

Aalto University

MS-E2177 - Seminar on Case Studies in Operations Research

Modelling of Paved Road Deterioration

Project Plan

Team Members:

Jonni Järvinen (Project manager)

Markus Junttila

Veera Wilkki

Oskar Heikkilä

Konsta Lahtinen

Returned: March 5, 2026

Contents

1	Background	3
2	Objectives	3
3	Tasks	4
4	Schedule	5
5	Resources	5
5.1	Human resources	5
5.2	Data resources	6
5.3	Software and tools	6
5.4	Hardware	6
6	Risks	7

1 Background

The Finnish Transport Infrastructure Agency (Väylävirasto), FTIA for short, is a Finnish government agency responsible for constructing and maintaining the public roads, railways and waterways in Finland. The road network under their management totals approximately 78 000 km, of which 53 000 km are paved roads, which carry 95% of all road traffic.

The importance of good road conditions is obvious. The general functioning of society requires huge amounts of goods and people to travel by car from point A to B safely every day. Deteriorating road conditions directly affect the safety of travel, in addition to having significant economic costs. Badly maintained roads increase the risk of accidents, force traffic to slow down, and damage cars more frequently.

The FTIA aims to mitigate these issues by regularly monitoring road conditions throughout the network, analyzing the collected data, and assigning repairs and other procedures as efficiently as possible. The goal is to make travel conditions as good as possible, for as many people as possible, while also maintaining passable travel routes everywhere.

Road conditions are currently measured with two indicators: IRI for evenness and URA for rut depth. Measurements are made between 1-4 years, depending on the road class and traffic volume. Other data, such as pavement type and information about past repairs, are also available. There is historical data for over 20 years.

The modelling and prediction of future road conditions is fairly difficult, and the models in use at FTIA are relatively simple and are not capable of utilizing all the historical data available. External factors that are difficult or simply cannot be measured effectively are also a big problem. These include climate change, the freeze-thaw cycle every year, changes in heavy traffic capacity legislation, and regional differences in soil, weather etc. The quality of the predictions directly affects the cost efficiency of repair operations, which in turn decides the general condition and future sustainability of all roads.

The motivation for this project is that there is a need for more accurate predictive models to model future road conditions. We aim to accomplish this by making a completely new machine learning based model to utilize all relevant existing data.

2 Objectives

The main goal of our project is to make a completely new machine learning based model to predict the behaviour of IRI and URA values for all roads in the dataset as accurately as possible. The potential secondary goal would then be to assess the most damaged continuous sections for repair.

The dataset of 20+ years from FTIA has a variety of different measurement categories. The main variables to consider would be previous measurements of IRI and URA, volume of

regular and heavy traffic, pavement structure and maintenance history, and climate data. However, the amount of data in each category varies greatly between different road sections, and we know some data to be inaccurate or missing completely. This has to be taken into account when building the model.

Because we are making a new model, rather than improving upon an existing one, there is a need to try and compare between different approaches. Possible candidates would be tree-based ML models, like Random Forest, Gradient Boosting or XGBoost, or time-series models.

Finally, we have been asked to analyze, whether climate change, regional variation, or the 2013 76-ton reform for vehicle combinations have any observable effects in road deterioration.

We hope to deliver a comprehensive and accurate model to help FTIA make informed and efficient pavement management decisions in the future.

3 Tasks

The project begins with a literature review to identify relevant pavement deterioration modelling approaches, previous findings on climate and traffic load impacts, and suitable machine learning methods for infrastructure prediction. In parallel, the team will familiarize themselves with the available datasets, their structure, variables, and limitations in order to define a feasible modelling scope.

Next, the data will be cleaned and preprocessed by addressing missing values and inconsistencies, and by aligning time-series measurements. Relevant explanatory variables will then be constructed, including indicators describing deterioration rates, traffic loading, climate-related factors, regional characteristics, and the pre- and post-2013 heavy vehicle reform period. Exploratory data analysis will be conducted to identify key patterns, trends, and relationships in the data and to guide modelling decisions.

Following this, baseline and machine learning models will be developed to predict IRI and URA values. Model development will start with simple features and models, with more complex methods adopted if they lead to improved performance. Model performance will be evaluated using appropriate metrics, and the results will be compared to identify the most suitable approach. Additional analyses will be carried out to assess the effects of climate factors, regional variation, and increased vehicle weights on road deterioration.

An interim report will summarize project progress, completed work, any changes to scope or plan, and the updated risk assessment. The project concludes with a final report presenting the methodology, results, and key findings, supported by relevant literature.

4 Schedule

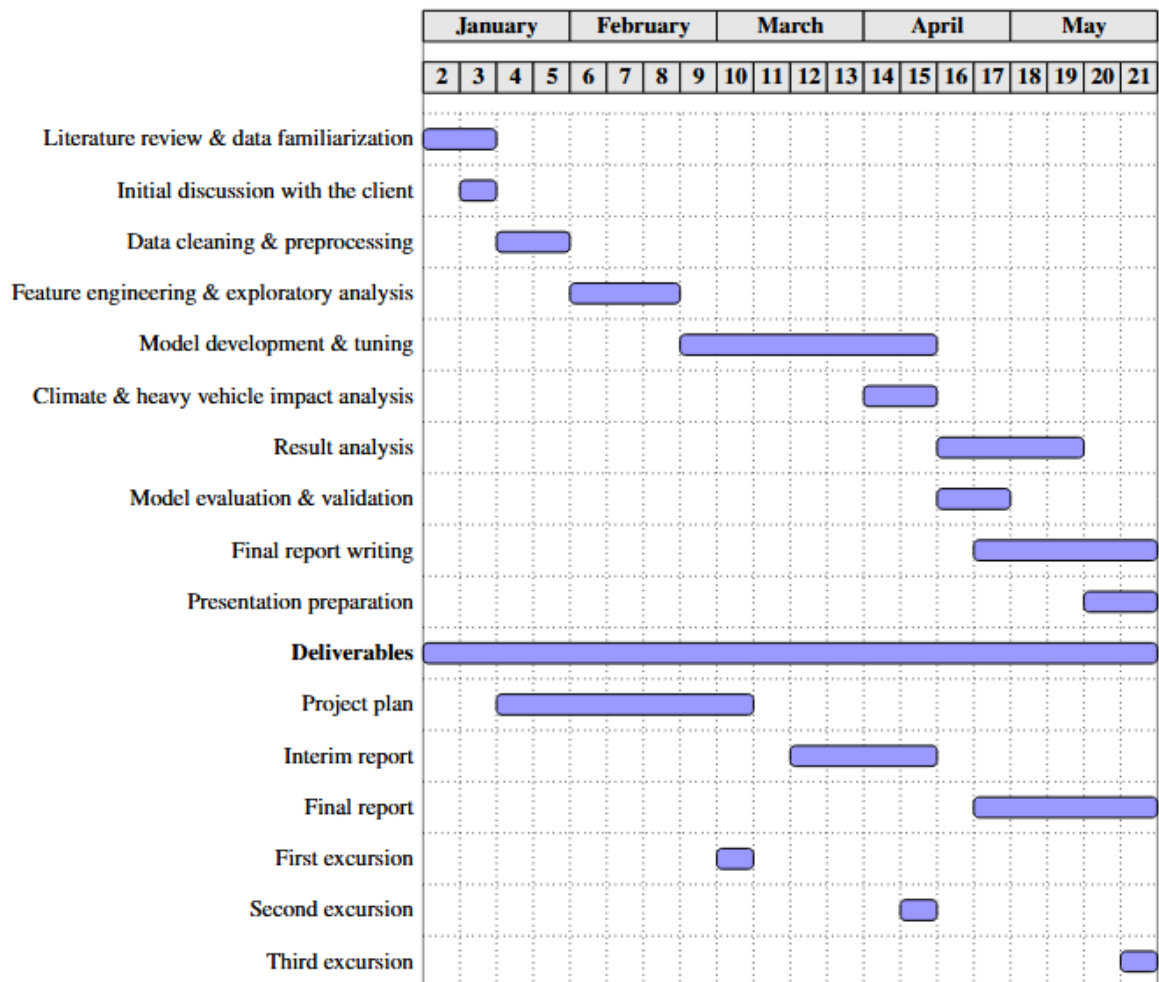


Figure 1: The initial schedule for the course.

The initial planned schedule for the project is presented in Figure 1. The blue bars represent the weeks during which the tasks are completed.

5 Resources

5.1 Human resources

Our project team consists of five students studying Mathematics and Operations Research, with some members coming from an engineering physics background. Our team has strong skills in data analysis and machine learning, which are essential for this project.

Our project manager is Jonni Järvinen, whose responsibilities include communicating directly with our contacts at Väylävirasto, scheduling team meetings, as well as handing in project deliverables. The remaining responsibilities are divided among the rest of the team according

to primary focus areas. Konsta Lahtinen is responsible for data engineering, including data preprocessing and feature engineering. Markus Junttila leads model implementation, while Oskar Heikkilä is responsible for model evaluation and validation. Veera Wilkki oversees the literature review and theoretical background. The designated person for each area ensures that jointly agreed objectives are met and that the work is well-documented. However, these roles do not represent strict divisions of work: all team members contribute to multiple areas of the project.

Our contacts at Väylävirasto are leading expert Vesa Männistö and expert Susanna Suomela, who will provide guidance throughout the project. The project team will meet with them approximately once or twice a month, or whenever necessary.

5.2 Data resources

Our primary data resource for this project is the Velho road dataset. The data is provided to us directly by our contacts at Väylävirasto in Excel form. The dataset contains a large number of parameters, most importantly IRI and rut depth. Other technical parameters are also provided, such as pavement information (type, width, thickness, etc.) and dates of maintenance and rehabilitation treatments.

Exploration of climate effects on road deterioration requires climate data. The Finnish Meteorological Institute (Ilmatieteen laitos) has open datasets that could be used for this purpose.

5.3 Software and tools

The main programming language used in this project is Python, which supports various libraries that can be used for data analysis and modelling. These include scikit-learn, TensorFlow and XGBoost, among others.

We use Jupyter Notebook for interactive coding and documenting our work. Git is used for version control as well as collaboration, so that up-to-date versions of code are always available for all team members.

We use LaTeX for reporting, i.e., creating project deliverables. More specifically, Overleaf allows for collaboration on the same file. Meeting notes and relevant articles are shared through Google Drive.

5.4 Hardware

We expect to implement the project mostly using our personal computers, but may use Aalto University's computing resources, if necessary. However, the aim is to avoid the need for additional computing resources so that deployment of our final model is as easy as possible for Väylävirasto.

6 Risks

The main risks assessed for this project are listed in Table 1. The risks and their effects are defined, and their probabilities and impacts are estimated on a discrete scale [Low, Medium, High]. The last column shows how to mitigate the risks.

Table 1: Main risks related to the project.

Risk	Effect	Probability	Impact	Mitigation
Data quality inconsistencies	Measurement errors or missing maintenance records may bias model training and reduce reliability of results.	High	High	Preprocessing; detect unrealistic jumps; different data validation methods.
Time limitations	Incomplete implementation or limited experimentation.	Medium	High	Divide tasks clearly; set weekly checkpoints.
Model performance remains insufficient with available data	Despite preprocessing and model tuning, the available data may not contain enough signal to build a reliable predictive model.	Medium	High	Perform exploratory data analysis early; test baseline models first; document limitations clearly; consider redefining target or narrowing scope if needed.
Incorrect problem formulation	Prediction target or time horizon may not align with project objectives.	Low	High	Define target and horizon clearly.
Team coordination issues	Scheduling conflicts, illness, or uneven workload distribution may delay progress and reduce overall project quality.	Medium	Medium	Agree on clear roles and responsibilities; maintain regular communication and backup plans for key tasks.
Overfitting of machine learning models	Model performs well on training data but poorly on unseen data.	Medium	Medium	Use cross-validation; start from basic models before complex models; avoid "black box" models.