Technical Specification Double-side Page

1. **TECHNICAL SCOPE:** Summarize the mock-up devised during the EXPLORE phase: how have you addressed the challenge/Theme Challenges and tackled with its requirements and data. Include a diagram

In our challenge, there were a total of 4 requirements, of which at least one had to be fulfilled. The four requirements were (1) Forecast of total heating consumption, (2) Heating demand for individual points in the distribution network and the dynamic plan for heating production, (3) Increased customer satisfaction concerning ambient conditions and public venues, (4) Input data for strategic planning of the heating network and its redesign. The biggest challenge in this task was the tiny database we had available. The data for the heat meter was recorded at 20-minute intervals in the period 16.11.2021 00:00 to 16.11.2021 16:40. The data for the pressure sensor in a 3-minute interval in the period 16.11.2021 00:00 to 16.11.2021 18:54 and finally the data for the local heating network in an interval of 20 minutes in the period 16.11.2021 00:00 to 16.11.2021 20:00. Further data such as upcoming events refer to the year 2022 and therefore cannot be logically linked to the existing data set. A total amount of energy of 0.071 was consumed at the heat meter in the above period. Due to the lack of data documentation, it is not clear what measured value (kWh, GJ, MJ, m^2) this value represents. Therefore, the assumption is made that this value represents m^2 to enable a price calculation with the respective tariffs. The homepage of the weather data was only available in Latvian and posed a hurdle, but the temperature data for the station "Riga University" was downloaded and inserted anyway. For the GIS data we needed online access, which we didn't received yet.

Due to these difficult conditions and the expected requirements, we decided to use existing frameworks in order to achieve maximum success with the existing data and at the same time show what a future project could look like.

For a fair implementation of requirement (1), the database was too small to deliver a high forecast quality. Nevertheless, we used a simple form of time series forecasting based on the heat meter data to show how the heat consumption develops after the end of the time series (+5 hours). In addition, an upper and lower bound is output to show where the upper and lower limits lie. Further we visualized the different temperature comfort levels.

(2) Since we only have one single measuring point, it was not possible to forecast the heat consumption for several end points.

(3) In order to create more transparency and customer satisfaction for the heat consumer, it is beneficial to provide the customer with insight into their heat consumption. In this mock-up, we used the available data to create the key figures and diagrams for the current heat consumption as well as the price to be paid and the price of the tariff.

(4) For the strategic planning and deployment of the plants, it is advantageous to know the internal key figures and to define KPIs with which to compare the heating networks and evaluate whether an improvement could be achieved.

2. **ALGORITHMS, TOOLS AND CONCLUSIONS:** Detail the algorithms and tools identified to accomplish the challenge/Theme Challenges. Show clear understanding of the used REACH dataset/s and addressed challenge/Theme Challenges.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951981

A cloud provider was chosen as the framework. The data provided was an Excel file with several spreadsheets. These were first split into individual CSV files before being uploaded into a container specific to this use case. Cloud provider service was used to create the database schema and finally another service was used to check whether the data was correctly recognized with an SQL query. Due to the small database, it was not possible to perform machine learning in a Jupyter Notebook instance. However, we still wanted to meet the requirements of the challenge and decided to use a basic machine learning model. The forecasting of the time series is only a basic model and only shows the possibilities of where the journey can go.

In the future, we hope to receive much more data in order to be able to fulfil all requirements with a perfected machine learning model and a high forecasting quality for the heat demand. Data understanding and the inclusion of challenges is described in detail in point 1 - "Technical Scope".

3. **SCALABILITY AND FLEXIBILITY OF THE SOLUTION:** Discuss whether the solution can truly cope with humongous and increasing datasets and how flexible it is to adapt to other related domains

Due to the cloud-based solution approach and the characteristics of a cloud, scalability and multiplicability are fully guaranteed. Any number of datasets can be stored, analyzed, cleaned and prepared for the ML pipeline. All the steps can also be automated, such as data collection with a REST API or the automated training of ML models every hour, every day or every week.

4. **DATA GOVERNANCE AND LEGAL COMPLIANCE:** Describe the security level of the proposed solution, i.e., how authentication, authorisation policies, encryption or other approaches are used to keep data secure. Explain how will be compliant with the current data legislations concerning security and privacy (e.g. GDPR).

The cloud providers have very high data security standards which are provided to the user. Starting with the role and authorization principle of the services up to the server-side encryption of the received data and the monitoring of all cloud activities. For example, it is not possible to execute services or access the database if no authorization role has been assigned for the new database beforehand. The databases also have versioning and a read-only setting. The security can thus be significantly increased by the administrator with the necessary knowledge and can also be logged in order to comply with the GDPR guidelines and even to be able to present them. In Germany in particular, meter data from the electricity, water and heat sectors is critical infrastructure and therefore requires the highest level of protection. For example, the data may only be hosted on servers in Germany and the Federal Network Agency may ask for security or historical data, which must then be provided immediately. Non-compliance can result in high contractual penalties.

5. **QUALITY ASSURANCE AND RISK MANAGEMENT:** Describe the quality process planned for the final product. Technologically, which are the potential risks in all the phases of the project (design of the solution, development, testing, deployment…) and indicate mitigation plans to still fulfil the challenge/Theme Challenges and data provider requirements.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951981

In order to develop a final product, a workshop with the data provider is essential. If the final product is also to be deployed immediately as a test phase, it is advisable to select a heating network operator as the data provider. During the workshop, it must be checked how the data is currently collected, where it is stored and what the current IT infrastructure looks like. After the fundamental services have been implemented, a number of added-value services can be offered and implemented. The potential risks lie in the lack of a data basis, no interface for data communication and the unwillingness of data providers to use a cloud solution. However, all the risks can be clarified in a workshop and solutions can be defined. The biggest risk, however, is when no data is yet received via heat meters and no data basis is available. In this case, the final product can be largely implemented through test data or a larger batch, but it would have no relation to reality and would only become a project instead of a final product.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951981