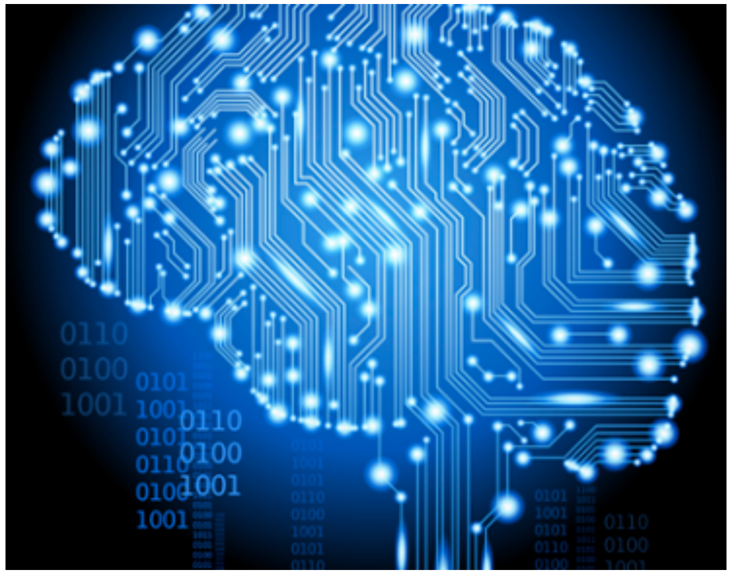
University of Strathclyde Engineering

Time-series analysis and prediction.

Bartosz Lewandowski
Supervisor: Dr. Robert Atkinson



Overview

Problem description & Motivation

The aim of this project is to develop a predictive model, that will indicate future river levels in Scotland. The project requires collecting river level and weather information data followed by researching and utilising number of machine learning algorithms.

Indicating future rival levels will help to predict floods. Thus, improve situational awareness of the authorities and allow them take appropriate action in case of a disaster. For instance, evacuate citizens nearby the endangered area.

Objectives

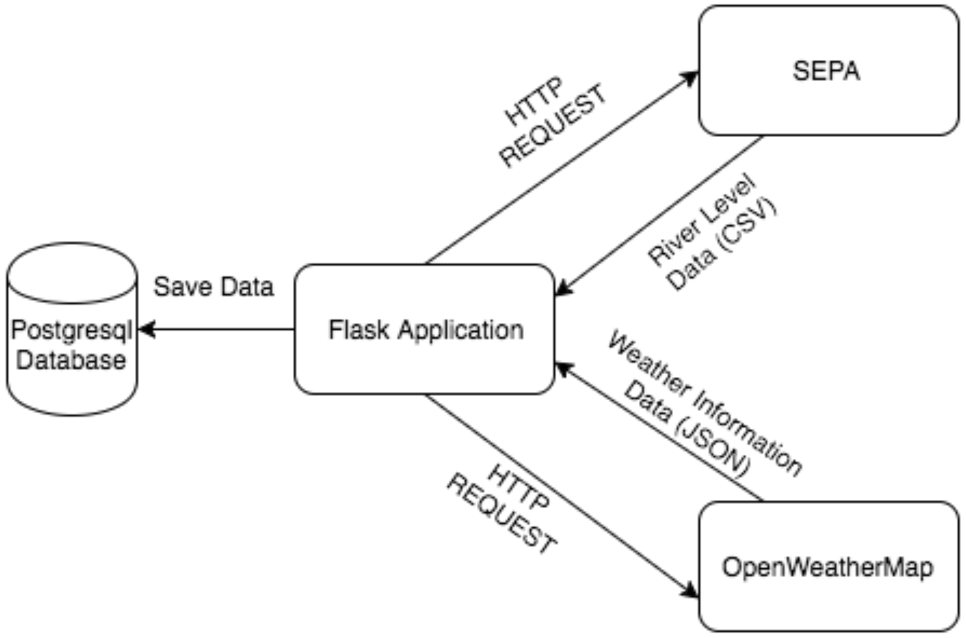
- Implement a Postgresql database for storing river level and weather information data.
- Create a data harvesting application that will poll data from SEPA website and OpenWeatherData api.
- Research and utilise machine learning algorithms in Python.
- Develop a predictive model to indicate future river level data.
- Host solutions as a Flask web application, display the prediction using Highcharts javascript library.

Collecting Data

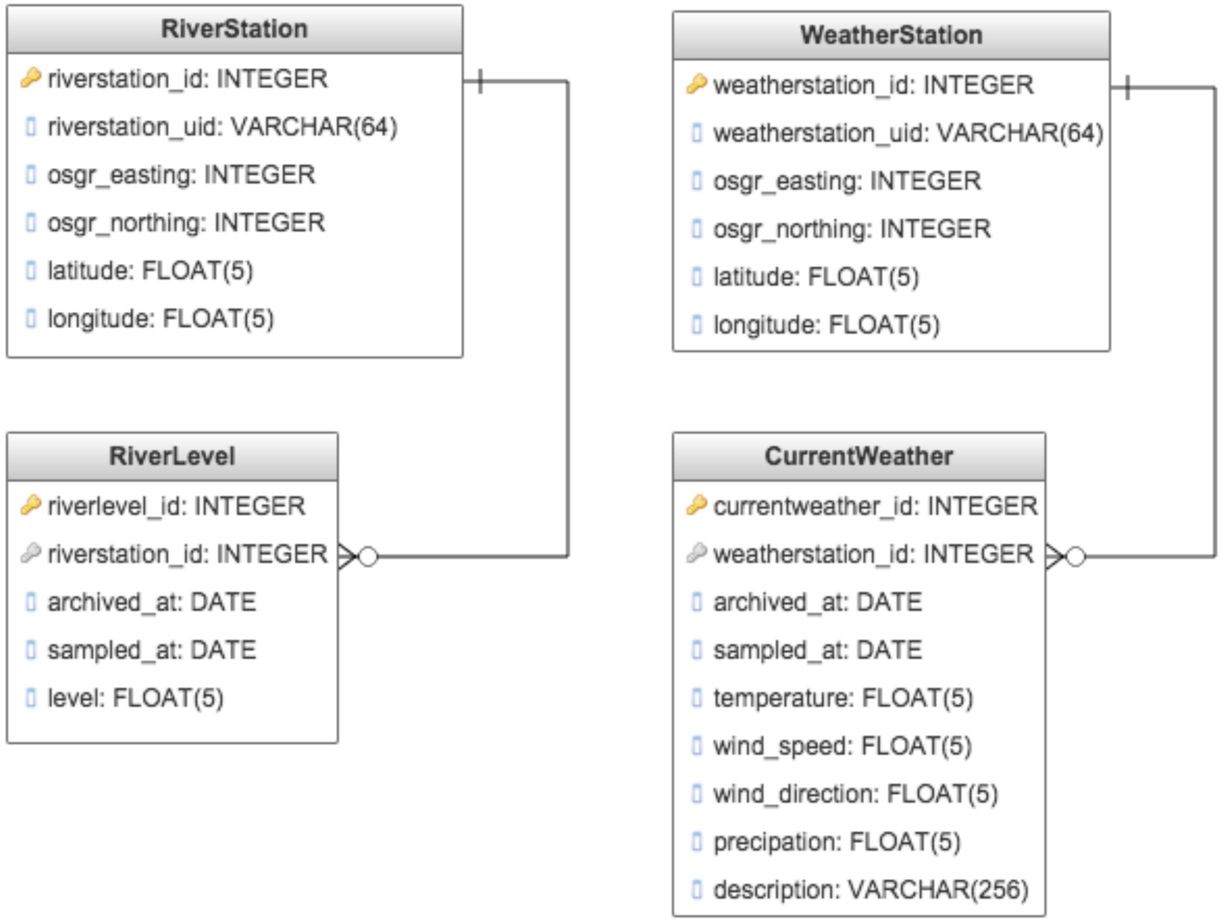
University of Strathclyde Engineering

Application

- Data Harvesting application is implemented using Flask micro-framework for Python.
- A scheduler sends http requests to both SEPA and OpenWeatherMap using a http client.
- The data is then processed and stored in a Postgresql database using an ORM - SQL Alchemy.



Database Tables



RiverStation
- riverstation_id: INTEGER
- riverstation_uid: VARCHAR(64)
- osgr_easting: INTEGER
- osgr_northing: INTEGER
- latitude: FLOAT(5)
- longitude: FLOAT(5)

WeatherStation
- weatherstation_id: INTEGER
- weatherstation_uid: VARCHAR(64)
- osgr_easting: INTEGER
- osgr_northing: INTEGER
- latitude: FLOAT(5)
- longitude: FLOAT(5)

RiverLevel
- riverlevel_id: INTEGER
- riverstation_id: INTEGER
- archived_at: DATE
- sampled_at: DATE
- level: FLOAT(5)

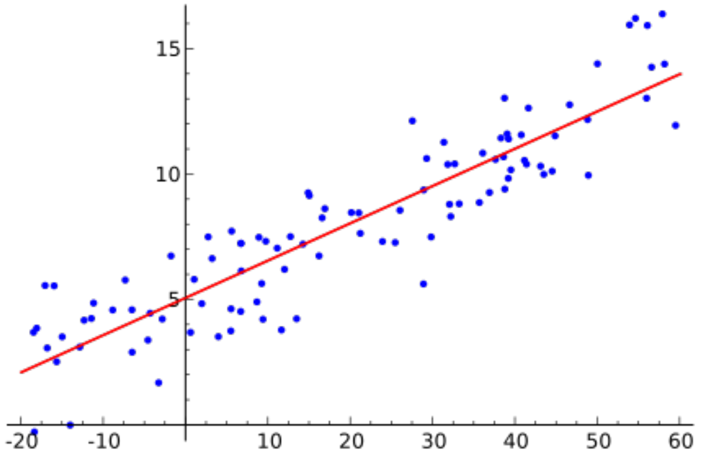
CurrentWeather
- currentweather_id: INTEGER
- weatherstation_id: INTEGER
- archived_at: DATE
- sampled_at: DATE
- temperature: FLOAT(5)
- wind_speed: FLOAT(5)
- wind_direction: FLOAT(5)
- precipation: FLOAT(5)
- description: VARCHAR(256)

Machine Learning

University of Strathclyde Engineering

Overview of the solution

Predictive model will be developed by solving a regression problem using supervised machine learning techniques.

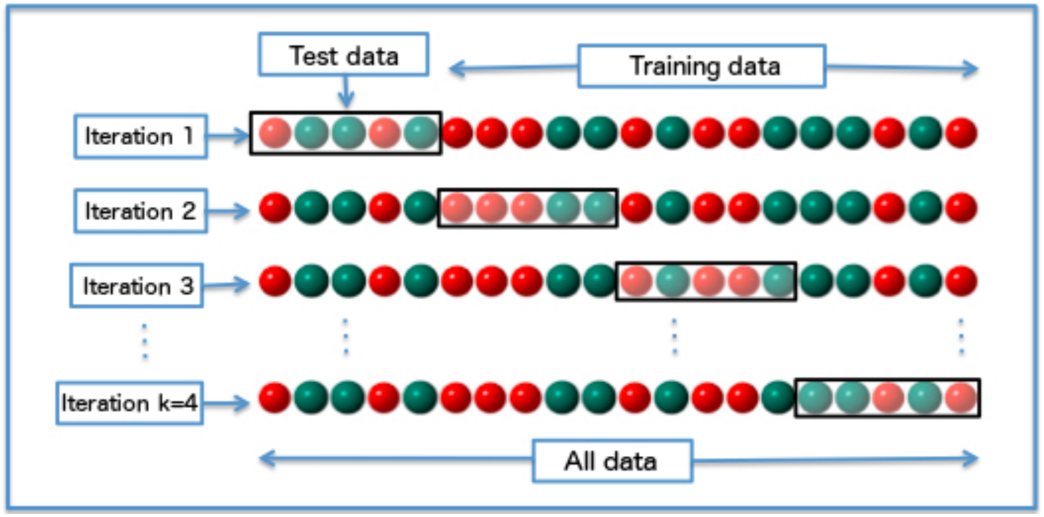


Process

- Preprocessing of data.
- Split the dataset using k-fold cross validation method.
- Choose an algorithm and create a predictive model.
- Evaluate model's performance on the test dataset.
- Repeat for another algorithm and compare results.

Python packages

- Scikit-learn
- NumPy
- SciPy
- matplotlib
- Pandas



Machine Learning Algorithms

	Pros	Cons
Linear Regression	• Good for first look at the data • Easily understandable model • Not likely to overfit	• Unable to model complex relationships • Unable to capture nonlinear relationships without tranforming the data
Support Vector Machines	• Can model complex, non linear relationships • Robust to noise • Popular in stock market analysis	• Good kernel function required • Model parameters are hard to interpret • Requires significant memory and processing power
Neural Network	• Extremely powerful • Can model very complex relationships • No need to understand the data	• Likely to overfit • Long training time • Requires significant computing power • Model is unreadable
K-Nearest Neighbours	• Simple, yet powerful • No training required (“lazy” learner) • Naturally handles regression	• Expensive and slow to predict new instances • Meaningful distance function required • Performs poorly on high-dimensional datasets.

Reference: googleplus and Hamilton, L.D. (2014) Machine learning algorithm cheat sheet - Laura Diane Hamilton. Available at: http://www.lauradhamilton.com/machine-learning-algorithm-cheat-sheet (Accessed: 28 November 2016).

Timeline

Semester 1

- Familiarisation with Python syntax
- Implement Data Harvesting Application
- Familiarisation with Python's Machine Learning and data manipulation libraries.

Semester 2

- Research different Machine Learning algorithm.
- Application of different algorithms to the data.
- Selecting the best performing model.
- Integration of the model into Flask application and displaying predictions using Highcharts JavaScript library.

**Report writing will be done along each step of the project.*