# Master Thesis Project Plan

Detecting Transport Hubs Using TensorFlow

Author:

Mahmoud Sukkar

Supervisor:

Shafiq Ur Rehman

January 31, 2018

# Contents

# 1 Introduction

This project plan is written towards the fulfillment of Master Thesis in Electronics for Robot and Control master program at Umeå University. The thesis holds the title "Detecting Transport Hubs using TensorFlow" and held at Scania. This plan givens an overview of project description, problem definition and research methods along with planned activities in order to meet the requirement of the project.

## 1.1 Project Overview

Transportation has become a vital part in people's ordinary life. Looking at it from the single perspective of connecting suppliers with consumers, we can realize partially how big and complex the transportation network is. In addition, it will be beneficial if we could find a pattern or usages for different nodes within the transportation network. Those interesting nodes are the transport hubs where vehicles stop for some purposes. It could be a fuel station, supermarket or even cargo terminal.

Scania has more than 300 thousand connected vehicles that keep sending location data and it is our concern to analyze the data and extract good information that helps in identifying potential transport hubs. A good classification of transport hub may infer the usage of the vehicle that uses it since vehicle movements will be between similar or related transport hubs.

## 1.2 Purpose and goal

The goal behind detecting transport hubs is to get a deeper knowledge of how Scania trucks are being used which can make it easier to optimize transport flows and developing a better understanding of costumer's needs and preferences.

## 1.3 Deliverables

Trained transport hub classifier using machine learning algorithms to be delivered and evaluated against ground truth data.

# 2 Plan for the project

In this section, project plan is presented during different phases of the project, followed by a calendar that illustrated activities and milestones to achieve.

## 2.1 Before start

- A pre-study to be concluded in this phases to get an overview of solutions followed to solve the classification problem.

- Reading documentations of frameworks that will be used in this Thesis such as HopsWork and TensorFlow.
- Implementations of small examples will be done along the pre-study to grasp a better understanding of the problem.
- Getting familiar with the provided data and discover the most relevant data to the problem.

## 2.2 During the project

Following tasks to be held in order during the thesis project:

- Extract good and relevant dataset from the database.
- Construct a binary classifier model for simplified problem statement e.g. (Fuel related transport hubs).
- Iteratively testing and evaluating the binary classifier.
- Scaling up the classifier to include more classes.
- Iteratively testing and evaluating the classifier.
- Acceptance of project by all supervisors and examiners.
- Construct a predictive model for truck usage if time permits.

It should be noted also that thesis report writing and regular meeting with my supervisor shall take a place.

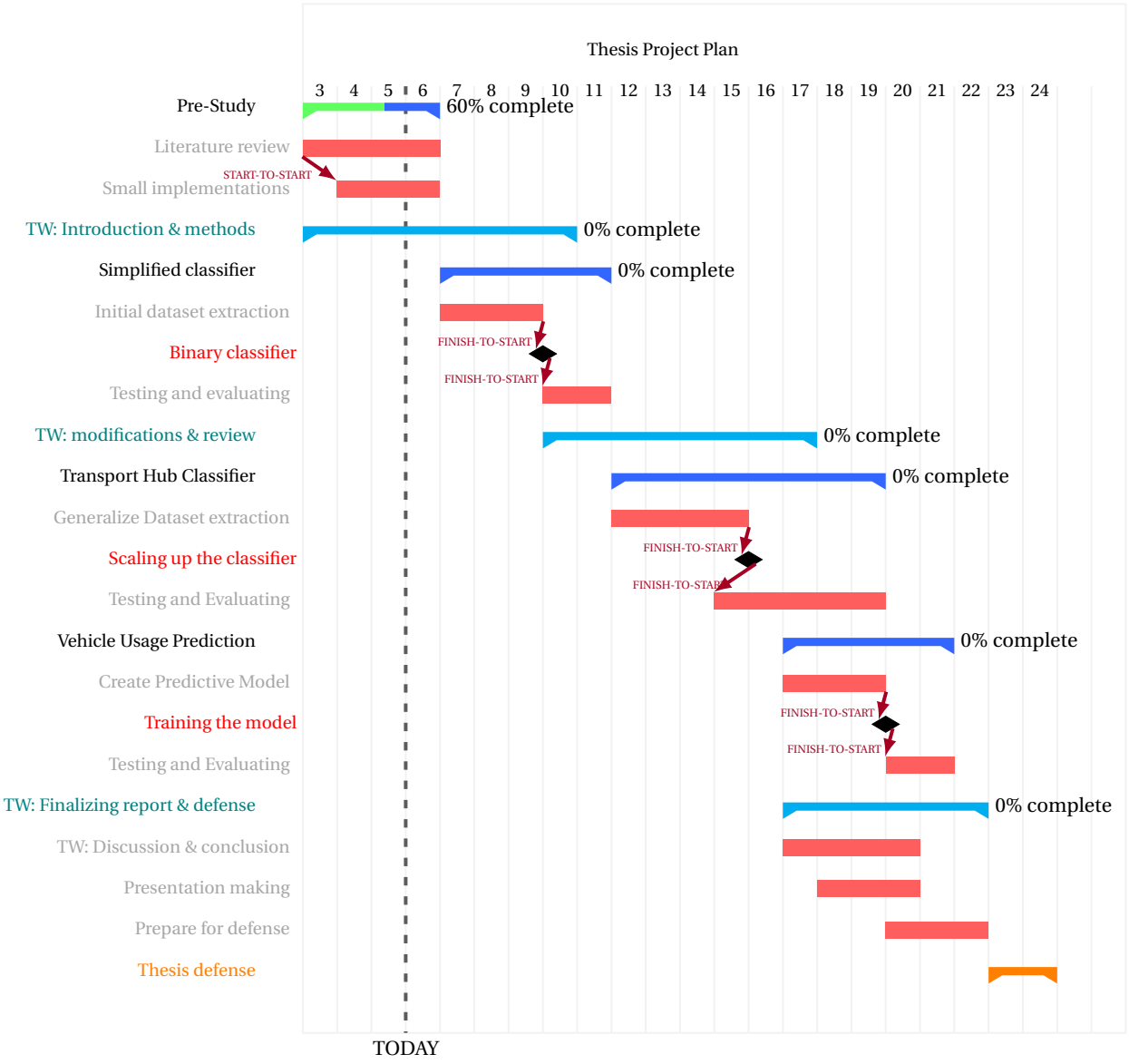
## 2.3 After the project

After all deliverables and requirements are met, following activities will be done:

- Evaluating results.
- Writing Discussion/conclusion part of the thesis.
- finalizing and reviewing the thesis report.
- Preparing presentation for thesis defense.
- Master Thesis defense.

## 2.4 Project plan schedule

The following chart represents preliminary Thesis plan starting from week 3 until week 24. Milestones in project are shown in red font. The TW label refers to Thesis writing. It should be noted that each testing and evaluating phase is iterative. Also, Plan will be modified upon regular meetings with Thesis supervisor.



# 3 SWOT Analysis

SWOT analysis is presented in the following table to demonstrate different aspects of the project and its learning outcome as shown in the following table.

| | | |
|---|---|---|
| | • Familiarity with TensorFlow<br>• Good overview of related literature | • Lack of knowledge in cluster computing<br>• Lack of knowledge in Scala |
| • Powerful computational resources<br>• Good external supervision | • Apply deep learning on big data<br>• Learn new skills e.g. Scala | • Hands-on cluster computing<br>• Learn new skills e.g. Scala |
| • Irrelevance of dataset<br>• Unreliable ground truth. | • Extract more relevant features from the current data.<br>• Come up with good dataset and consistent ground truth. | • Extracting poor dataset for training the classifier. |

# 4 Used applications

Scania's HopsWorks clustered framework will be used throughout the project. Following is a list of tools and frameworks to work with:

- TensorFlow: Open-source python library for machine learning.
- Zeppelin: Web-based notebook within the HopsWorks framework that supports coding in Scala, python and other programming languages.
- Apache Spark: an open-source cluster computing that allows dealing with big data efficiently and manipulating databases.
- Apache Kafka: an open-source stream processing platform that helps in creating pipelines for real-time applications.
- SQL Databases: Knowledge of SQL is essential in the project in order to manipulate and extract desired information.