

# Arrival Time Prediction On Buses in Linköping

---

-

**Max Lund, Simeon Jackman, Per Lindström, Elias Alesand, Jakob  
Lundberg, Sebastian Callh**

Handledare : Matthias Tiger  
Examinator : Cyrille Berger

## Upphovsrätt

Detta dokument hålls tillgängligt på Internet - eller dess framtida ersättare - under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida <http://www.ep.liu.se/>.

## Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a period of 25 years starting from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: <http://www.ep.liu.se/>.

## Sammanfattning

Abstract.tex

# Författarens tack

Acknowledgments.tex

# Innehåll

<b>Sammanfattning</b>	<b>iii</b>
<b>Författarens tack</b>	<b>iv</b>
<b>Innehåll</b>	<b>v</b>
<b>1 Introduction</b>	<b>2</b>
1.1 Motivation . . . . .	2
1.2 Aim . . . . .	2
1.3 Research questions . . . . .	2
1.4 Delimitations . . . . .	3
<b>2 Theory</b>	<b>4</b>
<b>3 Methods</b>	<b>5</b>
3.1 Preprocessing . . . . .	5
<b>4 Results</b>	<b>6</b>
<b>5 Discussion</b>	<b>7</b>
5.1 Results . . . . .	7
5.2 Method . . . . .	7
5.3 The work in a wider context . . . . .	8
<b>6 Conclusion</b>	<b>9</b>

3Arrival Time Prediction On Buses in LinköpingArrival Time Prediction On Buses in LinköpingArrival Time Prediction On Buses in Linköping



# 1 Introduction

The introduction shall be divided into these sections:

## 1.1 Motivation

[scigen]

This is where the studied problem is described from a general point of view and put in a context which makes it clear that it is interesting and well worth studying. The aim is to make the reader interested in the work and create an urge to continue reading.

## 1.2 Aim

What is the underlying purpose of the thesis project?

## 1.3 Research questions

This is where the research questions are described. Formulate these as explicit questions, terminated with a question mark. A report will usually contain several different research questions that are somehow thematically connected. There are usually 2-4 questions in total.

Examples of common types of research questions (simplified and generalized):

1. How does technique X affect the possibility of achieving the effect Y?
2. How can a system (or a solution) for X be realized so that the effect Y is achieved?
3. What are the alternatives to achieving X, and which alternative gives the best effect considering Y and Z? (This research question is normally broken down in to 2 separate questions.)

Observe that a very specific research question almost always leads to a better thesis report than a general research question (it is simply much more difficult to make something good from a general research question.)

The best way to achieve a really good and specific research question is to conduct a thorough literature review and get familiarized with related research and practice. This leads to

ideas and terminology which allows one to express oneself with precision and also have something valuable to say in the discussion chapter. And once a detailed research question has been specified, it is much easier to establish a suitable method and thus carry out the actual thesis work much faster than when starting with a fairly general research question. In the end, it usually pays off to spend some extra time in the beginning working on the literature review. The thesis supervisor can be of assistance in deciding when the research question is sufficiently specific and well-grounded in related research.

## **1.4 Delimitations**

This is where the main delimitations are described. For example, this could be that one has focused the study on a specific application domain or target user group. In the normal case, the delimitations need not be justified.





## 2 Theory

The main purpose of this chapter is to make it obvious for the reader that the report authors have made an effort to read up on related research and other information of relevance for the research questions. It is a question of trust. Can I as a reader rely on what the authors are saying? If it is obvious that the authors know the topic area well and clearly present their lessons learned, it raises the perceived quality of the entire report.

After having read the theory chapter it shall be obvious for the reader that the research questions are both well formulated and relevant.

The chapter must contain theory of use for the intended study, both in terms of technique and method. If a final thesis project is about the development of a new search engine for a certain application domain, the theory must bring up related work on search algorithms and related techniques, but also methods for evaluating search engines, including performance measures such as precision, accuracy and recall.

The chapter shall be structured thematically, not per author. A good approach to making a review of scientific literature is to use *Google Scholar* (which also has the useful function *Cite*). By iterating between searching for articles and reading abstracts to find new terms to guide further searches, it is fairly straight forward to locate good and relevant information, such as [test].

Having found a relevant article one can use the function for viewing other articles that have cited this particular article, and also go through the article's own reference list. Among these articles one can often find other interesting articles and thus proceed further.

It can also be a good idea to consider which sources seem most relevant for the problem area at hand. Are there any special conference or journal that often occurs one can search in more detail in lists of published articles from these venues in particular. One can also search for the web sites of important authors and investigate what they have published in general.

This chapter is called either *Theory*, *Related Work*, or *Related Research*. Check with your supervisor.



## 3 Methods

### 3.1 Preprocessing

The data for this project was delivered by Östergötatrafiken. The whole dataset is over 300GB in size. To be able to feed the data into our algorithms preprocessing had to be done.

#### Data structure

The provided data consists of one file per day over 90 day. Each day has an approximate size of 5GB. Within the data there are over 20 event-types that represent the state of a bus during a day. For our project we have used four types of events and discarded data with other events. The four events are:

**ObservedPositionEvent:** Gets triggered every second, contains the GPS data of a given bus.

**EnteredEvent:** Gets triggered when the bus is within a certain distance to a bus-station. Is used to split the journey into segments.

**JourneyStartedEvent:** Gets triggered when the bus is assigned a new journey. Is used to determine which line a bus is currently serving.

**JourneyCompletedEvent:** Gets triggered when the bus has completed a journey. Is used as a flag to determine when a journey has ended.

#### Loading data

We have used jupyter notebooks as the main way to develop our code to pre-process. Data operations such as filtering are done with the python pandas package.



## **4 Results**

This chapter presents the results. Note that the results are presented factually, striving for objectivity as far as possible. The results shall not be analyzed, discussed or evaluated. This is left for the discussion chapter.

In case the method chapter has been divided into subheadings such as pre-study, implementation and evaluation, the result chapter should have the same sub-headings. This gives a clear structure and makes the chapter easier to write.

In case results are presented from a process (e.g. an implementation process), the main decisions made during the process must be clearly presented and justified. Normally, alternative attempts, etc, have already been described in the theory chapter, making it possible to refer to it as part of the justification.



## 5 Discussion

This chapter contains the following sub-headings.

### 5.1 Results

Are there anything in the results that stand out and need be analyzed and commented on? How do the results relate to the material covered in the theory chapter? What does the theory imply about the meaning of the results? For example, what does it mean that a certain system got a certain numeric value in a usability evaluation; how good or bad is it? Is there something in the results that is unexpected based on the literature review, or is everything as one would theoretically expect?

### 5.2 Method

This is where the applied method is discussed and criticized. Taking a self-critical stance to the method used is an important part of the scientific approach.

A study is rarely perfect. There are almost always things one could have done differently if the study could be repeated or with extra resources. Go through the most important limitations with your method and discuss potential consequences for the results. Connect back to the method theory presented in the theory chapter. Refer explicitly to relevant sources.

The discussion shall also demonstrate an awareness of methodological concepts such as replicability, reliability, and validity. The concept of replicability has already been discussed in the Method chapter (3). Reliability is a term for whether one can expect to get the same results if a study is repeated with the same method. A study with a high degree of reliability has a large probability of leading to similar results if repeated. The concept of validity is, somewhat simplified, concerned with whether a performed measurement actually measures what one thinks is being measured. A study with a high degree of validity thus has a high level of credibility. A discussion of these concepts must be transferred to the actual context of the study.

The method discussion shall also contain a paragraph of source criticism. This is where the authors' point of view on the use and selection of sources is described.

In certain contexts it may be the case that the most relevant information for the study is not to be found in scientific literature but rather with individual software developers and open

source projects. It must then be clearly stated that efforts have been made to gain access to this information, e.g. by direct communication with developers and/or through discussion forums, etc. Efforts must also be made to indicate the lack of relevant research literature. The precise manner of such investigations must be clearly specified in a method section. The paragraph on source criticism must critically discuss these approaches.

Usually however, there are always relevant related research. If not about the actual research questions, there is certainly important information about the domain under study.

### **5.3 The work in a wider context**

There must be a section discussing ethical and societal aspects related to the work. This is important for the authors to demonstrate a professional maturity and also for achieving the education goals. If the work, for some reason, completely lacks a connection to ethical or societal aspects this must be explicitly stated and justified in the section Delimitations in the introduction chapter.

In the discussion chapter, one must explicitly refer to sources relevant to the discussion.



## 6

## Conclusion

This chapter contains a summarization of the purpose and the research questions. To what extent has the aim been achieved, and what are the answers to the research questions?

The consequences for the target audience (and possibly for researchers and practitioners) must also be described. There should be a section on future work where ideas for continued work are described. If the conclusion chapter contains such a section, the ideas described therein must be concrete and well thought through.