**Research document**

***SUE project***

A blue and white logo

Description automatically generated

|  |
| --- |
| **Datum : 17-06-2024** |
| **Versie : 1** |
| **Status : Finished** |
| **Auteur : Dani Berkvens, Murphy Minneboo, Edouard Lagesse, Lianne Louwers** |

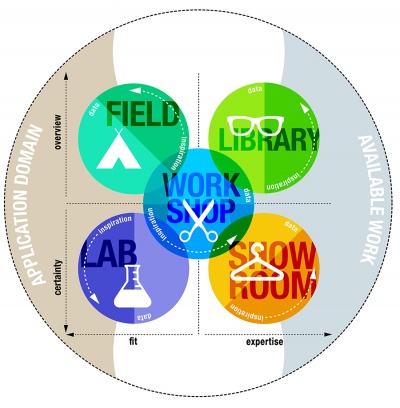


Table of contents

[Introduction 3](#_Toc169513063)

[1. Research questions, findings and answers 3](#_Toc169513064)

[1.1 Sub-question 1 3](#_Toc169513065)

[1.1.1 Available product analysis (literature study) 3](#_Toc169513066)

[1.1.2 Literature study 3](#_Toc169513067)

[1.1.3 Expert interview 3](#_Toc169513068)

[1.1.4 Document analysis (bekijken documenten die je hebt gevonden) 4](#_Toc169513069)

[1.1.5 Observation: 4](#_Toc169513070)

[1.1.6 Answer to sub-question 1 5](#_Toc169513071)

[1.2 Sub-question 2 5](#_Toc169513072)

[1.2.1 literature study 5](#_Toc169513073)

[1.2.2 expert interview 6](#_Toc169513074)

[1.2.3 SWOT-analysis 6](#_Toc169513075)

[1.2.4 EDA 7](#_Toc169513076)

[1.2.5 Answer to sub-question 2 8](#_Toc169513077)

[1.3 Sub-question 3 8](#_Toc169513078)

[1.3.1 expert interview 8](#_Toc169513079)

[1.3.2 available product analysis 8](#_Toc169513080)

[1.3.3 best practices 8](#_Toc169513081)

[1.3.4 EDA 9](#_Toc169513082)

[1.3.5 peer review 9](#_Toc169513083)

[1.3.6 co-reflection 9](#_Toc169513084)

[1.3.7 expo 10](#_Toc169513085)

[1.3.8 Answer to sub-question 3 10](#_Toc169513086)

[1.4 Sub-question 4 10](#_Toc169513087)

[1.4.1 Literature study 10](#_Toc169513088)

[1.4.2 Best practices 11](#_Toc169513089)

[1.4.3 Peer review 11](#_Toc169513090)

[1.4.4 Co-creation 12](#_Toc169513091)

[1.4.5 Answer to sub-question 4 12](#_Toc169513092)

[1.5 Sub-question 5 12](#_Toc169513093)

[1.5.1 Expert interview 12](#_Toc169513094)

[1.5.2. SWOT-analyse 13](#_Toc169513095)

[1.5.3 Answer to sub-question 5 13](#_Toc169513096)

[2. Conclusion 14](#_Toc169513097)

[Answer to the main research question 14](#_Toc169513098)

[3. Recommendation 15](#_Toc169513099)

[4. Bibliography 16](#_Toc169513100)

[5. Attachments 18](#_Toc169513101)

### Introduction

# Research questions, findings and answers

## Sub-question 1

What are the fundamental components and procedures characterizing the transfer and labeling of workloads between data centers regarding wind and solar energy generation?



### 1.1.1 Available product analysis (literature study)

The project group has looked at different comparable already available products during a literature study.

Findings of the components are:

* a resource scheduling and a flexible workload management to effectively integrate wind power (Wang et al., 2020).
* Potential of computational load shifting between data centers to improve renewable energy use (Rahmani et al., 2022).
* It’s necessary to have actually data from real time datacentre (Sheme et al., 2018).
* It’s important to have a datacentre in a wide variation in solar energy depending on the time of year, which naturally influences the amount of solar energy supplied (Sheme et al., 2018).
* the performance specification by servers as capacities, how fast the servers handle workloads is important (Song et al., 2022)
* The inefficient consumption of energy in datacenters can be mainly allocated to the following two components: energy loss and energy waste. Energy loss is the part of power lost during transport and conversion. Energy waste is the energy that is spent in the main task in data centers without any useful output. (Mastelic et al., 2014).
* CPU’s are a responsible for up to 60% of the total energy consumption of compute nodes (Dayarathna et al., 2016).
* A proposed model for multilabel classification of weather data using a deep learning model. With three layers: a single input layer that receives the weather variables, hidden layers where data is transformed and the output layer which computes in the final results of the model. (Doreswamy, 2018)
* Researchers from Rutgers University has developed Parasol and GreensSwitch, a software that running over real datacenters. The aim is reducing the total data centres coasts by properly managing workloads and available energy sources for maximum benefits.

### 1.1.2 Literature study

**Used method/process**

The project group has looked at different sources. Like CBS (Centraal Bureau voor Statistiek) and ResearchGate. At ResearchGate they have searched for some variables like green energy and renewable energy to find some important literature. Also google is used to find some background information about the project for understanding this project. Like information about datacentres, green renewable energy and workloads.

### 1.1.3 Expert interview

**Used method/process**

Every Monday we have a meeting with SUE, where we ask questions about the project. Every Thursday we have a content meeting with Joris van der Straten. We got feedback from both and used this for improvement.

**Findings**

The content coach gave us some advice for building the model in Visual Studio Code and to use Github, where we can show the model to SUE. SUE gave us some advice about the lay-out of the model and the content coach about helped figuring out which variables were helpful and what open data source to use. Firstly, two smaller first models were made in excel, after having this reviewed, the project group used visio code and Python to build the model. The content coach gave some help and feedback with building the model in Python.



### 1.1.4 Document analysis (bekijken documenten die je hebt gevonden)

**Used method/process**

We have read different documents and made a summary of the best literature.

The project group has looked at different comparable already available products during a literature study.

Findings of the components are:

* That it's important to have a resource scheduling and a flexible workload management to effectively integrate wind power (Wang et al., 2020).
* Solar power is generated by catching solar energy on a solar panel, the transformer converts this direct current into an alternating current, that is usable. (Zonneplan B.V., 2024)
* Wind energy is generated by having a wind turbine stand in the wind. The wind hitting the blades makes the turbine spin. This spinning generates low voltage power, a transformer transforms this to high voltage to be used on the power net. (Windcentrale, z.d.)
* Potential of computational load shifting between data centers to improve renewable energy use (Rahmani et al., 2022).
* It’s necessary to have actual data from real time datacentres (Sheme et al., 2018).
* It’s important to note that datacentres have a wide variation in solar energy depending on the time of year, which naturally influences the amount of solar energy supplied (Sheme et al., 2018).
* the performance specification by servers as capacities, how fast the servers handle workloads, is important (Song et al., 2022)
* The inefficient consumption of energy in datacenters can be mainly allocated to the following two components: energy loss and energy waste. Energy loss is the part of power lost during transport and conversion. Energy waste is the energy that is spent for the main tasks in data centers without any useful output. (Mastelic et al., 2014).
* CPUs are responsible for up to 60% of the total energy consumption of compute nodes (Dayarathna et al., 2016).
* A proposed model for multilabel classification of weather data using a deep learning model. With three layers: a single input layer that receives the weather variables, like, hidden layers where data is transformed and the output layer which computes in the final results of the model. (Doreswamy et all., 2018)
* Researchers from Rutgers University have developed Parasol and GreensSwitch, a software that they have running in real datacentres. Their aim is reducing the total data centres costs by properly managing workloads and available energy sources for maximum benefits.

### 1.1.5 Observation:

**Used method/process**

We went with our project group and content coach to SUE. We got an idea of their work environment, and we used some feedback and reflection method that they used.

**Findings**

SUE is an open-minded organization that likes to cooperate with other organizations and professionals. By using their methodology, we were able to better understand and adapt our working methods.

### 1.1.6 Answer to sub-question 1

What are the fundamental components and procedures characterizing the transfer and labeling of workloads between data centers regarding wind and solar energy generation?

The transfer and labeling of workloads between data centers regarding wind and solar energy generation involve several fundamental components and procedures:

It is crucial to have resource scheduling and flexible workload management for effectively integrating wind power and leveraging computational load shifting to enhance renewable energy use. Real-time data from data centers is necessary, and the location of a data center significantly impacts solar energy variation, depending on the time of year. Performance specifications of servers, including how fast they handle workloads, are also essential. Addressing energy loss and waste is critical, as energy loss occurs during transport and conversion, while energy waste happens during tasks without useful output. CPUs are major energy consumers in data centers, responsible for up to 60% of the total energy consumption. Additionally, proposed models for multilabel classification of weather data using deep learning and software like Parasol and GreenSwitch for managing workloads and energy sources demonstrate innovative approaches to optimizing renewable energy use. Expert feedback emphasized the importance of proper model building and variable selection, while observations of collaborative work environments contributed to refining the project’s approach.

Overall, these components and procedures ensure effective management and optimization of workloads based on renewable energy generation.

## Sub-question 2

What specific criteria and parameters define the ideal situation for optimizing the labelling process between data centres, considering the open weather data, energy source consumption, and energetic penalties.



### 1.2.1 literature study

The project group has looked at different comparable already available products during a literature study.

Findings of the components are:

* That it's important to have a resource scheduling and a flexible workload management to effectively integrate wind power (Wang et al., 2020).
* Solar power is generated by catching solar energy on a solar panel, the transformer converts this direct current into an alternating current, that is usable. (Zonneplan B.V., 2024)
* Wind energy is generated by having a wind turbine stand in the wind. The wind hitting the blades makes the turbine spin. This spinning generates low voltage power, a transformer transforms this to high voltage to be used on the power net. (Windecentrale.nl, z.d.)
* Potential of computational load shifting between data centers to improve renewable energy use (Rahmani et al., 2022).
* It’s necessary to have actual data from real time datacentres (Sheme et al., 2018).
* It’s important to note that datacentres have a wide variation in solar energy depending on the time of year, which naturally influences the amount of solar energy supplied (Sheme et al., 2018).
* the performance specification by servers as capacities, how fast the servers handle workloads, is important (Song et al., 2022)
* We have also found some images describing the growth of (green) energy usage by datacenters. We also found an image describing some percentual energy losses when using solar panels. (Attachment 1 different variants of energy consumption, attachment 2 data centre energy consumption, attachment 3 DC generation losses).
* The inefficient consumption of energy in datacenters can be mainly allocated to the following two components: energy loss and energy waste. Energy loss is the part of power lost during transport and conversion. Energy waste is the energy that is spent for the main task in data centers without any useful output. (Mastelic et al., 2014).
* CPUs are responsible for up to 60% of the total energy consumption of compute nodes (Dayarathna et al., 2016).

1.2.2 expert interview

In a series of meetings with our client SUE and coach Joris, several key criteria and parameters were identified to optimize the labelling process between data centres, considering open weather data, energy source consumption, and energetic penalties. The discussions highlighted the need for clarity and precision in the dashboard visuals, including determining whether tables or graphs are preferred, verifying the integrity of 0 values in the data, and creating a help button for user guidance. The stakeholders emphasized incorporating energy penalties and city distances into the dashboard, ensuring meaningful data presentation from the outset, and ordering city selections by the benefit of transfer. Additionally, feedback was provided to refine the project's scope, objectives, and research questions, with an emphasis on the specificity of models and labelling systems.

Joris also stressed the importance of a structured development approach, including committing code to GitHub, maintaining clean and organized code files, and validating the dashboard through rigorous testing. A focus on practical implementation was advised, starting with a simple model in Excel before expanding it, using GitHub for collaboration, and ensuring thorough documentation for potential future developers. The meetings underscored the necessity of aligning the project with SUE's operational environment, leveraging open data sources, and continuously improving the model based on real-time data and feedback.

Key takeaways included the need for a comprehensive and adaptable solution, the validation and integration of energy consumption data, and the importance of stakeholder engagement throughout the project to ensure that the final product meets organizational needs and can be seamlessly integrated into SUE's workflow.

### 1.2.3 SWOT-analysis

|  |  |  |
| --- | --- | --- |
|  | **Helpful** | **Harmful** |
| **In ternal** | **Strengths**  -The dashboard has a high adaptability  -The weather data is live  -The dashboard has been validated  -The dashboard was made specifically for SUE  -The dashboard can easily be taken over by another group or SUE  -The dashboard is backed up with research and documentation | **Weaknesses**  -The formula is not specific enough  -The variables are not specific enough  -The energetic penalty is not specific enough  -The amount of green energy generated is not specific enough  -Some variables and data about the locations is missing  -The dashboard is not adapted to SUE's environment yet  -There might be missing variables due to missing knowledge/insight |
| **external** | **Opportunities**  -Specifying the formula further  -Specifying the location data  -Specifying energetic penalty further  -Implementation into SUE's environment  -Collaborate with comparable solutions | **Threats**  -Different other solutions are being created (as seen in research)  -Most of the research is now pushed because of regulations, might these be lessened, the progress might slow or come to a halt. |



### 1.2.4 EDA

To optimize the current labelling process, we looked at the different API’s available to get weather data. We wanted at least two different variables to use in our labelling model namely a variable for the amount of sunshine and some sort of wind speed measurement. These two were chosen because by looking at the green energy sources we decided to first look at the most used green energy sources which are solar and wind. This meant that we wanted variables that would influence the energy generated by solar panels as well as wind turbines and the most common indicators for this are the amount of sunshine per day for solar panels and the average wind speed per day for the wind turbines.

After some research we found an easy-to-use API which also had the variables we were looking for regarding the solar panels and the wind turbines. (Open-meteo, 2024)

For the future the ideal situation would be to further build on the number of variables used in the labeling process and make it more complex. This would mean that the accuracy of the labeling model goes up and the outcome of the labeling process would be optimized to the best of its ability.

(Attachment 4 API requested data)

### 1.2.5 Answer to sub-question 2

What specific criteria and parameters define the ideal situation for optimizing the labelling process between data centres, considering the open weather data, energy source consumption, and energetic penalties.

The ideal situation for optimizing the labelling process between data centers involves a comprehensive approach that integrates real-time weather data, efficient resource scheduling, and workload management. It also requires continuous stakeholder engagement, adaptability to various environments, and the inclusion of detailed location-specific data. By addressing energy loss and waste, defining energetic penalties, and continuously improving the model, data centers can enhance their sustainability and operational efficiency with regards to the labelling process.

## Sub-question 3

What are the key features of an effective solution aimed at optimizing the labelling system between the data centers with a focus on green energy utilization/generation and the energetic penalty?



### 1.3.1 expert interview

**Used method/process**

The project group had several meetings with SUE to further discuss the key features. These meetings were held throughout the process of making the solution for SUE.

**Findings**

From the meetings the group found several key features. The first features discussed were: gathering data from a weather database using an API, adding a formula into the model for green energy generation and energetic penalty, making a dashboard using Python and adding information about the different locations.

1.3.2 available product analysis

**Used method/process**

Through literature studies, the project group has looked at several comparable solutions. Some of these were summarized and/or added to the findings below for insight into these methodologies or findings from said literature.

**Findings**

* The project group found multiple comparable solutions that already had a system or model in place that included labelling weather data. These are the solutions in short: a deep learning model for multi label classification of weather data. (Doreswamy et all., 2018)
* Research about what algorithm to use for an AI solution for energy efficiency in datacentres (Venkataswamy et all., 2023)
* A model used for increasing CPU resource utilization on active nodes to increase energy efficiency in datacentres. (Venkataswamy et all., 2023)
* Research exploring the potential of utilizing weather forecast data for optimizing energy consumption in data centers. (Verma 2023)
* A deep reinforcement learning job scheduler that uses renewable energy generation (Venkataswamy et all., 2023)
* A deep learning model that predicts the distribution of green energy and CPU workload in different locations and assigns workload accordingly to reduce brown energy usage (Gao et all., 2020)
* A report about a model to maximize datacenter hardware economic life to lower emissions and reduce cost. (Fenn 2021)

### 1.3.3 best practices

Common ways of validating a dashboard are to recreate it with the exact same dataset or snapshot of time when working with real time data. A snapshot contains the information for a certain time and date. For our Greenlabel dashboard we created a snapshot of a certain time and date of the dashboard. This makes us able to validate the dashboard in another environment. Most of the time when a dashboard is created in PowerBI for example, the validation needs to be done in another environment like tableau or Python with the exact same data either real time or static.

Our Greenlabel dashboard is made in Python which means we can validate it in a program like PowerBI based on a snapshot we explained before. To make the validation more unbiased we asked a part-time project member to do this process as he is less familiar with our code and dashboard. The reason for this is to let him try and make the exact same visuals and measures/calculations with the same data and see if the answers are the same as the initial dashboard. If the values are one and the same this means that the calculations are being performed accordingly, if not then there could be a problem with the initial code/dashboard.

We chose to make a snapshot to validate our dashboard, instead of connecting to the same API. This choice was made with certain things in mind like the connection that the API makes with the Python script and PowerBI. This API connection could possibly create a problem that would otherwise not be there, the other reason is to make the validation process easier, quicker and more reliable with the use of a snapshot.



### 1.3.4 EDA

**Used method/process**

For the energetic penalty we have done some research into what this could encompass. The energetic penalty we made consists of two parts. One part is the energy needed to transfer a certain amount of data (GB). The other part is the distance the data needs to travel between locations and the energy consumption this takes into account. These two parts put together form the energetic penalty we have in use now in the dashboard.

**Findings**

To optimize the energetic penalty, it is important to add more variables as well. Added variables could be the difference between locations regarding the hardware they use and the effectiveness of that hardware as well as the specific length of wire between the locations instead of a geographical line which is straight and does not represent reality. These are a few examples on how to make the energetic penalty more accurate/complex to add to the effectiveness of the whole labelling system.



### 1.3.5 peer review

**Used method/process**

The project group gave and got feedback internally from SUE, coaches and other group members.

**Findings**

From the meetings the group found several key features. The first features discussed were: gathering data from a weather database using an API, adding a formula into the model for green energy generation and energetic penalty, making a dashboard using Python and adding information about the different locations.

### 1.3.6 co-reflection

**Used method/process**

The project group did several meetings with SUE to show the progress made, discuss next steps and gain feedback.

**Findings**

By doing these meetings as mentioned above the group was able to create a goal, scope, research question, the model and the documentation around these subjects.

### 1.3.7 expo

**Used method/process**

The project group first made two different models in excel to get feedback on the features from SUE and other learning partners. When the model features were approved the group made the green label model in Python and presented this throughout the process.

**Findings**

By doing these presentations the model has become what it is today. All the feedback has been either written out in notes or immediately been implemented.

### 1.3.8 Answer to sub-question 3

What are the key features of an effective solution aimed at optimizing the labelling system between the data centers with a focus on green energy utilization/generation and the energetic penalty?

Key features include gathering data from a weather database using an API, incorporating a formula for green energy generation and energetic penalty, creating a dashboard in Python and including information about different locations. Various comparable solutions were identified, such as deep learning models for multi-label classification of weather data, AI algorithms for energy efficiency, CPU resource utilization models, and deep reinforcement learning job schedulers that use renewable energy generation. Best practices for validating a dashboard involved recreating it with the same dataset or a snapshot in a different environment, such as PowerBI, to ensure the calculations and visualizations align, highlighting the importance of unbiased validation. The energetic penalty was defined through research, encompassing the energy required to transfer data and the distance between locations, with suggestions to enhance accuracy by considering hardware differences and realistic wiring distances. Feedback from learning partners and SUE helped refine features, leading to the establishment of goals, the scope of the project, and the development of the model. Presentations and expos throughout the process allowed for continuous improvement based on feedback, ultimately shaping the Greenlabel model and dashboard into its current form.

## Sub-question 4

How can this labelling system be validated and tested to verify that the labels are assigned as intended by the labelling system?



### 1.4.1 Literature study

To validate data or a dataset there are a few common testing techniques. These include:

* **Range Checking:** Verifies that values fall within a specific range. For instance, if a field is meant to store an age, it ensures the value is between 0 and 100.
* **Type Checking:** Ensures that the data type is correct. For example, if a field should contain a date, this check confirms the value is indeed a date rather than a string or number.
* **Format Checking:** Confirms that the data adheres to a specific format. For instance, if a field is meant for an email address, it verifies that the value matches the standard email format.
* **Consistency Checking:** Ensures data consistency across fields or records. For example, if one field contains a country and another field contains a city, it verifies that the city is located in the specified country.
* **Uniqueness Checking:** Ensures that values are unique within a field. For example, if a field is supposed to contain unique user IDs, it verifies that no duplicate IDs exist.
* **Existence Checking:** Verifies that required fields are not null. For example, if a field must contain a value, it ensures that the value is not missing.
* **Referential Integrity Checking:** Ensures that data values reference existing values in related tables. For instance, if a field contains a foreign key, it verifies that this key exists in the referenced table.

These techniques check if the data is clean and has a certain degree of quality to it. If some of these checks contain errors, they can be corrected by a script or adjusting the data source.

using this method, we found that there are various techniques that can help us validate the dashboard and the data/code for our project. These include checks on tidiness and consistency for example.

### 1.4.2 Best practices

Common ways of validating a dashboard are to recreate it with the exact same dataset or snapshot of time when working with real time data. A snapshot contains the information for a certain time and date. For our Greenlabel dashboard we created a snapshot of a certain time and date of the dashboard. This makes us able to validate the dashboard in another environment. Most of the time when a dashboard is created in PowerBI for example the validation needs to be done in another environment like tableau or Python with the exact same data either real time or static.

Our Greenlabel dashboard is made in Python which means we can validate it in a program like PowerBI based on a snapshot we explained before. To make the validation more unbiased we asked a part-time project member to do this process as he is less familiar with our code and dashboard. The reason for this is to let him try and make the exact same visuals and measures/calculations with the same data and see if the answers are the same as the initial dashboard. If the values are one and the same this means that the calculations are being performed accordingly, if not then there could be a problem with the initial code/dashboard.

We choose to make a snapshot to validate our dashboard, instead of connecting to the same API. This choice was made with certain things in mind like the connection that the API makes with the Python script and powerBI. This API connection could possibly create a problem that would otherwise not be there, and the other reason is to make the validation process easier, quicker and reliable with the use of a snapshot.

**Findings:**

In this method common ways of validating a dashboard with either a snapshot or a real time data connection was found. It was also found that validating in another coding environment with the same data can help prove the validity of the data and thus the dashboard itself, that is if the values align.



### 1.4.3 Peer review

The dashboard validation has been reviewed by our project members as well as our content coach, who is a professional in this field. During the review they looked at the validation code and if it represented the same steps in both coding environments. The dashboard that we made was created in VisualStudioCode with Python and various libraries for visualizations and lay-out changes. The validation was meant to find out if the calculations and transformations that were done in Python are correct. To validate this, we made the same calculations and transformations in power bi with DAX and M code. The outcome of the PowerBI calculations and transformations were the exact same as the calculations and transformations in Python which means that the dashboard we made shows correct information according to the calculations we made.

**Findings:**

The validation process is more efficiently done with other members of the group and outside influences. Having the validation process checked by others also helps prove the validity of the initial validation process itself.



### 1.4.4 Co-creation

The validation process was done by collaboration with various users including our project members, our teachers and our part time group members. The part-time group members helped us make the first iteration of the validation in which we came to the conclusion that the API call is not accurate, because the data from the Python script did not align with the data loaded in PowerBI. After investigating this misaligned between the data we saw that the API call was requesting data from a column in a table that did not exist, but generated data anyway without generating an error.

**Findings:**

The collaboration helped us find some quarks in our Python script as well as in the validation process.

### 1.4.5 Answer to sub-question 4

How can this labelling system be validated and tested to verify that the labels are assigned as intended by the labelling system?

The labeling system can be validated and tested through a structured approach involving multiple techniques: Implement common data validation techniques such as range, type, format, consistency, uniqueness, existence, and referential integrity checks to ensure data quality. Best practices in dashboard validation include recreating the dashboard with the same dataset or a snapshot, validating in a different environment (e.g., PowerBI) to ensure calculations and visualizations match, and using snapshots to avoid API-related inconsistencies and ensure a stable dataset.

Conducting internal reviews with project members and experts to verify validation steps and comparing results across different tools (Python and PowerBI) to confirm accuracy. The collaboration involves engaging various stakeholders in the validation process and addressing issues identified during collaborative validation efforts to ensure a robust system. By combining these methods, the labeling system can be thoroughly validated to ensure that labels are assigned as intended and the data is valid, thereby enhancing the reliability and accuracy of the product.

## Sub-question 5

1. What steps are necessary to implement the proof of concept within the framework of SUE's operational environment?



### 1.5.1 Expert interview

In meetings with SUE, several steps were identified as necessary to implement the proof of concept within the framework of SUE's operational environment. Key points of discussion focused on the dashboard's adaptability, live weather data integration, and thorough validation. The dashboard, designed specifically for SUE, needs to be easily transferable to other teams or SUE personnel, supported by comprehensive research and documentation.

Several weaknesses were identified, including the need for more specific formulas, variables, and details about the energetic penalty and green energy generation. Stakeholders noted missing data about locations and the need to adapt the dashboard to SUE's specific environment. Addressing these gaps involves specifying formulas further, enhancing location data and refining the energetic penalty.

Opportunities for improvement include collaborating with comparable solutions, specifying the formulas, location data, and energetic penalties in greater detail. The group should also focus on using energy penalties in dashboard selections and graphs, incorporating city distance data, and ensuring the dashboard shows valid information from the start. Cleaning up and organizing code files and creating Data Flow Diagrams for better understanding and maintenance were also recommended.

Overall, the meetings underscored the importance of a structured and methodical approach, focusing on thorough documentation, clean code, specific and detailed formulas, and ensuring the solution is able to be well-integrated into SUE's operational environment for effective implementation and future scalability.

### 1.5.2. SWOT-analyse

|  |  |  |
| --- | --- | --- |
|  | **Helpful** | **Harmful** |
| **In ternal** | **Strengths**  -The dashboard has a high adaptability  -The weather data is live  -The dashboard has been validated  -The dashboard was made specifically for SUE  -The dashboard can easily be taken over by another group or SUE  -The dashboard is backed up with research and documentation | **Weaknesses**  -The formula is not specific enough  -The variables are not specific enough  -The energetic penalty is not specific enough  -The amount of green energy generated is not specific enough  -Some variables and data about the locations is missing  -The dashboard is not adapted to SUE's environment yet  -There might be missing variables due to missing knowledge/insight |
| **external** | **Opportunities**  -Specifying the formula further  -Specifying the location data  -Specifying energetic penalty further  -Implementation into SUE's environment  -Collaborate with comparable solutions | **Threats**  -Different other solutions are being created (as seen in research)  -Most of the research is now pushed because of regulations, might these be lessened, the progress might slow or come to a halt. |

### 1.5.3 Answer to sub-question 5

To successfully implement a system at SUE, several weaknesses need to be addressed. The current formula and variables lack specificity, impacting the precision of calculations. Details regarding the energetic penalty and the amount of green energy generated are insufficiently defined. Additionally, there is missing location-specific data critical for accurate system design. The dashboard is not yet adapted to SUE's environment, posing a significant integration barrier. Furthermore, there might be missing variables due to gaps in knowledge or insight. Addressing these issues through detailed research and refinement will enhance the system's accuracy, reliability, and seamless integration at SUE.

# Conclusion

### Answer to the main research question

How can a statistical model be made for labeling data centers regarding the weather data, energy source consumption and regarding the energetic penalty?

An effective solution would collect data from weather databases via APIs, use formulas for green energy generation and energetic penalties, and provide a dashboard with location-specific information. Incorporating comparable solutions, such as deep learning models and AI algorithms for energy efficiency could possibly be beneficial.

Best practices involve validating the dashboard by recreating it in different environments and ensuring unbiased and accurate results. Validation techniques should include common data checks (range, type, format, consistency, uniqueness, existence, and referential integrity) and internal reviews.

Engaging stakeholders and collaboratively addressing issues ensures a robust system. By integrating real-time data, detailed specifications, and continuous validation, the model will effectively optimize energy use and reduce energetic penalties in data centers, ensuring reliability and efficiency.

# Recommendation

Firstly, looking at the weaknesses and opportunities from the SWOT-analysis, there are several recommendations to make. First of all, the formula, variables, location specific information and green energy generated should be further specified and researched more in depth. By doing this the calculations will be more accurate, more trustworthy and have more accurate labels. Secondly by going more in depth, the next project group might find variables or research that was out of scope or simply missed by the current project group. Going further into the SWOT-analysis, the implementation at SUE is a valid opportunity that will cancel out another weakness of the system, as it is right now, the dashboard cannot be fully integrated by SUE, which is a (great) weakness in this specific scenario. Thirdly, the fact that multiple comparable solutions or projects are being made/have been made, is a threat for this specific solution’s survivability and viability, however this can possibly be turned into a big opportunity. By collaborating or sharing information with the makers/stakeholders/owners of these solutions, more insights can be gained, multiple solutions or parts of solutions can be combined, achieving a (near) optimal solution. The research that has already been done can contribute to making this happen and could also be used to get some ideas of what to implement or add into or besides the solution for SUE. Something that cannot be influenced by SUE or the next project group, but should still be kept in mind, are regulations. The comparable solutions are mainly pushed by government regulations requiring datacenters to use more green energy i.e. using less brown energy. If these regulations get changed or taken away entirely, this might have big implications for this or other solutions. The same goes for the cost of different energy sources, windmills and solar panels, but great changes in the balance between these prices are not expected in the coming years.

# Bibliography

Centraal Bureau voor de Statistiek. (2021, 23 februari). Productie groene stroom met 40 procent gestegen. Centraal Bureau Voor de Statistiek. <https://www.cbs.nl/nl-nl/nieuws/2021/08/productie-groene-stroom-met-40-procent-gestegen>

Dayarathna, M., Wen, Y., Fan, R.: Data center energy consumption modeling: a survey. IEEE Commun. Surv. Tutorials 18(1), 732–794 (2016)

Dayarathna, M., Wen, Y., Fan, R.: Data center energy consumption modeling: a survey. IEEE Commun. Surv. Tutorials 18(1), 732–794 (2016)

Dutch Data Center Association. (2024, 1 mei). Duurzaamheid & Energie - Dutch Data Center Association. <https://www.dutchdatacenters.nl/duurzaamheid-energie/>

Free Weather API. (z.d.). Open Meteo. <https://open-meteo.com/>

Jaggi, P. (2018, 3 augustus). How to Calculate Solar Energy Generation. <https://www.linkedin.com/pulse/how-calculate-solar-energy-generation-puneet-jaggi/>

Hoe werkt een windmolen? (z.d.). Windcentrale. <https://www.windcentrale.nl/blog/hoe-werkt-een-windmolen/>

Hoe werken zonnepanelen? Van zon naar stroom - Zonneplan. (z.d.-a). Zonneplan. <https://www.zonneplan.nl/kenniscentrum/zonnepanelen/werking>

How can you determine the energy output of wind turbines? (z.d.). Linkedin. <https://www.linkedin.com/advice/1/how-can-you-determine-energy-output-wind-turbines>

Mastelic, T., Oleksiak, A., Claussen, H., Brandic, I., Pierson, J.M., Vasillakos, A.V.: Cloud computing: survey on energy efficiency. ACM Comput. Surv. 47(2) (2014). <https://doi.org/10.1145/2656204>

Milieu Centraal. (z.d.). Hoe werken zonnepanelen. <https://www.milieucentraal.nl/energie-besparen/zonnepanelen/hoe-werken-zonnepanelen/>

Rahmani, R., Moser, I., & Cricenti, A. (2022). Inter-continental Data Centre Power load balancing for renewable energy maximisation. Electronics, 11(10), 1564. https://

doi.org/10.3390/electronics11101564

Sheme, E., Holmbacka, S., Lafond, S., Lučanin, D., & Frashëri, N. (2018). Feasibility of using renewable energy to supply data centers in 60°north latitude. Sustainable Computing, 17, 96–106. https://doi.org/10.1016/j.suscom.2017.10.017

Song, J., Zhu, P., Zhang, Y., & Yu, G. (2022). Versatility or validity: A comprehensive review on simulation of Datacenters powered by Renewable Energy mix. Future Generation Computer Systems, 136, 326–341. https://doi.org/10.1016/j.future.2022.06.008

Venkataswamy, V., Grigsby, J., Grimshaw, A. S., & Qi, Y. (2023). RARE: Renewable Energy aware resource management in datacenters. In Lecture notes in computer science (pp. 108–130). <https://doi.org/10.1007/978-3-031-22698-4_6>

Wang, P., Cao, Y., & Ding, Z. (2020). Flexible Multi-Energy Scheduling Scheme for data center to facilitate wind power integration. IEEE Access, 8, 88876–88891. <https://doi.org/10.1109/access.2020.2990454>

Doreswamy, N., Gad, I., & Manjunatha, B. R. (2018). Multi-label Classification of Big NCDC Weather Data Using Deep Learning Model. In Communications in computer and information science (pp. 232–241). <https://doi.org/10.1007/978-981-13-1936-5_26>

Radu, L. (2017). Green Cloud Computing: A literature survey. Symmetry, 9(12), 295. <https://doi.org/10.3390/sym9120295> p58-p145

itrenew the case for circularity data center impact report. (z.d.). opencompute.org. Geraadpleegd op 24 mei 2024, van <https://www.opencompute.org/documents/itrenew-the-case-for-circularity-data-center-impact-report-ocp-v11-pdf>

Gao, J., Wang, H., & Shen, H. (2020). Smartly Handling Renewable Energy Instability in Supporting A Cloud Datacenter. International Parallel And Distributed Processing Symposium (IPDPS). <https://doi.org/10.1109/ipdps47924.2020.00084>

VERMA, S. (2023). Optimizing Data Center Energy Efficiency: Insights from Weather Pattern Analysis. Indian Institute of Information And Technology.

# Attachments

Attachment 1 different variants of energy consumption

Afbeelding met tekst, schermopname, Lettertype, diagram

Automatisch gegenereerde beschrijving

Attachment 2 data centre energy consumption

Afbeelding met tekst, schermopname, Perceel, lijn

Automatisch gegenereerde beschrijving

Attachment 3 DC generation losses

Afbeelding met tekst, schermopname, Lettertype

Automatisch gegenereerde beschrijving

Attachment 4 API requested data

A screenshot of a computer

Description automatically generated