright. *What is a good accuracy score in machine learning?* URL: https://stephenallwright.com/good-accuracy-score/#:~:text=There%5C%20is%5C%20a%5C%20general%5C%20rule,6%5C%25%5C%20and%5C%2070%5C%25%5C%20%5C%2D%5C%20OK. (accessed: 23.11.2022).

[61] Joos Korstanje. *The F1 score*. URL: https://towardsdatascience.com/the-f1-score-bec2bbc38aa6. (accessed: 22.11.2022).

[62] scikit learn. *sklearn.metrics.f1*$_score$. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html. (accessed: 22.11.2022).

[63] Siladittya Manna. *Evaluation Metrics Part 3*. URL: https://medium.com/the-owl/evaluation-metrics-part-3-47c315e07222. (accessed: 22.11.2022).

[64] Katrina Wakefield. *A guide to the types of machine learning algorithms and their applications*. URL: https://www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html#:~:text=There%20are%20four%20types%20of,%2Dsupervised%2C%20unsupervised%20and%20reinforcement.. (accessed: 22.11.2022).

[65] Shailaja Gupta. *Pros and cons of various Machine Learning algorithms*. URL: https://towardsdatascience.com/pros-and-cons-of-various-classification-ml-algorithms-3b5bfb3c87d6. (accessed: 22.11.2022).

[66] scikit learn. *sklearn.model$_selection.GridSearchCV$¶*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. (accessed: 22.11.2022).

[67] Prateek Majumder. *Gaussian Naive Bayes*. URL: https://iq.opengenus.org/gaussian-naive-bayes/. (accessed: 11.1.2023).

[68] scikit learn. *sklearn.naive$_bayes.GaussianNB$*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html. (accessed: 7.1.2023).

[69] Onkar Manjrekar and et. al. "Identification of flow regime in a bubble column reactor with a combination of optical probe data and machine learning technique". In: *Chemical Engineering Science* (2019). DOI: https://www.researchgate.net/publication/332248436_Identification_of_flow_regime_in_a_bubble_column_reactor_with_a_combination_of_optical_probe_data_and_machine_learning_technique.

[70] Czako Zoltan. *SVM and Kernel SVM*. URL: https://towardsdatascience.com/svm-and-kernel-svm-fed02bef1200. (accessed: 7.1.2023).

[71] Carlos Domínguez García. *Visualizing the effect of hyperparameters on Support Vector Machines*. URL: https://towardsdatascience.com/visualizing-the-effect-of-hyperparameters-on-support-vector-machines-b9eef6f7357b. (accessed: 7.1.2023).

[72] scikit learn. *1.4. Support Vector Machines*. URL: https://scikit-learn.org/stable/modules/svm.html. (accessed: 7.1.2023).

[73] scikit learn. *1.10. Decision Trees*. URL: https://scikit-learn.org/stable/modules/tree.html. (accessed: 7.1.2023).

[74] scikit learn. *sklearn.ensemble.RandomForestClassifier*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. (accessed: 7.1.2023).

[75] Saul Dobilas. *XGBoost: Extreme Gradient Boosting — How to Improve on Regular Gradient Boosting?* URL: https://towardsdatascience.com/xgboost-extreme-gradient-boosting-how-to-improve-on-regular-gradient-boosting-5c6acf66c70a. (accessed: 7.1.2023).

[76] dmlc XGBoost. *XGBoost Tutorials*. URL: https://xgboost.readthedocs.io/en/stable/tutorials/model.html. (accessed: 7.1.2023).

[77] Samuele Mazzanti. *SHAP Values Explained Exactly How You Wished Someone Explained to You*. URL: https://towardsdatascience.com/shap-explained-the-way-i-wish-someone-explained-it-to-me-ab81cc69ef30. (accessed: 22.11.2022).

[78] Richard Dear and et al. "Developing an adaptive model of thermal comfort and preference". In: *ASHRAE Transactions 1998* 104.1 (1998), pp. 145–167. DOI: https://escholarship.org/uc/item/4qq2p9c6.

A Cleaning

Distributions of the features in the final models are seen in the following.

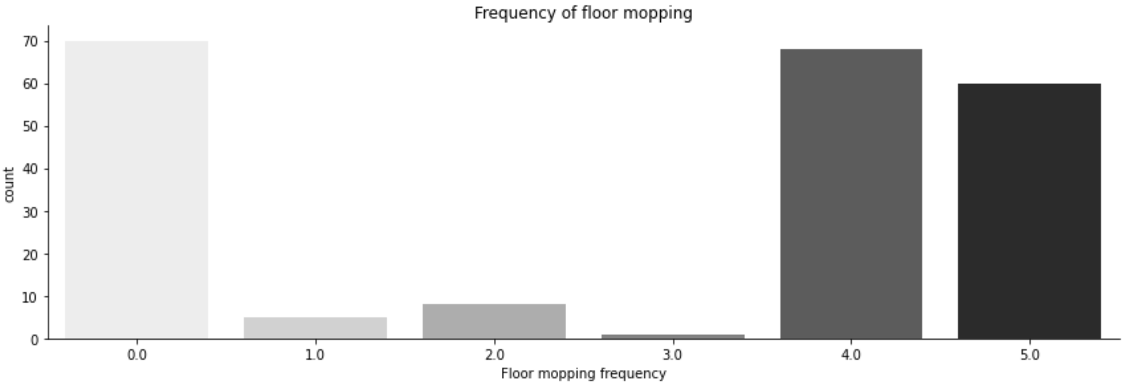
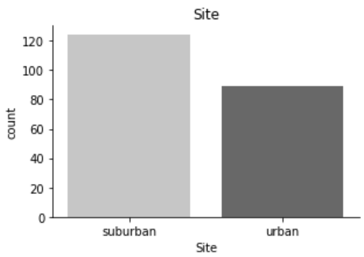
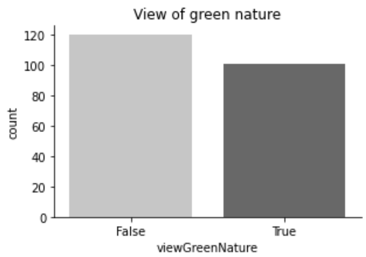


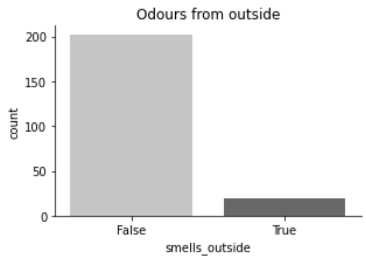
Figure A.1: Frequency of floor mopping: (0) never, (1) less frequent, (2) once a month, (3) every second week, (4) once a week, (5) 2-4 times a week, (6) every day.



(a) Site

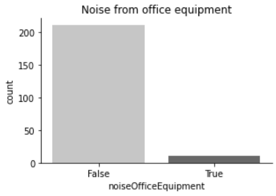


(b) View of nature

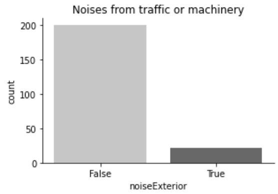


(c) Odours from outside

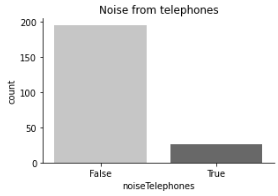
Figure A.2: Exterior



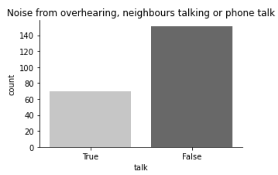
(a) Office equipment



(b) Exterior

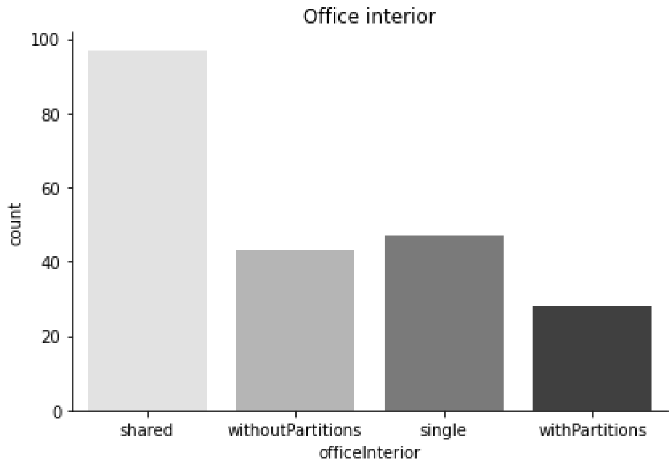


(c) Telephones

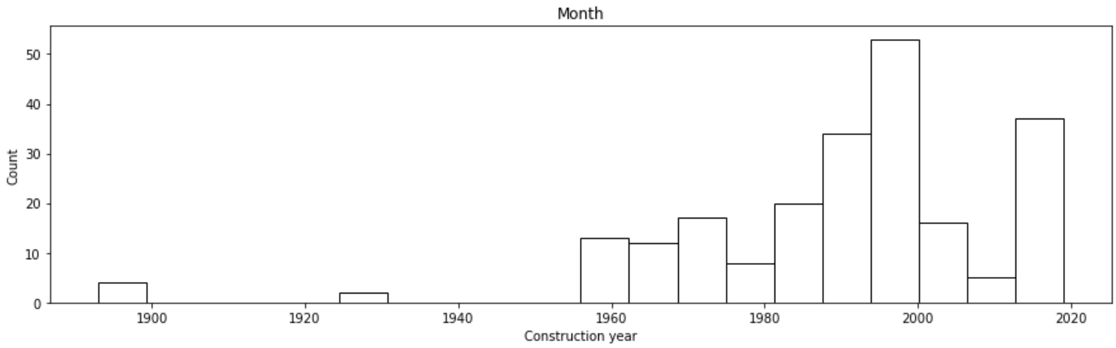


(d) Talk

Figure A.3: Noise

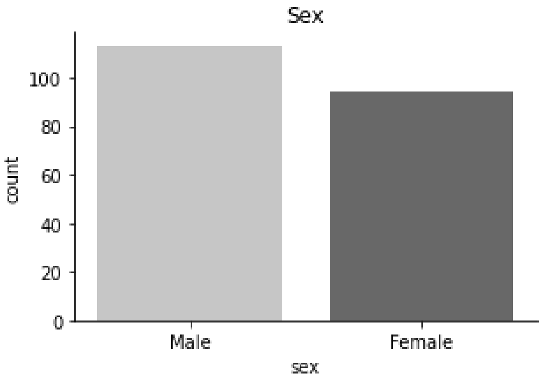


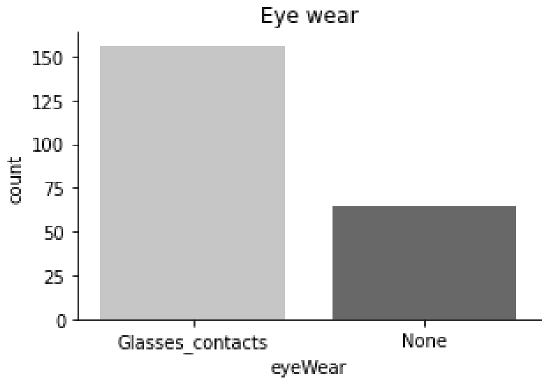
(a) Office interior



(b) Construction year

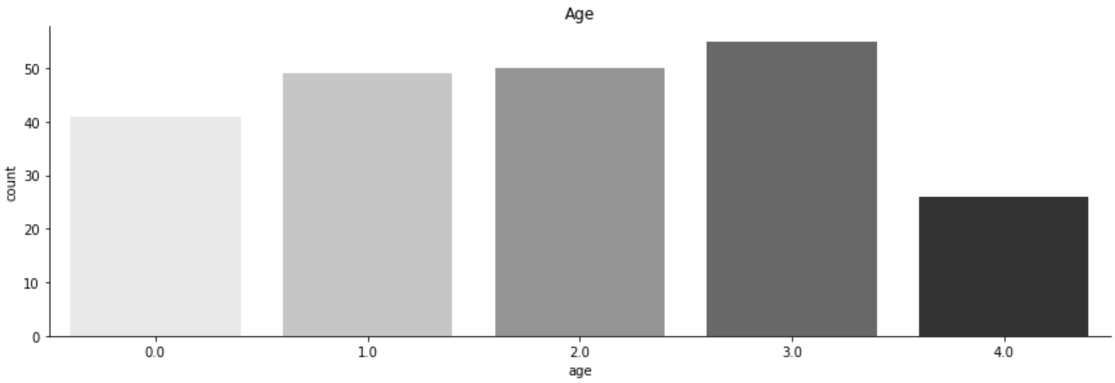
Figure A.4: Office design





(a) Sex

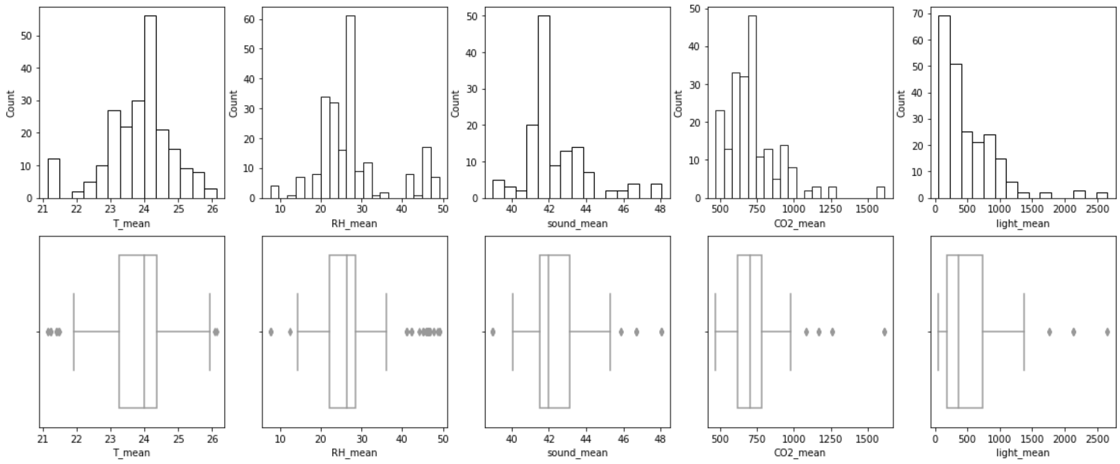
(b) Eye wear



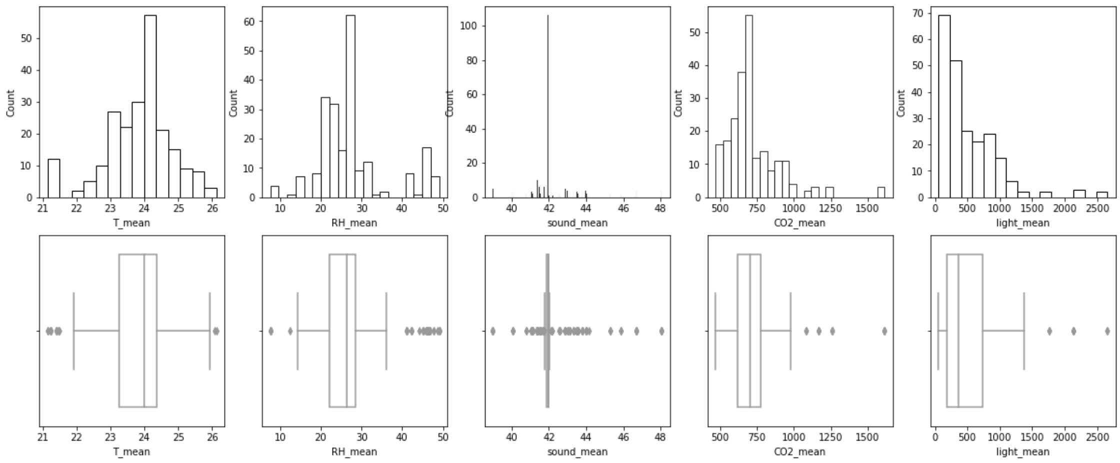
(c) Age: (0) 20-29 years, (1) 30-39 years, (2) 40-49 year, (3) 50-60 years, (4) >60 years.

Figure A.5: Occupant

In Figure A.6 histograms and box plots of the IEQ measurement means are seen before and after missing values are replaced by the median.



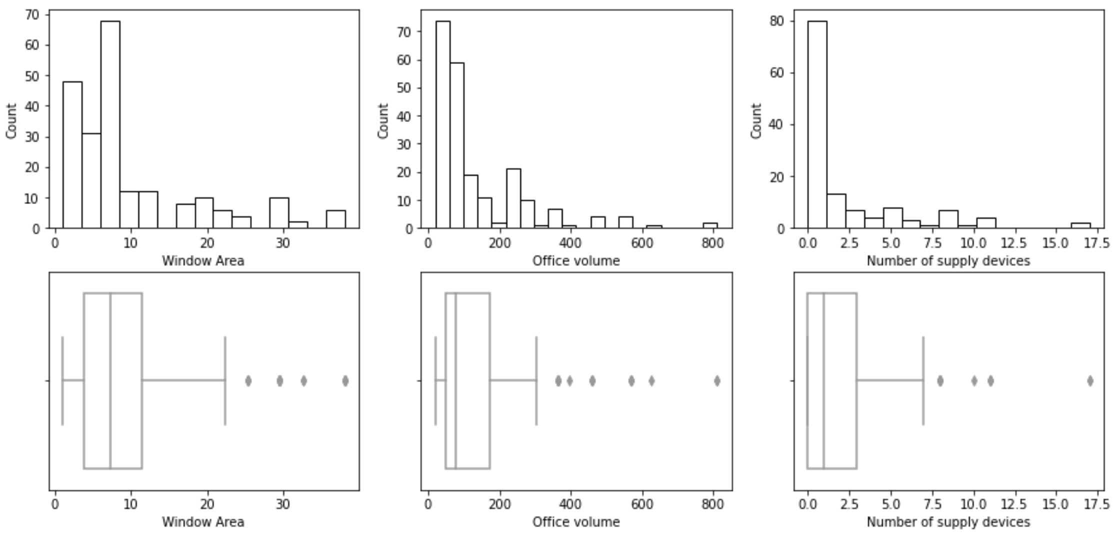
(a) Before replacing missing values



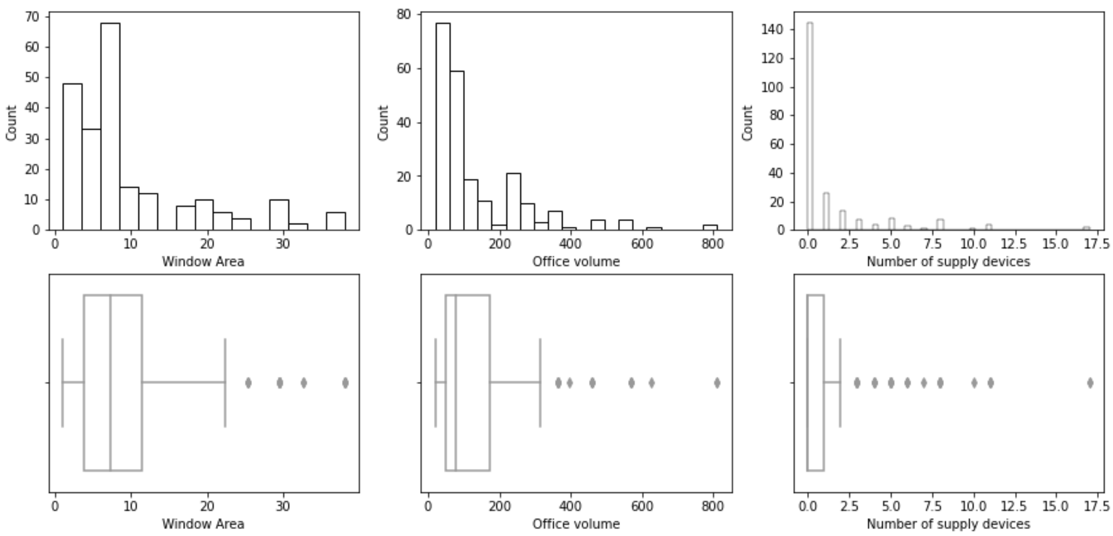
(b) After replacing missing values

Figure A.6: Mean measurements

In Figure A.7 histograms and box plots of the window areas, office volumes and a number of supply devices seen before and after missing values are replaced. The missing window areas and office volumes were replaced through KNN. The missing numbers of supply devices were replaced by 0.



(a) Before replacing missing values



(b) After replacing missing values

Figure A.7: Measurements

B Feature selection

B.1 Filter selection

The following tables are showing the results of the correlation-type statistic tests performed during the filter selection process. Only the cells of the features where the correlation statistic is appropriate are filled in. Where the appropriate association statistic was calculated, but it wasn't proven to be significant ($p > 0.05$), a (-) is seen. In the "Action" columns the features are marked with a drop or keep dependent on the decision made. In some cells the action is followed by an "imp." or a "corr.". The "imp." stands for importance, and means that the action has been made based on the feature's importance in the model. The "corr." stands for correlation, and means that the feature was dropped because it was highly correlated with another feature.

B.1.1 Indoor climate

Table B.1

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Light					
lightFlicker			-	5.43	drop
lightReflections			-	3.18	drop
lightNoTask			-	2.75	drop
lightTooDark			0.08	2.46	drop, imp.
lightTooBright			0.01	7.50	keep
Occupant daylight		-	0.02	0.45	Drop
Light mean	4.33	-0.13			keep

Table B.2

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Windows					
Occupant orientation			0.09	7.46	drop
Office orientation			0.021	3.40	drop
Window area			0.02	2.68	drop
Window Shade Percentage		-	0.12	6.88	drop
Operable Windows	-	-			drop

Table B.3

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
General info					
Locattion			-	8.90	drop
Month	-	-			drop
Renovation Year		-	0.00	9.19	drop
Construction Year	9.16	0.21			drop

Table B.4

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Building envelope					
Facade type			0.03	30.93	drop
Exterior wall type			0.09	12.69	drop
Insulation type			-	-	drop

Table B.5

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Office surfaces					
Wall finish			0.012463	1.149290	drop
Ceiling finish			0.054877	5.409335	drop, corr.
Suspended ceilings			0.051444	6.087745	drop, corr.
Floor: carpet			0.014757	6.000951	drop
Floor: wood			-	2.075898	drop
Floor: plastic			0.00199	8.420173	drop
Cracked window paint			0.001990	25.632649	drop
Colour or art on walls			–	1.638494	drop
Area of deterioration on floor	-	-			drop

Table B.6

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Office setup/facilities					
Office Interior			-	5.06	drop
Persons in office			0.06	7.72	drop
Lounging area			-	5.36	drop
Plants or water features			0.06	3.32	drop
Office volume	-	-			keep, imp.

Table B.7

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Cleaning					
Desk wash frequency		**0.14**	–	14.01	drop
Environmental cleaning materials	-	-	0.03	1.57	drop
Floor mopping frequency		0.23	0.08	27.70	keep
Vacuming frequency		0.15	0.06	14.55	drop
Window washing frequency		0.18	0.12	12.83	drop

Table B.8

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Occupant					
sex			0.10	7.99	keep
age		-	0.02	2.22	drop
job			-	17.72	drop
eyeWear			0.05	7.51	drop
smoker			0.05	4.06	drop
distToWindow		-	-	3.97	drop
hoursatdesk			-	0.93	drop

Table B.9

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
HVAC					
Cooling type			-	0.85	drop
airHeat			0.07	3.91	drop
radiatorHeat			-	0.88	drop
Supply air type			0.04	4.88	drop
Return air type			-	5.33	drop
Building ventilation type			0.04	17.91	drop
CO2 mean	8.94	-0.18			keep
RH mean	-	-			keep
T mean:	-	-			Keep
Number of space heaters in use	-	-			drop
Number of supply devices	7.70	0.21			keep

Table B.10

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Exterior					
Site			0.01	2.68	drop
View to road and buildings			0.10	0.10	drop
View to green nature			0.03	1.60	keep
Heavy traffic			0.02	5.04	drop
Garbage dumpsters			0.07	4.24	drop
Power plants			-	2.42	drop
Construction activities			0.071315	17.16	drop
Emergency generators			0.056152	18.86	drop
No outdoor contaminant sources			0.01	5.00	drop

Table B.11

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Noise					
Office equipment			0.05	6.32	drop, imp.
Telephones			0.02	6.79	keep
Talk			0.01	8.49	keep
Exterior			-	12.22	drop
sound mean	-	-			drop

Table B.12

Indoor climate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Odours					
Outside			-	6.70	drop
Carpet			-	3.57	drop
Cleaning products			-	4.38	drop
Tech			0.04	11.14	drop, imp.
Perfume			-	1.673	drop
Food			-	6.42	drop
Tobacco smoke			0.02	9.80	drop, imp.

B.1.2 Performance

Table B.13

Performance	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Light					
lightFlicker			-	0.65	drop
lightReflections			-	0.37	drop
lightNoTask			0.04	0.19	drop
lightTooDark			-	1.40	drop
lightTooBright			-	4.52	drop
Occupant daylight		-	0.01	0.17	drop
Light mean	-	-			Keep

Table B.14

Performance	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Windows					
Occupant orientation			-	12.11	Drop
Office orientation			0.01	6.67	drop, imp
Window area	-	-			drop
Window Shade Percentage	-	-			drop
Operable Windows	-	-			drop

Table B.15

Performance	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi2	Action
General info					
Locattion			-	7.134	drop, corr.
Month	-	-			drop, corr.
Renovation Year		-0,21	0.170	4.417	drop, imp.
Construction Year	-	1.19			Keep

Table B.16

Performance	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi2	Action
Office surfaces					
Wall finish			0.06	1.300	drop
Ceiling finish			0,07	5.29	drop
Suspended ceilings			0.02	4.58	drop, corr.
Floor: carpet			-	6.60	drop
Floor: wood			0.09	5.93	drop, imp.
Floor: plastic			-	4.518	drop
Cracked window paint			-	9.39	drop
Colour or art on walls			-	3.03	drop
Area of deterioration on floor	-	-			drop

Table B.17

Performance	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Building envelope					
Facade type			0.08	14.83	drop
Exterior wall type			-	7.57	drop
Insulation type			-	9.11	drop

Table B.18

Performance	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Office setup/facilities					
Office Interior			0.07	6.61	keep
Persons in office		-0.11	0.08	5.54	drop
Lounging area			0.01	3.77	drop, imp.
Plants or water features			0.14	1.49	drop, imp.
Office volume	-	-			drop

Table B.19

Performance	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Cleaning					
Desk wash frequency		**0.15**	0.14	25.00	drop
Environmental cleaning materials		-	0.04	10.01	drop
Floor mopping frequency		0.19	0.06	26.80	keep
Vacuming frequency		0.16	0.06	25.53	drop
Window washing frequency		0.21	0.12	33.57	drop

Table B.20

Performance	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared	Action
Occupant					
sex			0.02	6.60	keep
age		0.03	0.05	7.95	keep
job			0.02	22.73	drop
eyeWear			0.08	4.11	keep
smoker			-	4.30	drop
distToWindow			-	8.28	drop
hoursatdesk		-	0.02	0.75	drop

Table B.21

Performance	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
HVAC					
Cooling type			0.06	0.89	drop, corr.
airHeat			-	4.60	drop
radiatorHeat			-	1.25	drop
Supply air type			0.10	4.18	drop
Return air type			0.115	7.58	drop
Building ventilation type			0.06	9.36	drop
CO2 mean	3.18	-0.18			keep
RH mean	-	-			keep
T mean:	-	-			keep
Number of space heaters in use	-	-			drop
Number of supply devices	-	0.20			Keep

Table B.22

Performance	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Exterior					
Site			-	2.78	drop
View to road and buildings			0.04	0.25	drop
View to green nature			0.06	2.84	drop
Heavy traffic			0.10	2.49	drop, imp.
Garbage dumpsters			0.06	5.22	drop, imp.
Power plants			-	4.46	drop
Construction activities			-	12.58	drop
Emergency generators			0.04	5.38	drop, imp.
No outdoor contaminant sources			-	8.16	drop

Table B.23

Performance	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Noise					
Office equipment			-	9.22	drop
Telephones			0.06	13.08	drop
Talk			0.07	21.59	keep
Exterior			0.09	14.17	drop, imp.
sound mean	-	-			drop

Table B.24

Performance	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Odours					
Outside			0.07	6.92	drop, imp.
Carpet			-	8.49	drop
Cleaning products			0.02	3.55	drop, imp.
Tech			0.10	17.25	drop
Perfume			-	4.30	drop
Food			-	3.42	drop
Tobacco smoke			-	4.75	drop

B.1.3 Office-related symptoms

Table B.25

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Light					
lightFlicker			-	0.160256	drop
lightReflections			-	3.111	drop
lightNoTask			-	0.308	drop
lightTooDark			-	7.392308	keep
lightTooBright			-	6.001012	drop
Occupant daylight		-	0.020	0.039	Drop, imp
Light mean	-	-			keep

Table B.26

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Windows					
Occupant orientation		-	0.0151	3.062	drop
Office orientation			0.046	4.212	drop
Window area	-	-			drop
Window Shade Percentage	-	-			drop
Operable Windows	-	-			drop

Table B.27

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
General info					
Locattion		-	-	0.708	drop
Month	-	-			drop
Renovation Year			0.006	0.834	drop
Construction Year	6.146	-0.158			keep

Table B.28

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Building envelope					
Facade type			0.006	4.763	drop
Exterior wall type		-	-	2.561	drop
Insulation type		-	-	0.163	drop

Table B.29

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Office surfaces					
Wall finish				0.19	drop
Ceiling finish			-	6.75	drop, corr.
Suspended ceilings			0.01	1.71	drop, corr.
Floor: carpet			0.01	1.83	drop
Floor: wood			-	0.05	drop
Floor: plastic			0.02	3.85	drop
Cracked window paint			0.01	2.46	drop
Colour or art on walls		-	-	0.04	drop
Area of deterioration on floor	-	-			drop

Table B.30

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Office setup/facilities					
Office Interior			0.001	1.20	drop
P ersons in office			0.001	0.89	drop
Lounging area			0.02	2.52	drop
Plants or water features			0.02	0.15	drop
Office volume	-	-			drop

Table B.31

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Cleaning					
Desk wash frequency		-	-	1.69	drop
Environmental cleaning materials		-	-	0.82	drop
Floor mopping frequency		-0.16	-	8.29	keep
Vacuming frequency		-	-	1.43	drop
Window washing frequency		-	0.01	1.97	drop

Table B.32

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Occupant					
sex			0.04	1.50	keep
age		-0.128	-	4.00	keep
job			0.00	5.65	drop, imp
eyeWear			-	0.065	drop
smoker			-	0.00	drop
distToWindow			-	0.14	drop
hoursatdesk			-	0.26	drop

Table B.33

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
HVAC					
Cooling type			0.02	2.55	drop
airHeat			-	0.00	drop
radiatorHeat			-	0.34	drop
Supply air type			0.03	1.82	drop
Return air type			0.05	1.02	drop
Building ventilation type			-	0.04	drop
CO2 mean	4.84	-			keep
RH mean	-				keep
T mean:	5.79	-0.12			keep
Number of space heaters in use	-				drop
Number of supply devices	-				keep

Table B.34

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Exterior					
Site			0.02	0.89	drop
View to road and buildings			-	0.02	drop
View to green nature			0.01	0.78	drop
Heavy traffic			-	0.41	drop
Garbage dumpsters			-	0.15	drop
Power plants			-	0.23	drop
Construction activities			0.02	0.00	drop
Emergency generators			-	3.47	drop
No outdoor contaminant sources			0.02	0.01	drop

Table B.35

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Noise					
Office equipment			0.02	12.00	keep
Telephones			0.02	3.22	drop
Talk			0.03	10.55	keep
Exterior			-	9.72	drop
sound mean	-	-			drop

Table B.36

Office-related symptoms	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi2	Action
Odours					
Outside			-	8.94	drop
Carpet			-	1.23	drop
Cleaning products			0.05	0.31	drop
Tech			-	8.08	drop
Perfume			-	1.56	drop
Food			-	4.38	drop
Tobacco smoke			0.06	0.87	drop

B.1.4 Single office-related symptoms: Dry or irritated eyes

Table B.37

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Light quality					
lightFlicker			0.02	0.80	drop
lightReflections			0.01	3.84	drop
lightNoTask			0.01	0.18	drop
lightTooDark			-	1.79	drop
lightTooBright			0.01	8.90	drop
Occupant daylight		-	-	0.23	drop
Light mean	-	-			keep

Table B.38

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Windows					
Occupant window orientation			0.01	5.55	drop
Office window orientation			0.01	2.51	drop
Window area	-	-			drop
Window shade percentage	-	-			drop
Operable windows	-	-			drop

Table B.39

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
General info					
Location			-	3.67	drop
Month	-	-			drop
Renovation year		-	-	0.58	drop
Construction year	-	-			drop

Table B.40

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Office surfaces					
Wall finish			0.013	0.22	drop
Ceiling finish			0.03	2.98	drop
Suspended ceiling			0.02	1.92	drop
Floor: carpet			0.01	1.92	drop
Floor: wood			-	0.56	drop
Floor: plastic			0.02	2.74	drop
Cracked window paint			0.01	7.35	drop
Colour or art on walls			0.01	0.09	drop
Area of deterioration on floor	6.21	-			drop

Table B.41

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Office facilities/setup					
Office Interior			0.01	3.72	drop
Persons in office:			-	2.60	drop
Lounging area			0.02	2.85	drop
Plants or water features			0.02	0.17	drop
Office volume	-	-			keep, imp.

Table B.42

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Cleaning					
Desk wash frequency		-	-	2.29	drop
Environmental cleaning materials		-	-	0.03	drop
Floor mopping frequency		-	0.04	4.29	drop
Vacuming frequency		-	0.00	1.99	drop
Window wash frequency		-	0.01	1.36	drop

Table B.43

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Occupant					
sex			0.05	1.16	keep
age		-	0.01	1.00	drop
job			-	4.28	drop
Eye wear			-	1.05	keep, imp.
smoker			-	0.94	drop
distToWindow		-	-	0.38	drop
hoursatdesk		-	-	0.02	drop

Table B.44

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
HVAC					
Cooling type			0.01	0.01	drop
airHeat			0.00	2.42	drop
radiatorHeat			-	0.07	drop
Supply air type			-	0.01	drop
Return air type			-	2.13	drop
Building ventilation type			0.01	5.78	drop
CO2 mean	3.80	-			keep
RH mean	5.40	-			keep
T mean	6.96	-0.13			keep
Number of space heaters	-	-			drop
Number of supply devices	-0.13	-			drop

Table B.45

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Exterior					
Site			0.03	1.55	drop
View to road and buildings			-	0.01	drop
View to green nature			0.01	0.93	drop
Heavy traffic			-	0.00	drop
Garbage dumpsters			-	0.37	drop
Power plants			-	0.20	drop
Construction activities			-	1.63	drop
Emergency generators			-	5.15	drop
No outdoor contaminant sources			-	0.41	drop

Table B.46

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Noise					
Office equipment			0.09	23.92	drop
Telephones			0.01	2.96	drop
Talk			0.02	7.47	drop
Exterior			0.03	16.17	drop
Sound mean	-	-			drop

Table B.47

Dry or irritated eyes	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Odours					
Outside			0.01	7.20	drop
Carpet			0.01	0.70	drop
Cleaning products			-	0.18	drop
Tech			0.02	6.77	drop
Perfume			-	3.88	drop
Food			-	8.16	drop
Tobacco smoke			-	0.64	drop

B.1.5 Single office-related symptom: Headache

Table B.48

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Light quality					
lightFlicker			-	0.65	drop
lightReflections			-	0.37	drop
lightNoTask			0.04	0.19	drop
lightTooDark			-	1.40	drop
lightTooBright			-	4.52	drop
Occupant daylight		-	0.01	0.17	drop
Light mean	-	-			drop

Table B.49

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Windows					
Occupant window orientation			-	0.04	drop
Office window orientation			0.02	9.14	drop
Window area	-	-			drop
Window shade percentage	-	-			drop
Operable windows	-	-			drop

Table B.50

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
General info					
Location			0.06	5.43	drop
Month		-0.14			drop
Renovation year		-	0.03	1.99	drop
Construction year	7.60	-0.14			keep

Table B.51

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Office surfaces					
Wall finish			-	0.02	drop
Ceiling finish			-	-	drop
Suspended ceiling			0.04	4.94	drop
Floor: carpet			0.04	5.13	drop
Floor: wood			-	0.22	drop
Floor: plastic			0.02	6.86	drop
Cracked window paint			0.01	3.74	drop
Colour or art on walls			-	0.22	drop
Area of deterioration on floor	5.77	-			drop

Table B.52

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Office facilities/setup					
Office Interior			0.03	0.52	drop, imp
Persons in office:			-	0.10	drop
Lounging area			-	5.37	drop
Plants or water features			0.01	0.00	drop
Office volume	-	-			drop

Table B.53

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Cleaning					
Desk wash frequency		-	0.01	1.04	drop
Environmental cleaning materials		-	0.01	0.05	drop
Floor mopping frequency		-	0.00	3.11	drop
Vacuming frequency		-	-	0.63	drop
Window wash frequency		-	0.02	1.91	drop

Table B.54

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Occupant					
sex			0.01	4.61	keep
age		-0.17	0.02	7.58	keep
job			-	4.89	drop
Eye wear			0.04	0.39	drop
smoker			-	0.22	drop
distToWindow		-	-	0.11	drop
hoursatdesk		-	-	0.24	drop

Table B.55

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
HVAC					
Cooling type			0.02	0.01	drop
airHeat			0.02	0.01	drop
radiatorHeat			-	0.03	drop
Supply air type			0.01	0.34	drop
Return air type			-	1.26	drop
Building ventilation type			0.01	1.56	drop
CO2 mean	8.79	0.14			keep
RH mean	-	-			drop
T mean	-	-			drop
Number of space heaters	-	-			drop
Number of supply devices	-	-0.13			drop, imp

Table B.56

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Exterior					
Site			0.01	1.77	drop
View to road and buildings			0.04	0.01	drop
View to green nature			-	0.59	drop
Heavy traffic			0.01	0.00	drop
Garbage dumpsters			-	2.55	drop
Power plants			-	1.40	drop
Construction activities			-	0.00	drop
Emergency generators			0.01	8.40	drop
No outdoor contaminant sources			-	0.68	drop

Table B.57

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Noise					
Office equipment			0.03	8.39	drop, imp.
Telephones			-	2.19	drop, imp.
Talk			0.03	1.90	drop, imp.
Exterior			0.01	1.97	drop, imp.
Sound mean	-	-			drop

Table B.58

Headache	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Odours					
Outside			0.03	9.36	drop, imp
Carpet			0.01	0.77	drop, imp
Cleaning products			-	0.19	drop
Tech			0.06	5.71	drop, imp
Perfume			-	3.36	drop
Food			-	1.77	drop
Tobacco smoke			0.03	6.71	drop, imp.

B.1.6 Single office-related symptoms: Tiredness or fatigue

Table B.59

Tiredness or fatigue	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Light quality					
lightFlicker			-	0.35	drop
lightReflections			-	0.14	drop
lightNoTask			0.03	0.25	drop
lightTooDark			-	2.59	drop
lightTooBright			0.01	2.45	drop
Occupant daylight		-		0.00	drop
Light mean	-	-			keep

Table B.60

Tiredness or fatigue	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Windows					
Occupant window orientation			0.02	6.03	drop
Office window orientation			-	5.36	drop
Window area	-	-			drop
Window shade percentage	-	-			drop
Operable windows	-	-			drop

Table B.61

Tiredness or fatigue	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
General info					
Location			0.03	2.41	drop
Month	-	-			drop, corr.
Renovation year		0.18	0.02	6.43	drop, imp.
Construction year	-	-			keep

Table B.62

Tiredness or fatigue	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Office surfaces					
Wall finish			-	0.02	drop
Ceiling finish			0.00	0.39	keep
Suspended ceiling			0.03	1.53	drop
Floor: carpet			0.00	3.93	drop
Floor: wood			-	0.29	drop
Floor: plastic			-	4.02	drop
Cracked window paint			-	0.72	drop
Colour or art on walls				0.01	drop
Area of deterioration on floor	5.30	0.18			drop

Table B.63

Tiredness or fatigue	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Office facilities/setup					
Office Interior			-	0.43	drop
Persons in office:			-	0.29	drop
Lounging area			-	0.07	drop
Plants or water features			-	0.16	drop
Office volume	-	-			drop

Table B.64

Tiredness or fatigue	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Cleaning					
Desk wash frequency		-	0.01	0.59	drop
Environmental cleaning materials		-	-	0.06	drop
Floor mopping frequency		-	-	0.62	drop
Vacuming frequency		-	-	0.81	drop
Window wash frequency		-	-	0.90	drop

Table B.65

Tiredness or fatigue	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Occupant					
sex			-	1.91	drop
age		-0.15	0.02	6.00	keep
job			-	1.04	drop
Eye wear			0.06	0.98	keep
smoker			0.02	0.36	drop
distToWindow		-	0.04	0.12	drop
hoursatdesk		-	0.03	0.03	drop

Table B.66

Tiredness or fatigue	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
HVAC					
Cooling type			-	0.14	drop
airHeat			0.03	1.6	drop
radiatorHeat			-	0.41	drop
Supply air type			-	0.01	drop
Return air type			-	0.65	drop
Building ventilation type			0.02	1.71	drop
CO2 mean	-	-			keep
RH mean	-	-			drop, imp.
T mean	3.94	-0.12			keep
Number of space heaters	-	-			drop
Number of supply devices	-	-			drop

Table B.67

Tiredness or fatigue	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Exterior					
Site			0.03	0.63	keep
View to road and buildings			0.04	0.02	drop
View to green nature			-	0.02	drop
Heavy traffic			-	1.31	drop
Garbage dumpsters			0.04	1.28	drop
Power plants			-	0.66	drop
Construction activities			-	0.94	drop
Emergency generators			0.01	4.68	drop
No outdoor contaminant sources				0.21	drop

Table B.68

Tiredness or fatigue	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Noise					
Office equipment			0.04	15.95	keep
Telephones			-	3.63	keep
Talk			-	6.77	keep
Exterior			-	6.23	keep
Sound mean	-	-			drop

Table B.69

Tiredness or fatigue	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Odours					
Outside			0.01	9.16	keep
Carpet			-	0.07	drop
Cleaning products			0.03	0.25	drop
Tech			-	3.64	drop
Perfume			0.01	2.31	drop
Food			0.04	4.17	drop
Tobacco smoke			-	1.59	drop

B.1.7 Single office-related symptoms: Difficult to concentrate

Table B.70

Tiredness or fatigue	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Light quality					
lightFlicker			-	0.65	drop
lightReflections			0.03	5.17	drop
lightNoTask			-	0.19	drop
lightTooDark			-	0.104655	**drop**
lightTooBright			0.04	15.22	drop
Occupant daylight		-	0.03	0.35	drop
Light mean	-	-			keep

Table B.71

Difficult to concentrate	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Windows					
Occupant window orientation			-	0.46	drop
Office window orientation			0.01	2.86	drop
Window area	-	-			drop
Window shade percentage	-	-			drop
Operable windows	-	-			drop

Table B.72

Difficult to concentrate	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
General info					
Location			-	2.25	drop
Month	-	-		1.12	drop
Renovation year		-	-		drop
Construction year	5.72	-0.11			keep

Table B.73

Difficult to concentrate	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Office surfaces					
Wall finish			0.013	0.02	drop
Ceiling finish			0.01	0.49	drop
Suspended ceiling			0.00	2.22	drop
Floor: carpet			0.01	0.84	drop
Floor: wood			0.01	0.01	drop
Floor: plastic			0.02	3.85	drop
Cracked window paint			-	3.74	drop
Colour or art on walls			-	0.07	drop
Area of deterioration on floor	6.64	0.17			drop

Table B.74

Difficult to concentrate	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Office facilities/setup					
Office Interior			0.02	0.52	drop
Persons in office:			-	1.50	drop
Lounging area			0.02	1.60	drop
Plants or water features			-	0.13	drop
Office volume	-	-			drop

Table B.75

Difficult to concentrate	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Cleaning					
Desk wash frequency		-	0.05	0.31	drop
Environmental cleaning materials		-	-	0.17	drop
Floor mopping frequency		-	0.04	1.28	drop
Vacuming frequency		-	-	0.24	drop
Window wash frequency		-	-	0.05	drop

Table B.76

Difficult to concentrate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Occupant					
sex			-	1.55	keep
age		-0.25	0.06	15.43	keep
job			-	0.68	drop
Eye wear			-	3.79	drop
smoker			0.00	1.95	drop
distToWindow		-	-	0.11	drop
hoursatdesk		-	0.01	0.12	drop

Table B.77

Difficult to concentrate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
HVAC					
Cooling type			0.01	-	drop
airHeat			0.00	0.10	drop
radiatorHeat			-	0.00	drop
Supply air type			-	0.49	drop
Return air type			-	0.08	drop
Building ventilation type			0.02	1.56	drop
CO2 mean	10.23	0.14			keep
RH mean	-	-			drop
T mean	-	-			drop
Number of space heaters	-	-			drop
Number of supply devices	-	-			drop

Table B.78

Difficult to concentrate	ANOVA F value	Kendall's Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Exterior					
Site			-	1.05	drop
View to road and buildings			-	0.01	drop
View to green nature			-	0.59	drop
Heavy traffic			0.01	0.06	drop
Garbage dumpsters			-	0.42	drop
Power plants			-	0.10	drop
Construction activities			-	-	drop
Emergency generators			0.07	5.95	drop
No outdoor contaminant sources			-	0.55	drop

Table B.79

Difficult to concentrate	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Noise					
Office equipment			-	4.16	keep
Telephones			0.03	21.81	keep
Talk			0.08	20.94	keep
Exterior			0.06	18.46	keep
Sound mean	-	-			keep

Table B.80

Difficult to concentrate	ANOVA F value	Kendall’s Tau, tau	Mutual information, I(x;y)	Chi-squared Chi2	Action
Odours					
Outside			-	3.29	drop
Carpet			0.04	0.23	drop
Cleaning products			0.01	0.19	drop
Tech			-	-	drop
Perfume			0.01	0.23	drop
Food			0.01	3.87	drop
Tobacco smoke			-	2.66	drop

B.2 Correlation plots

In the following, heat maps are seen of the correlation between the features of each model. Here, 1 means that they are completely positively correlated while - means that they are completely inversely correlated.

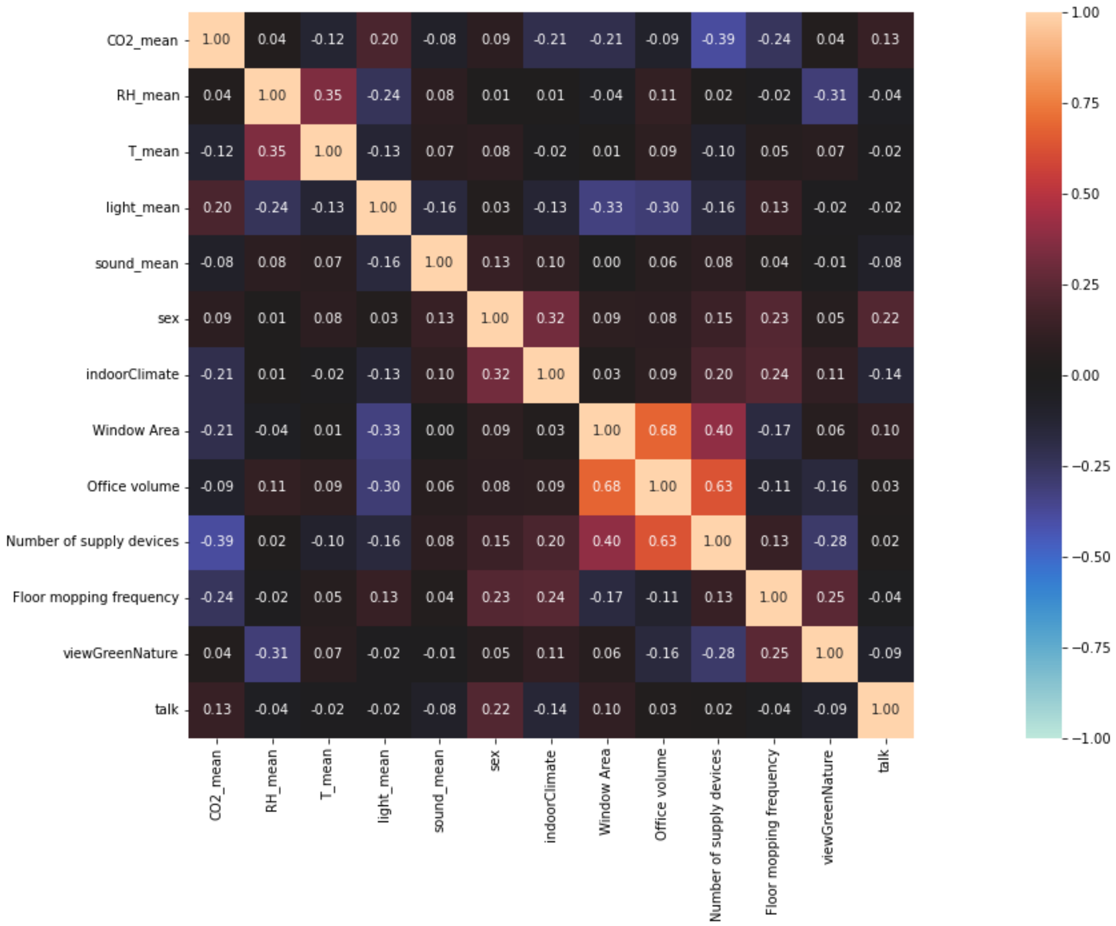


Figure B.1: Correlation of features in the indoor climate model.

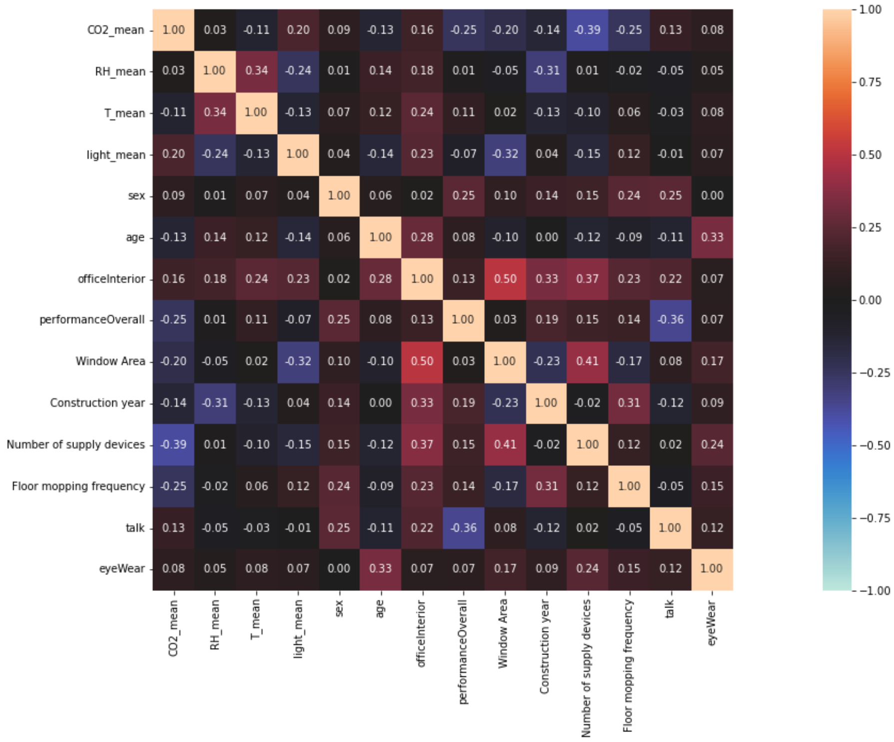


Figure B.2: Correlation of features in the performance model.

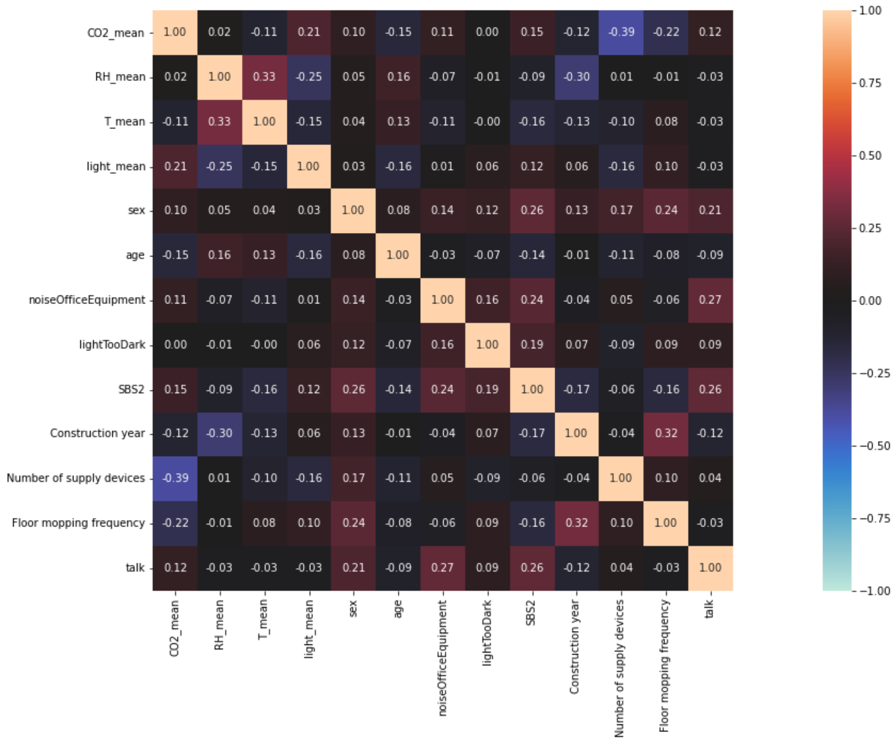


Figure B.3: Correlation of features in the office-related symptoms model.

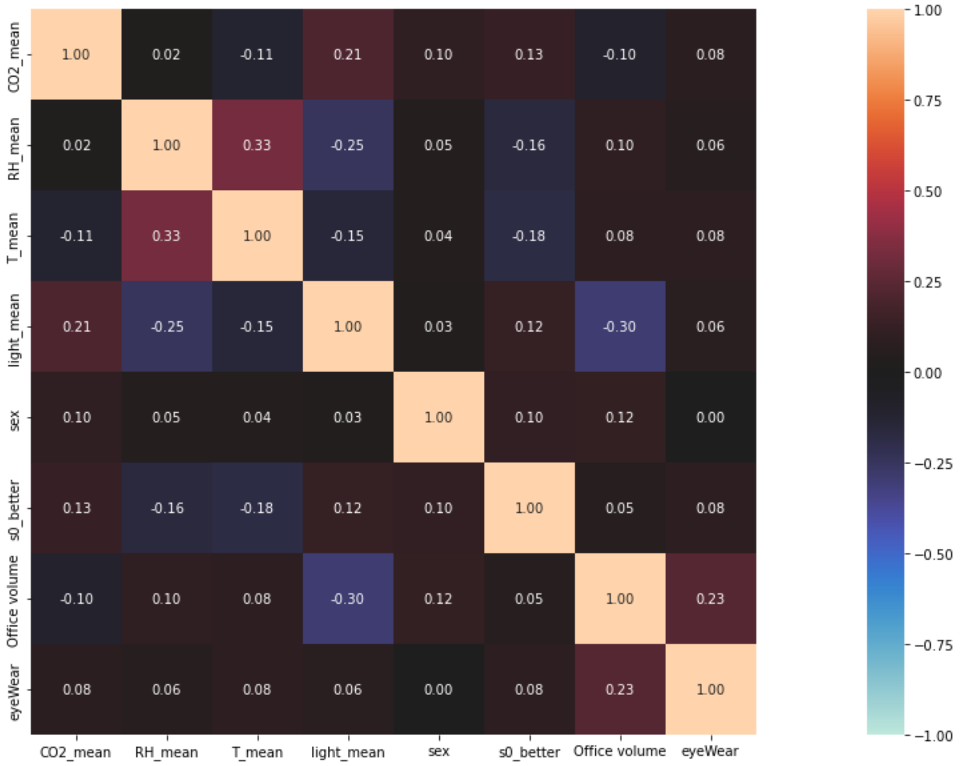


Figure B.4: Correlation of features in the single symptom model: Dry or irritated eyes.

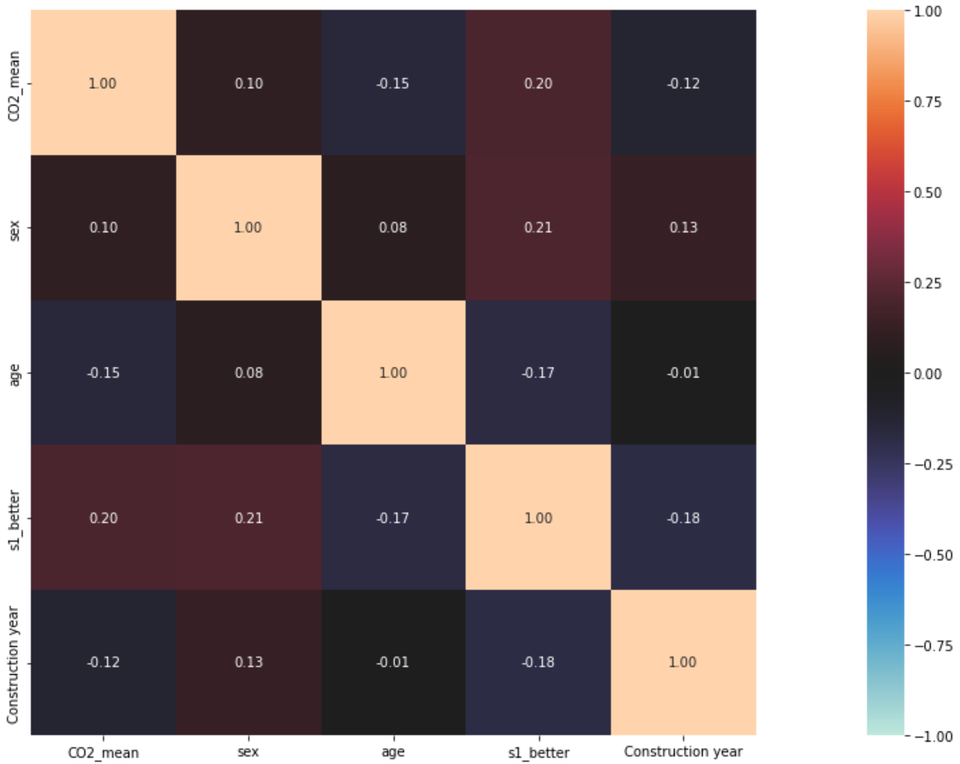


Figure B.5: Correlation of features in the single symptom model: Headache.

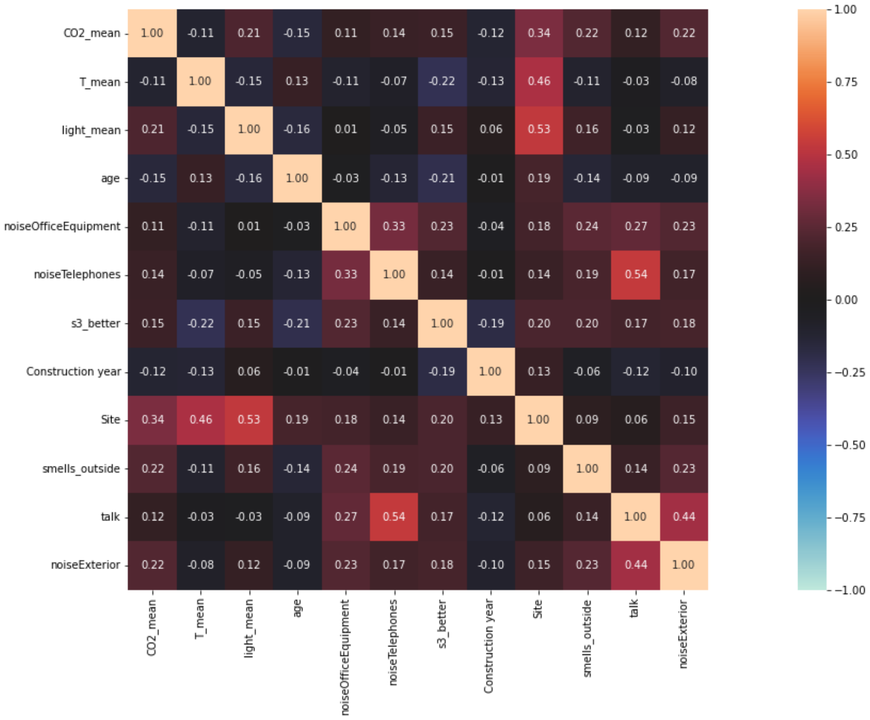


Figure B.6: Correlation of features in the single symptom model: Tiredness or fatigue.

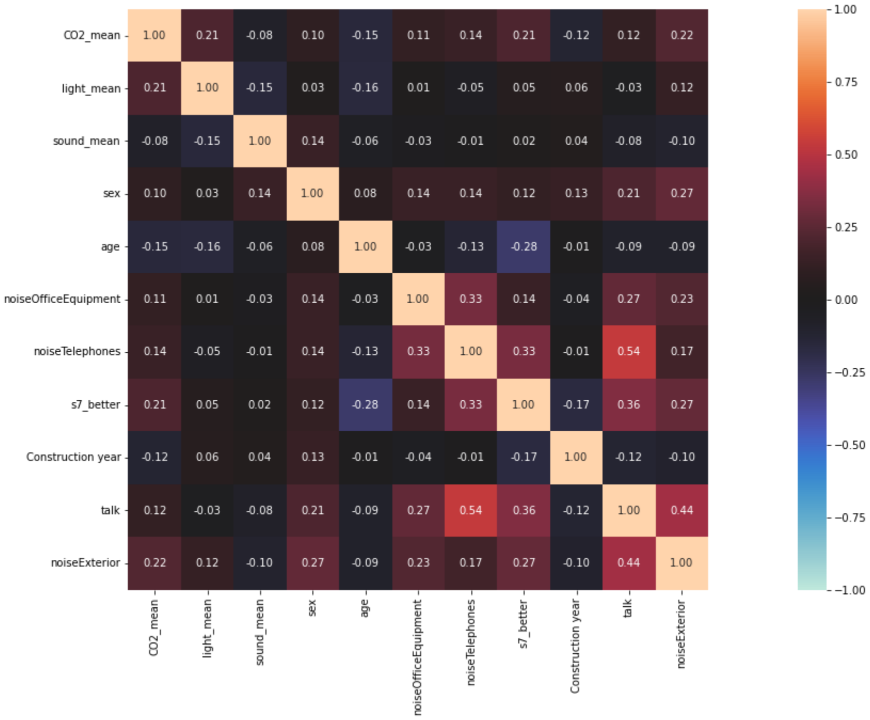
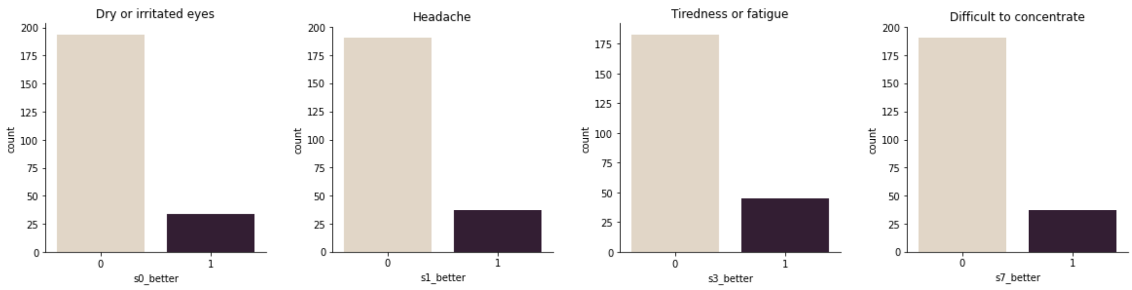


Figure B.7: Correlation of features in the single symptom model: Difficult to concentrate.

C Model selection and tuning: Single symptom models

C.1 Single symptom targets

The four building-related symptoms predicted in the single-symptom models are ”dry or irritated eyes”, ”headache”, ”tiredness or fatigue” and ”difficulty concentrating”. The balance of each symptom target class is seen in Figure C.1. Here, 1 are observations where the occupant experienced the symptom and felt better after leaving the offices, during the last four weeks.



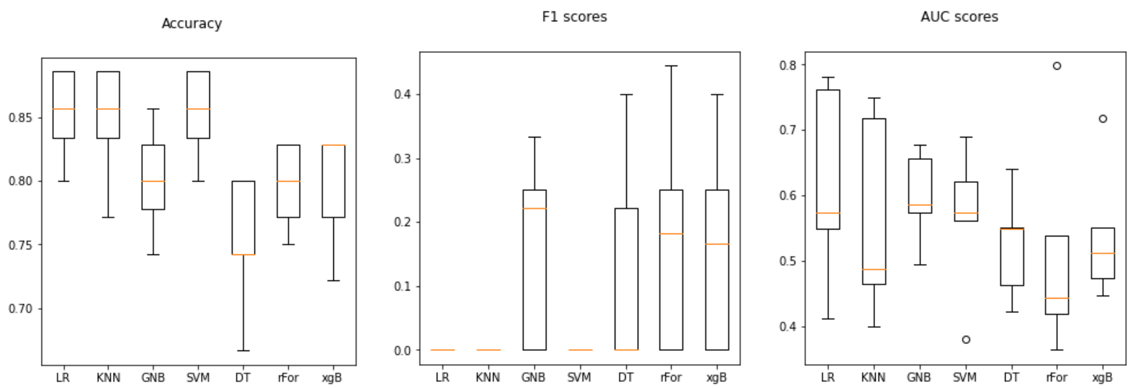
(a) Dry or irritated eyes (b) Headache (c) Tiredness or fatigue (d) No concentration

Figure C.1: Target balance in single symptom models.

In the following box plots of performance in each fold are shown for each of the single-symptom target models. The median is marked with orange.

C.2 Model comparison

Symptom: Dry or irritated eyes

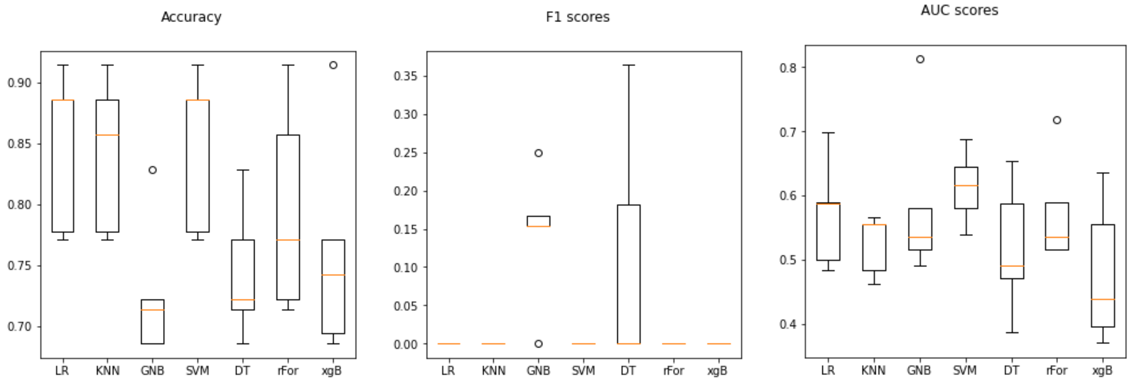


(a) Accuracy (b) f1 score (c) ROC curve and AUC

Figure C.2: Symptom: Dry or irritated eyes

In Figure C.2 it is seen that while the Gaussian Naive Bayes model has the highest median f1 score, the 25th percentile is f1 = 0. This means that while the model performs best in some folds, it is unable to predict anything in other folds. The second highest median f1 score is found using the Extreme Gradient Boost model. While still being a fairly low-performing model, it is able to predict something in all folds. Both the Gaussian Naive Bayes and the Extreme Gradient Boost model are thus tuned to further reduce over-fitting and improve model performance.

Symptom: Headache

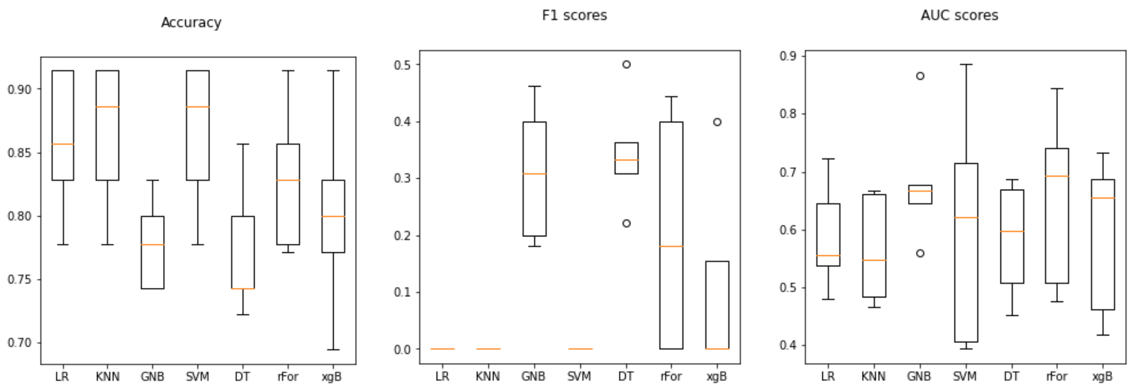


(a) Accuracy (b) f1 score (c) ROC curve and AUC

Figure C.3: Symptom: Headache

As it is seen in Figure C.3b, all models have low f1 scores. The model is, thus, very dependent on the tuning. The random forest algorithm, while being similar to the decision tree model, is known to be less prone to over-fitting and has more tuning parameters. Both models are, thus, tuned to find the best model.

Symptom: Tiredness or fatigue

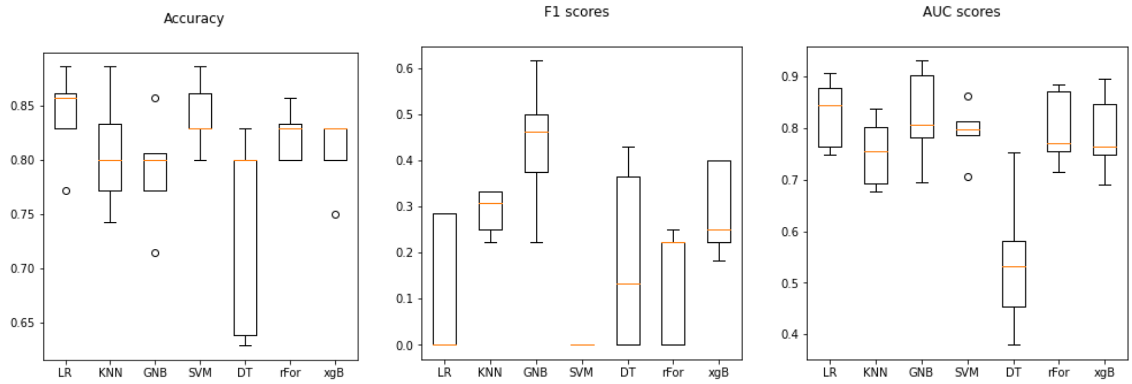


(a) Accuracy (b) f1 score (c) f1 score

Figure C.4: Symptom: Tiredness or fatigue

In Figure C.4 it is seen that the Gaussian Naive Bayes model has the highest median f1 score while maintaining an acceptable minimal f1 score. The Gaussian Naive Bayes model is, thus, tuned to further reduce over-fitting and improve model performance.

Symptom: Difficult to concentrate



(a) Accuracy (b) f1 score (c) ROC curve and AUC

Figure C.5: Symptom: Difficulty concentrating

Both the Gaussian Naive Bayes and the tree-based models are tuned to see if a better performance can be achieved.

C.3 Feature analysis

The SHAP summary plots of the prediction of the single office-related symptoms, are seen in the following. The more positive the SHAP values are, the more likely the feature is to influence the model to predict that the occupant experiences the symptom.

Symptom: Dry or irritated eyes

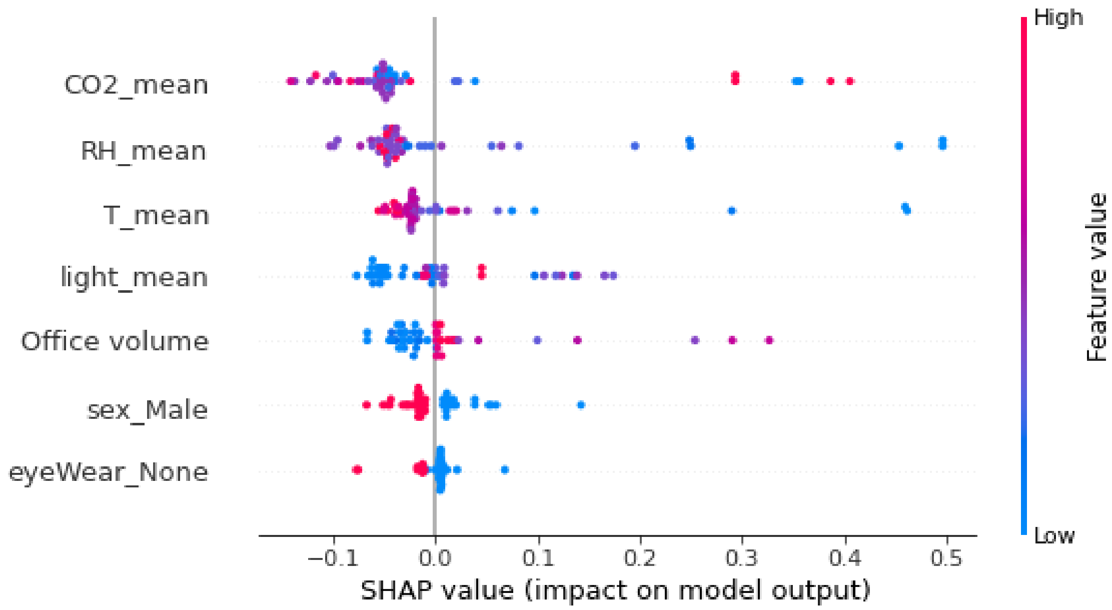


Figure C.6: SHAP summary plot of single symptom model: Dry or irritated eyes

The model shows that low relative humidity and temperatures are causing the model to predict that the occupant is experiencing dry or irritated eyes. The most influential values are very low or very high CO_2 concentration.

Symptom: Headache

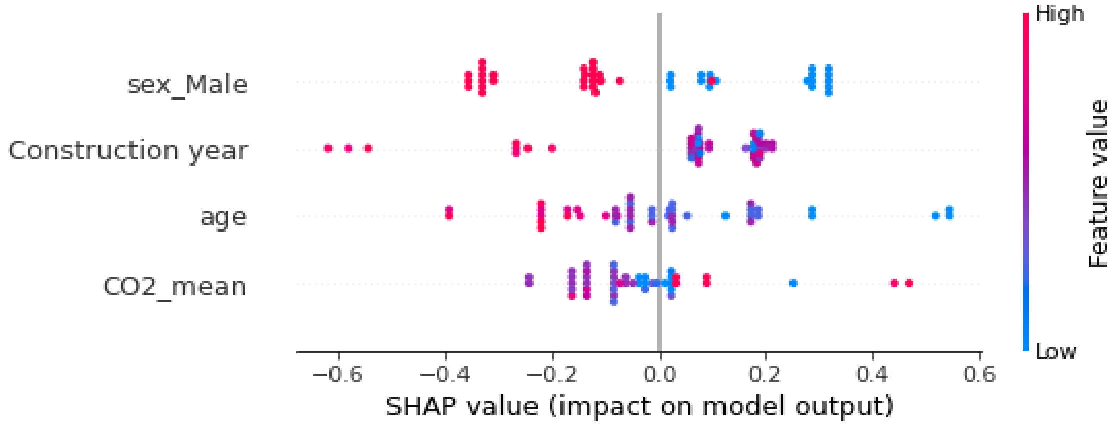


Figure C.7: SHAP summary plot of single symptom model: Headache.

In Figure C.7 it is seen that women and the youngest age group are more likely to experience office-related headaches. Furthermore, it is seen that high CO_2 concentrations are likely to cause headaches, while newer buildings are likely to lead to fewer office-related headaches.

Symptom: Tiredness or fatigue

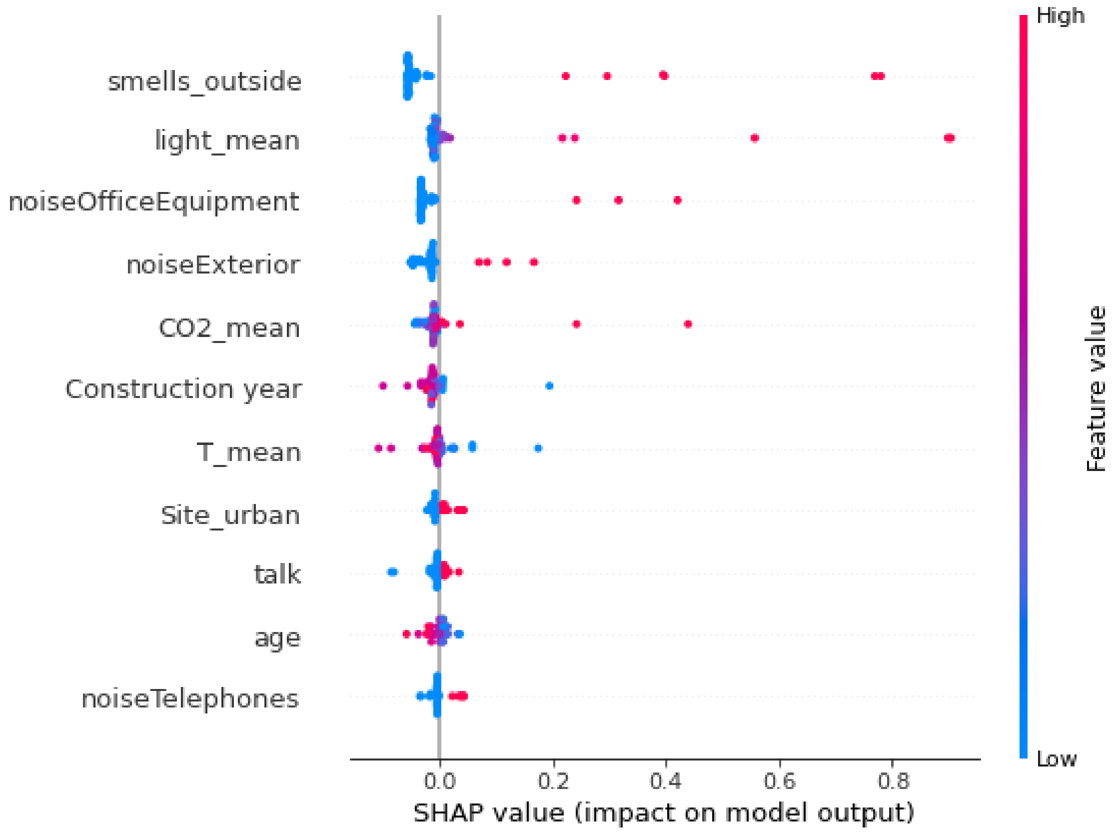


Figure C.8: SHAP summary plot of single symptom model: Tiredness or fatigue.

All noise-related features have an influence when it comes to the prediction of tiredness and fatigue.

Regarding the exterior, it is seen that an urban site of the building will be more likely cause the symptom than a suburban site. Equally, smells from the outside are a big factor.

Symptom: Difficult to concentrate

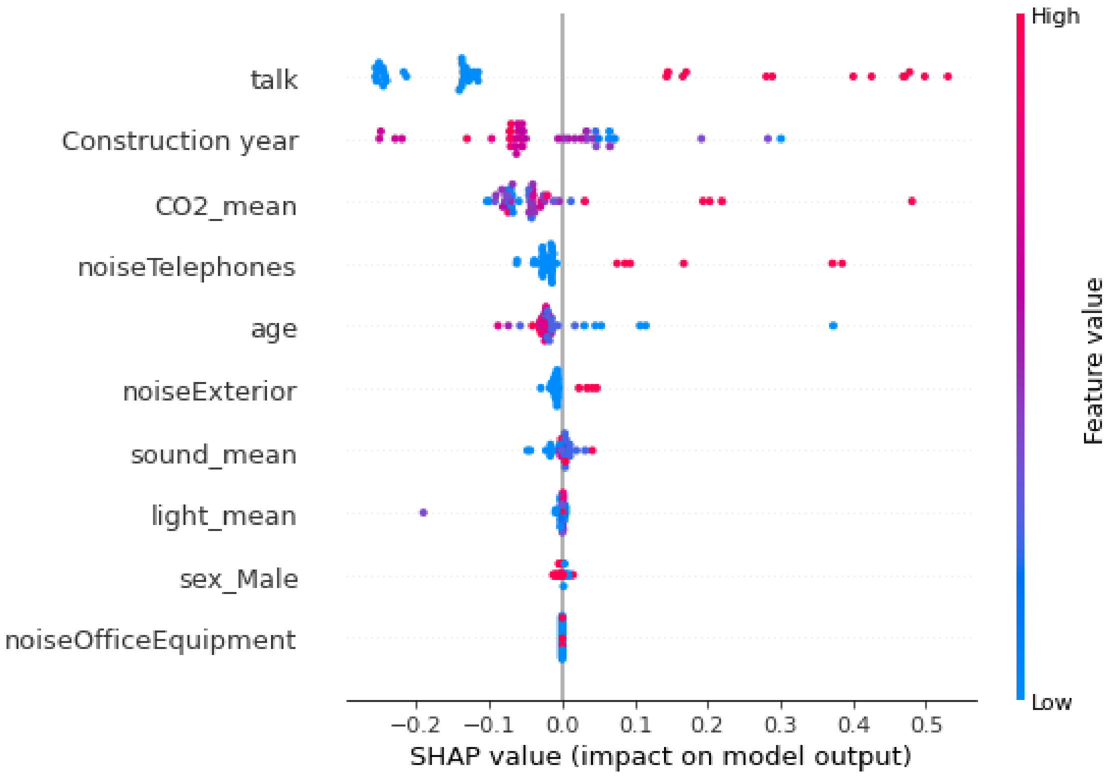


Figure C.9: SHAP summary plot of single symptom model: Difficult to concentrate.

The model shows that background noise-related features are big factors in regard to the occupant's difficulty concentrating.

We also see that older buildings and high mean CO_2 concentration are likely to cause difficulty with concentration.

Regarding the occupant-specific features, we see that young people are more likely to struggle with concentration in the office setting.