

## 0.1 Problem formulation

### 0.1.1 Application of ML Problem

The current winter has shown that slippery weather might come as a surprise. However, when slippery conditions are expected during the day, a service called Liukastumisvaroitus (Slipping warning) sends subscribers a text-message.

According to the website, the slippery condition is identified by humans. Though one could think that these dangerous weather conditions can be predicted without human knowledge, but to what level of accuracy? This makes it a great target for testing machine learning as an application.

Ideally the machine learning application could predict, given the current weather conditions, whether the slippery warning would be raised.

The data point is going to be a daily observation of weather data, with the additional slippery warning parameter. In other words, a single data point represents the weather conditions of a day. Data includes all data points (daily observations) from around November 2013, as the earliest records of slipping warnings are from then.

Concluding the parameters,

- Potential features for the application could be *Precipitation amount*, *Air temperature* and *Snow depth*. All of these properties are numerical and are easily measureable.
- The label of the application is going to be whether the *slippery warning* would be raised, with values 0/1.

### 0.1.2 Data sources

The slipping warning service offers an API for historical data analysis. Some 600 warnings have been issued in total since November 2013. As the slipping warning service data only consists of a timestamp and the city issued, training the machine learning algorithm to account for the current weather conditions needs more data to work with.

Thus, historical weather data from the Finnish Meteorological Institute is combined. Although FMI offers hourly historical data, in this project, the plan is to use the daily aggregated weather recordings for simplicity. In total, there are some 3000 daily weather reports since November 2013

Combined together, these data sources will be used to train and validate the machine learning algorithm.