FERDINAND PORSCHE
FERN FH

Predicting and improving indoor air quality based on IoT sensor data and machine learning

Bachelorarbeit

eingereicht von: **Manuel Berger**
Matrikelnummer: 51910535

im Fachhochschul-Bachelorstudiengang Wirtschaftsinformatik (0470)
der Ferdinand Porsche FernFH

zur Erlangung des akademischen Grades eines

Bachelor of Arts in Business

Betreuung und Beurteilung: Prof. Dr. Tom Gross

Wiener Neustadt, September 2022

FERDINAND PORSCHE FERN FH

Ehrenwörtliche Erklärung

Ich versichere hiermit,

1. dass ich die vorliegende Bachelorarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Inhalte, die direkt oder indirekt aus fremden Quellen entnommen sind, sind durch entsprechende Quellenangaben gekennzeichnet.
2. dass ich diese Bachelorarbeit bisher weder im Inland noch im Ausland in irgendeiner Form als Prüfungsarbeit zur Beurteilung vorgelegt oder veröffentlicht habe.

München, 14.09.2022

Manuel Berger

Unterschrift

Die Kapitel 1 bis 4 der vorliegenden Bachelorarbeit wurden im Rahmen der Lehrveranstaltung „Bachelor Seminar 1“ eingereicht und am 31.01.2022 als Bachelorarbeit 1 angenommen.

Abstract: Predicting and improving indoor air quality based on IoT sensor data and machine learning

Indoor air quality is affected by many variables, such as temperature, humidity, particulate matter, or gases like carbon dioxide. Considering that European citizens spend, on average, 90% of their time indoors (cf. EPA 2021, WHO 2013, 9), the indoor air we breathe must be healthy and free of pollutants.

Even though most people are aware of the harmful health effects caused by indoor air pollution, many do not take the necessary steps to prevent them or to improve indoor air quality (cf. Osagbemi, Adebayo and Aderibigbe 2009).

Based on this background, a prototype was developed to help occupants of a room to improve indoor air quality. Therefore, a system was built to record crucial indoor air quality parameters. Sensor measurements were recorded every 10 seconds over five months, from May 2022 until September 2022. The focus was on the carbon dioxide concentration, a parameter easy to measure that increases when humans exhale air (cf. Palmer, 2015) into the room and that decreases when windows or doors are opened.

One crucial question was whether machine learning could reliably predict indoor air quality parameters such as CO2 to notify room occupants at the right moment, for example, before a parameter exceeds its limit or in further development to improve heating costs in winter. Therefore a machine learning model was trained with the historical recordings of CO2 concentrations and covariates such as the state of open doors and windows.

The other, more important question was whether the system's suggestions could help room occupants to improve indoor air quality significantly and efficiently in a user-friendly and practical way. As a result, a notification- and recommendation service, a web application, and a voice interface were developed.

Three methods were used to find answers: Historical forecasting to test the performance and accuracy of the machine learning model, a user survey to receive feedback about practicality and user-friendliness, and an experiment to evaluate whether the system can help improve indoor air quality. Based on previously recorded sensor data, historical forecasting showed a prediction accuracy of 94.58%. The user survey concluded that 96.9% of participants found the system's suggestions user-friendly and practical. The experiment showed that indoor air quality significantly improved thanks to the system's notifications and recommendations.

The results show an excellent potential of the system and an overall positive acceptance of users towards IoT devices and sensor data helping to improve indoor air quality.

Keywords:

Indoor air quality, Machine learning, IoT, CO2, sensors, AWS, Virtual Assistant, Cloud

Table of Contents

1. Introduction

1.1 Background and motivation

When we think of air pollution, we often only consider the outdoor environment, such as emissions from combustion engines, pollution from industrial buildings, or windblown dust. As a result, we tend to forget about the air quality indoors, where we spend most of our time.

According to a report from the U.S. Environmental Protection Agency (EPA) and a study from the WHO, most U.S. and European citizens spend, on average, 90% of their time indoors, which is almost 22 hours a day (cf. EPA 2021, WHO 2013, 9). Unfortunately, this comes with a range of health risks. One of these risks is poor indoor air quality, where some pollutants' concentration is often 2 to 5 times higher than outdoors (cf. EPA 2021). That often is by many factors, such as high carbon dioxide levels, combustion byproducts like particulate matter, or various volatile organic compounds from personal care products or house cleaners. Indoor air may also contain allergens such as dust, pet dander, and mold, or it may just be too humid or have a temperature causing discomfort (cf. WHO 2013).

Exposure to these pollutants can cause headaches or fatigue. In addition, it can be associated with irritation of the eyes or nose and even cause respiratory diseases like asthma, heart diseases, or cancer in the long term (cf. EPA 2021).

A study from 2008 revealed that despite the majority being aware of harmful health effects caused by indoor air pollution, many still engage in high-risk behaviors leading to it. They concluded that raising awareness of indoor air pollution remains one of the most pragmatic ways to prevent and reduce indoor air pollutants (cf. Osagbemi, Adebayo and Aderibigbe 2009).

Therefore, as indoor air quality has such an impact on human health and creating awareness plays a vital role, the motivation of this study is to find a way to help people improve indoor air quality with the help of modern technology that interacts with them. Furthermore, it is to determine whether it is possible to predict and improve indoor air quality reliably. The main idea is that a self-learning system will warn about any high concentration of certain pollutants and recommend efficient actions (e.g., opening a window only for the time necessary) for improvement before pollutants reach a level where they impose a health risk.

1.2 Overview of Indoor Air Quality

This part gives a brief introduction and overview of several parameters relevant to indoor air quality. Chapter 3.1 later provides a more detailed explanation of these parameters and related health effects.

Two critical parameters of indoor air quality are temperature and humidity. Germany's Federal Environmental Agency recommends maintaining a temperature between 20 and 23 degrees and keeping the relative humidity between 30 to 65 percent for healthier living (cf. BfS and UBA 2015, 7-8).

As humans, we breathe approximately 11.000 liters of air daily at rest breath. When doing sports, this number increases (cf. Wells 2012). So, we must make sure that each liter of the air we inhale is not harming our health and is free of pollutants. For example, we emit carbon dioxide (CO2) into the room when exhaling, which is considered a pollutant on its own (cf. Von Pettenkofer 1858). A too high concentration of it can lead to headache, fatigue, and sleepiness (cf. Azuma, et al. 2018). Therefore, it is necessary to ventilate the rooms regularly, ideally multiple times a day (cf. BfS and UBA 2015).

Several toxic gases appear indoors, like carbon monoxide (CO), nitrogen dioxide (NO2), or benzene, released by gas kitchen stoves, cigarette smoke, or candles. Some of them are even carcinogenic (cf. WHO 2010). Therefore, it is vital to find sources of these pollutants and remove or limit them (cf. EPA 2021).

Particulate matter (PM) are particles in the air that vary in size and shape, such as dust, smoke, and drops of liquid. There are various categories of particulate matter: Coarse particles that usually get stopped in our lungs (PM10), fine particles (PM2.5), and ultrafine particles (PM1) (cf. UBA, Feinstaub 2021). Inhaling them may impact the health of our respiratory system, heart, and other organs (cf. Du, et al. 2016). Some particles could even reach our brain and directly damage neurons (cf. Kim, et al. 2020). In addition, they might cause cognitive decline, and there's evidence that particulate matter is a risk factor for dementia (cf. Kim, et al. 2020). Opening a window next to a busy road during rush hour may increase the levels of particulate matter indoors, as these air pollutants enter the building (cf. Miller, et al. 2017).

Volatile organic compounds are organic chemicals emitted as gases into the air. They could come from personal care products, perfumes, air fresheners, or cleaning agents (cf. EPA, Volatile Organic Compounds' Impact on Indoor Air Quality 2021). It would be best to ventilate after using products causing a surge in these pollutants. Inhaling such organics could damage your liver, kidney, and central nervous system (cf. EPA, Volatile Organic Compounds' Impact on Indoor Air Quality 2021). In addition, some organics cause cancer in animals, while some may even cause cancer in humans (cf. EPA, Volatile Organic Compounds' Impact on Indoor Air Quality 2021).

Besides the parameters mentioned above, there are also pollutants of natural origin. These include, for example, radon, pet dander, pollen, and mold (cf. EPA 2021). Most of them are harmful to our lungs and cause shortness of breath, coughing, headaches, and fatigue (cf. EPA 2021). Radon, for instance, is a human carcinogen, and according to the WHO, a leading cause of lung cancer (cf. WHO 2021). It is the consequence of the radioactive decay of natural uranium. Moreover, it appears almost everywhere indoors, as radon-containing soil air enters buildings from the subsoil (cf. EPA, Radionuclide Basics: Radon 2021).

To briefly summarize this introduction, many different parameters are essential to good and healthy indoor air quality. It shows us that it is not enough to maintain only the temperature and humidity at a comfortable level and open a window occasionally. It is vital to keep a close eye on the bigger picture, including pollutants, as some may cause severe damage to human health.

1.3 Objective, research question, and hypothesis

There's an objective, a research question, and a hypothesis. Specific methods must prove the latter. The motivation is to bring awareness about indoor air quality to people through modern technology and reduce concentrations of harmful pollutants before they impose a health risk. Therefore, it would be necessary to predict the Development of specific pollutants to provide recommendations with the least effort in time and at the right moment before. A self-learning system based on machine learning will be the foundation for reaching this objective.

Objective

First, it must be evaluated if a system based on machine learning and IoT sensor data can reliably monitor and predict indoor air quality. One of such a system's key features is that predictions would be unreliable at first. However, they will gradually improve based on growing monitoring data, previous successes, and errors. The primary data source will be sensors installed indoors. Although, outdoor air parameters like weather forecasts may benefit the predictions and should be investigated as the outdoor environment and weather will impact indoor air quality.

As a result, the objective of this study is to evaluate if a self-learning system can reliably predict indoor air quality and provide relevant and user-friendly suggestions to improve indoor air quality significantly and efficiently.

Research Question

Considering the objective, the research question to be answered is as follows:

> "Is it possible to reliably predict indoor air quality based on machine learning and IoT sensor data to provide applicable and user-friendly suggestions to the occupants of a room to significantly and efficiently improve indoor air quality?"

Hypothesis

The following hypotheses can be derived from the research question:

> A self-learning system based on machine learning and IoT sensor data can reliably[1] predict indoor air quality parameters.
>
> A self-learning system can provide practical and user-friendly[2] suggestions to the occupants of a room to significantly[3] and efficiently[4] improve indoor air quality.

Now, the goal, the research question, and the hypotheses are clear. Therefore, the following chapters will provide some theoretical background, explain how to solve the problem, verify the two hypotheses and finally answer the research question.

---

[1] Reliable: a prediction is at least 75% equal to or close to the value that occurred.

[2] User-friendly: the user is satisfied with the result and the system's usability (e.g., expressed by a survey).

[3] Significant: comparison of data over a certain period once with and once without activated suggestions.

[4] Efficient: energy efficiency (e.g., heating), but also improved ventilation times or amount of user prompts.

2. State of the science

This chapter analyzes the state of the science and research already conducted in this direction. Furthermore, there's a brief introduction to consumer electronics available on the market.

2.1 Studies and approaches on predicting and improving indoor air quality

Using Deep Learning and Sensor Data for analyzing indoor air

Previous studies have shown that it is possible to predict indoor air quality parameters using sensor data and machine learning.

In an experiment from Korea posted in the MDPI journal, a team of students used six air quality variables (CO_2, Dust, Temperature, Humidity, Light, VOC) from sensor data they collected periodically over almost seven months. The sensors were inexpensive but relatively accurate. They applied three different machine learning models to compare results: The linear regression model, the GRU model, and the LSTM model, while the latter two are deep learning models (cf. Ahn, et al. 2017). These are briefly explained in chapter 3.4.3.1. It turned out that the GRU model obtained the highest accuracy when comparing the estimations of the system with real data. The GRU model delivered an accuracy of 84.69% and showed remarkably similar tendencies, as visible in **Figure 1**. For comparison, the LSTM model of 70,13% and the linear regression model 60.96% (cf. Ahn, et al. 2017).

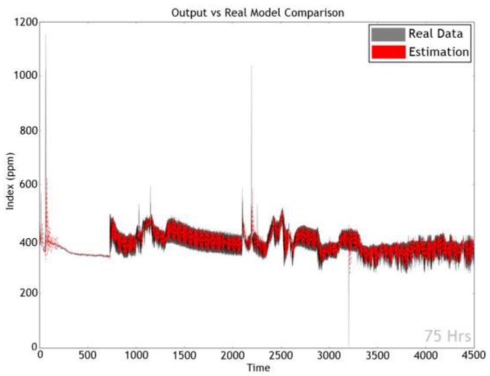


Figure 1: Analysis of dust data (Ahn, Shin, Kim, & Yang, 2017).

Another research proved similar, where machine learning algorithms forecasted CO_2 concentrations with sensor data from a campus classroom. Interestingly, the machine learning-based control improved the energy consumption of the classroom's ventilation/HVAC system by 51.4% compared to classical ON/OFF control. At the same time, it preserved thermal comfort. In addition, the machine learning-based control has the advantage that it can deal with temporal-dependent processes and flexibly adjusts the fan speed, for example, when there is more CO_2 emitted into the room (cf. Taheri and Razban 2021).

Indoor air quality improvement during the COVID-19 pandemic

Relevant to the current times, studies have also shown a link between CO2 concentrations and COVID-19 transmission indoors. A study from the TU Berlin stated that every person is emitting CO2 and aerosols (e.g., when exhaling). If an infected person is in the room, virus-laden aerosols emit into the air. They can cause transmission and increase everyone else's risk of getting infected with SARS-CoV-2. It was possible to calculate an approximate aerosol concentration based on CO2 concentrations, which are comparably easy to measure (cf. Hartmann and Kriegel 2020).

A high air change rate can decrease both CO2 concentrations and aerosol concentrations. Therefore, the study concluded that CO2 is a good indicator for measuring the room's ventilation system efficiency. Finally, this can lower the risk of infections, as poorly ventilated rooms may increase the risk of COVID-19 transmission through aerosols (cf. Hartmann and Kriegel 2020).

These studies already show how machine learning predicts and improves indoor air quality. They also show how specific parameters are related to each other. The following subchapter will show some parallels to weather predictions.

2.2 Machine learning used in weather forecasting

Even though weather forecasting depends on more parameters than indoor air quality, it might be interesting to see how machine learning can help predictions.

Current weather forecasts rely heavily on complex physical models and extensive numerical simulations run on supercomputers that are millions of times more powerful than our average desktop computers. The high cost and energy demand (cf. Kar, Mukhopadhyay and Deb Sarkar 2022).

Using machine learning models may be more straightforward and run on almost any computer, as they are more resource-friendly than traditional physical models. For example, a research team from Tennessee Tech University used real weather data (e.g., temperature, humidity, and pressure) from Nashville, USA. They concluded that machine learning models could compete with traditional models and accurately predict the next day's temperature for our day-to-day lives (cf. Jakaria, Hossain and Rahman 2020).

One essential part of weather forecasting is 'Nowcasting,' which usually predicts meteorological conditions over the next one to two hours. However, engineers at Google's DeepMind recently made a promising advance in that field, using a type of machine learning called generative modeling to nowcast weather for the next 90 minutes. Moreover, their model provided fast and accurate short-term predictions, beating existing methods (cf. Ravuri, et al. 2021).

Considering the current state of science, the strengths and advantages of machine learning become indisputable.

2.3 Consumer electronics for improving Indoor Air Quality

Consumer electronics for improving indoor air quality have become more prevalent in recent years. A few examples are listed below:

Smart Air Quality Monitors

Smart air quality monitors measure several parameters like temperature, humidity, or VOC. Based on that, they create an air quality score, making users aware to eventually open windows or remove sources of pollution. In addition, they usually come with a mobile app and support voice assistants (cf. Eve Systems GmbH 2018).

Smart AC controls

Another option are smart AC controls that adapt the cooling of the AC based on weekly schedules, daytime, or outdoor parameters like weather data for lowering energy consumption and creating a healthier indoor environment (cf. tado GmbH 2018).

Smart Radiator Thermostat

Smart thermostats detect open windows or stop heating when nobody is at home. They also consider factors like the weather forecast to reduce the heating based on sunlight and outdoor temperatures (cf. tado GmbH 2018).

Smart Air Cleaners

Smart air purifiers come with PM-sensors for detecting air pollution. They can filter the air from particles, reducing pollen allergens, pollution from the outside, or unpleasant smells, such as cigarette smoke (cf. Xiaomi 2020).

To summarize, this chapter provided an overview of the state of science in predicting and improving indoor air quality. In addition, it gave an insight into how machine learning can be used to forecast specific air quality parameters, and it provided a concise list of available consumer electronics that help us live in a healthy and sustainable home.

3. Theoretical framework

The theoretical framework provides a detailed background to the single aspects related to this research. First, different important air quality parameters, sources, and health effects will be explained in detail. After that, recommended and effective measures to reduce and prevent indoor air pollution will be introduced, and last. Still, not least, the available technology from sensors to machine learning gets explained.

3.1 Important indoor air quality parameters and their impact on health

Temperature

Temperature is one of the most critical indoor air quality parameters, as it can cause thermal discomfort and negatively influence the perceived air quality (cf. Toftum, et al. 2002). Moreover, considering that the average time spent indoors is 90% (i.a., due to office work or studying) (cf. WHO 2013, 9), it is essential to find the right temperature to increase our performance and productivity. It turns out that a temperature between 21 and 22 °C has a positive impact on our performance, whereas temperatures above 23 and 24 °C cause a decrease. Therefore, the determined temperature for reaching the highest productivity is around 22 °C (cf. Seppanen, Fisk and Lei 2006).

The WHO housing and health guidelines recommend keeping the indoor temperature at least at a minimum of 18 °C. Temperatures below 18 °C can increase respiratory and cardiovascular mortality risks such as asthma or high blood pressure (cf. WHO 2018, 34-36).

Relative Humidity (RH)

The relative humidity, expressed as a percentage (%), is the amount of water vapor present in the air. Air can only hold a certain amount of water vapor depending on temperature and pressure (cf. Benda 1999, 70). Therefore, the relative humidity would be 100% if the air reaches the maximum amount of H2O molecules it can hold. Furthermore, warm air can absorb more humidity than cold air (cf. Benda 1999, 70).

Dry air, for example, leads to irritation of the eyes due to a reduced tear film quality, an increased frequency of eyelid blink, and discomfort. It also decreases the hydration state of the skin (cf. Nienaber, et al. 2021). In addition, it causes a drying out of the mucous membranes in the respiratory system (e.g., nose and throat), which are necessary for the body's natural defense mechanisms (cf. Nienaber, et al. 2021). Relative humidity also impacts our productivity, sleep, and stress. For example, for test persons in an environment of 30 – 60 %RH, 25% fewer stress symptoms were recorded than for persons in drier conditions (cf. Nienaber, et al. 2021).

Humidity also significantly impacts pathogens and pollutants, such as mold infestation, dust mites, aerosols, and viruses. For example, a mold infestation increases the risk of coughing, infections of the respiratory systems and can cause allergic asthma. However, there is usually no risk for fungal growth if the relative humidity stays below 70 to 80 % (cf. Nienaber, et al. 2021). Also, the optimal conditions for dust-mite contamination are between 70 and 80 %RH. Keeping the relative humidity below reduces that risk (cf. Nienaber, et al. 2021).

In addition, relative air humidity also influences the inactivation time of viruses. For example, an influenza virus survives for longer in an environment of 15 – 40 %RH. On the other hand, poliomyelitis viruses have their most extended survival times above 80 %RH (cf. Nienaber, et al. 2021).

These examples show the importance of relative humidity to human health in various aspects. Therefore, scientific literature that focuses on human health recommends a range of 40 – 60 %RH as optimal for indoor environments (cf. Nienaber, et al. 2021).

Carbon Dioxide (CO2)

The Munich chemistry professor Max von Pettenkofer has proven in his book from 1858, "Über den Luftwechsel in Wohngebäuden," (which means, "About the air exchange in residential buildings") that humans are the most significant indoor air polluters. Based on that, he suggested not exceeding a maximum carbon dioxide concentration of 1.000 ppm CO2, the so-called Pettenkofer number, as he assumed that the outside concentration of CO2 was about 500 ppm (cf. Von Pettenkofer 1858).

The German Federal Environmental Agency keeps that early proposed value of 1.000 ppm as a guideline for naturally ventilated rooms. A hygienic evaluation rates a CO2 concentration below 1000 ppm as harmless, a concentration between 1.000 and 2.000 ppm as hygienically conspicuous, and above 2.000 as unacceptable (cf. UBA, Gesundheitliche Bewertung 2008). As a reference, the CO2 concentration of the air outdoors has about 400 ppm (cf. European Environment Agency 2019).

Research finds that a high CO2 concentration harms human performance and decision-making. For example, a CO2 concentration of about 2.500 ppm, a typical concentration found in many buildings, decreases the performance to an almost dysfunctional level. Already CO2 levels between 600 and 1.000 ppm can worsen the general condition of some people, e.g., asthma patients. CO2 concentrations ranging from 1.000 – 2.000 ppm negatively impact the ability to focus, mental concentration, and attention (cf. Satish, et al. 2011).

It is also worth mentioning that CO2 also plays a role in oxidative damage, such as cell death, DNA mutation frequency, and the number of DNA lesions (cf. Ezraty, et al. 2011).

Radon

As mentioned in the introduction, radon is a human carcinogen and a leading cause of lung cancer (cf. WHO 2021). The radon concentrations in the air are expressed in becquerels per cubic meter (Bq/m3). One becquerel is equal to one radioactive decay per second (cf. Health Canada 2017, 8).

Radon is formed from natural uranium in the ground and rocks (Bundesamt für Strahlenschutz, 2021). Outdoors, radon is generally not a health issue. However, radon exposure mainly occurs indoors, as the gas enters buildings, e.g., through pores in hollow-block walls and gets trapped within the walls (cf. WHO 2021). The Federal Office for Radiation Protection in Germany published that the annual mean value indoors in Germany is 50 becquerel per cubic meter on average (cf. BfS 2019, 24). Furthermore, findings from medical examinations show a measurable increase in the risk of lung cancer from a concentration of 100 becquerels per cubic meter (cf. Axelsson, Andersson and Barregard 2015).

Particulate matters in different diameters

Particulate matter is generated primarily by human activity and is emitted, for example, from automobiles, power plants, or industries (cf. EPA, Particulate Matter (PM) Basics 2021). However, according to the Federal Environmental Agency in Germany, the dominating source of particulate matter in urban areas is road traffic. It includes the pollution from engines, the brake- and tire abrasion, and the whirling up of dust from the road surface (cf. UBA 2021).

Indoor sources for particulate matter are, for example, burning candles, fireplaces, or cigarette smoking (cf. EPA 2022), including electronic cigarettes (cf. Fernández, et al., 2015). Although, pollution from outside could enter indoors when opening the windows (cf. EPA 2021). At homes with a ventilation system, efficient air filters can significantly reduce particle pollution indoors (cf. Lawrence Berkeley National Laboratory n.d.).

Natural sources of PM include sea salt, dust, or pollen, but pollution can also originate from wildfires and volcanic ash. Indoors also airborne mold spores can occur (cf. Stats NZ 2018).

Particulate matter is usually measured in µg/m3, a unit for measuring pollutants in the air (cf. Defra 2005, 1). Such particles can lead to short-term health effects such as irritation of the eye, nose, and throat and cause coughing or shortness of breath (cf. New York State Health Department 2018). We separate particulate matter into three main groupings:

PM10 particles (coarse particles)

PM10 are inhalable particles that are 10 micrometers and smaller. These are usually pollutants like dust, pollen, or mold (cf. EPA, Particulate Matter (PM) Basics 2021).

Such particles are small enough to pass through the nose and throat and enter the lungs (cf. UBA, Feinstaub 2021). There they can cause respiratory and cardiovascular health issues (cf. Hamanaka and Mutlu 2018).

The WHO guidelines recommend not to exceed an annual mean of 15 µg/m3 and a 24 hour mean of 45 µg/m3 (cf. WHO 2021).

PM2.5 particles (fine particles)

PM2.5 are inhalable particles that are 2.5 micrometers and smaller. These are usually pollutants like combustion particles, organic compounds, or metals (cf. EPA, Particulate Matter (PM) Basics 2021). PM2.5 can penetrate the bronchi and alveoli and worsen medical conditions like asthma and heart diseases (cf. New York State Health Department 2018).

The WHO guidelines recommend not to exceed an annual mean of 5 µg/m3 and a 24 hour mean of 15 µg/m3 (cf. WHO 2021).

PM1 particulates (ultrafine particles)

PM1 are inhalable particles that are 1 micrometer and smaller. These ultrafine particles could be diesel exhaust particles or particles caused by cooking or wood-burning in indoor environments (cf. Shen, et al. 2017). These particles can travel to the deepest area of the lungs and enter the bloodstream (cf. UBA, Feinstaub 2021). From there, these particles could potentially spread to organs, like the brain, causing neurodegenerative conditions (cf. Nephew, et al. 2021).

Volatile organic compounds (VOC)

Volatile organic compounds are gases that emit from solids or liquids like perfumes, hair sprays, household products, paints, glues, or permanent markers into the air (cf. EPA, Volatile Organic Compounds' Impact on Indoor Air Quality 2021). Just to have named a few, thousands of products emit VOC.

Studies from the EPA show that concentrations from some organics average up to 5 times higher indoors than outdoors. Symptoms related to a high indoor concentration of VOC are, for example, headache, nausea, fatigue, or dizziness. Some people even experienced visual disorders and memory impairment (cf. EPA, Volatile Organic Compounds' Impact on Indoor Air Quality 2021).

It is complex to measure and distinguish all these different single VOCs, as many volatile organic compounds have a similar chemical structure. Therefore, a measurement parameter for the total volatile organic compounds (TVOC) was developed, summarizing these individual values (cf. Meyer 2021, 3).

In a paper published by the German Federal Environmental Agency, a TVOC concentration below 200µg/m3 causes no irritation or impairment of well-being.

However, concentrations from 200 – 3.000µg/m3 can cause discomfort, from 200 – 25.000µg/m3 headaches are possible, and TVOC concentrations above 25.000µg/m3 cause headaches and may lead to other neurotoxic effects. Therefore, the agency recommends not to exceed a TVOC value of 300 µg/m3 indoors (cf. Seifert 1999, 271).

Carbon Monoxide (CO) and Nitrogen Dioxide (NO2)

Several other gases are influencing indoor air quality, such as:

Carbon monoxide (CO)

It is an odorless, tasteless, and colorless gas that arises from the incomplete combustion of carbon-containing fuels (cf. Prabjit Barn, Fong and Kosatsky 2016). Typical sources of danger are gas stoves, tobacco smoke, or leaking chimneys (cf. EPA 2021). It impairs oxygen uptake in humans and animals and may affect the central nervous system (cf. UBA 2021).

According to the German Federal Environmental Agency, the highest 8-hour average value of a day may not exceed 10 mg/m3 (cf. UBA 2021).

Nitrogen dioxide (NO2)

Nitrogen oxides are caused by combustion engines and plants for coal, oil, and gas. It is an irritant gas that affects the eyes and damages mucosal tissue throughout the respiratory tract (cf. UBA 2021). In high concentrations, nitrogen dioxide can cause shortness of breath, bronchitis, and pulmonary edema. The annual limit value for the protection of human health is 40 µg/m^3 as an annual mean. However, within an hour, 200 µg/m^3 must not exceed more than 18 times (cf. UBA 2021).

Air Change Rate

Despite all the parameters mentioned above, the air change rate is vital in maintaining healthy indoor air quality. It is the rate of outdoor air replacing indoor air (cf. EPA 2022). In addition, the European Standard EN 12831 specifies a minimum air exchange rate of 0,5 times per hour for indoor spaces (cf. European Committee for Standardization 2017). The importance of air change rate was also previously mentioned in the study related to CO2 and aerosols (cf. Hartmann and Kriegel 2020).

Now, as many pollutants and their consequences are known, the upcoming subchapter introduces possible measures to reduce indoor air pollutants and improve indoor air quality.

3.2 Measures for improvement of indoor air quality

There are several ways to improve indoor air quality and to counter harmful pollutants:

Ventilating

One of the most critical indoor air quality improvement measures is letting fresh air into the room. It reduces humidity to avoid mold growth (cf. UBA 2017) and lowers the concentrations of CO2 and other harmful pollutants (cf. UBA 2021).

It is recommended to ventilate the rooms several times a day with a wide-open window. Ideally, two opposite windows are opened to create a strong airflow. Each ventilation should last for at least 20 to 30 minutes in summer, whereas 5 to 10 minutes are sufficient in winter (cf. BMUV 2020). It is necessary always to ventilate when water vapor is created, e.g., at showering and cooking or when drying laundry (cf. BMUV 2020). On the other side, outside air tends to be dry in winter, which is why ventilation may result in low relative humidity indoors. Therefore, constantly tilted windows should be avoided (cf. UBA 2019).

Modern office buildings usually have ventilation systems that ventilate the room automatically. These ventilation systems must be regularly maintained and cleaned by qualified personnel. Also, the filters need to be exchanged depending on the outdoor air pollution (cf. UBA 2017).

Regular cleaning

At the same time, it is crucial to keep it clean indoors, as it can reduce dust and allergens such as animal dander. For example, vacuum cleaning helps prevent dust from settling for a long time. In addition, vacuum cleaners must have an additional filter, e.g., a HEPA filter, to prevent the vacuumed dust from returning into the room's air (cf. Harvard Health Publishing 2021).

Harvard Health Publishing recommends in an article to vacuum carpets and area rugs at least twice a week. Other recommendations are to regularly clean items such as beddings and clear clutter, as it traps dust (cf. Harvard Health Publishing 2021).

Eliminate sources of pollution

It might also be effective in identifying and removing sources of pollution (cf. EPA 2021). For example, EPA determined that homes with gas stoves have around 50 percent higher concentrations of NO2 than homes with electric stoves (cf. EPA 2016). Other sources of pollution are candle burning, smoking, incense sticks, certain home cleaners, and care products (cf. EPA 2021).

Air Cleaners

If it is not possible to control the source of the problem, e.g., to counter allergens like pollen or pet dander, air cleaners might be a viable solution (cf. EPA 2021).

According to a recommendation from EPA, there are various highly effective air cleaners on the market, helping to remove such particles. However, they state that air cleaners are generally not designed to remove gaseous pollutants. Therefore, it is essential to note that air cleaners will not eliminate all air pollutants in your home. They are also not a solution for mold. To solve a mold problem, the source of moisture must be identified and removed first (cf. EPA 2018).

Plants

An article in the journal *"Trends in Plant Science"* recently suggested that plants could be a solution to remove pollutants and improve indoor air quality through stomatal uptake (e.g., CO2) and non-stomatal deposition (e.g., oxygen) efficiently and sustainably. However, not every plant is suitable indoors to remove air pollutants, as some plants may have scented leaves or flowers releasing VOC (cf. Brilli, et al. 2018).

The article highlights that science should urgently screen for the optimal-performing plant species in indoor environments, as they could be a win-win strategy for sustainably improving indoor air quality (cf. Brilli, et al. 2018).

Previous research found several plants to effectively reduce pollutants such as VOCs, CO2, and even particulate matter. Plants that sustainably improve indoor air quality are, for example, species like the hedera helix, the chrysanthemum × morifolium, the dieffenbachia compacta as well as the epipremnum aureum. One species reduced formaldehyde, one of the most common VOCs, over 10 hours by about 98% (cf. Aydogan and Montoya 2011).

Figure 2: Example of an epipremnum aureum (Photo by Olena Shmahalo on Unsplash)

Knowing possible measures to reduce air pollutants is one part, although reliably sensing air quality parameters and pollutants are vital to becoming aware of the problem. The following subchapter lists sensors to measure different parameters.

3.3 Available IoT sensors and pitfalls

There are cheap but reasonably accurate sensors available to measure indoor air quality parameters such as the concentrations of TVOC, CO2, or particulate matter. In the following list, some of the available sensors, their specifications, and pitfalls are listed:

Measuring Temperature

The DS18B20 produced by Maxim Integrated is a reliable, digital temperature sensor used in thermometers and thermally sensitive systems. It measures temperatures from -55 °C to +125 °C and has an accuracy of ±0.5°C at a range of -10°C to +85°C (cf. Maxim Integrated Products, Inc. 2019).

Measuring Humidity and Pressure

One available sensor that digitally measures humidity and pressure combined is the BME280 from BOSCH Sensortec. It is designed for low power consumption and ideally works for home automation control and climate monitoring (cf. Bosch Sensortec GmbH 2021).

The humidity sensor has a response time of 1s and an accuracy of ±3 %RH. In addition, it's equipped with a temperature sensor whose output is used for temperature compensation of the pressure and humidity sensors (cf. Bosch Sensortec GmbH 2021).

Measuring Carbon Dioxide (CO2)

A cheap way of measuring the CO2 concentration indoors would be the so-called estimated CO2 (eCO2) or CO2 equivalent (CO2eq). As the name says, the CO2 concentration is only estimated based on easier-to-measure air quality parameters such as VOCs or H2. The Adafruit SGP30 sensor, for example, calculates the eCO2 value based on the H2 concentration. However, it is worth mentioning that this sensor must be calibrated against known sources and that an eCO2 sensor is not a "true" CO2 sensor (cf. Adafruit Industries 2021).

A more accurate but more expensive "true" CO2 sensor is the Sensirion SCD30, which uses infrared to detect carbon dioxide in the air. It enables more precise and stable monitoring. The CO2 sensor has a response time of 20 seconds and has a CO2 measurement range from 0 – 40.000 ppm (cf. Sensirion AG 2020).

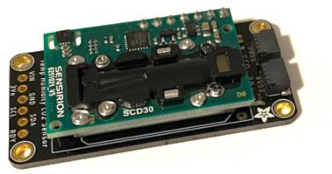


Figure 3: SCD30

Measuring Particulate Matter

Figure 4: PMS5003

The PMS5003, a digital universal particle concentration sensor, is suitable for measuring particulate matters. It can measure PM10, PM2.5, and PM1.0 concentrations using laser scattering to radiate suspending particles in the air. With the collected scattering light later, the curve of scattering light and change with time is obtained. Then the microprocessor can calculate the equivalent particle diameter and the number of particles (cf. Adafruit Industries 2021).

Measuring Volatile Organic Compounds (VOC)

Figure 5: SGP30

The Adafruit SGP30, which was mentioned above, detects not only H2 but also a wide range of VOCs. Based on that, it returns the Total Volatile Organic Compound (TVOC) concentration within a range of 0 to 60,000 parts per billion (cf. Adafruit Industries 2021).

If a humidity sensor is available, the sensor can more accurately calculate the TVOC concentration based on the percentage of the relative humidity (humidity compensation) (cf. Adafruit Industries 2021).

Measuring Carbon Monoxide (CO) and Nitron Dioxide (NO2)

To measure gases such as CO or NO2, the analog sensor MiCS-6814 from SGX Sensortech is a practical option. It can detect carbon monoxide, nitrogen dioxide, ethanol, hydrogen, ammonia, methane, propane, and isobutane (cf. SGX Sensortech n.d.).

It consists of three sensors: a RED sensor for measuring reducing gases (e.g., CO), a OX sensor for measuring oxidizing gases (e.g., NO2), and an NH3 sensor for measuring ammonia (cf. SGX Sensortech n.d.).

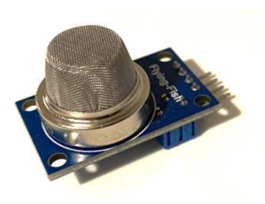


Figure 6: MQ135

Another affordable, analog air quality sensor is the MQ135 from AZ-Delivery, suitable for detecting nitrogen oxides, alcohol, or benzene (cf. AZ-Delivery Vertriebs GmbH n.d.).

Both sensors are ideal for monitoring indoor air quality and detecting gas leaks.

Now that it becomes clear that sensing most indoor air parameters is technically feasible, after monitoring, the topic of machine learning moves into focus. The next part will deliver a general introduction to this topic.

3.4 Introduction to Machine Learning

The examples mentioned in chapter 2 already gave insights into how machine learning could help predict and improve indoor air quality. For example, thanks to machine learning, it is possible to forecast the Development of specific parameters like CO2 concentrations for the upcoming hours to recommend necessary actions before a person in the room will notice any early health effects (e.g., fatigue).

In addition, for example, when ventilating the room, machine learning might be the right tool to determine the perfect amount of time a window has to stay open. That is to improve indoor air quality parameters and heating costs or the number of required interactions by the users. Maybe, opening the windows every two hours will suffice, instead of ventilating the room every hour. These are only a few examples of how this technology could help prove the hypothesis.

A few terms and ideas need to be explained to get a brief understanding of how machine learning works.

Definition of machine learning

Machine learning is a part of artificial intelligence that enables a system to learn automatically and gradually improve its accuracy. It is based on substantial amounts of data and algorithms for making predictions or decisions without being explicitly programmed (cf. Google Inc. n.d., cf. UC Berkeley School of Information 2020).

According to the UC Berkeley School of Information, typical supervised machine learning algorithms consist of three main parts (cf. UC Berkeley School of Information 2020):

- A decision process: The algorithm will estimate a pattern in the data and can make a prediction.
- An error function: A method that measures if the prediction was correct to assess the accuracy.
- And an optimization process: The algorithm will look at the missed prediction to improve the next guess' accuracy.

Types of machine learning

As stated by UC Berkeley, many machine learning models either work with datasets influenced by humans or not. For example, data can be labeled or have certain feedback (cf. UC Berkeley School of Information 2020).

Nvidia mentions in an article several types of machine learning models:

Supervised learning

Supervised learning uses pre-labeled and classified datasets with the help of users to train algorithms by showing how accurate their performance is. An example of a supervised learning model is, for instance, the prior mentioned linear regression model in which the use of known parameters predicts the result (cf. NVIDIA Corporation 2018).

Unsupervised learning

In the case of unsupervised learning, no data is pre-labeled or classified. Instead, the algorithm searches the data and finds specific patterns on its own (cf. NVIDIA Corporation 2018).

Semi-supervised learning

Semi-supervised learning is a perfect compromise between supervised and unsupervised learning. A small amount of pre-labeled and classified datasets enables machine learning algorithms to label unlabeled data on their own and finally to make independent decisions (cf. NVIDIA Corporation 2018).

Reinforcement learning

Reinforcement learning uses a reward and punishment system for training an algorithm. Nvidia compares it to classical video games, where players earn badges or complete a level when defeating a bad guy or mechanisms like "game over" when stepping into a trap. These things help the player get better (cf. NVIDIA Corporation 2018).

A machine learning model based on reinforcement learning uses the same principles, as the algorithm learns from its own experience through feedback, trial, and error.

Deep Learning

Compared to the models mentioned above, deep learning is a newer field of machine learning. More precisely, deep learning is a sub-field of machine learning (cf. UC Berkeley School of Information 2020).

A deep learning model automatically learns from raw data without explicit instructions or human feedback. By analyzing data and extracting valuable features, it automatically finds structure. While it can learn without human supervision, deep learning models can also work with structured data. The backbone of deep learning are artificial neural networks. These networks are trying to simulate the human brain by mimicking how biological neurons communicate (cf. IBM Cloud Education 2020).

Two deep learning models were used in one of the studies mentioned prior:

LSTM and GRU

According to a paper from Google, Long Short-Term Memory (LSTM) is a specific recurrent neural network (RNN) that works more accurately than conventional RNNs (Recurrent Neural Network). LSTM models have a 'memory.' Therefore, they can maintain information for extended periods and work better with long-range dependencies (cf. Sak, Senior and Beaufays 2014).

IBM explains that a recurrent neural network (RNN) is a type of artificial neural network which uses sequential data. RNNs are usually used for language translation, speech recognition, and natural language processing (cf. IBM Cloud Education 2020).

The GRU further develops the LSTM, requiring fewer resources and parameters (cf. IBM Cloud Education 2020).

This brief introduction will help further refine the general idea of how machine learning could be used to prove the hypothesis.

3.5 Predicting Indoor Air Quality based on data from the past

The learning is: For a machine learning model to reliably predict parameters like CO2 or humidity, it needs the training dataset to learn from (cf. Ahn, et al. 2017).

For the research mentioned in chapter 2, the students tracked indoor air quality parameters three times a day (sunrise, afternoon, and sunset) for almost seven months. Using the GRU model, they got an accuracy of 84.69% for their predictions (cf. Ahn, et al. 2017).

The other study used data collected over three months during the fall session at a university campus classroom, and it consisted of 13.003 weather-related values. Again, they could get accurate CO2 level predictions based on the MLP model (Multilayer perceptron) (cf. Taheri and Razban 2021).

Therefore, choosing a suitable test environment/room is necessary and equipping it with the required sensors for measuring parameters like temperature, humidity, CO2, TVOC, and gases. Ideally, indoor air monitoring and data collection should start at an early stage of this research and be prioritized, storing the recordings in a database.

	Indoor Air Quality
PK	UniqueID
	timestamp
	temperature
	humidity
	co2
	window-opened

Figure 7: Example of database table

If the indoor air quality is recorded every 5 minutes over two months, there are already 17.280 data entries available. Although to detect, for example, an opened window and the associated effects, it might make sense to record indoor air quality every 2 minutes instead, leading to 43.200 data entries after two months. Over time, more data is collected to train the system and gradually improve accuracy. Different models (e.g., linear regression model, GRU model) should be tested on the same dataset and compared for their performance.

It is necessary to reliably predict data and determine which suggestions can be made based on specific data. Finally, the interaction with the user must be implemented. The following and last subchapter of this theoretical framework will give an overview and clear guidance about how systems and humans should ideally communicate and interact based on principles from Human-Computer Interaction.

3.6 Usability and Human-computer interaction

Reliable predictions for different indoor air quality parameters are created based on accurate sensor data and machine learning. Then, based on those predictions, people in the room will receive relevant actions at a given time for improving indoor air quality. Communication between the system and the user could, for example, happen through visual effects (e.g., a red light symbolizing low indoor air quality) or acoustic signals (e.g., a voice speaking). But also, an app (e.g., a website) or a skill/action for a voice assistant with which the user can interact is a possible solution.

In this case, the most crucial part for improving indoor air quality parameters like CO2 or humidity is users interacting with the system and following the recommendations. Therefore, disciplines like human-computer interaction and user experience play a vital role in a successful design, concept, and implementation.

Human-computer interaction focuses on the design and use of technology. More precisely, it focuses on the interaction between human and computer technology (cf. Brey and Søraker 2009). In HCI (Human-Computer Interaction), the users must be perfectly supported through technology, considering their strengths and weaknesses (cf. Gross 2021).

Therefore, usability plays in human-computer interaction a crucial role. It's defined by three essential characteristics every computer technology should have (cf. Moreno, et al. 2013):

Effectiveness: "Did the user reach the target accurately and completely?"

Efficiency: "Did the user reach the target with low effort?"

Satisfaction: "Is the user satisfied with the software when using it?"

When designing software, user orientation plays a vital role. It includes (cf. Yu, Gu und Ostwald 2012):

Attention: e.g., that the user focuses on the critical CO2 concentrations.

Perception: warning the user with gradually increasing light or sound.

Memory: e.g., recognizing a specific sound related to a pollutant.

Learning: the web-app and voice skill/action awakens the spirit of discovery

Reading, speaking, and listening: the system could explain actions either by voice or show a message as text.

And problem-solving: The user should know what to do, e.g., open a window and not be overwhelmed with actions.

Then, we must think of technologies that could be helpful in the case of an indoor air quality system that can provide recommendations and advice. Some technologies could be:

A smart LED lamp: If e.g., the CO2 concentration exceeds 1.000 ppm, the light could gradually turn from its natural color to red. Once the CO2 concentration improves, the light could return to its standard color.

A speaker: When specific actions are needed, a speaker could play a warning sound or have an artificial voice give an instruction.

A voice assistant: An even better option would be a voice assistant. It allows interaction with the user. The user could then even ask on their own if the indoor air quality is okay.

A mobile or desktop app: An app or even a website might be another solution, showing all parameters on a dashboard with graphs. Also, an app could send push notifications to the user's phone or desktop when actions are needed.

After the technology, another critical factor is interaction. It is separated into four types (cf. Vega-Barbas, et al. 2018):

Instruction: The user could instruct the system to, e.g., mute all recommendations for the next hour, not disturbing the user.

Conversation: A voice assistant could dialog with the user in terms of a voice assistant. For example, "The CO2 concentration is currently at 528 ppm. Do you want to know more?"

Manipulation: Eventually, inside an app like a website, the user could drag and drop certain graphs and activate or deactivate specific values shown.

And exploration: The user checks and explores the voice skill/action or clicks through the web application to get to know the system.

Several methods are available when implementing interactive systems like an intelligent indoor air quality monitor. Besides creating a concept or a prototype that finally leads to a state-of-the-art system, it is essential to let the users give regular feedback to get valuable insights and findings on how to improve the software for the future.

Especially in this case, when the system can precisely predict the indoor air quality and give efficient suggestions on how to improve specific parameters, it's all for nothing if the user does not interact with it.

Designing the User Interface

Furthermore, it is worth mentioning the eight golden rules of interface design from Ben Shneiderman, an American computer scientist, and professor at the University of Maryland (cf. Shneiderman, et al. 2016):

- Strive for consistency: There should be consistency (e.g., menus, prompts, layout) in similar situations, such as checking the value of a specific pollutant.
- Seek universal usability: Frequent or experienced users should be able to use shortcuts.
- Offer informative feedback: The user should receive feedback for every interaction.
- Design dialogs to yield closure: Users should always know what their action has led to, e.g., when the user tells the system to mute, it should show a confirmation.
- Prevent errors: The user should be helped to prevent mistakes. E.g., when wondering why the graph shows no historical CO2 data, the unticked checkbox could be highlighted.
- Permit easy reversal of actions: A user should always be able to reverse steps, e.g., when muting the system, there also must be an unmute.
- Keep users in control: Users need to feel that they are in full control. Therefore, the system must behave as expected.
- Reduce short-term memory load: The interface must be as simple as possible and avoid the user remembering any unnecessary information, e.g., even a message "open the window in 6 hours and 42 minutes" is redundant.

Keeping all the above suggestions in mind (usability, user orientation, interaction, and Shneiderman's eight golden rules of interface design) will make it possible to implement a user-friendly system that users would like to interact with.

4. Differences, Methodology and Architecture example

This chapter summarizes the overall approach and describes what differentiates this work from existing approaches. Furthermore, this chapter explains the methodologies used to prove the hypotheses.

4.1 Differences to existing work

This research differentiates from existing work by focusing on the user. Despite predicting indoor air quality and recommending possible steps, its outcome depends entirely on user interaction and actions.

This research focuses mainly on indoor environments with natural ventilation and daily habits. It aims to improve indoor air quality through creating awareness.

4.2 Methodology

Several methods are required to test whether the hypothesis is true or not. Therefore, the research design consists of a concept (proving the technical feasibility), a prototype (demonstrating the reliability of predictions), an experiment (proving the significant and efficient improvement in air quality), and a user survey (indicating the usability).

Concept

A concept will be the foundation for building a prototype. It must describe a feasible technical architecture with the underlying technology and the interactions between the system and the user. Besides a general description, an architecture diagram and rough wireframes of the possible user interface are part of it. Finally, it will provide a reliable picture of the efforts needed to implement the system and serve as orientation and guidance throughout the development phase.

Prototype

Then, based on the concept, a prototype is built covering the minimum required features to prove the hypothesis and answer the research question.

Therefore, such a system must monitor the indoor air quality using sensors, predict specific parameters through machine learning, and improve those parameters (e.g., concentrations of pollutants) via interaction with users to fulfill its task.

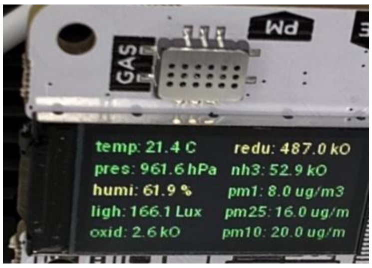


Figure 8: Sensor results on a Raspberry Pi with an Enviro+ board

Such a system will consist of the following:

a sensor node to obtain the air quality

a data storage to store the sensor data

a machine learning server to forecast indoor air quality parameters

and a user interface (e.g., a web app or voice-assistant) to communicate and interact with the users for indoor air quality improvement

One feature of this method is that predictions can already be compared with real data to determine the average accuracy of the predictions. Then, it is possible to answer whether the first hypothesis is true, "A self-learning system based on machine learning and IoT sensor data can reliably predict indoor air quality parameters."

Additional methods are necessary to evaluate the prototype to answer the second hypothesis.

Experiment

The system will collect data over two separate periods, e.g., two weeks each, to evaluate the prototype. In the first period, the users will get recommendations through the system. In the second period, the system remains silent. After all, this creates two data groups where the data from the second period is the control group.

Later, when comparing the groups for specific parameters (e.g., average CO2 level throughout the day, heating), it is possible to determine whether the system can significantly and efficiently improve indoor air quality.

Besides significance and efficiency, the practicality and user-friendliness need to be evaluated.

Survey

Finally, a survey can answer the last question and show if users perceive the system's suggestions as user-friendly and practical. Therefore, a group of independent people gets first introduced and familiar with the system's user interface and recommendations and later answer a survey of how they think of it in terms of usability.

After both the experiment and the survey, it will be possible to answer whether "a self-learning system can provide practical and user-friendly suggestions to the occupants of a room to significantly and efficiently improve indoor air quality."

5. Results and Analysis

5.1 Concept and Architecture

This section explains a basic concept and architecture and serves as a template for implementing a prototype. This concept considers the available and relevant technology, technical feasibility, pitfalls, and workarounds.

Sensors and collecting data

A single-board computer (SBC) such as the Raspberry Pi or Arduino Uno is a suitable low-cost solution for operating different sensors. These sensors can then be plugged directly onto the board (CO_2, VOC) or connected wirelessly through WiFi, Bluetooth, or Zigbee (contact sensors, outdoor temperature sensor) to deliver periodic readings.

Figure 9: Raspberry Pi 4B, Raspberry Pi Zero and Arduino Uno (Photo by Harrison Broadbent on Unsplash)

Periodically (e.g., every 10 seconds), a dataset of sensor readings is created and sent to a Cloud service, where it is stored securely with the benefit of automated backups. Therefore, the increasing amount of sensor data doesn't need to be managed on the SBC's internal memory, and the risk of data loss due to technical malfunction or other factors is minimal. At the same time, the development efforts can be focused on user interaction and the user interface rather than server infrastructure or storage management.

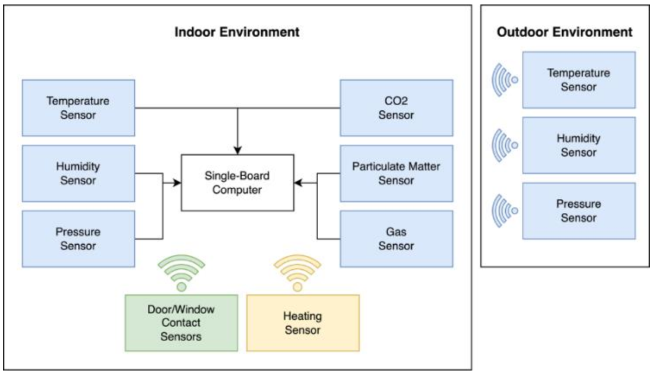


Figure 10: Setup of Board and Sensors

For the Internet of Things (IoT), the MQTT protocol has become a standard for lightweight publish/subscribe messaging between the IoT devices and the Cloud service (cf. OASIS, 2022). Moreover, IoT-related Cloud services are available at a low cost. Some examples are AWS IoT, Google Cloud IoT, Azure IoT, and IoT Solutions from IBM, offering a variety of services.

Figure 11: Visualization of Data Transfer

Time-Series forecasting with Machine Learning

Based on sensor data collected over a more extended period, a machine learning model is trained to forecast the time series of specific indoor air quality parameters like temperature or relative humidity.

As mentioned in chapter 2.1, two particular kinds of recurrent neural networks delivered promising results for predicting indoor air quality parameters in an experiment in Korea: The LSTM model had an accuracy of 70,13%, and the GRU model provided an accuracy of 84.69% (cf. Ahn, et al. 2017). Therefore, the LSTM and the GRU models are great for forecasting time series, particularly indoor air quality parameters.

Several open-source libraries for machine learning are available, making applying RNN models much more accessible.

Two of the most common libraries are PyTorch from Meta AI and TensorFlow from the Google Brain Team (cf. Statistics and Data, 2022). Both are suitable for time-series forecasting and work with Python. Another library worth mentioning is Keras, which acts as an interface for TensorFlow enabling rapid experimentation with deep neural networks like LSTM or GRU. Finally, darts from Unit8 is one more Python library suitable for this project to forecast time series quickly and easily.

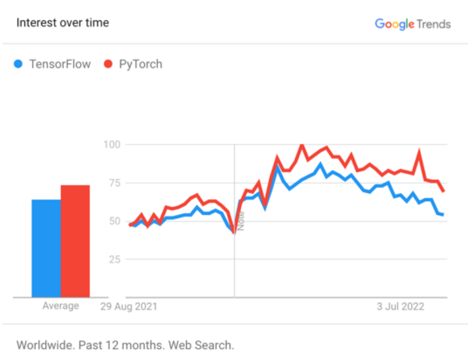


Figure 12: Google Trends comparision between TensorFlow and PyTorch

The CO2 concentration indoors would be an ideal parameter to forecast, as it's relatively independent of weather, summertime, or wintertime. It certainly increases with people in the room exhaling air and decreases when windows are open to ventilate fresh air into the room.

Based on the CO2 level prediction of the upcoming 10 or 15 minutes, occupants in the room will be notified before certain limits (e.g., 1.000 ppm) are reached. These forecasts may vary depending on different factors, e.g., how many people are in the room, what activities they do (reading a book versus running on the treadmill), but especially on covariates such as open windows and doors.

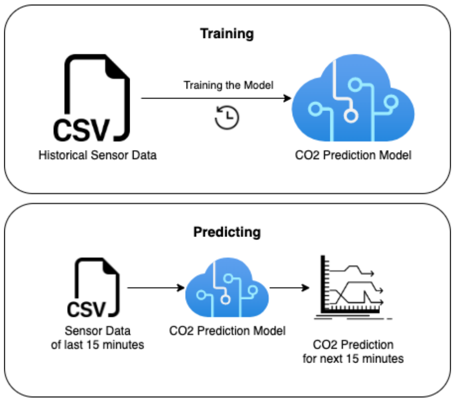


Figure 13: Visualization of model training and prediction

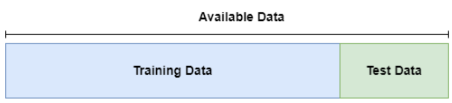


Figure 14: Training Data vs. Test Data

The historical sensor data is split into two parts: Training data and test data. The training data is the more extensive set, based on which the model is trained to find meaningful patterns for its predictions. The test data is the smaller set used to evaluate the model's performance and to optimize it afterward.

Notifications and Recommendations

The system focuses on people in the room, independent of any mobile or desktop app requiring prior download and installation. This way, the system can guarantee that every person in the room affected by bad indoor air quality is made aware of the problems.

First, users are warned. Therefore, a colorful light informs them of any indoor air quality parameter (e.g., CO2) exceeding its limit. If CO2 levels grow above a certain threshold, one or more lights in the room turn red. Red is often associated with warnings and alarms, so users already might assume that something is wrong (cf. Cherry, 2020). For example, a possible threshold for CO2 concentration indoors is 1.000 ppm, the so-called Pettenkofer number (cf. Von Pettenkofer 1858). Once the CO2 level falls below 500 ppm, the red lights will turn off, and the user knows that the indoor air quality is good again.

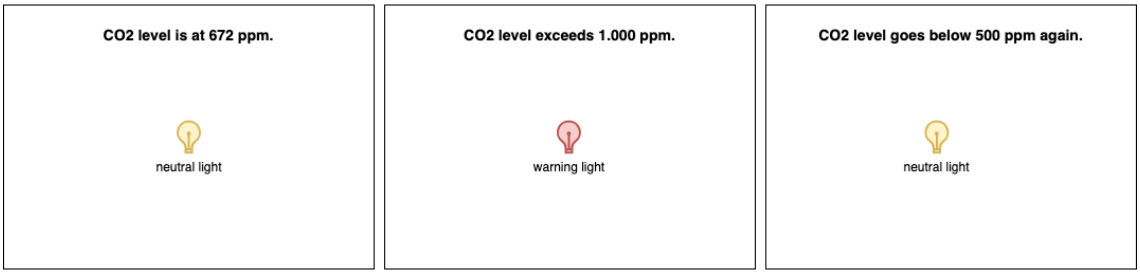


Figure 15: Concept for Notifications through light

Secondly, acoustic signals will give an additional notification. Push messages can be sent to a virtual assistant like Amazon Alexa or Google Home, playing an individual message through the automated voice. For example, if the CO2 concentration exceeds 1.000 ppm, the virtual assistant could read the statement "the level of CO2 is too high." Once it falls below 500 ppm, the message "the level of CO2 is good again" will be played.

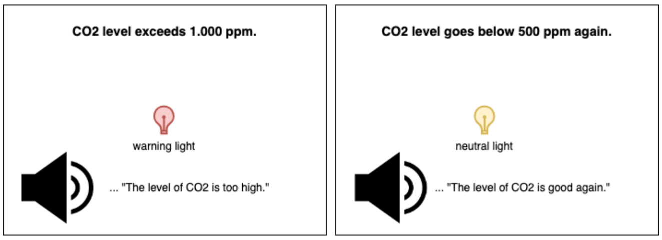


Figure 16: Concept for Notifications through sound

Additionally to the notification, the system recommends the user some actions to improve the current issue efficiently. For example, "open the window" or "open the balcony door." This way, a user would receive precise advice on influencing a specific indoor air quality parameter for the better. However, if, for example, the concentration of particulate matter rises due to outdoor traffic, the advice could be quite the opposite. For example, instead of opening the window, the system might recommend closing it.

User Interface

Finally, a user interface should allow users to check indoor air quality parameters. Such UI could be represented as a web application where the users can access live- and historical data. Or through a voice interface, where users talk to their virtual assistant. Both versions have different advantages, as the web application provides a quick and structured overview of indoor air quality. In contrast, the voice interface offers a more natural user interaction through speech, independent of any mobile phone or desktop application, to all users in the room.

The wireframe on the side visualizes a rough overview of a possible web application. It consists of two parts: the live data and the historical data.

On the top, the user can find the most important indoor air quality parameters such as temperature, humidity, CO2, and the predicted value for the CO2 level in 10 or 15 minutes. Additionally to that, there is information about open doors and windows. The historical data below shows the parameter's Development of the last minutes, hours, or the last day.

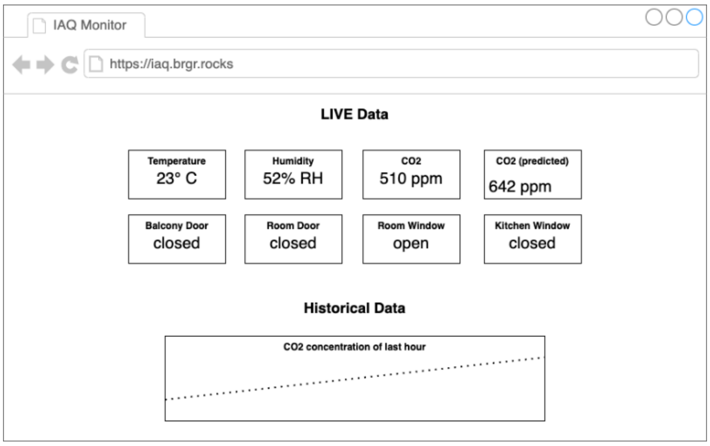


Figure 17: Basic wireframe of the web application

As shown in Figure 18, a simple voice interface enables users to ask for quick information on indoor air quality parameters. Such a feature is provided with a smart speaker inside the room. Users then solely interact through speech, independent of any desktop or mobile device.

If a user asks, "What's the CO2 level?" the virtual assistant will gather the information and answer, "The current CO2 level is 540 ppm."

There will be an intent for most parameters such as CO2, temperature, humidity, particulate matter, the state of the doors and windows, and more. In Amazon Alexa, an intent represents an action that fulfills a user's spoken request (Amazon.com, Inc., 2022).

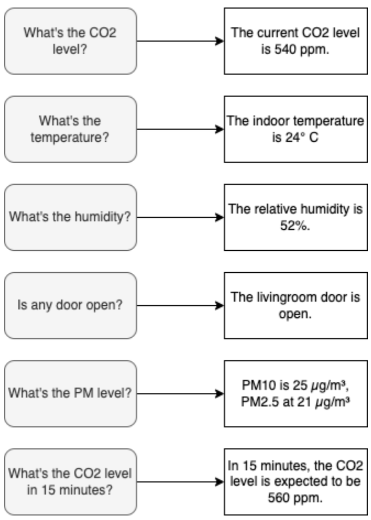


Figure 18: Concept for the Voice Interface

5.2 Prototype

A well-suited parameter for a possible prototype is the CO2 concentration measured in ppm (parts per million). Therefore, based on the concept above, a prototype was developed. The prototype consists of a single-board computer, multiple sensors to measure indoor air quality, the capability to forecast the Development of CO2 indoors, a notification and recommendation service, a web application, and a voice interface.

Board and Sensors

A Raspberry Pi 4B was chosen as a single-board computer, and all the necessary sensors were connected.

For measuring CO2, a Sensirion SCD30 was plugged via cable to the board. This CO2 sensor uses infrared to detect carbon dioxide. It had to be calibrated with the outdoor air for about five days for precise measurements, using the Automatic Self-Calibration of the SCD-30. The outdoor air usually has a CO2 concentration of around 400 ppm. Additionally, the CO2 sensor also is capable of measuring temperature and humidity.

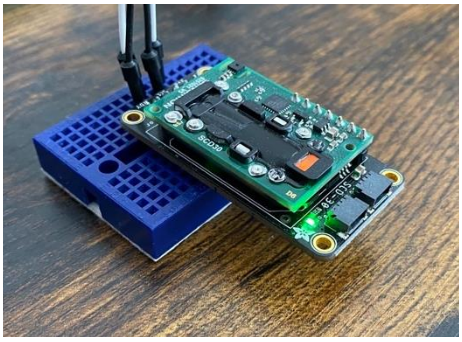


Figure 19: Sensirion SCD30 connected to board

Also, different sensors like a particulate matter sensor (PMS5003), a TVOC sensor (SGP30), and a gas sensor were connected to the board (MiCS-6814). Even though the board has many different sensors connected, the prototype focuses mainly on measuring and improving CO2 concentrations.

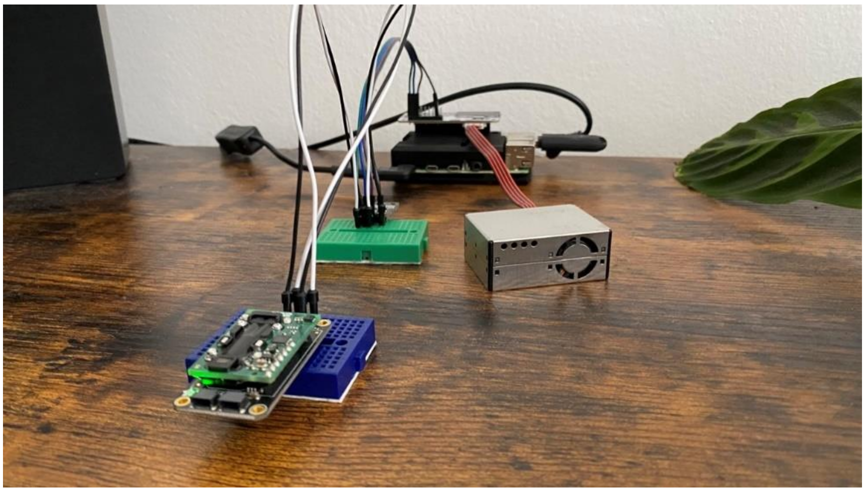


Figure 20: Raspberry Pi connected to all relevant sensors

A Zigbee gateway (ConBee II) connected to the Raspberry Pi enables wireless communication between some sensors and the board.

For example, contact sensors on doors and windows use Zigbee for wireless data transfer, a protocol requiring low power consumption (cf. Farahani, 2008). In addition, the contact sensors are operated with a simple button cell (CR1623), promising a battery life of two years (Aqara, 2022).

The contact sensors are installed on the living room window, the living room door, the balcony door, and the kitchen window.

In addition, there is one wireless temperature, humidity, and air pressure sensor in the living room (indoors) for reference to the temperature and humidity sensor already connected via cable to the board and one outdoors on the balcony for measuring the outdoor pressure, humidity, and temperature (PHT).

Figure 21: Contact sensor at window

Figure 22: Temperature, humidity and pressure sensor

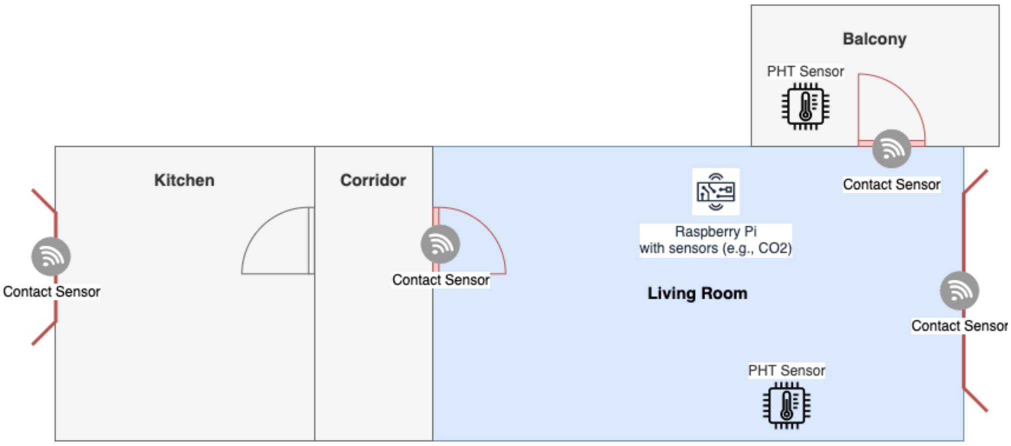


Figure 23: Setup of sensors accross different rooms.

Sensor Measurements and Collecting Data

On the Raspberry Pi, a Python script gets the value of all sensors every 10 seconds. Each time, the data is sent from the Raspberry Pi via the MQTT protocol to the Cloud, or more specifically to AWS IoT Core, a service for connecting IoT devices to other AWS services.

AWS IoT Core stores the live data in an AWS DynamoDB table (NoSQL database service). A REST API through AWS API Gateway later provides the live data from the DynamoDB table. Finally, the respective applications, like the virtual assistant or the web application, can show information on indoor air quality parameters.

AWS IoT Core additionally sends the data further to AWS IoT Analytics, which supports the preprocessing, storage, and analysis of IoT data without managing hardware or infrastructure. AWS IoT Analytics contains all historical data of the Indoor Air Quality monitoring since the 3rd of May 2022. In addition, content delivery rules regularly query and update individual files in AWS S3 (e.g., data of the last 15 minutes or day), which are later accessed by the web application and used for machine learning training and forecasting.

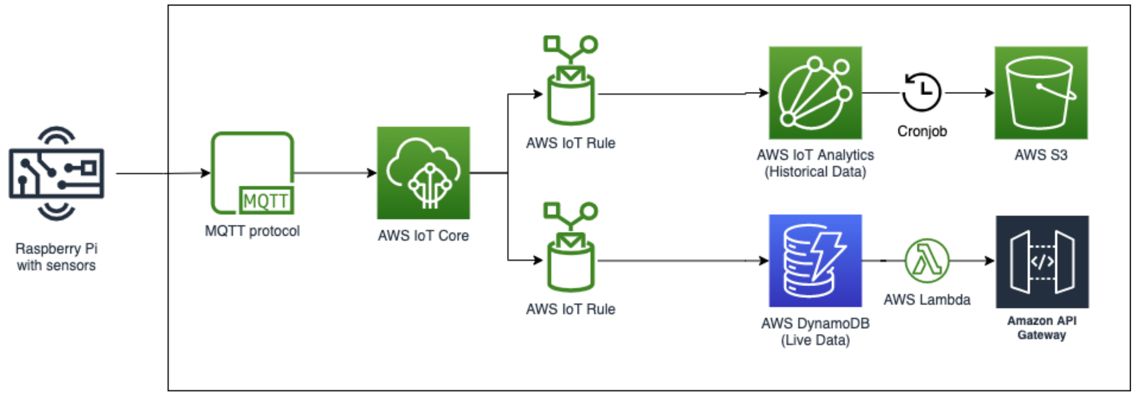


Figure 24: Architecture of data transfer and processing from the Raspberry Pi to AWS

Time-Series forecasting with Machine Learning

The open-source library Darts was used for time-series forecasting, as it supports an easy usage of past covariates. For example, an open door is a covariate to CO2. Even temperature and humidity can affect CO2 (cf. Unit8 SA, 2022).

As the model needed to be trained, the total historical data since the 3rd of May 2022 was downloaded from the Cloud Storage (AWS S3) and split into two sets: A training set and a test set. As mentioned in the concept, the training data is much larger than the test data.

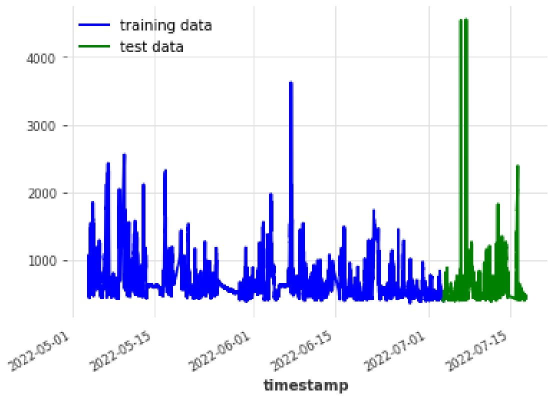


Figure 25: Full historical data of recorded CO2 concentrations, split into training and a test sets

The training set contains data for precisely three months, from the 3rd of May 2022, 16:12:40, until the 3rd of July, 16:12:40. The test set includes data for the two following weeks, from the 3rd of July, 16:12:40, until the 17th of July, 16:12:40. Every 10 seconds a measurement was sent from the Raspberry Pi to AWS IoT Core.

Once the data was ready, all prerequisites were made for training the model. Therefore with the BlockRNNModel class, the model gets instantiated using the following setting:

- "GRU" was the chosen model.
- The input and output chunks were set to a length of 90, meaning that the model always gets the data of the last 15 minutes to predict the future 15 minutes. That is because every 10 seconds, a measurement is recorded. Fifteen minutes are 900 seconds. Therefore 90 times 10 seconds are 15 minutes.
- The model got trained on the dataset 60 times (epochs).
- The GPU of the computer was used for faster training / better performance.

```
model_co2 = BlockRNNModel(
    model="GRU",
    input_chunk_length=90,
    output_chunk_length=90,
    n_epochs=60,
    model_name="GRU",
    pl_trainer_kwargs={
      "accelerator": "gpu",
      "gpus": [0]
    },
    force_reset=True,
)
```

Figure 26: Screenshot of Python code for instantiating the model

Finally, the model got trained with the training data, which takes up to several hours, depending on the computer's processing power.

```
model_co2.fit(series=train_series_co2, past_covariates=train_covariates, verbose=True)
```

Figure 27: Screenshot of Python code for training the model

For the prototype, the model received all CO2 values from the training data with past covariates, including measurements of the living room door, the balcony door, the living room window, the kitchen window, and the humidity and temperature indoors and outdoors.

Once the training is done, the model can predict future CO2 values, as shown in Figure 28. Finally, the test data comes into action to measure the model's performance and accuracy, which is explained in section 5.3, proving the hypothesis that "A self-learning system based on machine learning and IoT sensor data can reliably predict indoor air quality parameters."

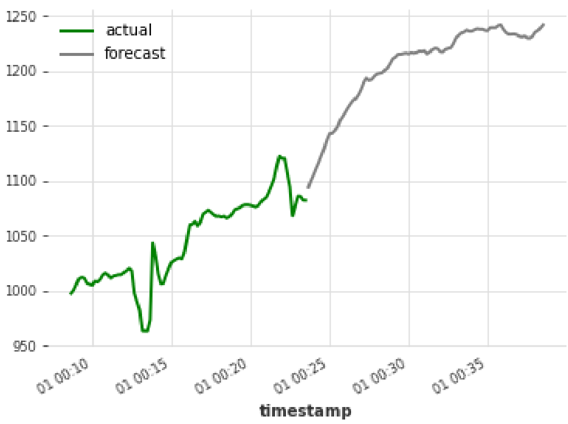


Figure 28: Actual CO2 concentration with forecast

The model predictions take only a few seconds compared to the model training. The Raspberry Pi then runs in parallel to the sensor data reading another Python script that periodically, every 30 seconds makes the predictions for the CO2 concentration in 15 minutes, which is also sent to AWS IoT Core for further processing.

Notifications and Recommendations

One essential feature of the system to interact with the users is the notifications- and recommendations service. The prototype uses visual and acoustic warnings to notify users about unhealthy CO2 levels.

Once the CO2 concentration or prediction of CO2 exceeds a specific limit, 1.000 ppm, the room's occupants will be warned. The visual warning is realized through Smart LEDs (Philips Hue) in the living room turning red. The color red is well suited for this, as it instantly grabs people's attention and is associated with danger (cf. Cherry, 2020).

Figure 29: Living room without warnings

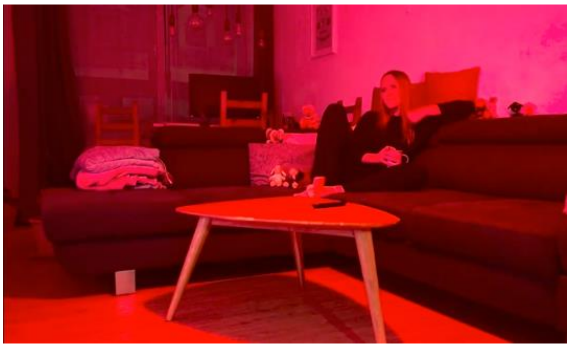


Figure 30: Living room with warnings

Figure 30: Amazon Echo Dot

Additionally, the room's occupants are warned acoustically through the smart speaker (Amazon Echo Dot), saying a sentence like: "Please open the balcony door, the level of CO2 is too high". This phrase also includes the recommendation to solve the issue with the CO2 concentration by simply opening the balcony door. Ideally, the users consider the system's warnings and open the balcony door or windows to improve indoor air quality. Then, depending on different factors (e.g., outdoor temperature, pressure, or wind speed), the room needs to be ventilated for a few or more minutes with fresh air until the CO2 concentration drops below 500 ppm. 500 ppm is sufficient enough not to warn the users too often and to keep the CO2 levels sustainably below 1.000 ppm.

Web Application

A web application accessible for mobile- and desktop devices via the URL https://iaq.brgr.rocks provides an additional interface for users to interact with the system and to check on the most important indoor air quality parameters. The web application consists of a Dashboard page split into current- and historical data.

The live data is on top, showing the current values for temperature, humidity, CO2 concentration, the predicted CO2 concentration, states of doors and windows, whether they are open or closed, and outdoor temperature, humidity, and pressure.

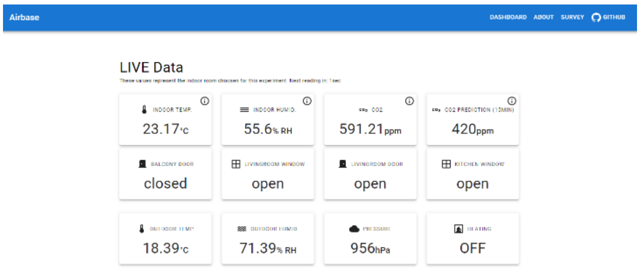


Figure 31: User Interface of the Web Application

The application automatically updates every 30 seconds requesting the data from the provided REST API. Then, values are shown based on the system's latest sensor readings sent by the Raspberry Pi to AWS IoT Core.

Additionally, if a parameter exceeds its limit, it is highlighted red, showing a message in the tooltip when hovering over the info icon. For example, the limit for the CO2 concentration is equal to the notification service set to 1.000 ppm.

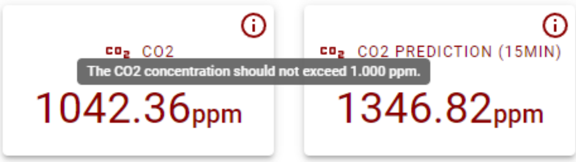


Figure 32: Warning state within the UI

One additional parameter the live data shows that is irrelevant for the prototype is the heating in percent.

Furthermore, below the live data, the historical data is shown. This way, a user can get an overview of the CO2 concentration's Development over the last day, hour, or 15 minutes. With this, it is directly visible what effect a specific action had, for example, a drop in the CO2 levels when opening a window. It is also visible how long it was necessary to wind a room until the CO2 concentration dropped below 500 ppm.

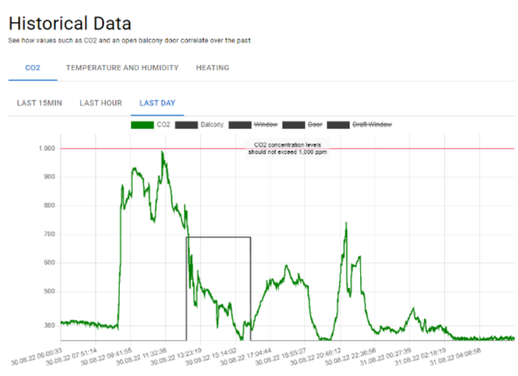


Figure 33: Historical Data of the Web Application

In the image of figure 34, the green line represents the CO2 concentration, the red line the limit of 1.000 ppm, and the black line represents the open balcony door.

Voice Interface via Amazon Alexa

In contrast to the web application, an Amazon Alexa Skill provides a basic voice interface for users to get essential information about the current indoor air quality through speech. As mentioned in the concept, this doesn't require any mobile- or desktop device. Therefore, all occupants in a room can interact with the system and request information about indoor air quality naturally.

Figure 34: Smart Speaker Echo Dot

The Alexa Skill provides responses for the following indoor air quality parameters: CO2 concentration and its prediction, indoor and outdoor temperature, humidity, the open/close state of relevant doors and windows, particulate matter, and the concentration of gases and heating.

To get information, for example, about the CO2 concentration, the user would ask the skill, "What's the CO2 level?" Then, Alexa would trigger a request to the REST API, gathering data on the current indoor air quality and ultimately answering the user, "The current CO2 level is at 540 ppm."

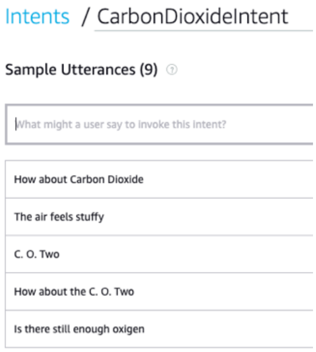


Figure 35: Alexa Developer Console

5.3 Proving the Hypotheses

Historical Forecasting

The first hypothesis that needs to be proven is: "A self-learning system based on machine learning and IoT sensor data can reliably predict indoor air quality parameters."

So-called historical forecasting can evaluate the performance of the trained model and verify its predictions' accuracy. Moreover, it can simulate forecasts on historical data. Therefore, the obtained test data was used to predict a specific time frame into the future, iteratively, from the beginning of the test data until the end. Finally, there are two graphs: the CO2 concentration of the original test data and the predictions of it. Both charts, the test data, and the historical forecasts are then compared.

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

Figure 36: Formula for calculating MAPE

One of the most commonly used metrics to measure a model's performance or accuracy is the so-called Mean Absolute Percentage Error (MAPE). It is the mean of all absolute percentage errors between the predicted and actual values. It is easy to understand and compare (cf. Vandeput, Forecast KPIs: RMSE, MAE, MAPE & Bias, 2019).

Another metric is the Mean Absolute Error (MAE), which provides the mean absolute difference between the actual and the predicted value. Compared to MAPE, it gives an absolute value instead of a percentage (cf. Vandeput, Forecast KPIs: RMSE, MAE, MAPE & Bias, 2019).

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

Figure 37: Formula for calculating MAE

When it comes to time-series forecasting, a MAPE score below 20% is considered as good. Likewise, a MAPE score below 10% is very good.

Predicting 15 minutes into the future with the trained GRU model provided an overall MAPE score of 5.42%. Therefore, the model's predictions were 94.58% accurate, which is a very good result. As a second metric, the MAE score was 34.82 (ppm).

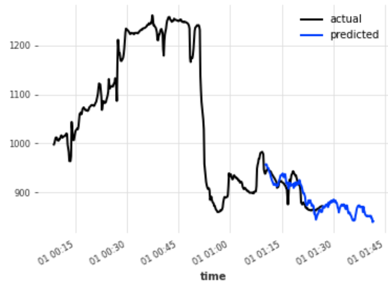


Figure 38: Actual vs. Predictions

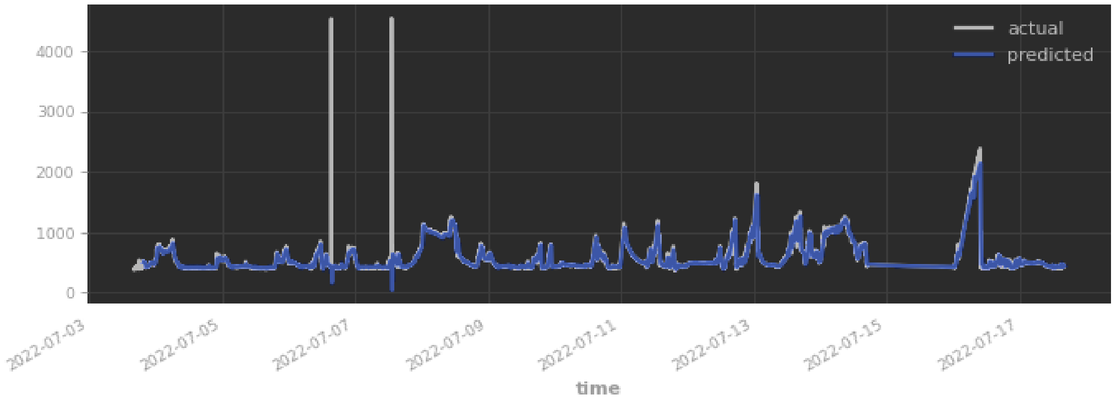


Figure 39: Historical forecasting with a 15 minutes forecast horizon, using the trained GRU model

The further the model predicts into the future, the less accurate its predictions become. For example, forecasting the next 30 minutes resulted in a MAPE score of 8.59% with an MAE of 108 (ppm). However, it has to be mentioned that the model was specially trained to predict the next 15 minutes, not further.

Finally, the first hypothesis can be proven, as the model provides very good and reliable predictions on CO2 concentration indoors.

User Survey

"A self-learning system can provide practical and user-friendly suggestions to the occupants of a room" was part of the second hypothesis that had to be proven.

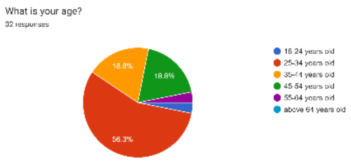


Figure 40: Age distribution of survey participants

A user survey was the best method for answering the question about practicality and user-friendliness.

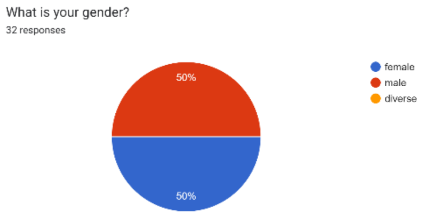


Figure 41: Gender distribution of survey participants

Therefore, 32 people were invited to participate. First, they were introduced to the system, including the thesis background, the notifications and recommendations through light and sound, the web application, and the voice interface.

In the end, the participants were asked to fill in voluntarily and anonymously an online survey with questions about their awareness of health risks related to bad indoor air quality, their tech savviness, and their feedback and perception of the overall system.

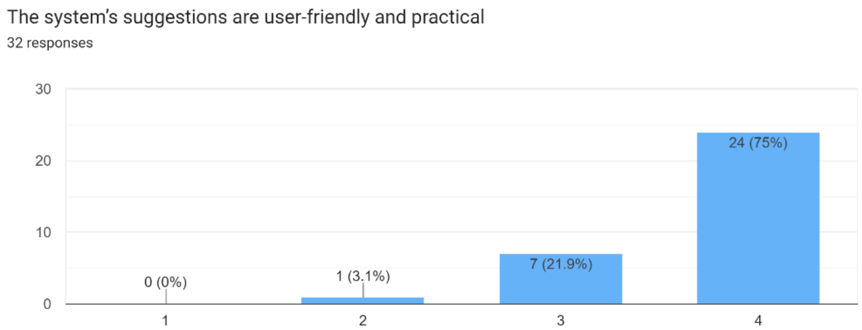


Figure 42: Survey results about user-friendliness and practicality

In the shown graphs of Figures 43, 44, 45, and 46, 1 means strongly disagree, 2 means disagree, 3 means agree, and 4 means strongly agree. The survey concluded that 96.9% of respondents found the system's suggestions user-friendly and practical. Moreover, 87.5% would be ready to install such a system at home. On the other hand, 18.7% perceived the system's visual or acoustic suggestions as irritating.

While most participants were aware that high CO_2 concentration could negatively impact mental concentration, 25% weren't aware that exposure to indoor air pollution could cause respiratory diseases like asthma, heart diseases, or cancer in the long term.

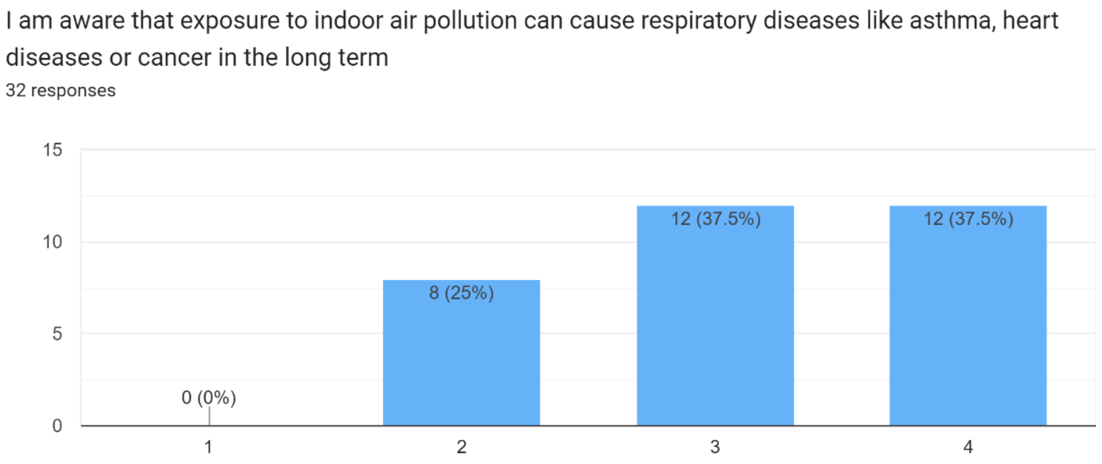


Figure 43: Survey results about awareness of health risks

All respondents agreed that the system could help create awareness and improve indoor air quality.

Another important finding for further developing such a system is that 28.1% of users worry about IoT data privacy and security. Therefore, investing efforts into security and user privacy protection is essential.

Considering the results of this survey, most users found the system practical and user-friendly. Therefore, an essential part of the second hypothesis is proven.

The survey also offered a free text field for additional feedback, which is roughly covered in the outlook of chapter 6.

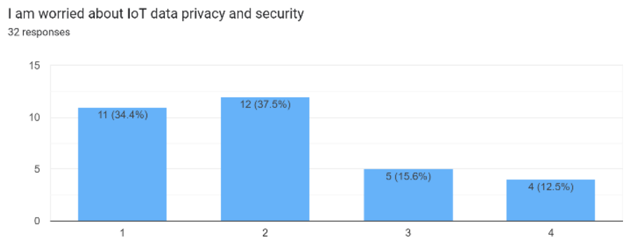


Figure 44: Survey results about data privacy

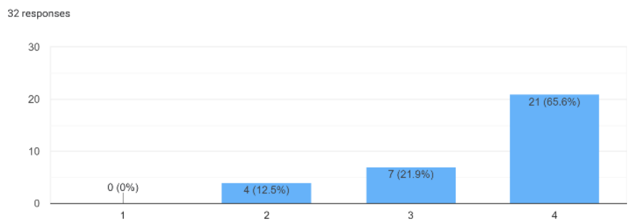


Figure 45: Survey results about readiness to install such system at home

Experiment

The complete second hypothesis stated that "A self-learning system can provide practical and user-friendly suggestions to the occupants of a room to significantly and efficiently improve indoor air quality."

The user survey already answered the first part, that the system could provide practical and user-friendly suggestions to the occupants of a room. Although, "if the suggestions can help to significantly and efficiently improve indoor air quality" is still to be answered.

Therefore, an experiment was carried out. For a week, from Monday the 25th of July until Sunday the 31st of July, a person had been in the room for activities such as working from home, eating lunch, watching TV, talking on the phone, reading books, and doing sports. However, during the first week, the system didn't provide suggestions when the CO2 concentration exceeded 1.000 ppm. This phase is called Variant A.

The following week, from the 1st of August until the 7th of August, the experiment was continued. Only this time, the system provided visual and acoustic warnings and suggestions when the CO2 levels or their respective predictions exceeded 1.000 ppm. This second phase is called Variant B.

There is a clear difference when comparing Variant A (without suggestions) to Variant B (with suggestions).

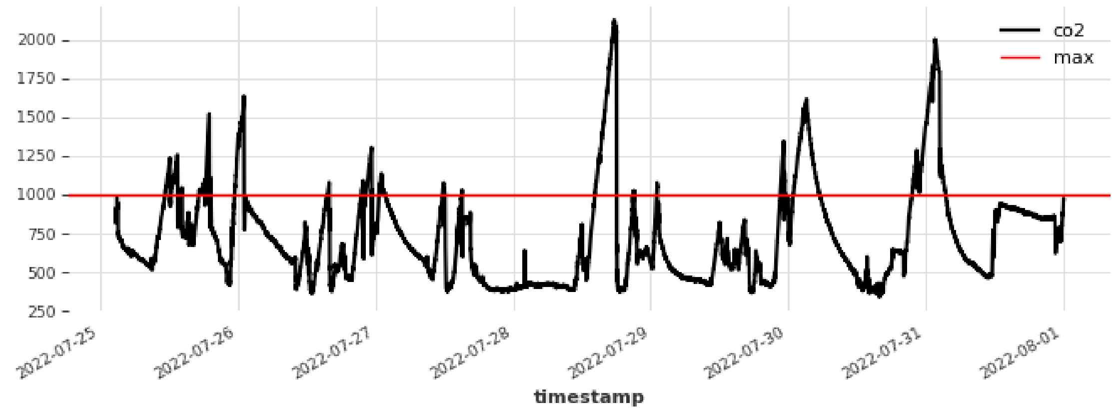


Figure 46: Results of Variant A (without suggestions)

Variant A had an average/mean CO2 concentration of 722.88 ppm. The maximum measured CO2 concentration during the test period was 2.124 ppm. Also, the CO2 concentration was for 8.621 measurements above 1.000 ppm, which relates to 23.95 hours when CO2 concentration was above its limit (1 measurement every 10 seconds) within this test phase of seven days.

In contrast, Variant B had a significantly better result. The average/mean CO2 concentration was at 566.69 ppm with a maximum of 1451.34. When looking at the graph, the peak of 1451.34 ppm only lasted for a few seconds as somebody accidentally exhaled above the sensor. The best metric is the number of measurements the CO2 level exceeded 1.000 ppm, which happened in Variant B only 21 times, equal to 3.5 minutes compared to almost 24 hours in Variant A.

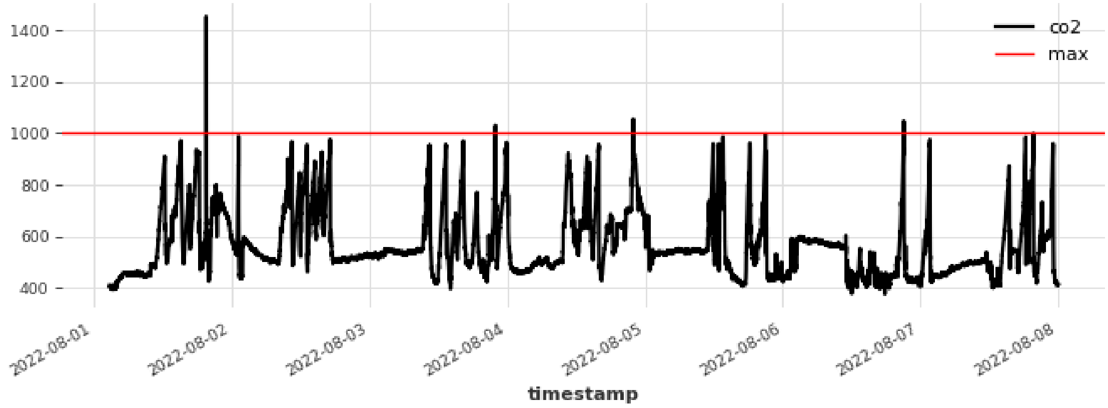
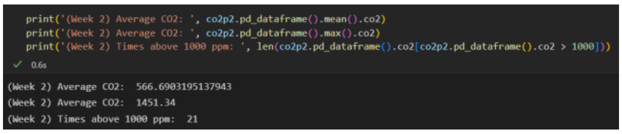


Figure 47: Results of Variant B (with suggestions)

Therefore, the second hypothesis is also proven, as the system's suggestions can help to significantly and efficiently improve indoor air quality. It's significant because Variant A and Variant B show a clear difference.



```
print('(Week 2) Average CO2: ', co2p2.pd_dataframe().mean().co2)
print('(Week 2) Average CO2: ', co2p2.pd_dataframe().max().co2)
print('(Week 2) Times above 1000 ppm: ', len(co2p2.pd_dataframe().co2[co2p2.pd_dataframe().co2 > 1000]))
✓ 0.6s
(Week 2) Average CO2:  566.6903195137943
(Week 2) Average CO2:  1451.34
(Week 2) Times above 1000 ppm:  21
```

Figure 48: Screenshot of IDE/Python code

And efficient because the user isn't warned too often, although still sufficient to keep the CO2 concentration lasting below 1.000 ppm. Figures 48 and 49 make the difference visible. Furthermore, Figure 50 shows a screenshot of the relevant data processing in the Python IDE PyCharm.

6. Conclusion and Outlook

Conclusion

Based on the concept and architecture, a prototype was developed. Sensors connected to a Raspberry Pi board collected data from the 3rd of May 2022, securely stored in the Cloud (AWS).

Based on the test data, historical forecasting has proven that the model's predictions for the upcoming 15 minutes are 94.58% accurate. The accuracy proves that the system can reliably predict indoor carbon dioxide concentrations / indoor air quality parameters.

Additionally, the user survey showed that the overall system, including the web application, the voice interface, and the notifications/recommendations service, is perceived as user-friendly and practical. As a result, 96.9% out of 32 participants voted for it.

Furthermore, an experiment showed that the system's warnings and suggestions help the user significantly and efficiently improve indoor air quality, more precisely, CO2 concentration. In conclusion, both hypotheses were proven.

Outlook

The experiment showed a positive effect on indoor air quality and proved a valid reason why such systems are a valuable help in improving indoor air quality. Furthermore, users could be explicitly supported with warnings and suggestions once specific indoor air quality parameters are below or above a particular value, especially in buildings without automatic ventilation systems.

Therefore, further development of this system would be an excellent investment. Additional parameters, despite the CO2 concentration, can be added in the future. Also, the system could provide more precise warnings and recommendations and consider users' worries about data protection and privacy. One idea would be a complete system, with sensors, speakers, microphones, and lights all in one.

During the user survey, many participants expressed their wish to have such a system in their homes via the free text field. Some comments were, “Great product! Will definitely consider getting one installed”, or “Would like to have one at home.” Additionally, users provided excellent suggestions for improvement. One recommendation was to add “color grading” to the visual warnings, meaning that the lights slowly, instead of abruptly, turn from their natural light color to red depending on the CO2 level. Others wished for better indicators when the air quality is regular or poor, e.g., when asking the virtual assistant. Another request was to add a settings page to the web applications, where users can turn off/on certain notifications or change modes.

Finally, this thesis shows great potential and demand for such systems.

References

Adafruit Industries. (2021, November 15). *Adafruit SGP30 TVOC/eCO2 Gas Sensor.* Retrieved January 2022, from https://cdn-learn.adafruit.com/downloads/pdf/adafruit-sgp30-gas-tvoc-eco2-mox-sensor.pdf

Adafruit Industries. (November 2021). *PM2.5 Air Quality Sensor.* Abgerufen am January 2022 von https://cdn-learn.adafruit.com/downloads/pdf/pm25-air-quality-sensor.pdf

Ahn, J., Shin, D., Kim, K., & Yang, J. (2017). Indoor Air Quality Analysis Using Deep Learning with Sensor Data. *Sensors, 17*(11).

Amazon.com, Inc. (2022, 9 7). *Create Intents, Utterances, and Slots*. Retrieved from Alexa Skills Kit: https://developer.amazon.com/en-US/docs/alexa/custom-skills/create-intents-utterances-and-slots.html

Aqara. (2022, 9 7). *Aqara Door and Window Sensor*. Retrieved from Aqara: https://www.aqara.com/us/door_and_window_sensor.html

Axelsson, G., Andersson, E. M., & Barregard, L. (2015). Lung cancer risk from radon exposure in dwellings in Sweden: how many cases can be prevented if radon levels are lowered? *Cancer Causes Control, 26*(4), 541-547.

Aydogan, A., & Montoya, L. D. (2011). Formaldehyde removal by common indoor plant species and various growing media. *Atmospheric Environment, 45*(16), 2675-2682.

AZ-Delivery Vertriebs GmbH. (n.d.). *MQ-135 Gas Sensor Modul.* Retrieved January 2022, from https://cdn.shopify.com/s/files/1/1509/1638/files/MQ-135_Gas_Sensor_Modul_Datenblatt_AZ-Delivery_Vertriebs_GmbH_94c8b153-6f8e-4696-8cfa-0d8ba832d603.pdf?v=1608535104

Azuma, K., Kagi, N., Yanagi, U., & Osawa, H. (2018). Effects of low-level inhalation exposure to carbon dioxide in indoor environments: A short review on human health and psychomotor performance. *Environment International, 121*(1), 51-56.

Benda, G. J. (1999). *Indoor Air Quality Case Studies Reference Guide.* USA: Fairmont Press.

BfS. (2019). *Strahlung und Strahlenschutz.* Salzgitter: Bundesamt für Strahlenschutz.

BfS, & UBA. (2015, March). GESÜNDER WOHNEN — ABER WIE? Praktische Tipps für den Alltag.

BMUV. (2020, June 8). *Richtiges Lüften und Heizen*. Retrieved from https://www.bmuv.de/themen/gesundheit-chemikalien/gesundheit/innenraumluft/richtiges-lueften-und-heizen

Bosch Sensortec GmbH. (2021, October). *BME280 - Data sheet.* Retrieved January 2022, from https://www.bosch-sensortec.com/media/boschsensortec/downloads/datasheets/bst-bme280-ds002.pdf

Brey, P., & Søraker, J. H. (2009). Philosophy of Computing and Information Technology. In D. M. Gabbay, & P. W. Thagard, *Philosophy of Technology and Engineering Sciences* (pp. 1341-1407). North-Holland.

Brilli, F., Fares, S., Ghirardo, A., De Visser, P., Calatayud, V., Muñoz, A., . . . Menghini, F. (2018). Plants for Sustainable Improvement of Indoor Air Quality. *Trends in Plant Science, 23*(6), 507-512.

Bundesamt für Strahlenschutz. (2021, 10 06). *What is radon?* Retrieved from https://www.bfs.de/EN/topics/ion/environment/radon/introduction/introduction.html

Cherry, K. (2020, 9 13). *The Color Psychology of Red*. Retrieved from Verywell Mind: https://www.verywellmind.com/the-color-psychology-of-red-2795821

Defra. (2005). *Particulate Matter in the United Kingdom.* Retrieved from Department for the Environment, Food and Rural Affairs: https://uk-air.defra.gov.uk/library/assets/documents/reports/aqeg/Particulate_Matter_in_The_UK_2005_Summary.pdf

Du, Y., Xu, X., Chu, M., Guo, Y., & Wang, J. (2016). Air particulate matter and cardiovascular disease: the epidemiological, biomedical and clinical evidence. *Journal of thoracic disease, 8*(1), 8-19.

EPA. (2016). *Integrated Science Assessment (ISA) for Oxides of Nitrogen – Health Criteria.* Washington, DC: U.S. Environmental Protection Agency.

EPA. (2018). *Guide to Air Cleaners in the Home.* USA: U.S. Environmental Protection Agency.

EPA. (2021, September 30). *Carbon Monoxide's Impact on Indoor Air Quality*. Retrieved from https://www.epa.gov/indoor-air-quality-iaq/carbon-monoxides-impact-indoor-air-quality

EPA. (2021, March). *Improving Indoor Air Quality*. Retrieved from U.S. Environmental Protection Agency: https://www.epa.gov/indoor-air-quality-iaq/improving-indoor-air-quality

EPA. (2021, September). *Indoor Air Quality*. Retrieved from United States Environmental Protection Agency: https://www.epa.gov/report-environment/indoor-air-quality

EPA. (2021, May). *Particulate Matter (PM) Basics*. Retrieved from https://www.epa.gov/pm-pollution/particulate-matter-pm-basics

EPA. (2021, July). *Radionuclide Basics: Radon*. Retrieved from https://www.epa.gov/radiation/radionuclide-basics-radon

EPA. (2021, September). *Volatile Organic Compounds' Impact on Indoor Air Quality*. Retrieved from https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality

EPA. (2022, January 11). *How much ventilation do I need in my home to improve indoor air quality?* Retrieved from https://www.epa.gov/indoor-air-quality-iaq/how-much-ventilation-do-i-need-my-home-improve-indoor-air-quality

EPA. (2022, January 22). *Indoor Particulate Matter*. Retrieved from https://www.epa.gov/indoor-air-quality-iaq/indoor-particulate-matter

European Committee for Standardization. (2017). EN 12831. *Energy performance of buildings - Method for calculation of the design heat load*, 58.

European Environment Agency. (2019, December 5). *Trends in atmospheric concentrations of CO2 (ppm), CH4 (ppb) and N2O (ppb), between 1800 and 2017*. Retrieved from https://www.eea.europa.eu/data-and-maps/daviz/atmospheric-concentration-of-carbon-dioxide-5

Eve Systems GmbH. (2018, November). *eve room - Reviewers Guide.* Retrieved from Eve: https://www.evehome.com/sites/default/files/2018-11/RG_Eve_Room_DE_20180927_vr.pdf

Ezraty, B., Chabalier, M., Ducret, A., Maisonneuve, E., & Dukan, S. (2011). CO2 exacerbates oxygen toxicity. *EMBO Reports, 12*(4), 321-326.

Farahani, S. (2008). Chapter 2 - ZigBee/IEEE 802.15.4 Networking Examples. *ZigBee Wireless Networks and Transceivers*, 25-32.

Fernández, E., Ballbè, M., Sureda, X., Fu, M., Saltó, E., & Martínez-Sánchez, J. (2015). Particulate Matter from Electronic Cigarettes and Conventional Cigarettes: a Systematic Review and Observational Study. *Curr Environ Health, 2(4)*, 423-429.

Google Inc. (n.d.). *What is machine learning?* Retrieved January 2022, from https://cloud.google.com/learn/what-is-machine-learning

Gross, T. (2021, October 5). Retrieved January 2022, from https://www.uni-bamberg.de/hci/

Hamanaka, R., & Mutlu, G. (2018). Particulate Matter Air Pollution: Effects on the Cardiovascular System. *Front Endocrinol, 9*, 680.

Hartmann, A., & Kriegel, M. (2020). *Risk assessment of aerosols loaded with virus based on CO2-concentration.* Berlin: Technical University of Berlin.

Harvard Health Publishing. (2021, February 15). *Easy ways you can improve indoor air quality*. Retrieved from https://www.health.harvard.edu/staying-healthy/easy-ways-you-can-improve-indoor-air-quality

Health Canada. (2017). *Guide for Radon Measurements in Residential Dwellings.* Ottawa: Health Canada.

IBM Cloud Education. (2020, May 1). *Deep Learning*. Retrieved from https://www.ibm.com/cloud/learn/deep-learning

IBM Cloud Education. (2020, September 14). *Recurrent Neural Networks*. Retrieved from https://www.ibm.com/cloud/learn/recurrent-neural-networks

Jakaria, A., Hossain, M. M., & Rahman, M. (2020). Smart Weather Forecasting Using Machine Learning. *ArXiv, abs/2008.10789*.

Kar, M., Mukhopadhyay, J., & Deb Sarkar, M. (2022). *South Asia and Climate Change: Unravelling the Conundrum.* India: Routledge.

Kim, H., Kim, W.-H., Kim, Y.-Y., & Park, H.-Y. (2020). Air Pollution and Central Nervous System Disease: A Review of the Impact of Fine Particulate Matter on Neurological Disorders. *Frontiers in Public Health, 8*.

Lawrence Berkeley National Laboratory. (n.d.). *Minimizing Airborne Particles Indoors*. Retrieved January 2022, from https://svach.lbl.gov/hvac-air-filtration/

Maxim Integrated Products, Inc. (2019). *DS18B20 - Programmable Resolution 1-Wire Digital Thermometer.* Retrieved January 2022, from https://datasheets.maximintegrated.com/en/ds/DS18B20.pdf

Meyer, C. (2021). *Overview of TVOC and Indoor Air Quality.* Tokyo: Renesas Electronics.

Miller, S. L., Facciola, N. A., Toohey, D., & Zhai, J. (2017). Ultrafine and Fine Particulate Matter Inside and Outside of Mechanically Ventilated Buildings. *International journal of environmental research and public health, 14*(2), 128.

Moreno, A. M., Seffah, A., Capilla, R., & Sánchez-Segura, M.-I. (2013). HCI Practices for Building Usable Software. *Computer, 46*(4), 100-102.

Nephew, B., Torres, A., Nephew, G., Lu, S., Cali, R., & Howell-Munson, A. (2021). *The Role of Ultrafine.* Nova.

New York State Health Department. (2018, February). *Fine Particles (PM 2.5) Questions and Answers*. Retrieved from https://www.health.ny.gov/environmental/indoors/air/pmq_a.htm

Nienaber, F., Rewitz, K., Seiwert, P., & Müller, D. (2021). *Einfluss der Luftfeuchte auf den Menschen.* Aachen: RWTH-EBC.

NVIDIA Corporation. (2018, August 2). *SuperVize Me: What's the Difference Between Supervised, Unsupervised, Semi-Supervised and Reinforcement Learning?* Retrieved from https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/

OASIS. (2022, 09 06). Retrieved from MQTT.org: https://mqtt.org/

Osagbemi, G., Adebayo, Z., & Aderibigbe, S. (2009). Awareness, attitude and practice towards indoor air pollution (Iap) amongst residents of Oke. *The Internet Journal of Epidemiology, 8*(2).

Palmer, B. (2015, 5 19). *Do We Exhale Carbon*. Retrieved from NRDC: https://www.nrdc.org/stories/do-we-exhale-carbon

Prabjit Barn, P., Fong, D., & Kosatsky, T. (2016). Carbon monoxide detectors: A powerful, underused way to protect lives. *British Columbia Medical Journal, 48*(1), 49. Retrieved from https://bcmj.org/bccdc/carbon-monoxide-detectors-powerful-underused-way-protect-lives

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., . . . Mohamed, S. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature, 597*, 672–677.

RKI. (2019). Richtwerte für Stickstoffdioxid (NO2) in der Innenraumluft. *Bundesgesundheitsblatt, 62*, 664-676 .

Sak, H., Senior, A., & Beaufays, F. (2014). *Long Short-Term Memory Recurrent Neural Network Architectures.* USA: INTERSPEECH.

Satish, U., Fisk, W., Mendell, M., Eliseeva, K., Hotchi, T., Sullivan, D., . . . Teng, K. (2011). Impact of CO2 on human decision making and productivity. *12th International Conference on Indoor Air Quality and Climate 2011, 2*.

Seifert, B. (1999). Richtwerte für die Innenraumluft. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 270-278.

Sensirion AG. (2020, May). *Datasheet Sensirion SCD30 Sensor Module.* Retrieved January 2022, from https://www.sensirion.com/fileadmin/user_upload/customers/sensirion/Dokumente/9.5_CO2/Sensirion_CO2_Sensors_SCD30_Datasheet.pdf

Seppanen, O., Fisk, W. J., & Lei, Q. H. (2006). *Room temperature and productivity in office work.* Helsinki, Berkeley: Helsinki University of Technology, Lawrence Berkeley National Laboratory.

SGX Sensortech. (n.d.). *The MiCS-6814 is a compact MOS sensor with three fully independent sensing elements on one package.* Retrieved January 2022, from https://www.sgxsensortech.com/content/uploads/2015/02/1143_Datasheet-MiCS-6814-rev-8.pdf

Shen, G., Gaddam, C. K., Ebersviller, S. M., Vander Wal, R. L., Williams, C., Faircloth, J. W., . . . Hays, M. D. (2017). A laboratory comparison of emission factors, number size distributions and morphology of ultrafine particles from eleven different household cookstove-fuel systems. *Environmental science & technology, 51*(11), 6522–6532.

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., & Elmqvist, N. (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction: Sixth Edition.* Pearson Education.

Statistics and Data. (2022, 09 06). *Statistics and Data*. Retrieved from https://statisticsanddata.org/data/most-popular-machine-learning-libraries/

Stats NZ. (2018, October 18). *Natural sources of particulate matter*. Retrieved from https://www.stats.govt.nz/indicators/natural-sources-of-particulate-matter

tado GmbH. (2018, August). *Smarter heating with your Climate Assistant.* Retrieved January 2022, from tado GmbH: https://www.tado.com/t/app/uploads/2018/08/tado%C2%B0_SalesSheet_V3_EN.pdf

Taheri, S., & Razban, A. (2021). Learning-based CO2 concentration prediction: Application to indoor air quality control using demand-controlled ventilation. *Building and Environment*.

Toftum, J., Reimann, G., Foldbjerg, P., Clausen, G., & Fanger, P. O. (2002). *Perceived air quality, thermal comfort, and SBS symptoms at low air temperature and increased radiant temperature.* Lyngby: Technical University of Denmark.

UBA. (2008). *Gesundheitliche Bewertung.* Germany: Umweltbundesamt.

UBA. (2017). *Anforderungen an Lüftungskonzeptionen in Gebäuden.* Berlin: Umweltbundesamt.

UBA. (2017, November 7). *Richtig Lüften, Schimmelbildung vermeiden*. Retrieved from https://www.umweltbundesamt.de/themen/gesundheit/umwelteinfluesse-auf-den-menschen/schimmel/richtig-lueften-schimmelbildung-vermeiden

UBA. (2019, April 9). *Tipps für richtiges Lüften*. Retrieved from https://www.umweltbundesamt.de/tipps-fuer-richtiges-lueften

UBA. (2021, June 18). *Feinstaub*. Retrieved from https://www.umweltbundesamt.de/themen/luft/luftschadstoffe-im-ueberblick/feinstaub#undefined

UBA. (2021, 10 5). *Feinstaub-Belastung*. Retrieved from https://www.umweltbundesamt.de/daten/luft/feinstaub-belastung#herkunft

UBA. (2021, May 25). *Kohlenmonoxid*. Retrieved from https://www.umweltbundesamt.de/themen/luft/luftschadstoffe-im-ueberblick/kohlenmonoxid

UBA. (2021, Dezember 22). *Richtig Lüften in Schulen*. Retrieved from https://www.umweltbundesamt.de/richtig-lueften-in-schulen#warum-ist-ein-regelmassiger-luftaustausch-in-klassenzimmern-grundsatzlich-wichtig-und-in-der-pandemie-umso-mehr

UBA. (2021, October 05). *Stickstoffdioxid-Belastung*. Retrieved from https://www.umweltbundesamt.de/daten/luft/stickstoffdioxid-belastung#gesundheitliche-wirkungen

UC Berkeley School of Information. (2020, June 26). *What Is Machine Learning (ML)?* Retrieved from https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/

Unit8 SA . (2022, 09 06). *Time Series Made Easy in Python*. Retrieved from https://unit8co.github.io/darts/

Vandeput, N. (2019, 7 5). *Forecast KPIs: RMSE, MAE, MAPE & Bias*. Retrieved from Towards Data Science: https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d

Vandeput, N. (2019, 7 5). *Forecast KPIs: RMSE, MAE, MAPE & Bias*. Retrieved from Towards Data Science: https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d

Vega-Barbas, M., Pau, I., Augusto, J. C., & Seoane, F. (2018). Interaction Patterns for Smart Spaces: A Confident Interaction Design Solution for Pervasive Sensitive IoT Services. *IEEE Access, 6*, 1126-1136.

Von Pettenkofer, M. (1858). *Über den Luftwechsel in Wohngebäuden.* München: Cotta.

Wells, G. (2012). *Superbodies: Peak Performance Secrets From the World's Best Athletes.* USA: Collins.

WHO. (2010). *WHO guidelines for indoor air quality: selected pollutants.* Copenhagen: WHO Regional Office for Europe.

WHO. (2013). *Combined or multiple exposure to health stressors in indoor built environments.* Bonn: WHO Regional Office for Europe.

WHO. (2018). *WHO Housing and health guidelines.* Geneva: World Health Organization.

WHO. (2021, September 22). *Ambient (outdoor) air pollution*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

WHO. (2021, February). *Radon and health*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/radon-and-health

Xiaomi. (2020). *Mi Air Purifier 3C*. Retrieved January 12, 2022, from https://www.mi.com/global/mi-air-purifier-3c/

Yu, R., Gu, N., & Ostwald, M. J. (2012). *Explore patterns of human-computer interactions in parametric design environments.* Taiwan: National Taiwan University Press. Retrieved January 2022, from Informatics 131 Human-Computer Interaction: https://www.ics.uci.edu/~kay/courses/131/Slides4up.pdf

Table of Figures