Valtteri Airaksinen

Project plan

TTV20SAI

Customer Project 1

Fall 2025

KAMK • University of Applied Sciences

Table of Contents

1 Project Introduction

The goal of this project is to collect data from the FMI interface and prepare it for integration into a neural network pipeline, which I am developing as part of my final thesis. The purpose is to create clear documentation and flexible functions that can be used to get data easily from FMI database.

2 Project plan

The basic idea for this project emerged when I first started planning the integration of weather data into a neural network. Initially, I thought obtaining the data would be a simple task, but I quickly realized that understanding the documentation would take more time than expected.

Therefore, I will be solving these issues in this project and at the same time I can fulfill the course requirements and receive credit for this course.

2.1 The objectives of the project

In this project, I will be creating different functions that fetch data from FMI -database and save this data to azure blob storage. I will also be doing some calculations myself, since not all features are directly available from FMI -database. Lastly, I will be using observed and forecasted data in the prediction models' pipeline to calculate changes in effective temperature sum and then placing it in the data given to the prediction model.

2.1.1 Collecting data from FMI -database

First step is writing flexible functions to fetch data from FMI -database and upload it in azure blob storage. The plan is to create an azure function app that runs once a day first collecting all necessary data and shaping it into the desired format and then storing this data in an azure blob storage so it can be later used to train better models or something else.

2.1.2 Calculating missing features

The base data that was used to train the current prediction model has features that are not directly available from FMI -database and need to be calculated. This includes various min, max and mean values and vapour pressure.

2.1.3 Calculating and implementing effective temperature sum

Since the model will be predicting varying number of days to the future depending on how quickly the sample data is ready and sent to the model. Therefore, when calculating what the effective temperature sum is + 3 and + 6 days from the sample being taken from the field, we can use observation data to get the most accurate possible starting point and then use the forecasts to calculate from there on out. These calculated values will then be added to the data given to the prediction model.

2.2 Project scope and boundaries

The scope of this project is focused on the collection, preprocessing, and integration of weather data from the Finnish Meteorological Institute (FMI) interface for use in a neural network pipeline. The project aims to create reliable documentation flexible functions for fetching different types of weather data and preparing it for integration into machine learning models.

2.2.1 Project scope

The scope of this project is focused on collecting and preprocessing weather data from the Finnish Meteorological Institute (FMI) interface and uploading the data to an azure blob storage. The primary goal is to create comprehensive and user-friendly documentation and flexible functions that will allow for the easy retrieval of various types of weather data, including temperature, precipitation, wind speed, and snow depth. My second goal is to use this data to calculate missing features.

2.2.2 Project boundaries

While the project covers data collection, preprocessing, and integration for machine learning, there are certain boundaries that define the limits of the work. The project will only focus on weather data sourced from the FMI interface and will not explore other external weather data providers or APIs. The development of the neural network model itself is not a central objective of this project even though the data gathered in this project will be later used for that purpose.

2.3 Methods and approach

For this project, I will be creating a public GitHub repository that will serve as a central resource for the documented process of fetching different types of weather data from the Finnish Meteorological Institute (FMI) interface using Python. This repository will contain all the necessary documentation and code related to the course and the project, ensuring transparency and accessibility for future reference.

The main programming environment will be Visual Studio Code, which will be used for writing and testing the Python code. I will leverage the official FMI documentation to understand the API and data retrieval methods. Additionally, I will consult various online resources, including forums, troubleshooting websites, and large language models (LLMs), to resolve potential issues or refine my approach as needed. These resources will help address any challenges that may arise during the development process and provide guidance on best practices. For example, I plan to utilize resources like Stack Overflow for troubleshooting and advice, and ChatGPT as a language model for assistance in solving problems and refining code.

The project will focus on developing a well-documented codebase that details each step of the process, from fetching weather data to preprocessing it for machine learning use. This will include comprehensive comments in the code, examples of how to retrieve different types of data, and explanations of any issues encountered along the way and how they were solved.