Valtteri Airaksinen

Lessons learned

KAMK • University
of Applied Sciences

Table of Contents

1   Introduction

This document summarizes key lessons from developing a system to fetch, preprocess, and store weather data from the Finnish Meteorological Institute (FMI) for Valio's machine learning pipelines. Insights focus mainly on technical challenges.

2 Key Technical Lessons

2.1 API & Data collection

Understanding the FMI API and configuring access for the kriging models required a significant amount of effort. While some parts of the documentation were useful, much of it provided little guidance for accessing the customer-specific data I needed. Correctly using parameters for different models was particularly challenging, as there was no clear guidance on which parameters or coordinate grids corresponded to each model. During this process, directly contacting FMI proved highly effective, and with their support, I was able to obtain the necessary data. This experience highlighted the importance of combining independent exploration with proactive communication when working with complex or partially documented APIs.

2.2 Data preprocessing

Calculating features such as minimum, maximum, mean, and vapor pressure using Tetens formula was essential to ensure that the collected data matched Valio's historical dataset. Implementing these calculations as modular functions proved particularly valuable, as it simplified future expansions and ensured consistency across the dataset. This was a pretty straightforward process, but I found learning about Tetens formula and implementating it into a python function particularly intriguing.

2.3 Architecture & automation

The use of Azure Function Apps allowed for the automation of daily data collection as well as frequent updates of the effective temperature sum. This automation significantly improved the reliability and reproducibility of the data pipeline, ensuring that data was consistently available for downstream machine learning processes.

I personally find azure applications such as function apps quite complicated, and it took quite a lot of time to gather enough understanding to create these applications, I have previously created multiple web apps on azure and on function app which helped me tackle the issues I ran into.

3   General observations

On a general level, this project reinforced the importance of conducting thorough independent research and seeking help when encountering challenges. It also highlighted how complex I find Azure applications, emphasizing the need to further develop my skills in that area. Additionally, working with data from multiple sources and combining it for calculations—such as vapor pressure and future temperature sums—was both enjoyable and intellectually engaging, providing valuable hands-on experience in integrating and processing diverse datasets.

4 Conclusion

This project provided valuable insights into building a reliable data pipeline for weather data. It highlighted the importance of thoroughly understanding APIs, creating modular feature calculations, and automating processes with Azure Function Apps.

The experience reinforced the value of independent research, seeking help when needed, and integrating data from multiple sources. Overall, the project was both challenging and rewarding, providing a strong foundation for future work in data collection, preprocessing, using azure tools and machine learning.