FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Comparing Bus Travel Time Prediction Using AVL and Smart Card Data

Fernando Aparecido dos Santos Silva

Master in Information Engineering

Supervisor: Dr. João Pedro Carvalho Leal Mendes Moreira

Co-Supervisor: Dr. Luís Alexandre Moreira Matias

September 25, 2017

Comparing Bus Travel Time Prediction Using AVL and Smart Card Data

Fernando Aparecido dos Santos Silva

Master in Information Engineering

Approved in oral examination by the committee:

Chair: Dr. António Pedro Rodrigues Aguiar

External Examiner: Dr. Rui Camacho

Supervisor: Dr. João Pedro Carvalho Leal Mendes Moreira

September 25, 2017

Resumo

Como uma área de estudo importante, o Sistema de Transporte Inteligente emergiu como o preditor do tempo de viagem nas redes de autoestrada para fornecer informações aos passageiros em relação ao transporte público. O Sistema de Transporte Inteligente fornece tecnologias de análise, controle e comunicação através de aplicativos para melhorar a segurança, mobilidade e eficiência. As informações são constantemente compartilhadas por agências de trânsito e empresas que usam o Sistema de Transporte Inteligente para minimizar os problemas de trânsito como congestionamento, impacto ambiental e melhorar o gerenciamento de tráfego. Informações precisas sobre horários de chegada do autocarro ajudam os passageiros a planejar suas viagens além de fazê-los sentir-se mais satisfeito com o serviço de transporte. Um serviço eficiente de transporte de autocarro pode reduzir o uso do carro particular, o que reduz os problemas de excesso de tráfego e leva a uma redução dos tempos de espera dos passageiros no ponto de ônibus, o que, em última instância, leva a passageiros de ônibus mais satisfeitos.

O objetivo desta dissertação está relacionado ao uso de dados de cartão inteligente de passageiros e dados de sistema automático de viaturas fornecidos pela empresa de transporte portuguesa, Sociedade de Transportes Colectivos do Porto (STCP), Portugal. O objetivo é comparar e avaliar as previsões do tempo de viagem das linhas de autocarro, além de ajudar na confiança do serviço de autocarro portugues, na redução de atraso do tempo de viagem do autocarro, no aumento da lucratividade das empresas em suas rotas de autocarros a integrar os dados do sistema automático de viaturas e cartão inteligente de autocarros. À medida que os sistemas automático de viaturas usam o dispositivo do sistema de posicionamento global colocado no veículo para poder localizar sua posição e velocidade em toda a rota (Algueró, 2013). Os dados do cartão inteligente são gerados por um sistema de bilhética integrante que captura dados dos cartões de usuários de autocarros e comunica informações através do sistema para integrar dados em cada rota (Algueró, 2013). É importante entender qual é o impacto do número de pessoas na parada de autocarros no plano de horário de autocarros, bem como demonstrar uma análise preditiva do tempo de viagem do ônibus.

Os dados dos sistemas de localização automática de viaturas e bilhetica foram usados para comparar modelos preditivos para previsão do tempo de viagem de autocarros. A necessidade de comparar a previsão do tempo de viagem de autocarros a usar dados do cartão inteligente vem junto com o avanço da tecnologia, integração de disponibilidade dos dados e o uso desses dados na análise com algoritmos de machine learning. Esta análise reforça a competitividade das empresas públicas públicos de autocarros para garantir estrategicamente a fiabilidade do serviço de autocarros. A motivação para entender as tecnologias de cartões inteligentes tornou-se útil para as empresas de trânsito que utilizam isso para obter indicadores de transporte público. Além de realizar uma análise para entender as características dos dados,os dados também são usados para propósitos preditivos.

Para este estudo proposto foi usado alguns algoritmos como os modelos (Random Forests (RFs), Artificial Neural Networks (ANNs) and Support Vector Regression).

A aplicação de modelos preditivos que usam esses dados disponíveis para aplicar métodos de machine learning forneceu uma melhora média dos modelos preditivos de 10% em comparação com a avaliação do conjunto de dados do sistema automático de viaturas.

Abstract

As an important study area, Intelligent Transportation System (ITS) has emerged as the predictor of travel time on road networks in order to provide information to passengers regarding public transportation. ITS provides analysis, control and communication technologies through applications to improve safety, mobility and efficiency. Information is constantly shared by transit agency and companies using ITS to minimize transit problems as congestion, environmental impact and improve traffic management. Accurate information on bus arrival times helps passengers to plan their trips in addition to making them feel more satisfied with the transport service. An efficient bus transport service can reduce private car usage which reduces problems of excess traffic volume and leads to a reduction of passengers waiting times at bus stop which ultimately leads to more satisfied bus passengers.

The purpose of this dissertation is related to the use of smart card (SC) data from passengers and Automatic Vehicle Location (AVL) data provided by the Portuguese transport company, Sociedade de Transportes Colectivos do Porto (STCP), Portugal. The aim is to compare and evaluate travel time predictions of bus lines, in addition to helping Portuguese bus service reliability, reducing bus travel time delay, increasing company profitability across its bus routes by integrating archived bus AVL and SC data. As AVL systems use Global Positioning System (GPS) device placed in the vehicle to be able to locate its position and speed throughout the route (Algueró, 2013). The smart card data is an integrating ticketing system which captures data from the bus user cards and communicate information through the system to integrate data in each bus route (Algueró, 2013). It is important to understand what is the impact of the number of people at the bus stop on the Bus Schedule Plan (BSP), as well as demonstrating an predictive analysis of the bus travel time.

AVL and SC data were used to compare predictive models for bus travel time (TT) prediction. The need to compare the forecast of bus travel time using AVL and smart card data comes along with the advance of the technology, data availability integration and the use of these data on analysis with machine learning algorithms. This analysis reinforces the competitiveness of public bus transport companies to ensure strategically bus service reliability. The motivation to understand smart cards technologies has become useful for transit companies which make use of it to get indicators of public transportation. Besides performing an analysis to understand the characteristics of the data, the data is also used for predictive purposes.

For this proposed study we use some algorithms such as Random Forests (RFs), Artificial Neural Networks (ANNs) and Support Vector Regression (SVR) models.

Applying predictive models using these available data (AVL and SC) to apply ML methods provided an average improvement of the predictive models of 10% in comparison to the evaluation of the AVL dataset.

Keywords: Machine learning (ML), algorithms, smart card (SC), automatic vehicle location (AVL), global positioning system (GPS), travel time prediction (TTP) and bus schedule plan (BSP).

Acknowledgements

I have some people to be grateful about my academic path at Faculty of Engineering of the University of Porto. Thank God! He has been my strength so far and firstly my help comes from Him and Jesus Christ.

I would like to thank my supervisor and co-supervisor Prof. Dr. João Mendes Moreira and Dr. Luís Alexandre Moreira Matias who helped me to choose the theme of this thesis and advised me when was necessary and when I had questions about the research. I would also like to thank Proof. Dr. Antonio Pedro Aguiar, coordinator of Information Engineering master, who always kept the door of his office open for students and accepted me in this master course.

I would like to thank my fiancee Priscila Felix who is my partner, best friend, personal advisor and my love. She stayed by my side in every moments even when she went to Spain, her presence was noticed by her caring and support.

I could not forget my parents' support, Luzinete Silva and Manuel Silva, without them this objective would never be accomplished. They were more than I expected even as parents and they were my structure in Lusitanian lands. To my brother Lucas Silva who made part of my trajectory even living in Brazil he did to me what I will never forget. Thank you my brother!

A special thanks to my aunt Creuza Alves and Jose Belmiro (in memory). Aunt Creuza is like a mother for me and I would like to thank her because she always supports me whenever I am and makes the difference in my life by praying and being an example of woman. To my uncle Jose Belmiro Alves who left this world two months ago and he will never be forgotten by me. He always had positive words to tell me in hard times. One more time thanks uncle Guega.

I would like to thank my parents in law, Geraldo Felix and Nidi Felix who have supported me and Priscila during our stay in Portugal and I really appreciate the way that they did not measure their efforts to support us living in Portugal.

To my friends Borgine and Mailson, thank you very much. The time we shared in this course made me learn a lot with you two. From colleagues they became friends.

I am grateful to some friends that I have made in Portugal and one of them is Wellington Bernardes. He was always available to help and I will not forget this.

Last but not the least a big thanks to Tony Silva who shared some time spreading knowledge with me at the LIAAD building. He helped me going through machine learning for this specific research area in this dissertation and better understand R programming language.

Fernando Aparecido dos Santos Silva

“It always seems impossible until it’s done.“

Nelson Mandela

Contents

CONTENTS

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES

Abbreviations

AI	Artificial Intelligence
APC	Automatic Passenger Counting
ANN	Artificial Neural Networks
AVL	Automatic Vehicle Location
BSP	Bus Schedule Plan
GPS	Global Positioning System
ITS	Intelligent Transportation System
KNN	K-Nearest Neighbours Regression
LOWESS	LOcally WEighted Scatterplot Smoothing
ML	Machine Learning
MSE	Mean Squared Error
PFI	Permutation Feature Importance
RF	Random Forests
RBF	Radial Basis Function
RMSE	Root Mean Squared Error
SC	Smart Card
STCP	Sociedade de Transportes Colectivos do Porto
SVM	Support Vector Machine
SVR	Support Vector Regression
TT	Travel Time
TTP	Travel Time Prediction

Chapter 1

Introduction

In recent times, the public transport system has been the challenge of many companies around the world, not only the infra-structure, i.e. vehicle in good conditions, well-trained drivers and automated fare payment. The corresponding number of vehicles running is over 800 million on road worldwide (Ferreira et al., 2012). The increase of vehicles running on the main cities' roads worldwide has caused problems as traffic congestion and discouraged many people to use their private vehicles lately (Moreira-Matias et al., 2015). The use of buses has been an alternative to control the increase of vehicles running and traffic congestion, in addition it has caused a result of the recognition of public transport being a reliable alternative for the short-travel in these main cities recently.

The quality of the bus transport services offered by companies has been seen with more attention and care lately, as bus travel time reliability is more sensitive for passengers. A way of reducing negative experiences/feelings in passengers towards the transit system is through improved reliability, however such disappointments come along with a long period waiting at bus stops, crowding situation on buses or not to be able to board the bus (Carrel et al., 2013). The public transport system nowadays has been supported in its enhancements by some studies and through the access of technologies, which helps to improve service reliability (Shalaby and Farhan, 2004). Mendes-Moreira (2008) states that the technology has been important for transport companies through investments in high technology systems such as ticketing system, automatic passenger counting, operations software, multi-modal traveller information in the past three decades.

The city of Porto, as other cities worldwide, has been impacted by improvements with bus travel time prediction, providing real time information on user's device (i.e. smartphones and computers) to be a reliable transportation service, also corresponding to the travel time. The impact of the bus travel time reliability on passengers, such as for those who are on the bus or for those who wait for it to be on time at the bus stop. It requires constant evaluation, day-by-day, with some operational techniques, then discussion will be taken along of this dissertation. In addition to describing the machine learning algorithms used to build bus travel time prediction and frame the dataset in order to review the old structure, nevertheless ensure that the service is offered in a high quality.

1.1 Motivation

This study proposes to identify some proper machine learning models, which can be applied in the bus travel time prediction. The use of predictive algorithms adequate for our problem aims to optimize the results.

The aim of this dissertation is the use of ML models to predict bus travel time, using two different data sources. Considering the number of boarding and alighting passengers as well as the bus arrival and departures times that is related from Automatic Vehicle Location (AVL) and smart card (SC) data on vehicle in real time information. The passengers wait times and the variability of bus arrival times are conditioned on an accurate and improved implementation of schedule adherence at each bus stop (Kimpel, 2001).

Global Positioning System (GPS) and Automatic Passenger Counting (APC) are metric systems used to hold information about the bus travel, including passengers entry and exit. These electronic systems were the support for the proposed study in this dissertation that will be the baseline to begin the research with evaluation and measurements.

AVL data can be an advantaged guide for making decisions which come from advanced technology, proof of this, according to information from the U.S. Department of Transportation, the buses in the city of Chicago are fully equipped with the AVL system since 2004 (U.S. Department of Transportation, 2007). The development of the U.S public transport system has been present through the massive investment in technology and the implementation of AVL system on the fleets.

A demonstration sample of an AVL system using GPS data is shown in figure 1.1. It shows the route for bus line 600 since it leaves the terminal until the last stop.

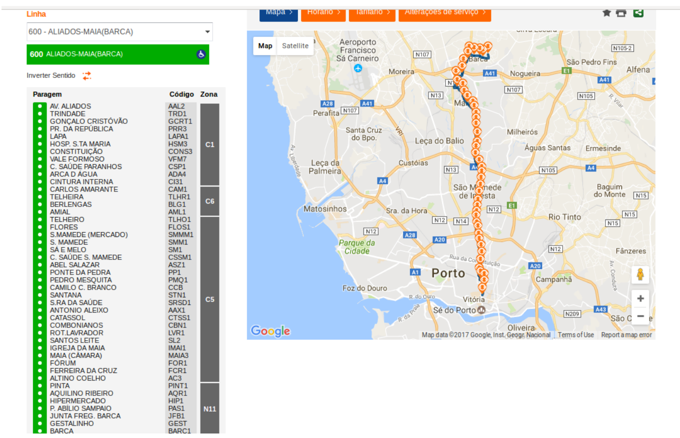


Figure 1.1: Bus route line 600 GPS location

Following the aforementioned description, according to the goal and target to be accomplished, more details will be presented in section 1.2, which involves the effective purpose in evaluating bus travel time prediction in public transport considering the impact of the passenger load at the bus stop in relation to the time a bus link has been affected.

1.2 Objective

To provide a more consistent travel time prediction (TTP) system which does not mean a system that is 100% perfect, but that is more accurate in every step of the process. We address the relevance of considering some events, i.e. the number of passenger at each bus stop. The performance is evaluated using the root mean squared error (RMSE). These aforementioned events will be considered as influences on forecasts and should also be important reason of delays that buses on their route go along to each stop, therefore the amount of passenger at each bus stop is another parameter to explain how accurate travel time is, in order to evaluate its relevance and average throughout the bus route. The excessive number of users on a bus or at a bus stop can cause an abrupt interruption passenger's use of the service, in addition the decrease of bus travel time reliability (Tirachini et al., 2013). This is a problem in relating to passenger overcrowding could affect the service reliability.

Firstly, the evaluation of accurate bus travel time prediction is related to the quality of dataset which informs the importance of attributes wherein it should capture data to become necessary information and all data to be trained and tested in the dataset by the following generation the prediction model.

With such a good data from the dataset, Bagchi and White (2004) say that bus travels can be built with smart card data to provide users a schedule where it can achieve their needs throughout the day by transportation companies. It is an advantage to have data which come from this type of technology, as it helps to support the process of implementation, development and analysis. Therefore, ITS is important to promote quality of bus public transport service and reflect the main modernization about the service.

1.3 Structure of this Dissertation

This dissertation contains five chapters, so throughout the proposed study and the baseline to comprehend more bus travel time prediction on the scope of the research field, which meets the needs of transportation companies and passengers. The increasing of more reliable bus service and travel time is a measure which ultimately leads to more satisfied bus passengers. In addition to the introduction, this dissertation contemplates four more chapters, according to the development of this study.

In Chapter 2 is shown a description about the start-of-art review for travel time prediction approach, whereas explains the proposed study and takes the opportunity to compare other studies. In addition, we will see, in details, the involved technology mechanism in capturing data from buses in circulation on city's road networks in Porto.

Chapter 3 settles down the main importance to promote the explanation about machine learning methods, models and algorithms that are applied to reach our goals or are used to provide some feedback and experience throughout the study from other researchers, therefore a description of the general and specific point of view from each models and algorithms were proposed. A mathematical explanation helps us to better comprehend the proposed algorithm over the idea and implementation on bus travel time prediction. Nevertheless, there will be proposed a particular analysis whereas results and graph will show up, not as definitive but as part of the comparative model.

In chapter 4 is proposed a more concise definition of each valuable algorithm, that better defines our goals without losing any perspective and providing an explanation about the choice of the problem, whereas it might be more than one predictive model. The suitable algorithm provides us more support over our analysis, therefore it particularly gives the idea through its applicable method on the dataset, even though some can be elected more suitable depending on the results and its applicable structure, in order to identify correlation with proposed data frame and objective.

Chapter 5 finalizes this dissertation with expecting result and helps us to conclude, after seeing some results in chapter 4, the contemplation of the idea and the problem is partially or totally solved by the conclusion on this method, algorithm and result figured out in the final part.

Chapter 2

State of the Art Overview: Bus Travel Time Prediction

Travel time prediction is a demanding study area, requiring proper data, technological tools access and, moreover, the effort to understand the behaviour of the transportation service across a state variable, in order to predicts its result considering some factors that can interfere the realization of operation system. Mendes-Moreira (2008) defines travel time prediction as being the time that a vehicle travels from a start point to its terminal, and this definition was used from the Turner et al. (1998). He also complements to clarify travel time prediction explaining some other definitions related to slack time which is the time a bus is stopped in a terminal between the final trip time and the beginning of a next trip. For the cycle time the author says that is the two consecutive times a bus is leaving from the same point in a time interval.

The ability to comprehend data, technological tools, and environment where we want to find solutions for specific problems raised in people's or companies' routine. It comes with the need to use machine learning and artificial intelligence as part of a learning model to ensure better results are being applied and comparing both. Machine learning has been present in many study areas consolidating its importance in employing learning methods in order to exemplify the problem-solving structure, solve them and show results. With historical data, as for example the studied in this dissertation, is necessary to train the methods. The training set and test are part of the aggregated data so they consist of a learning model and methods for trying to find a specific function which maps both input and output in a training dataset, as well as a second method is applied to evaluate the test dataset, in order to provide some feedback about whether the prediction is correct or not (Dahl et al., 2014).

Starting from a baseline study and considering bus travel time a problem for public transportation, we review the main idea of a public transport service problem given by its pattern to realize the contextual previous research. To relate a practical use, taking into account the routine service and the external factor of passenger that are distributed along a bus route at each stop and this distribution is unbalanced, which must influence the travel time for the bus schedule plan.

Moreira-Matias et al. (2015), in their project about improving public road transportation planning and monitoring, describes the general travel time prediction function as defined as following to be $TT_{(i,j)}$ the run time between two bus stops $b_i, b_j : j > i$ and the computation of travel time will be:

$$TT_{(i,j)} = \sum_{k=1}^{j-1} dwT_k + RT_{(k,k+1)} \tag{2.1}$$

The notation $RT_{(k,k+1)}$ is to represent the non-stop running time in a road segment between two consecutive bus stops $b_k, b_k + 1$, and the dwell time is represented by dwT_k at the bus stop bk which is the first bus stop.

Some other studies provide an understanding of the idea that passengers distribution along the bus route is considered the number of stop and the number of passenger distributed along the bus route. For us to have some understanding, whereas the kind of problem we are handling in this dissertation, Zhao et al. (2006) provide an explanation about the probability density function (p.d.f) for the random passengers arrival at the bus stop:

$$f_A(t) = SH^{-1}, \tag{2.2}$$

In this p.d.f (t) is a representation of the passenger's arrival at the bus stop, according to Bowman and Turnquist (1981) other alternative is the distribution of passenger's arrival time in a circumstance the bus schedule time is already known as shown in the following p.d.f:

$$f_A(t) = \frac{exp(U(t))}{\int_0^{SH} exp(U(t))d\tau} \tag{2.3}$$

Where $U(t) = \alpha E[w(t)]^b$ is the value of an arrival at time t. From the data we have "a" and "b" as constant. An expected waiting time for an arrival at time t is represented by the function $E[w(t)]$.

Previous studies using regression models were created by getting results in estimating predictions, they are used in average time, number of stops, time period of the day (i.e. independent variable), average transit, bus flow average, heavy vehicle proportion and travel distance (Yu et al., 2016).

Some complex models are applied for travel time prediction, in order to relate the most important model and result obtained. The authors' specifications have already been applied to bus travel time prediction problem in different approaches, however it has been a significant problem-solving in many circumstances that involves travel time prediction and its complexity.

These models are well employed for travel time prediction with accurate results. One of the most known approaches for travel time prediction is Regression which has the propose to infer the arrival times in a mathematical form using a set of independent variables, such as some regression algorithms can be easily applied for long-term travel time prediction (Moreira-Matias et al., 2015). The Support Vector Regression, K-nearest Neighbourhood and Artificial Neural Networks have

Table 2.1: Complex Regression Models Employed in Travel Time Prediction (Moreira-Matias et al., 2015)

Publication	Denomination	Description
Rosenblan' (Rosenblatt, 1958)	Artificial Neural Networks (ANN)	It uses multiple layers of neurons to obtain highly non-linear decision boundaries between values.
Nadaraya' (Nadaraya, 1964)	Kernel-based Regression	The goal is to establish a non-linear relation between a pair of random variables to estimate a conditional expectation between them.
Cover' (Cover and Hart, 1967)	K-Nearest Neighbours Regression (KNN)	This method uses a max. threshold ε which stands for the residual between the target function and any of the training samples. It is used to establish a confidence hyperplane to define the function which contains all these training samples.
Friedman' (Friedman and Stuetzle, 1981)	Projection Pursuit Regression (PPR)	The model consists of linearly combining non-linear transformation in the linear combination of explanatory variables.
Cleveland' (Cleveland, 1981)	LOcally WEighted Scatterplot Smoothing (LOWESS)	Non-parametric regression method that combines multiple classical regression methods in a KNN meta-model.
Drucker' (Drucker et al., 1997)	Support Vector Regression (SVR)	His method uses a max. threshold ε which stands for the residual between the target function and any of the training samples. It is used to establish a confidence hyperplane to define that function which contains all these training samples.
Breiman' (Breiman, 2001)	Random Forests (RF)	Random Forests model is a bagging-type ensemble method which employs decision tree induction where the split criteria is set using a randomly selected feature subset.

the ability to find out complex non-linear relationships from independent variables and the target one (Moreira-Matias et al., 2015).

According to Moreira-Matias et al. (2015) information displayed by advanced time system might reduce the high number of passenger at any bus stop for short-term travel time. That relates the use of the algorithm to predict in advance bus travel time at some bus stop and avoid unreliability of the transport system.

2.1 Technological Development on Bus Transport

Technology is everywhere since a small task that human beings want to do until big task, and provides tools, insights and knowledge. Beyond that, Intelligent Transport Service (ITS) is one of special system in referring communication technologies in addition to improving people's transportation mobility conditions and economic performance with transport improvements in Europe (Taylor, 2010). The ITS structure provides to mass transit companies the possibility to build mechanisms to support the transport system with GPS, AVL and smart card data. The perspective of improvements through the technology was considered new to this research area some years ago, however its emergence brought unprecedented advantage in the matter of public transport. The following contribution, on the European Union transport report Taylor (2010) explains some technology components that integrate ITS:

- Sensing technology: It consists a feed control system with infrastructure-based to gather information from a radar or cameras installed on a bus for example. Employing sensors have the same technology structure and components with this vehicle based data.

- Computational technologies: Inside the structure, model based process control and artificial intelligence are the two main reasons for the most expensive embedded technology.

- Long range and short range: Data exchange combined with several forms of wireless communication for the vehicles.

All of these technologies has as main objective to capture important data to be used as competitive advantage, or in a future research on the specific topic about public/private transportation. The facility of contribution on improvement of transport services with such technological devices inserted in a vehicle highlights how important services and technology must go together. The way these data sources has enriched vehicles operational planning and control to get more automated with the help of computational learning algorithms (Moreira-Matias, 2014). Delgado et al. (2012) also collaborate stating that GPS and AVL are the new information technologies to cooperate with such complex transportation system and holding control schemes. Automatic Passenger Counter (APC) system also has cooperated with such data from public transport and helped improve the service reliability nowadays. It seems to be more decisive for operational structure to have these transport companies providing this type of data to better serve their customers.

The GPS technology has been important since its discovering in the 20th century to drives and companies. They make use or produce this system, not only for privates vehicles where it can be utilized for transport companies which are interested in storing and accessing data through a device. Companies are also able to keep themselves informed about their vehicle fleets in a correct vehicle position. They can also locate their vehicles in real-time and have some other future perceptions about their business with these data (Moreira-Matias, 2014). Figure 2.1 shows the operation and implementation of the GPS data collection system.

With this technological implementation, the process of capturing all the data is performed in order to generate the GPS data to be transmitted by an AVL system, informing the position of the vehicle. In this case, a bus, as we can see in the image. Another antenna sends data to be stored in a data server through the internet and available on the control front-end and personal devices.

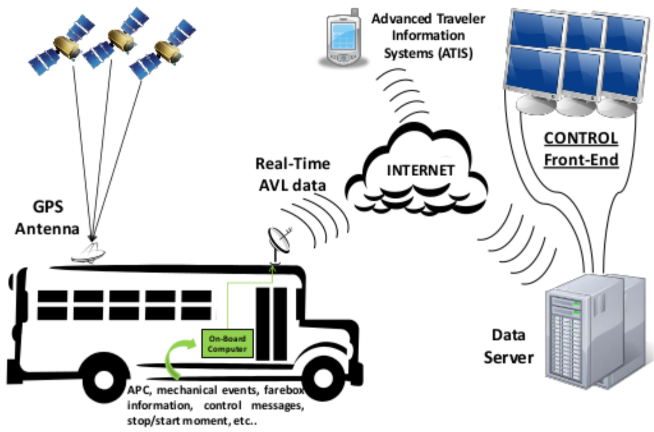


Figure 2.1: GPS and Automated Data Collection System image from (Moreira-Matias, 2014)

The technological development passes in collecting data in order to provide for people better service along with the discovering of new and advanced techniques. It is important to understand how automated tools have provided a good comprehension about service planning and operations management and avoided creating operational problems, in addition to developing a prior correctness strategy when problems come up (Cham, 2006).

Next section revises in more detail the reasons and importance of the use of AVL and smart card system as baseline in the dissertation and main system for capturing our data.

2.2 AVL and Smart Card System Cooperation for Bus Travel Time

This section describes the most important technologies (Automatic Vehicle Location and smart card) for supporting the realization of this study through the provision of bus travel time data, passengers boarding and alighting data. Cooperating with the need to establish strong operational structure to minimize the problems relating to the public transport service, as being necessary nowadays. To have good technologies behind, which support for the operational structure of a company a better way to visualize any inconsistency or future problems related to operations. In order to achieve problem-resolution as the substantial and large data coming from these parametric systems, there was much more what to do in regarding to travel time improvements for public transport.

What else relevant to say about these technologies is related to bus travel time prediction? You might have asked this question and there is much related explanation for their importance in bus travel time. Some of them are explained by researchers, regarding the efficiency of mobility system and enhancing travel times. Before collaborating in details with some of the importance of two transport technological systems used in this dissertation, we focus on the technical characteristic between AVL and smart card system. Moreira-Matias et al. (2015) explain that AVL and smart card are bus dispatch systems, while GPS is an AVL-based measurements system. Smart card system has an estimation-based technique about weight sensors where companies have been introduced in their bus fleets. The authors also contributes saying that these automated systems store bus location data with a type of sensors broadcasting values in a interval of 10 to 30 seconds, depending on which is the capacity rate for collecting data on these systems.

The smart card system consists in capturing data from a card which has in its part a chip and it is capable of processing, collecting and storing data or a memory chip which is capable of reading data from these cards, available with smart card system (Bagchi and White, 2004).

All information kept in an automated process combines smart card and the reading machine configured to read and collect data from these cards. Figure 2.2 shows us the smart card transaction data integration and passenger information flow since the user whose the card is validated by a reading device until the data storage.

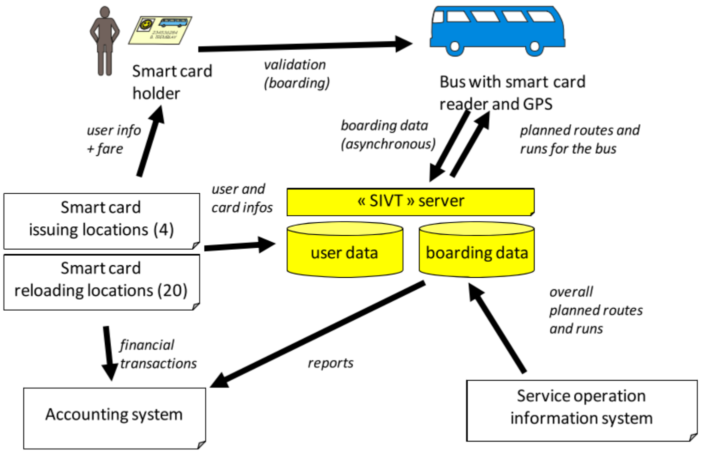


Figure 2.2: The operation of a Smart Card Information System (Pelletier et al., 2009)

It is important to ensure the data are stored in a secure way, preserving all passengers' data and reliability of the system. The users' validation data are not kept in the same database, the reason is to guarantee the security of all information traffic come through the system (Pelletier et al., 2009).

Bagchi and White (2004) also provide us some information about the overall collecting of data through smart cards:

- Linking data: It is possible to link a card to a person in a way of membership to relate most of travel done by each person who has this personalised card, and facilitate trend travel issues.

- Volume and scope of the data: Transport companies have to handle with large amount of data which comes from automated system. This happens because of the volume of people who have signed up for smart cards and a little data is lost by the amount of people who did not pay their fare with smart cards.

- Continuous information: This type of collection data collects information in a continuous way. These data will be collected in long periods of time that undertakes getting trip data from individuals. It provides more accurate long-term analysis and forecasting with most trend day and time used by passengers on a travel.

The authors explain that with smart cards, the need to make recording start from the human role whom a person presents his card to the reading machine, it initializes the pre-defined collecting of information variables, and always considering how many adepts to the modern transaction are using it.

AVL and APC also called automated data collection are mechanisms linked by devices to operate in a vehicle while in movement broadcasting localization, time, speed, passenger boarding and alighting. As in the previous section, we intend to describe the technological advances nowadays in using such AVL and SC systems. Thus, Cham (2006) contributes with the understanding about the relevance of smart card system which is also used for off-line analysis as the use of evaluating performance while APC is used for real-time applications. A first hierarchy level of the automated technological system is described in table 2.2:

According to Furth et al. (2003) archived data is composed by four key dimensions and supported by automated systems, as shown in table 2.2. These key dimensions are: fleet penetration and sample size, complete vs. exception data, level of spatial and temporal detail and data quality control.

In general to agree about the opportunity to use automated data collection system can do statistically valid analyses for the first moment on the reliability of the service (Furth, 2000; Kimpel et al., 2004). The cooperation between AVL and smart card system brought the understanding to manage bus travel time and its particularities in addition to comprehending the predictability of the service. To have advantage it is necessary for all these data to be built a framework with the use of machine learning models, where a prior decision should be made to relate uncertainties and to collaborate in improving many perspectives of transport service such as operations, performance monitoring, scheduling and planning.

Table 2.2: Spatial and Temporal Detail for Automated Data (Furth et al., 2003)

Level	**Description**	**Event-Independent**	**Event Records**	**Between-Stop Performance**
A	AVL without real-time tracking	Infrequent (typically 60 to 120s)	-	-
B	AVL with real-time tracking	Infrequent (typically 60 to 120s)	Each time point	-
C	APC or event recorder	-	Each time point	-
D	Event recorder with between-stop summaries	-	Each stop and between-stop events	Recorder events and summaries
E	Event-recorder / trip recorder	Very frequent (every second)	All types	All events, full speed profile

2.3 Leading to the Bus Service Reliability

As one of the main concerning of this study, the reliability of the public transport has a fundamental role for transit companies that provides for their customer, thus passengers are always expecting at least a good service supplied by transport companies. Levinson (1991) explains the meaning of providing reliable service as for example to be on the right schedule every time, to low the variance of high passenger loads and to keep uniform headways. Therefore, this implies offering reliable services, compared to their schedule time running according to what was proposed in the schedule, which means without delays or much anticipation. The transit industry considers buses on travel time schedule, when they arrive or depart from a time point within one minute window early or five minutes in a late schedule (Bates, 1986).

It is considerable some service adjustment of schedule, although reliability does not mean a bus to be on time all times, but to be on a right schedule for most of times. Unreliable transport service has the influence of affecting passengers into dissatisfaction with the system. Some problems have relation to wait time increase and travel time uncertainty (Strathman et al., 1998). The most relevant problems which affect bus variability and its travel time are related to congestion roads, traffic signals and passenger demand at stops, so traffic conditions are the most serious cause of low reliability (Ma et al., 2014).

When time variation is found in a balanced interval of time for a bus on a specific route, this might be considered a delay, some seconds or until one minute of adjustment because of some external factors that might affect a bus schedule plan. As we are focused on travel time and passenger demand, Kimpel (2001) describes, on his research work about transit service reliability and passenger demand a type of hypothetical relationship between demand and service quality for route level analysis which consists of:

$$D_r = f(SQ_r, SR_r, ...X_r) \tag{2.4}$$

$$SQ_r = f(D_r, SR_r, ...Y_r) \tag{2.5}$$

$$SR_r = f(D_r, SQ_r, ...Z_r) \tag{2.6}$$

These functions show that (D_r) represents the number of passenger boardings related to a specified route, and (SQ_r) is represented by the quantity of service provided on a route, while (SR_r) is the measure of service reliability, (X_r) is the vector to explain passenger demand, (Y_r) is another route vector to explain service quantity and (Z_r) is another route vector to explain transit service reliability of this route (Kimpel, 2001).

Some practices are considered to be made, in order to avoid problems related to public transport. These practices correspond to the bus travel time reliability. In a scenario that passengers tend to get to plan their way and time to go to a specific location by bus. They are given information through applications about next bus coming in advance of a real time bus. The bus travel time along of a route is more comfortable when given information about times and passengers get more satisfied with the service and can arrange their time in a schedule for their activity in a more planned way. According to Kimpel (2001), the reduction of the variability of bus arrival times and the decrease of average passenger wait times are consequences of an improved schedule adherence at bus stops providing benefits for passenger's schedule plan. The author also states that this strategy gives more regularity, reducing bus bunching and efficiently ensure that buses are being used in their capacity.

In an opposite scenario where buses are running in a short headways and random arrivals define characteristics of heavy demand for transport service and under these conditions, buses will be running loaded of passengers on the other hand a bus behind, trailing the loaded bus will be in a light loads and boardings causing bus bunching problems (Strathman et al., 1998).

In the Transit Capacity and Quality of Service Manual (3rd Edition) is described a number of factors which affect bus service reliability. In table 2.3 we can see a description of each one considering its factor on service to have an overview about the most caused problem on service reliability.

Offering a more reliable transport service helps to improve conditions for passenger of public transportation system which concerns to predict travel time and consistent availability of seats and space on buses for passengers to feel comfortable to board, so this is a measure of quality of the public transport service under the right conditions (Bates et al., 2001; Brownstone and Small,

Table 2.3: Some Factor Affecting the Reliability of Bus Service (Cham, 2006)

Factor	**Description**
Traffic conditions	For on-street, mixed-traffic operations, it includes traffic congestion, signal delays, parking, incidents, etc.
Road construction and track maintenance	Creates delays and may force detours
Vehicle maintenance quality	Influences the probability of breakdowns
Vehicle and staff availability	Involves the availability of vehicles and operations to operate scheduled trips.
Transit preferential treatment	Includes exclusive bus lanes and conditional traffic signal priority.
Schedule achievable	Reflects ability to operate under normal conditions and loads with sufficient recovery times to allow most trips to depart on-time.
Evenness of passenger demand	Describes loads between successive buses and from day-to-day.
Differences in operator driving skills	Involves route familiarity and schedule adherence (particularly in terms of early running).
Wheelchair ramp and ramp usage	Includes frequency of deployment and amount of time required.
Route length and number of stops	Relates to the exposure to events that may delay a vehicle.
Operations control strategies	Application of actions to counteract reliability problems as they develop.

2005; Golob et al., 1972; Prashker, 1979). Carrel et al. (2013) contributes saying that an improved reliability service can keep low passengers' negative experiences due to both long waiting times at the bus stop or not be able to get on a bus because of excess of passengers.

We can consider two perspectives in relation to the prediction of bus arrival time, one from the operator's side of the system and other from user's. Although, the relevance of both to give the understanding about its importance and relevance, each one will be considered it differently. Operators tend to be more concerned in predicting reliably vehicles route, while passengers are interested in predicting the remaining waiting time at a particular bus stop (Oshyani and Cats, 2014).

To make an effort to ensure conditions in order to provide passenger a more reliable service it means always propose improvements according what they are more sensitive to be dissatisfied with how this service is offered, combine improvement factors which might have influence on travel time prediction and assess the implementation in a short period of time.

In the next section we describe the two most used techniques for travel time prediction.

2.4 Approaches on Bus Travel Time Prediction

The contribution on travel time prediction has highlighted for years of research through studies where the practical evidences have particularly made changes in the transport system. Approaches and models have been applied during the implementation on the purpose to figure it out how travel time could get more structural, providing the correct model for different approaches. Overall, travel time prediction is an effective way to get future travel times on a trip where passengers will have knowledge about the bus schedule plan in an ahead trip.

The characterization of some linear and no-linear models can be used and verified by solving travel time prediction with its variable parameter, although the most used model for this purpose is the linear one. Some of these models will help us to understand the problem and solve them as part of understanding of the problem that verifies linear or non-linear functions as for example Artificial Neural Networks (ANNs). For regression models that are considered as conventional approaches with dependent variables characterizing its main function (Padmanaban et al., 2009).

The two other methods and models are related to travel time prediction analysis which are data-driven methods and model-based approaches. For data-driven methods are employed historical travel time, beyond the use of related variable such as occupancy, speed, day of week and time of the day (Zou et al., 2014; Zeng and Zhang, 2013; Zhang and Haghani, 2015). The model-based technique uses different model for capturing the dynamics of system through settling a mathematical relationship with variables (Kumar et al., 2014). Other used approaches are related to historical and real time information. The travel time prediction for historical approach involves the average travel time at a specified period of time and it is also the same period over different days, then on the other hand, real time approach is represented by future time in an interval travel time as being the same as the current travel time over next days (Padmanaban et al., 2009).

Li et al. (2011) help us to better understand the facts of public bus arrival time with figure 2.3, where the bus starts its trip from a specific bus stop until its final destination and shows the entire process as a trip.

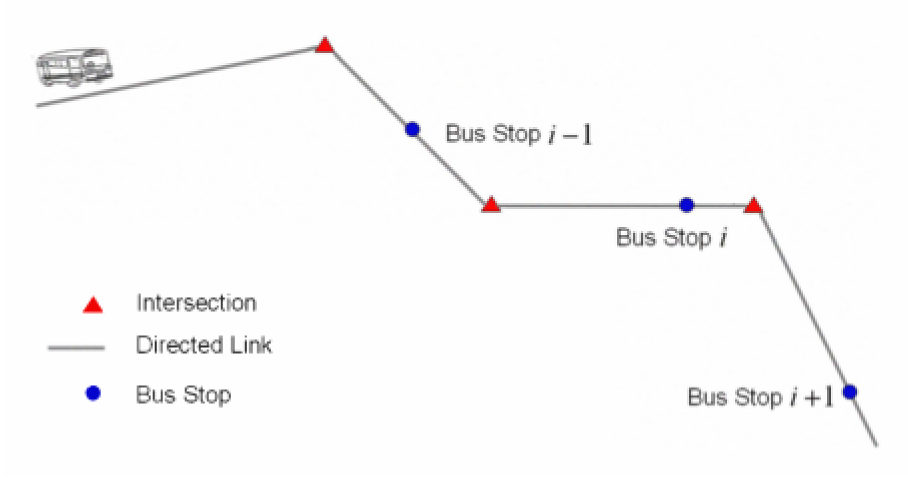


Figure 2.3: Bus Travel Route (Li et al., 2011)

Some researchers have developed some algorithms to provide real-time bus arrival information through a historical approach where this information might be displayed for passenger at bus stop electronic boards. Lin and Zeng (1999) were two of the researchers who used a historical approach to analyse bus travel time prediction. They used bus location data to assess what was the difference waiting time at real time stops, scheduled arrival time and current arrival times.

Other two researchers Shalaby and Farhan (2004), who also used bus arrival time prediction with automated data from AVL and APC. The authors developed two consistent Kalman filter model, which the first one predicts travel time and the second one predicts the dwell time, considering the number of passengers boarding and alighting the bus at each bus stop as a function of their approach.

Another approach developed by Moreira-Matias et al. (2016) intended to eliminate bus bunching for public transport in Porto, in real time, through an on-line learning proactive approach. This predictive method consists in using an automatic control framework to mitigate bus bunching and one of the causes of bus service unreliability. This kind of problem causes infrequency of buses on a route and unbalanced distribution along the bus schedule plan. The authors also present an effective action to automatically prevent bus bunching using machine learning techniques and methods to get valuable information from location based data.

In this dissertation, the proposed analytical approach will be employed with some machine learning (ML) models for our framework to evaluate the influence of the number of passengers on the bus arrival time on each bus route and bus travel time. The advantage of this approach, the short-term forecast, is to quantify input variables to get immediate bus travel time prediction. The combination of effort to try to figure it out how predictive travel time is, in the sequence of BSP provides us more certainty equivalent when accurate algorithms are applied in order to combine efficiency and accuracy to help relate the problem.

Chapter 3

Machine Learning Models and Approaches in Travel Time Prediction

A Review

During years, the combination of analysis through Machine Learning (ML) models and algorithms has been important to achieve expressive results on a variety of problems. A ML model is structured as part as the predictive analysis process and implementation related to data and a specified problem. ML concept is formulated on the idea that we build analytical models with automated software and with ability to learn interactively without periodic configuration or tuning it (Jou, 2017).

Some advantages in using ML methods, in a process of statistical analysis, are: the way of dealing with certain complex relationships between predictors in a large volume of data, the process of how each predictive model treats non-linear relationships with predictors and how it will treat noisy and complex data (Recknagel, 2001).

Machine learning has powerful techniques in relating predictive analysis, and Artificial intelligence (AI) is another study area frequently correlated to important methods allowing the application of prediction for travel time which is part of ITS. AI, ML and statistics were applied, in the Dahl et al. (2014) research work, in order to be categorized as search, optimization, classification or a combination of the methods. In the authors research work, they also state that the applicability of AI was given from the beginning by searching and optimizing car navigation to help drivers find their way. The potential of using ANN during identification, implementation and collection in complex and large amount of data is to be explained and investigated in more details in the next sections. Beyond ANN, others predictive models will be discussed such as Random Forests model, Linear regression model and Support Vector Regression that will be compared in our data set.

3.1 Predictive Methods and Analyses

Conducting predictive analysis of travel time is also a way to understand correlated approaches, which is performed in a set of characteristics through implementation of statistical methods. It can be useful to provide accurate information about buses travel time prediction, as mentioned on this dissertation. So historical approach is one of the based-models considering the structure of the database, by making use of historical data and a statistical analysis of travel time average by taking in consideration one bus route. Our approach focus is on the use of historical data, perceiving the output generated through some important features that will be relevant to consider as it is the verification analysis of passenger amount along bus route. To provide a fully understanding and some prior knowledge about ML, it is advisable to cover on the matter of learning methods and definitions which are explained below:

- Supervised Learning: The analysis is to predict the output of some component and these components give us the correct available feedback in a certain way, which confirms or provides the intuition about what correct output is. Beyond the prediction of a certain action which will provide certain output with the perception of description of the correct output in an algorithm (Russell and Norvig, 2013).

- Unsupervised Learning: It represents the lack of knowledge about the output. Reminding that an unsupervised learner might have relationships inside its perception in a supervised method. This ability to learn can predict future perception with previous percepts, but it cannot learn without a previous useful function and some restrictions (Russell and Norvig, 2013).

- Off-line Learning: This is a method by applying in a static dataset and it does not consider how dynamic the status and tracking of the target is on the data beyond their correlation (Jou, 2017).

- On-line Learning: This method is supposed to be applied in a live dataset, which is able to learn from the dataset and the use features considering the status and tracking of the data target. To be able to analyse this real data even with an enhanced ML models, in on-line learning is a difficult task because at the same time the data is coming in to be processed in a smart way and run with these on-line algorithms (Jou, 2017).

- Incremental Learning: A learning algorithm has an incremental method when any training data sample contains $e_1 .... e_n$, and it generates a list of hypotheses values $h_0 .... h_n$, so $h_i + 1$ will depend on h_i for the current e_i (Giraud-Carrier, 2000)

- Real-Time Learning: This is a somewhat similar method to on-line, however this method has the ability to process the data and use their features in real time with a up-to-date model while the previous sample is processed and the next one is arriving (Huang et al., 2006).

Two other consistent methods (Analytical and Inductive) can be applied for travel time prediction. These methods are the baseline intuition of better understanding different formulations of learning problems, taking into account some generalizations. Identifying features from observed training samples where exist distinction of positive and negative training sample is more applicable to the inductive method (Mitchell, 1997). A more detailed explanation about these two methods is demonstrated by Mitchell (1997), in which is possible to observe the formulation of difference between them:

- Inductive learning: A hypothesis space H is given by the learner, and that shall select a hypothesis as an output, generating a set of training samples $D = \langle (x_1, f(x_1)\rangle, ..., \langle x_n, f(x_n)\rangle)$ where $f(x_i)$ represents the target value of instance x_i. A consistent training sample is the output of the learner.

- Analytical learning: The same happens to the inductive learning where the hypothesis space H and training samples D are included with the learner input, but for the analytical learning the provided input generated by the a learner it is called domain theory B, which has intrinsically the background knowledge to be explained by observed training samples. For the consistent training samples D and domain theory B in which have a desired output learner and have hypothesis from h to H.

Mendes-Moreira (2008) contributes stating, in his research work, that the use of inductive learning techniques for prediction aims to learn a model from the data set, where this data set has V as input variables. It is only one output variable and through this model we are able to predict with new data frame.

Some other approaches, such as historical data and real time might be used in the implementation process, depending on your aim to assess bus travel time prediction which correspond to analyses processed in the dataset and following the admissibility of features data as well. In the next section, it will be addressed the importance of predictive models in this work, such as examples of ML models which should be used during the algorithm development explanation.

3.2 The Relevancy of Predictive Models

It was said at the beginning of the chapter 3 that is addressed the importance of predictive model, so we will detail their relevance in this dissertation, in order to present the most used models in this area of research. However, it will be necessary to compose a complete explanation for each model relating to the objective and the proposed analysis.

ML has two stages, one is the process of choosing a candidate model among those that might be applied to solve predictive problem, the other one is related to the parameters of the model where this learning process implemented in a data structure (Jin and Sendhoff, 2008). What is also important to say about these models is that they can be used in travel time prediction, without previous analysis or during the implementation process to address the traffic data (Altinkaya and Zontul, 2013).

The algorithms infer different ML models that can be useful to make sense of how relevant the information is. Who help us to understand this building process of algorithm learning and the use of relevance-based learning are Russell and Norvig (2013). The authors suggest an algorithm to attempt to find a simpler and consistent determination through observations where this determination P $\succ$ Q, so correlations among examples between P and Q.

An example of algorithm is shown in figure 3.1. It is an example of how to find the minimal consistent determination to represent a basic algorithm with a group of attributes (Russell and Norvig, 2013).

```
function MINIMAL-CONSISTENT-DET(E, A) returns a set of attributes
  inputs: E, a set of examples
          A, a set of attributes, of size n

  for i = 0 to n do
     for each subset A_i of A of size i do
        if CONSISTENT-DET?(A_i, E) then return A_i
---------------------------------------------------------------
function CONSISTENT-DET?(A, E) returns a truth value
  inputs: A, a set of attributes
          E, a set of examples
  local variables: H, a hash table

  for each example e in E do
     if some example in H has the same values as e for the attributes A
       but a different classification then return false
     store the class of e in H, indexed by the values for attributes A of the example e
  return true
```

Figure 3.1: Minimal consistent determination algorithm (Russell and Norvig, 2013)

ML provides the idea of looking for possible and alternative forms of assumptions in the data analysis development process, nevertheless the involvement includes getting insights from our training and test data, which makes to review any possible constraints might create any noise into data analysis. So the concept of ML learning as the research of parameters is described as the task through a large space of assumptions and defined by the representation of hypothesis, notwithstanding it is relevant to consider that the selection of hypothesis representation should be defined in the algorithm and all hypotheses space learned and developed on a computer program (Mitchell, 1997).

In different fields of research, ML predictive models have an important role by which they can be applied in different approaches, particularly for travel time prediction its models have been considered valuable in predicting travelling time of buses given by complex non-linear data relationship. It will be addressed, in the following sections, some relevant ML models, such as historical model, regression model, including artificial neural network model, support vector regression model, linear model, Kalman filtering-based model, and dynamic model, as a literature review and they have been applied to bus travel time prediction by researchers with successful results nowadays.

3.3 The Cross Validation Method

The cross validation is a method to validate a predictor which does not require any assumption. There are two process in using cross validation: the first one is used to learn or train the model and the second one is used to validate the model. This method is also used to evaluate and compare ML algorithm which is built every time using K-1 folds. In the remaining fold which is used as test set, and then the process is repeated K times, always considering the out-of-sample error as the mean squared prediction errors with k different folds (Viviano, 2016). Its validation technique has been used in different models in order to evaluate the results of a statistical analysis process that integrates in a data set and in the training step.

The two existing types of cross validation are methods used to implement the cross validation into data set, the first one is called exhaustive techniques and the second one is non-exhaustive techniques. The exhaustive technique is related to split the original data set into different training and validation samples besides learning and testing all possible formats, which consists of the leave-one-out method. The non-exhaustive technique is the opposite of the other one when not compute all the formats of splitting the original data set, which consists of the k-fold cross validation and holdout methods.

The leave one out cross validation consists in getting k=n, where n takes the value of the number of observations which means that n separate times. The holdout cross validation method has no benefits to use all the available data, if the use of half the data for the test set so the training set is built on half the data and there might get weak hypothesis but the opposite for the test set with 10% of the data, then a statistical change from the actual accuracy gets poor (Russell and Norvig, 2013).

The k-fold cross validation is used to get an accurate estimation using its technique which basically has the idea of each example shows double duty for the training data and test data. The first step is to split data into k equal subsets to perform k round of learning, thus the average test set score of the k rounds gets a better estimate rather than a single score (Russell and Norvig, 2013).

The k-fold cross validation has been proposed by researchers in the matter of predicting travel time based on historical data and it was used to partition data randomly into subsets, however Dahl et al. (2014) noticed to the use of this type of techniques in this analysis should be avoided because of the only travel time to forecast is the future.

Cross validation and sliding windows method are the two important methods to be validated with predictive models as described on this dissertation. In the next section we will explain about the sliding window method.

3.4 The Sliding Window Method

Sliding window is another important method used with predictive models that guarantees the use of the most recent data, it remains active in order to get the current ones. The two methods are equally important in the sliding windows technique, which is the fixed window and time-stamp based window. The fixed window contains data items arriving at a time for the most recent "n" items remain active for a fixed parameter "n". The times-tamp-based window consists in storing many data items arrive in bursts at a single step and for this technique the last "t" steps gets active for a fixed window parameter "t" (Braverman et al., 2012).

Contextualizing the sliding window technique is an effective method for stream data set where we can choose the most attractive part of data, which can vary with the proposed analysis and the data (i.e. in a financial data frame an assessment using sliding window can vary depending on the month of the year where the seasonality of the data is more biased than the other months). This might not have any generalization from your incoming data set and it still can match with your aim to find out and guarantee what are the most important and recent data.

Zhou et al. (2016) proposed in their research work, a sliding window ensemble framework which the aim was to determine the performance of the model used to predict passenger demand on bus services and estimate accuracy using time series forecasting error metric. The authors explain that the sliding window ensemble framework as being the use of three distinct predictive methods which learns from long, medium and short-term historical data. The combination of the three models based on the weighted ensemble where is applied to improve the prediction using a sliding window method. They also applied time-varying Poisson averaged models to be updated every 24 hours and a sliding window of 4 hours. The authors concluded that their implementation of sliding window ensemble was the best model in every shift and period with high accuracy of 78%.

The contribution for another research was described by Ferreira and Ruano (2009) in their research work about online sliding window methods. They provided a method to adapt an online sliding window and model parameters to be adjusted when new information is added, so the main idea is to find a termination criterion at every moment a new input and output standard data is appended and create a need to carry out a number of interactions in the training method and meet the termination criteria.

It is notable that the sliding window method has been implemented in different forms and models and its importance has been registered in many research works. It is described how sliding window is relevant in our bus travel time prediction analysis and its implementation steps in this dissertation.

3.5 Linear Regression Model

Linear regression model can be used to solve a regression problem where the dependency is known or assumed to be a linear function. Its efficient form of calculating optimal parameters with accuracy by using such as gradient descent as precise computing calculation (Russell and Norvig, 2013). It has been the models employed by Bin et al. (2006) to solve bus arrival time prediction.

In a straightforward definition we begin split linear regression in two analysis components, the first one is addressed to univariate linear regression and the second one is the multivariate linear regression. For the univariate one is defined with input x and output y which has the computation of $y = \omega_1 x + \omega_0$, where ω_0 and ω_1 are real coefficients to be used as learner (Russell and Norvig, 2013). The following formula defines w as a vector of $[\omega_0, \omega_1]$:

$$h_w = \omega_x + \omega_0. \tag{3.1}$$

In the process of finding the h_w that better fits a line to our data and gives the values of weights$[\omega_0, \omega_1]$ and minimizes the loss which is described by Russell and Norvig (2013) below:

$$Loss(h_w) = \sum_{j=1}^{N} L_2(y_j, h_w)) = \sum_{j=1}^{N} (y_j - h_w(x_j))^2 = \sum_{j_1}^{N} (y_j - (\omega_1 x_j + \omega_0))^2. \tag{3.2}$$

If we want to minimize finding the value of $w^* = argmin_w Loos(h_w)$, so we must sum the function until its partial derivative gets zero along with its respective ω_0 and ω_1:

$$\sum_{j=1}^{N} (y_j - (\omega_1 x_1 + \omega_0))^2 \tag{3.3}$$

$$\frac{\partial d}{\partial d\omega_0} \sum_{j=1}^{N} (y_j - (\omega_1 x_j + \omega_0))^2 and \sum_{j_1}^{N} (y_j - (\omega_1 x_j + \omega_0))^2 = 0 \tag{3.4}$$

It is important to consider that for linear regression models it might vary from different formulated function depending on the proposed problem solution and the form that its function might be addressed. Some functions help us to address the problem in a efficient way and build the resolution close to the optimal point. Below it is described by Russell and Norvig (2013) some of these function which make part of learning regression approach.

- Gradient descent is a function to minimize the loss, choosing a starting point in weight space (ω_0, ω_1) to move downward to find another point and this process is repeated until it converges to a minimum loss.

- Batch gradient descent is another function inside the univariate linear regression model, and its convergence must be achieved in only one global minimum, but α must be chosen in a really small value. Many steps will guarantee the cycle for training data and its each steps.

- Stochastic gradient descent is considered a single training point for once, it can be used for on-line or off-line setting. The cycle steps in the data are taking as many times as is needed. This function is faster than batch gradient, but even with a fixed α its convergence is not guaranteed.

Now, multivariate linear function is an extension of univariate linear regression where each example x_j is a vector element of n. So the function is represented by the hypothesis space form, described by authors Russell and Norvig (2013).

$$h_{s_\omega}(x_j) = \omega_0 + \omega_1 x_j, 1 + \ldots + \omega_n x_j, n = \omega_0 + \sum_i \omega_i x_j, i. \tag{3.5}$$

The representation of the best vector of weights w* can minimize squared-error loss with the following function:

$$w^* = argmin_w \sum_j L_2(y_j, w.x_j). \tag{3.6}$$

Multivariate linear models have been the focus for some research work in travel time prediction (Strathman et al., 1998; Yetiskul and Senbil, 2012) through different measures of services variability, as well as the amount of scheduled stops and variation of passenger activities.

We present a more complex than linear regression model in section 3.6. The Support Vector Regression model is frequently used by researchers in this research area.

3.6 Support Vector Regression Model

Support Vector Regression (SVR) model is another version of Support Vector Machine (SVM) which stands for regression problems while SVM stands for classification problems and their differences, in terms of calculation, are a few. These differences are noticed for the use of slack variable in which for SVM is used only one time for each training data while SVR uses two slack variable for each training data. SVM has the ability to implement a structural risk minimization inductive rule through learning machine to gain generalization with a number of learning patterns, so this number of observations must be limited to get good generalization (Basak et al., 2007).

The main practical use of SVR is related to the choice of function types as radial, kernel or loss function and the corresponding parameters besides establishing the model with its respective problem and SVR algorithm (Wang et al., 2009).

SVR is a powerful model for travel time forecast research because of its structure, considering predictive values with the use of parameters (C-values and kernel) and results have been demonstrated the effectiveness and efficiency of model with studies throughout the years. Dahl et al. (2014) suggest, in their research work about performing regression model for traffic scenarios, two important decisions steps must be discussed in the implementation process of SVR, which are the choice of the kernel might be implemented and the parameters to be used with kernel function.

Kernel functions explain about the function variations inside the SVM model, each kernel solves a specified problem. With kernels the implementation of a problem solution avoids choosing a transformations when we have a combination of infinite transformations, and the SVM model tries not to impose a specific shape to the data set minimizing with a loss function (Viviano, 2016).

Table 3.1: Kernel functions (Wang et al., 2009)

Kernel	**Function**
Linear	$K(x,x') = x * y$
Polynomial	$K(x,x') = [(x * x') + c]^d$
Radial Basis Function	$K(x,x') = exp - \frac{\|\|x-x'\|\|}{2\sigma^2}$
Sigmoid	$K(x,x') = tanh[v(x * x')] + c]$

The authors Wu et al. (2004) applied SVR in travel time prediction and performed for traffic data analysis, considering the model feasible and well performed for both analyses including for time series data. The implementation of authors' SVR model consisted in two approaches to calculating point measurement and link measurement. For the second link was considered route travel time is calculated its metrics with these two points, also using test vehicles or license plate matching. For the point-measurement approach was a travel time estimation using traffic data measurement by considering point-detection devices on the highway. In the Figure 3.2 we can see a representation of the authors' travel time prediction problem.

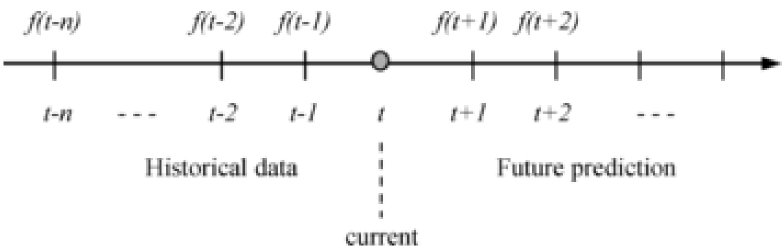


Figure 3.2: Travel time prediction assuming the current time from two points Wu et al. (2004)

Basically, the prediction of travel time has two different main approaches, which is the statistical models and the other one is data-driven methods and figure 3.2 represents statistical models characterized as data-driven using time series, speeds and volumes as input data. For example it is given historical travel time data f(t-1). f(t-2),..., and f(t-n) with t-1, t-2,...t-n, accordingly to data set it is possible to predict f(t+1), f(t+2),..., as the future values analysing historical data model (Wu et al., 2004).

The SVR approach which Wu et al. (2004) considered to solve travel time problem was the generic estimation function $f(x) = (\omega.\phi(x)) + b$, using a set of training data $\{(x1,y1),..,(x1,y1)\}$, so for each $xi \subset R^n$ which is the input space with the value $yi \subset \mathrm{R}$ for $i = 1,...,l$, and letter l is the size of the training data (Müller et al., 1997).

In the following section we will still see some of the common models implemented in travel time prediction, according to their baseline approach and specified implementation for each predictive model, one different problem was solved.

3.7 Kalman Filter Model

Kalman filter model was developed by Rudolf E. Kalman with the objective to be applied in continuous variables over time series [1]. The model is structured by the one of the Kalman filters' possible notation for the use of time series analysis describing Y as described below:

$$Y = \{Yt_1, Yt_2, ...\} \tag{3.7}$$

[1]Time series are data points listed in time sort or a sequence taken equally with certain point in time or date

Kalman filters have demonstrated a good analysis algorithm option for travel time prediction and researchers have given due importance through the potential algorithm to predict the future of dependent variables, moreover the recursive procedure updates its prediction whenever new observation are available to evaluate the data by minimizing the estimated error covariance (Chang et al., 2010). Shalaby and Farhan (2004) employed two Kalman filter algorithms for the prediction of bus running time along a specific route and dwell times variable in a data set composed of AVL and APC data. For this specified route an instant k+1, considering the first algorithm and the use of the last three days historical data of the bus route running time, and makes use of the previous bus on a current day with the instant k value. In the second algorithm employed by the authors, it is called Passenger Rate Prediction, and has been implemented in a similar historical data, but now they consider passenger arrival time.

According to Dahl et al. (2014), Kalman filter works well with observations presumed to be noisy measurements in addition to estimating future variables as traffic prediction, bus dwell time and arrival time. The authors also contribute that the Kalman filter method previous observations and estimations are calculated their future, so a new observation arrives, its averages will be used and with the updated estimate, the next estimations will be improved. The greatest importance of Kalman filter is the implementation of a predictor with the sort of correction of the estimation that is a great minimizer of the estimated error covariance.

The contribution of Kalman filter has been widely verified through much research, as it is presented in this dissertation, thereby bus travel time has granted with benefits with specific approaches that Kalman filter provides from deep studies in this research field. The two main general approaches have been developed with Kalman filter. They have been focused on using historical travel times data with estimation of trip time through the segment, current travel times provided by AVL system. The second approach is estimated by the bus travel time speed, travel distance data using historical speeds and real time speed data estimation (Gayah and Wood, 2016).

Despite Kalman filter is a well employed model for bus travel time prediction with many advantages in applying it for different problems, it has been successful in estimating travel time through working with uncertainty data, so this dissertation will be focused on other models in order to use three different algorithms which also worked well for this analysis.

3.8 Random Forests Model

Random Forests (RFs) model is characterized by being an ensemble[2] method, which also works for supervised or unsupervised learning. It is composed by techniques such as bootstrap and bagging, which are consistent techniques for better model performance. RFs have a type of combination of tree and each tree depends on the values of an independent random vector sample in the same distribution in the forest and for all trees (Breiman, 2001). The property of RFs in its analysis

[2]Ensemble method is represented by the combination of multiple models, however they are individual ones and RFs have a powerful prediction model.

process is that the increasing of value ntree makes the result converge to an optimal value (Mendes-Moreira, 2008).

According to Breiman (2001), RFs can improve the accuracy because of its randomness and minimization of correlation to gain strength while maintaining it. The author also says the randomness of input to be selected into RFs model consists of combining input values at each decision tree node for growing a better sample in terms of accuracy to compare with Adaboost model.

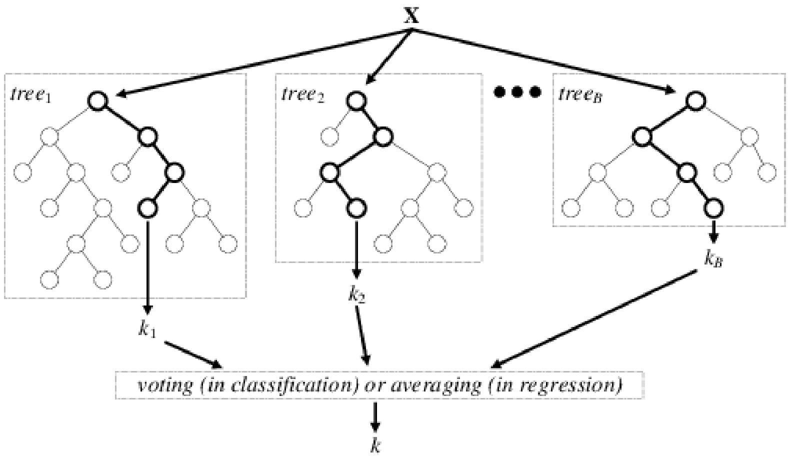


Figure 3.3: Random Forest decision tree structure (Verikas et al., 2016)

Figure 3.3 shows us the structure of the RFs decision tree for classification and regression. These decision trees take the x as data input for each tree and its dimension parameters, which will be treated as classification or regression. Also their data output are represented by k and after processing these data through the decision tree models.

The RFs approach for regression is designed by growing trees that depends on a random vector ϕ and for the tree predictor $h(x,\phi)$ gets numerical values, so these numerical output values are taken by assuming in a training data set to be independent for the distribution of the random vector Y and X (Breiman, 2001).

The RFs model has the decision tree method for building several trees, all with independent path to get a prediction and the average of the results by set of trees (Mendes-Moreira, 2008). Mendes-Moreira (2008) states that the bootstrap method and random feature selection is the implementation of the trees and it gets diversified from producing trees.

Elhenawy et al. (2014) applied RFs in their research work about "Random Forest Travel Time Prediction Algorithm using Spatio-temporal Speed Measurements" to model the relationship between predictors and their response corresponding to historical travel time. It was also used by the authors an algorithm, which aggregates the congestion probability matrix using spatio-temporal speed measurements to build feature vectors to be used as predictors.

As the RFs regression approach will be employed in chapter 4 of this dissertation, we will discuss in more detail about the model and the algorithm implementation to achieve our results, besides citing some other RFs approaches in the same research field.

3.9 Artificial Neural Networks Model

Artificial Neural Networks (ANNs) arose with the observation of the functioning of the biological system through its complex network structure of interconnected neurons. A simpler explanation is presented by Mitchell (1997) that ANNs are made up of an interconnected set of simple units, so each unit gets a number of real-valued inputs to produce a single real-valued outcome. In figure 3.4 it is possible to see how it is characterized by Feedforward Neural Networks' input data, neurons and layers.

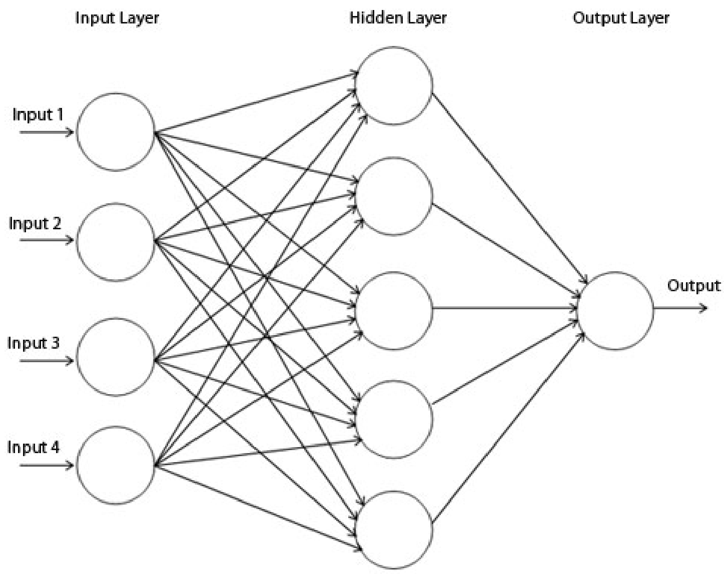


Figure 3.4: Aritificial Neural Network structure

As a complex and advanced model, ANNs provides a consistent and robust approach that has an approximating real-valued, discrete-valued and vector-valued target function, in addition to approaching certain types of problems, as learning to predict complex real-world sensor data , it is one of the most powerful and effective learning method for identifying errors in the training data sample (Mitchell, 1997).

One of the most used method in ANNs model is based on each unit called multilayer perceptron, which takes a vector of real-valued input to calculate a linear combination of these inputs, so outputs are 1 whether its results are greater than a threshold or it would be a negative value of -1 (Mitchell, 1997).

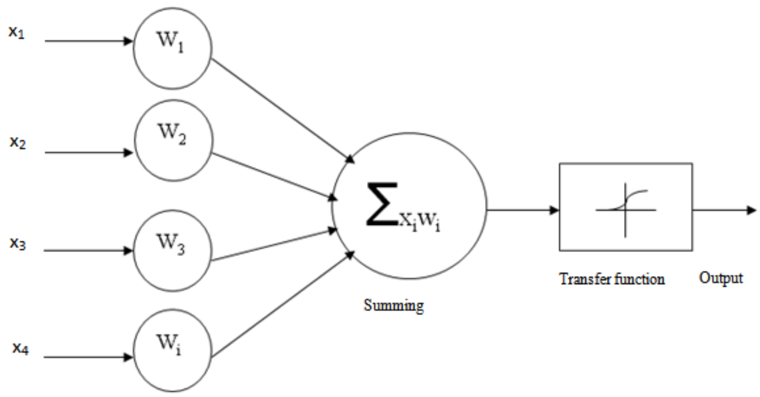


Figure 3.5: ANN Perceptron Structure (Kashid and Kumar, 2012)

In figure 3.5 are shown the inputs x1, x2, x3, x4 corresponding to their weights W1, W2, W3, W4 of the functions which relate all data into input and weight function (Kashid and Kumar, 2012).

Another method used by ANNs is called back-propagation which requires the activation function (linear or non-linear) to be differentiable, in addition to calculating gradient of the loss function related to their weights (Mitchell, 1997).

Mitchell (1997) presents the appropriate use of back-propagation method for the characteristics of the following problems:

- Instances are represented by many attribute-values. It starts by the definition of instances to be learned, the target function, as it can be described by a vector of predefined features. Input attributes might be highly correlated or even independent of one another for the specified characteristic. They also can be real values.

- The target function output might be discrete-valued, real-valued or a vector discrete attributes. These values from each output are real number from 0 to 1, for this case it represents the confidence to predict the corresponding steering direction. A single network can be trained to generate output for steering command and suggested acceleration by grouping the vectors, which encode the output predictions wanted.

- The training examples might have errors which ANNs learning methods are robust to reduce noise in the training data.

- Long training times are acceptable case because of network training methods usually require longer training times. These methods may take a few seconds to many hours it will depend on factors that might be introduced in the algorithm. Such factors might influence in the algorithm running time, which take into account the number of weights in the network, and the settings of several learning algorithm parameters and the number of training examples to be considered in an implementation analysis.

- Fast evaluation of the learned target function might be required, and ANNs learning times are long in assessing the learned network, however it relates the implementation for a following instance to be fast.
- The ability of humans to understand the learned target function is not so important, so the weights in the neural networks are difficult to comprehend for humans and learned rules are more communicative than learned neural networks.

Jeong and Rilett (2004) applied ANNs for bus arrival time prediction by integrating dwell time, schedule adherence and arrival time as input variables for each bus stop. They also used a number of hidden neurons through empirical analysis in order to train and learn functions. Their analysis using ANNs demonstrated to be much better than the implementation of model based for historical averaging and linear regression.

An enhanced ANNs model has been developed by Chien et al. (2002) which consists in predicting dynamic bus arrival time, so back-propagation was used in their research work. The motivation to apply the ANNs model was due to difficult learning process and application in an on-line mode. They also had to adjust the factor to modify the new input real-time data for travel time prediction.

Despite this explanation about ANNs, in chapter 4, ANNs will be introduced in the proposed ML models analysis and the results shown to achieve our proposed objective. We will approach one more time ANNs with our impression about the effective use of this model in this dissertation.

Chapter 4

Evaluation and Comparison of Predictive Models in the Proposed Analysis

We start this chapter by describing the structure of the dataset used to apply in our machine learning (ML) algorithms. The description of the datasets contain the attributes of Automated Vehicle Location (AVL) data and smart card (SC) data.

Afterwards, in section 4.2, we shall introduce how our datasets were prepared, explaining which software was used to apply them in a standard of ML algorithm for each predictive model.

In section 4.3, it was applied an average travel time method to better understanding our dataset in terms of travel time. In this case we use observations from AVL dataset for bus line 805.

The two important techniques applied in order to select importance variable and frame the datasets and for the importance of more recent data. These two techniques are permutation importance measure and sliding windows method, described in sections 4.4 and 4.5.

This description is used to address the ML models and algorithms (Random Forests, Artificial Neural Networks and Support Vector Regression) used to achieve our goals in this dissertation. Some of the models have already been described, in chapter 3, in a general way, but often addressing bus travel time prediction as common approach of these models and methods. Although we introduced some of these models as part of our current analysis, which seeks to find the best way to compare and evaluate the models with the datasets we have.

We start this chapter by highlighting some of the methodology used in this study, considering models, algorithms, methods and approach. The focus on building a better comprehension about the proposed analysis, it makes a great opportunity to provide through evaluation and comparison of predictive models inserted in the context of assessing the results, which is proposed in section 4.9, provide us the best results.

In general, this chapter is used to address the results according to predictive models in addition to explaining in detail how was the implementation method, and all the data preparation until the combined effort to describe the results of the analysis as clear as possible.

4.1 Structure of the Dataset

The dataset was supplied by the Portuguese transport company, Sociedade de Transportes Colectivos do Porto (STCP). It was possible for the study development whereas this database was performed as AVL data and smart card data to examine travel time behaviours, that was difficult to get assumptions before the implementation of the ticketing system on buses for transportation companies. This sort of data source facilitated the access of information and preparation during a bus travel and gave the possibility of analysing different approaches for ITS.

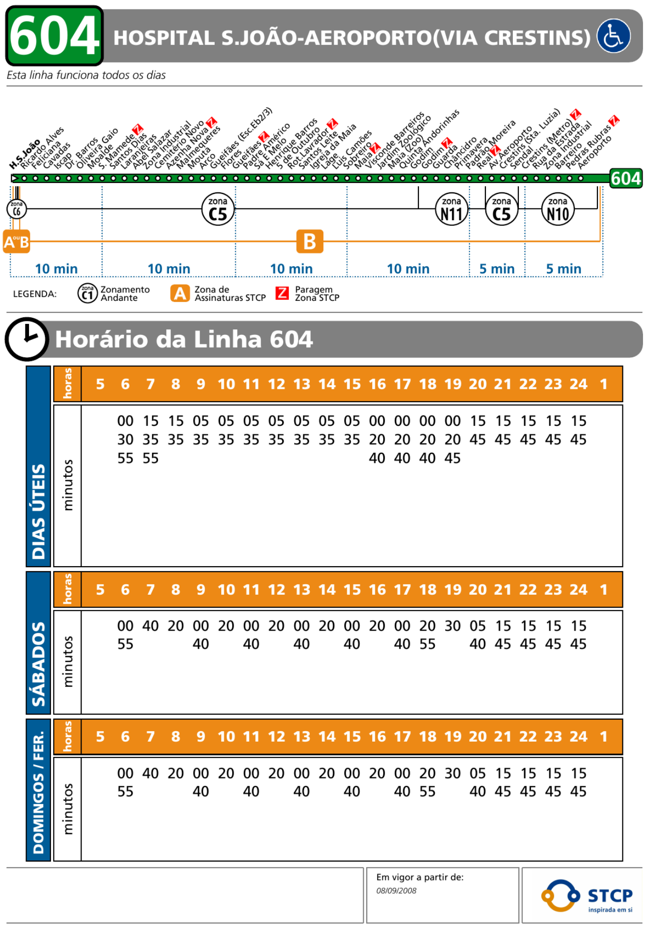
Concerning to the structure of the dataset, AVL is made up by variables (columns) and observations (rows) whether the variables are related to number of vehicle, travel date/time, number of bus line, number of shift, number of trip, number of stop order, direction and stop id. Some of these variables will not be relevant to the development of this dissertation. These variables can be described briefly as follows:

1. Vehicle number: The vehicle description number related to the bus it is a type of vehicle id.
2. Travel date/time: It is the date and time description related to each bus at each stop.
3. Bus line number: It reveals a line number in order to inform passengers which bus corresponds to its route.
4. Shift number: It corresponds to the bus shift during a travel day.
5. Trip number: It corresponds to the amount of journeys a bus does in one day.
6. Stop order number: It relates a specific number that corresponds to the bus stops order.
7. Direction: This attribute on a bus way is in order to provide us the direction, and it also corresponds to the number 1 or 2. The Number 1 is the way going and the number 2 is the way back.
8. Stop id: It is a terminology represented by an acronym that names all the bus stops with this name.

Each row presents us the action of bus lines through their routes, stop-to-stop, it also demonstrates that date and time from these actions are happening. The provided SC data show us almost the same attributes and features than the previous AVL data with some differences regarding number of passengers boarding and alighting at each bus stop and bus route. Due to the large amount of SC data, it was necessary to split it by bus line to consider a better first analysis of the structure of dataset.

The bus lines information contained in the AVL and SC data will help us understand, effectively the behaviour of each bus line on its route according to the travel time prediction and the effects on the bus schedule plan. The bus behaves according to its route and period of day, and number of passengers at each stop, so it will show us how travel time prediction could behave over time.

The data source describes information corresponding to one year (2010) of bus schedule. This (AVL) dataset contains a set of 8 bus lines, and they are 10M, 200, 201, 305, 401, 502, 600 and 805, that represents 10 percent of total lines managed by STCP. All data is structured with specified information for analysis in each determined bus line according to its route and schedule. That will be a good parameter to provide better estimation of future time. The figure 4.1 [1] is a time schedule structure for bus 604 that is possible to standardize and see the estimated time for this route at each stop.



horas	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	1
DIAS ÚTEIS (minutos)		00 30 55	15 35 55	15 35	05 35	05 35	05 35	05 35	05 35	05 35	05 35	00 20 40	00 20 40	00 20 40	00 20 45	15 45	15 45	15 45	15 45	15 45	
SÁBADOS (minutos)		00 55	40	20	00 40	20	00 40	20	00 40	20	00 40	20	00 40	20 55	30	05 40	15 45	15 45	15 45	15 45	
DOMINGOS / FER. (minutos)		00 55	40	20	00 40	20	00 40	20	00 40	20	00 40	20	00 40	20 55	30	05 40	15 45	15 45	15 45	15 45	

Figure 4.1: Estimated time information for line 604

[1] The figure was downloaded from the web-page http://s1.livrozilla.com/store/data/001715759_1-1701250b2ebcb5bee561664f253c5845.png

From this bus schedule plan, it is a simple board to show passengers an approximated time to be fulfilled by bus drivers. Sure, it is not the best accurate way to provide us with a travel time prediction, however it is another way to do it. This image is a representation of a timetable example used on Porto's road at a specified bus stop. There is one of this bus timetable at each bus stop in the Porto city as estimation of time for the specified stop, moreover it is possible to see another estimated time along the bus route, for example how much time a bus takes to go from each stop of its route on travel time estimation.

4.2 Data Preparation

The data preparation is one of the first steps for machine learning algorithm in order to set up the pre-analysis on a dataset before implementing algorithm. Understanding this dataset helps us formulate target questions, which will lead to the most desired and structured data, following some conductive experiments. The preparation of the dataset shows how our problem can be addressed, in addition to using the most effective predictive ML models to get a great solution.

Using R statistical programming language was processed the data preparation and used for the implementation of the predictive algorithms in our datasets. The process of getting data ready for an ML algorithm is part of three steps: selection, processing and transformation (Brownlee, 2013).

1. Selecting data is concerned with the choice of the subset of all available data that we want to work. In this step is also important to make assumptions about the data needed to solve our problem.

2. Processing data concerns the form that we can work with these data. There are other steps inside the data processing, which are formatting, cleaning and sampling.

3. Transforming data is a final step and more related to the algorithms that we are working with, because of their peculiarity in analysing the dataset according to our problem domain and knowledge. There are three common data transformations to be used in this step, which are scaling, decomposition and aggregation.

We work with two datasets, as described in section 4.1, AVL and smart card contain information about bus lines 600 and 805, so these are the bus routes to be used and compared in this dissertation, according to the data files (AVL and smart card). To get the assumption of these datasets, it was necessary to consider the AVL data to have more features than smart card data and also more observations (over 1 million for bus line 600 dataset). It was only necessary to have less than 32 trips and do the removal of trips bigger than 32. The inclusion of some new variable to better structure our data was the alternative way to separate some complex ones into two or more new variables, i.e. the date variable has a format ISO 8601[2] date and time. The date variable is

[2]This format applies dates in the Gregorian and times based on the 24 hours.

split into two other new variables called date and time, so from these two variables other three new ones show up as day of the month, day of the year and day of the week.

The only difference between the final AVL and smart card data frame is the presence of the number of boarding passengers and number of alighting passengers, these variables are zero in the AVL dataset. Afterwards, building these two final data frame with the bus line 600 and 805, an extra two data frames appear from these ones, merging AVL and smart card dataset of each bus line. How does this work? Simply, carrying out the comparison between datasets variables, which data does not contain in one of the joined datasets (the dataset that contains more variables) in our case the AVL is the main dataset.

4.3 Historical Data with Average Travel Time

The use of the historical average technique, on this dissertation, is only to represent the travel time average for some specific bus route, which means the bus travel time average from a specified day and time. This information will tell us, in terms of duration of travel time, we are only considering one-way trip and the average time spent by a bus during its trip along the route. The aim of this section is to provide a basic travel time estimate, considering these travel times (they might change throughout the bus route of the day) and will be used to inform the average travel time. This study also can be a future research, using this historical data with average travel time applying predictive models with some external data as traffic, weather and car speed to estimate the average travel time for each of the bus line or each of the time period as a baseline.

Gayah and Wood (2016) applied historical data average travel time representing observed travel times from stop 9 to stop 15 what shows in their research is that the expected travel time for the segment represents 590 seconds with standard deviation of 131 seconds for individual values.

Shalaby and Farhan (2004) developed a model for travel time prediction implementing and comparing the performance over the historical average model and the time lag recurrent ANNs model. Their model using Kalman filter showed an important improved performance in comparison with historical and regression model. Although, it is considered requirements for these models; real time data in order to update the estimates, so observation inputs in the time interval must be given. The statistical measures of travel time duration and average travel time for each day can be followed in the table 4.2.

Table 4.1 shows the travel time observations for 3 days, each one in a different period of the month, taking into account the weekday starts from 1 to 7 and the column bus departure time which means the time a bus leaves its starting point or a terminal. If we consider observations of each day and the total travel duration for day 75, we have an average of 1750 seconds by a complete travel from the start to the final stop. For the day 97 we have an average trip of 1764 seconds and for the day 141 we have an average travel of 1771 seconds, which is the approximation of the real bus travel time duration for line 805 which is about 1800 seconds according to the information on the webpage [3].

[3] http://www.stcp.pt/pt/viajar/linhas/?linha=805&sentido=0&t=horarios

Table 4.1: Sample Observations of 3 days Average Travel Time for bus line 805

Day of the month	**Day of the year**	**Weekday**	**Bus departure time**	**Travel duration (in sec)**	**Average travel time between bus stop (in sec)**
16	75	2	22219	1294	43.13
16	75	2	24311	1312	43.73
16	75	2	25336	1563	52.10
16	75	2	26722	1588	52.93
16	75	2	28207	1957	65.23
16	75	2	29724	1937	64.57
16	75	2	31099	2191	73.03
16	75	2	32543	1657	55.23
16	75	2	34228	1963	65.43
16	75	2	36028	2028	67.60
7	97	3	22238	1426	47.53
7	97	3	23897	1483	49.43
7	97	3	25335	1528	50.93
7	97	3	27005	1506	50.20
7	97	3	30525	1905	63.50
7	97	3	32525	2024	67.47
7	97	3	34198	1966	65.53
7	97	3	36030	1902	63.40
7	97	3	37808	1898	63.27
7	97	3	39649	1998	66.60
21	141	5	22224	1389	46.30
21	141	5	23893	1383	46.10
21	141	5	25365	1521	50.70
21	141	5	26706	1822	60.73
21	141	5	28210	1671	55.70
21	141	5	29656	1727	57.57
21	141	5	31105	2237	74.57
21	141	5	32505	2104	70.13
21	141	5	34192	1831	61.03
21	141	5	36007	2021	67.37

Table 4.2: Statistics for Data in Bus Travel Time

Day of the year	Travel duration (in sec)		Average travel by bus stop (in sec)	
	Standard deviation	Variance	Standard deviation	Variance
75	309.94	96064.88	10.33	106.73
97	243.93	59500.49	8.13	66.11
141	292.05	85294.26	9.73	94.77

The calculation takes the average formula where $t_2 - t_1$ is the time difference between two bus stops. The result will be presented in seconds and taken time, a bus spends to get from a point (a bus stop) to another. As referenced in section 4.2, our dataset contains 32 fixed trips along the day. For this travel time might contain less than the standard 32 trips that we chose in the data frame.

4.4 Applying the Permutation Importance Measure

As discussed in section 4.2, the data preparation before using an ML algorithm, that is quite necessary to use the most important variables in a predictive model in order to get efficient analyses and the best results. As we work with two different datasets and the use of the important variables via permutation with the RFs model, it made itself viable to compute a set of feature importance scores, based on the performance of the model with the specified feature values. The intuition of permutation importance variable was also used in our SVR algorithm.

This technique is usually applied to RFs, the Permutation Feature Importance (PFI) technique uses a trained model, a test dataset, and an evaluation measure to create a random permutation of a feature variable, where the performance of the model input data is evaluated on the dataset. It is also done to each of the feature variables, one at a time, and the technique returns a list of the feature variables and related to their importance scores (Bleik, 2015).

This type of technique was used to ensure that the model was applied with proper importance feature scores. The graphs 4.2 and 4.3 are the representation of relevance of the variables by their features to the predictive performance of RFs, in terms of how much the evaluation metric (Mean Square Error) deviates after permuting the values of these features for bus line 600 and 805 AVL datasets.

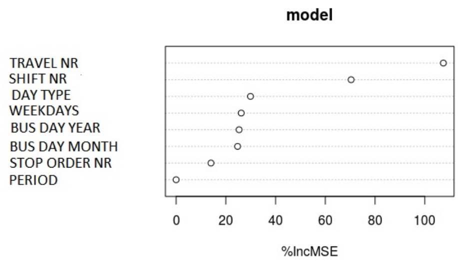


Figure 4.2: Variables of importance to the data of the bus line 600

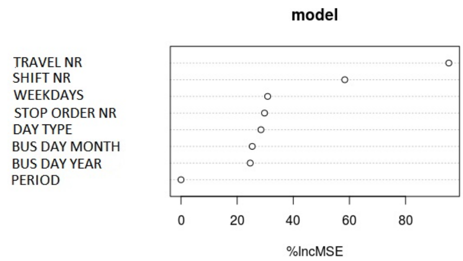


Figure 4.3: Variables of importance to the data of the bus line 805

Figures 4.2 and 4.3, we can see the importance variables implied by the main feature variable (in this case it is the time variable) with most high correlations after the use of the model. These variables with high correlation are travel number and shift number while period variable has demonstrated score equal to zero.

The importance score is determined by the reduction of the performance after randomly changing the feature values. Then, the evaluation metrics are used to measure how precise the predictions are. The PFI score is defined as pfi = Pb-Ps, where Pb is the base performance metric score and

Ps is the performance metric score after shuffling. However, the evaluation metric is used as an error/loss metric, then the score is determined as: pfi = -(Pb-Ps), therefore no matter which measure is chosen within the module, a higher value implies the feature is more important (Bleik, 2015).

Applying the permutation importance method, it was the alternative to be applied regarding our important feature variables and it was given what we want to know about the performance of a RFs algorithm fitted with this importance variables method.

4.5 Applying the Sliding Windows Method for the Proposed Machine Learning Models

The characteristic of our dataset as time series data, implicitly requires effective methods and techniques for framing it, and for the importance of recent data has to be used in an ML model. The sliding window method provided us a good point of view in restructuring the data set into a supervised learning problem and according to the target problem.

The structure of the sliding windows method in this study had a potential to discover in a fragmentation of the segment in the data set. As the analysis combines the use of sliding windows method with our proposed predictive model, so it was needed this combination with every segment of sliding windows applied to our data and after that to get the results.

About the implementation of sliding windows method, we chose to combine the training and test data within 15 days for each window stamp (window size equal to 10 days for training and five for test) in which only the first 10 days were the only ones not used as test data. The slide action of the sliding window method, in our case, infers for each part of the sequence-based window will move 5 days ahead in the data set. This sequence of data will be concluded until the last part of the window is tested in the end of time series data.

The analysis on implementing sliding windows method enabled us to later use the predictor as a result of the proposed ML models and conducted to a more efficient implementation of the effort in getting design predictive model, capable to predict and compare bus travel time into two different datasets.

In the next sections, the implementation and comparison of the proposed ML algorithms will be addressed through their results.

4.6 Experimenting with the Random Forest Algorithm

The Random Forests model used for this study was proposed by applying the regression decision trees in an analysis, which considers the dataset, variables and a relationship with the predictors. As RFs model is a fully non-parametric statistical method and requires no functional assumptions and some covariate non-relationship to the outcome response (Breiman, 2001).

Some considerations in implementing RFs algorithm have to be highlighted in some circumstances, once the dataset fits a classification or regression analysis on its structure. For our data, we are taking the bus time values for growing trees depending on this time vector.

As our regression problem is built in the RFs model, it can also be used to compare the prediction of bus travel time using the features vectors through all trees with our different datasets. We stated the target vector with the other predictor variables which contain necessary information to predict and after that compare this information with our RF algorithm. The structure of the algorithm improved with sliding window method, provided us a better understanding of how a time varying data should be assessed. With these structured windows of 15 days along our dataset, the evaluation through data using the sliding window, was applied during whole year (in this case the year 2010). This analysis was also determined by the importance variables which have more influence in our algorithm, as explained in section 4.4.

The proposed RFs algorithm was outlined to infer the bus travel time prediction starting by building out trees on the way a decision tree works, although each split has to combine each split with a small random subset of features variables. In this RFs algorithm, the number of trees was used as default equal to 500 and it was used for the number of predictors in a random sample at every split.

When we start by verifying our RF algorithm with first AVL data, after applying the predictor variables through RF model the coefficient of determination R^2, it was used to analyse the differences in one variable, it was explained by the difference in the other variables. Time is the main variable according to the objective to compare prediction using these two AVL and smart card datasets. Getting some results from the AVL for bus line 600 and 805 using the RF model, it was revealed a goodness of fit indicator for observations.

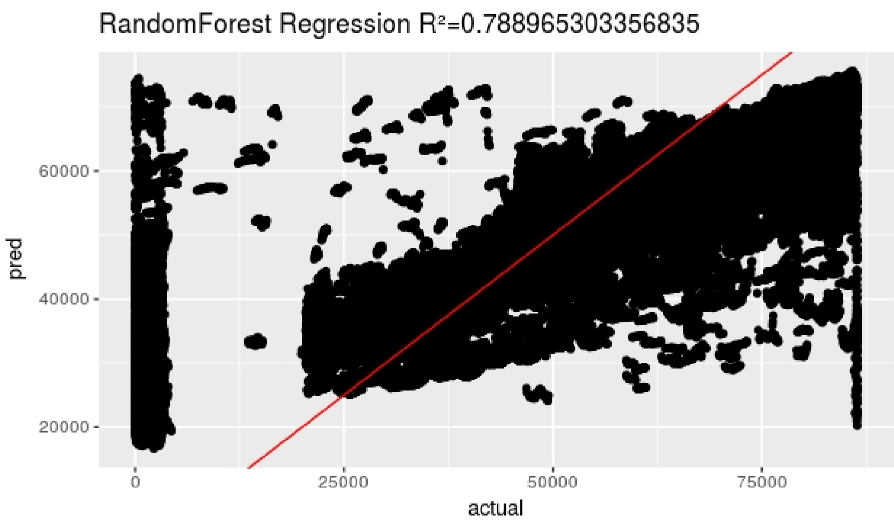


Figure 4.4: Regression test of AVL data with R^2 for bus line 600

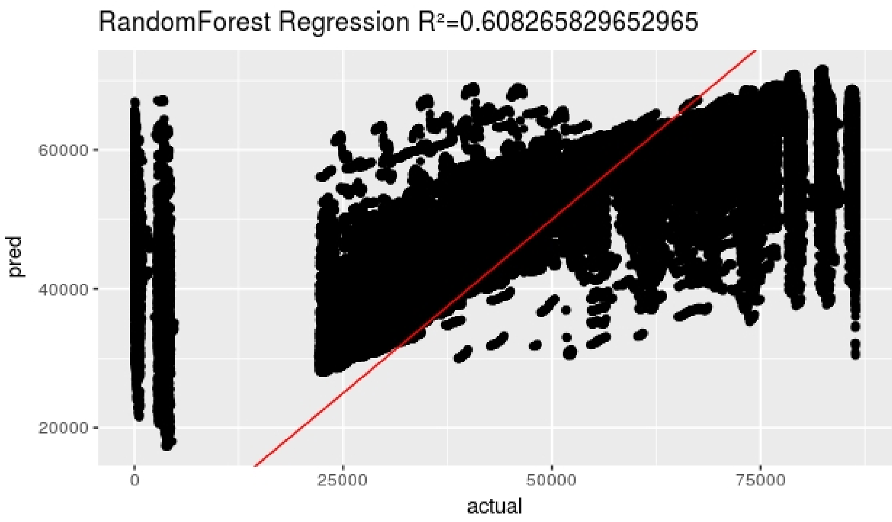


Figure 4.5: Regression test of AVL data with R^2 for bus line 805

As we can see, AVL dataset for bus line 600 has the highest R^2 with the RFs algorithm with 79%, while AVL data for bus line 805 gets 61%. It shows that even framing both datasets with the same variables and applying the RFs model in the same scenario the analysis represented a stronger evaluation of relationship among the dependent variables for bus line 600. It might also be explained by the high volume of data contained in this dataset, what can be an influence in this RFs regression model, taking into account the importance of the R^2 in a massive dataset.

When we employ our RF algorithm to the AVL and smart card dataset together, i.e. the AVL dataset joined with smart card dataset for each bus line, after applying the model with these data, we see an improvement related to R^2 through this model, as shown in the graphs 4.5 and 4.6.

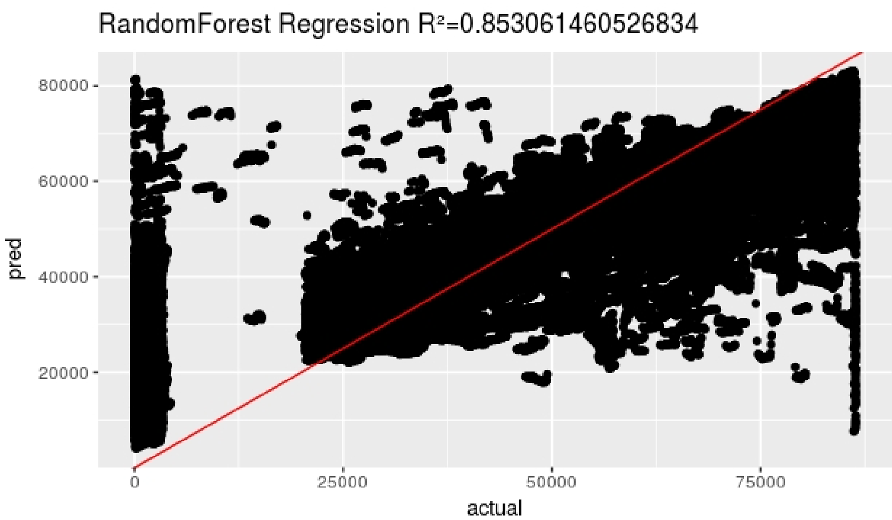


Figure 4.6: Regression test of AVL and smart card data together with R^2 for bus line 600

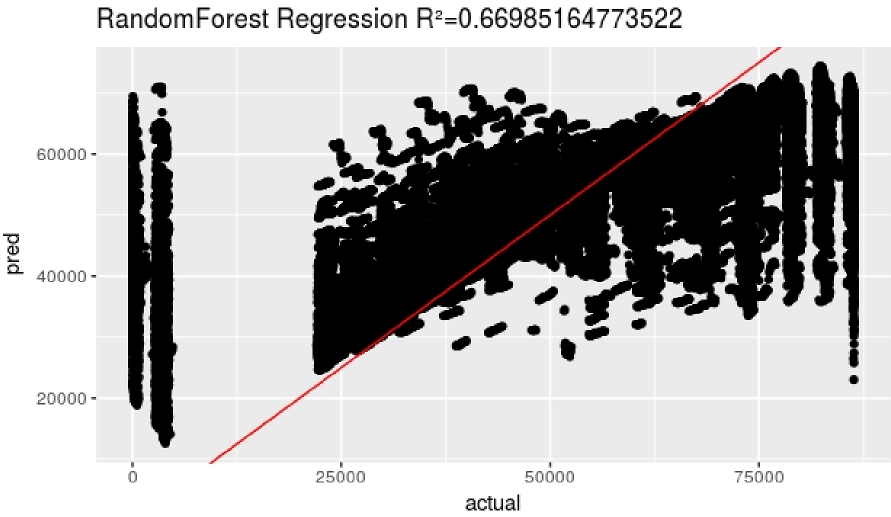


Figure 4.7: Regression test of AVL and smart card data together with R^2 for bus line 805

It is perceptible an increase of goodness of fit of the RFs algorithm in both datasets, as we can see, firstly the bus line 600, the model gets R^2 precision of 85% and for the bus line 805, the model gets 67%. Comparing with the RFs algorithm of AVL datasets (line 600 and 805) we see an increment for both bus lines.

Employing the RFs algorithm providing us a regression analysis to be precise, taking into account our goal and the characteristics of the dataset. The imputation of bus travel time prediction was carried out by building the RFs algorithm and applying the y argument for predictor variables to choose the best split option for this problem.

4.7 Experimenting with the Support Vector Regression Algorithm

The Support Vector Regression (SVR) is one of the models applied in the database with radial basis functions (RBF) intended to assess and compare results with prediction models. It can be used with sliding windows method for training and test data or also applied root mean squared error as performance measure for the model.

The relevance of SVR in travel time prediction area has been highlighted through many researchers due to feasibility of applying the model and its performance as well. In the Wu et al. (2004) research work, SVR was used to predict travel time for highway outperforming many other methods applied.

We used SVR with the RBF kernel. For each data frame, it was applied the same algorithm setting. Some tests were done in order to assess whether the dataset was used with correct data frame. We realised that SVR was more sensitive for some variable, indicating a type of scaling problem, but in this case, only one variable was unconsidered to be used in our algorithm. Passed this problem, the SVR algorithm modelled our data and it was used with the time-stamp-based window method.

The figures 4.8 and 4.9 represent the SVR algorithm using sliding window with training data equal to 10 days, test equal to 5 days and number of windows for 30 days ahead with the AVL and smart card data together, so it is a representation of SVR algorithm with a slice of sliding window for the dataset.

Now the comparison of SVR algorithm through the dataset was conducted by evaluating RMSE of model in each dataset. As an evaluation, which allows to verify how far our prediction data is from the regression line, as it was used with RFs algorithm, we can consider these results as part of evaluating how good was the SVR algorithm in the datasets, but in chapter 5 we will have an understanding in which predictive model gets better results.

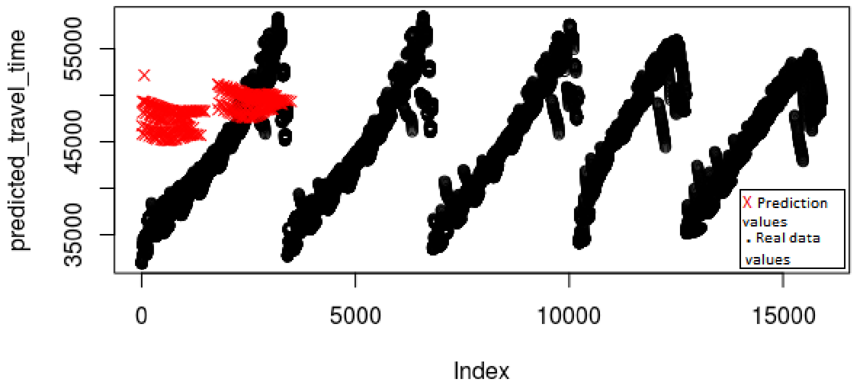


Figure 4.8: Real and predicted values for bus line 600 of SVR model

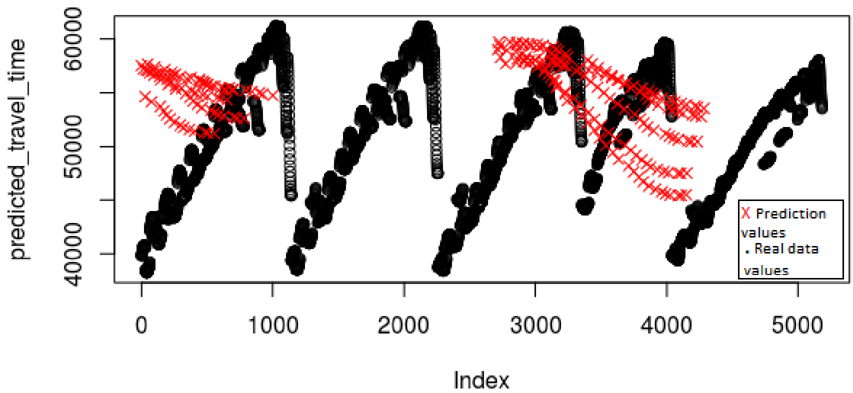


Figure 4.9: Real and predicted values for bus line 805 of SVR model

The graphs 4.8 and 4.9 show the behaviour of forecasts for a specific period of time which was considered the same period of time for both databases, that means the first 15 days of January only for a graphic representation of the SVR algorithm and predictions, so the index is represented by all input variables after training, testing and modelling the SVR prediction and the y-axis represents the travel schedule time in seconds. The black dots in the graphs represent the computational values of real data made up by SVR algorithm and the red cross represents prediction values taking into account the bus time and each prediction. The observations in the database is made up by the travel schedule at each stop, therefore the time variable was converted in seconds for better understanding the results.

Even considering to improve the performance of the SVR with the selection of the best parameters (called hyper-parameter optimization) applying a grid search, the improvements were minimal. In this grid search process, we trained many models for the different pairs of ε values and cost parameters the best pair. In our case, the values of ε and cost were: $\varepsilon = 0, 0.1, 0.2, ..., 1$ and $cost = 2^2, 2^3, 2^4, ...2^9$ to be used the tune method to train models. This means the training process was carried out for 88 models, which was a time-consuming process to calculate it.

In section 4.8, it will be approached the last proposed predictive model (Artificial Neural Networks). The same data structure was also used with the sliding windows method and using the RMSE metrics applied.

4.8 Experimenting with the Artificial Neural Networks Algorithm

Artificial Neural Networks (ANNs) have been applied to many research works, not only for bus travel time prediction. As ANNs are considered an advanced model, its implementation process involves normalization or scaling data, besides its capability to find answers for non-linear relationships even more to generate improved results using different travel time models (Park and Rilett, 1999).

The advantages of ANNs are related to deal with large amount of data, including the advanced process to train the data which requires more understanding of which manner the model must treat data and process in its structure. Besides some advantages, ANNs have the ability to train models themselves, so they are normally a kind of "black box" which they cannot show the nature of a relationship that are uncovered, so the main purpose of using ANNs is for predictions by providing an indicator of estimation as well (Gayah and Wood, 2016).

The proposed analysis used ANNs algorithm with the datasets to verify the effectiveness and efficiency of the results of an advanced ML model. In order to compare these results using the RMSE metrics with other proposed predictive models and providing a concise comparison.

Starting the ANNs model by building our algorithm, some care with our data was needed. Unlike the other proposed models that we have already applied and explained the implementation process and the results, ANNs have a particular way of process data which is specified from its predictive model structure. As we have time varying data, which contains each observation, it has a specified field to determine the date time variable that is described in our data frame, the date and time a bus passes at a bus stop. ANNs are more sensitive with some type of data, so it requires a data preparation, as well as we have already explained in section 4.2. Another factor of data consideration was raised by the need to check whether the dataset would have missing data and if it exists any missing data so it should be fixed in the dataset. Another data specification must be highlighted concerns to address data processing, which is to normalize the data before training our model, so this process allows the algorithm converge before a number of maximum iterations. The method chosen to normalize data in our work was scaling data. The process involves centring and scaling all data for multiplying by a constant c. After this process, we have data normalized to be in the intervals from 0 and 1 or even 1 and -1 which makes our ANN model give better results.

The parameters of ANNs that were used in this analysis to create our structure of ANNs contain layers and neurons. Many different format of layers and neurons were used in order to get the best analysis and result for the dataset. At first time, we considered as an attempting to test our algorithm from a beginning ANNs structure, which consists in establishing the layers and neurons that better fits for our problem by assessing the results. Firstly the parameters of representation of the model, consisted in applying the algorithms to the dataset, which has 9 variables for AVL data, and 11 variables for AVL and smart card data together. Then it is considered for the ANNs structure, the data input variables with 2 hidden layers, containing 5 and 3 neurons performing a regression analysis. After many tests, it was considered a more adjustable ANN algorithm structure, where we can find a more concise structure, that provided us a better result, and taking into both datasets. For final ANNs algorithm, it was chosen a structure model of 2 hidden layers with 3 and 1 neurons for the datasets. A fast cross validation was used, applying a for loop with the ANN algorithm. The 10 fold cross validated RMSE was applied for our ANN algorithm which returns the results shown in section 4.9.

As we can see in figures 4.10 and 4.11, both datasets (AVL and SC data together) are representing the implementation model of the chosen ANN structure, after that we can say that our training algorithm has converged and the model is ready to be used.

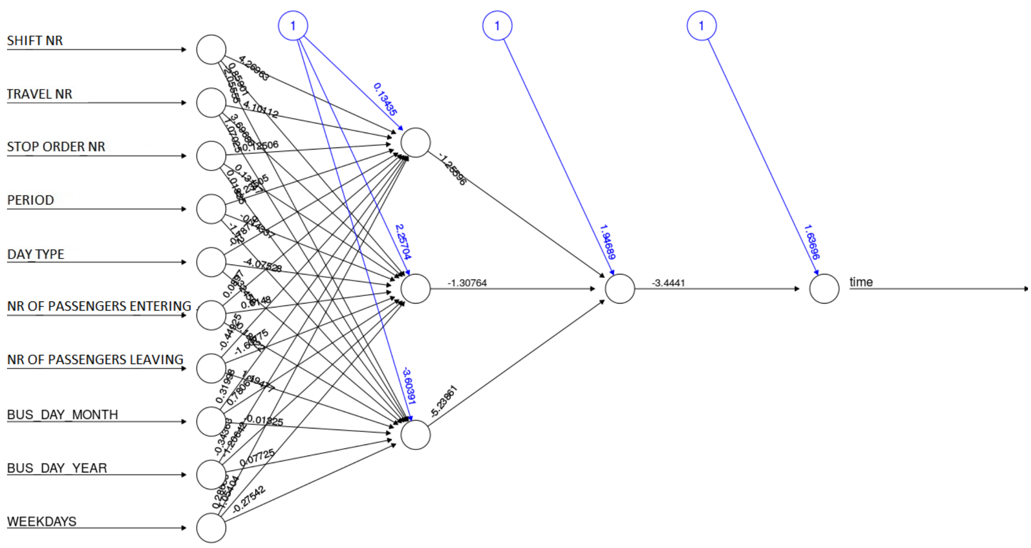


Figure 4.11: Graphical representation of the ANN model for bus line 805

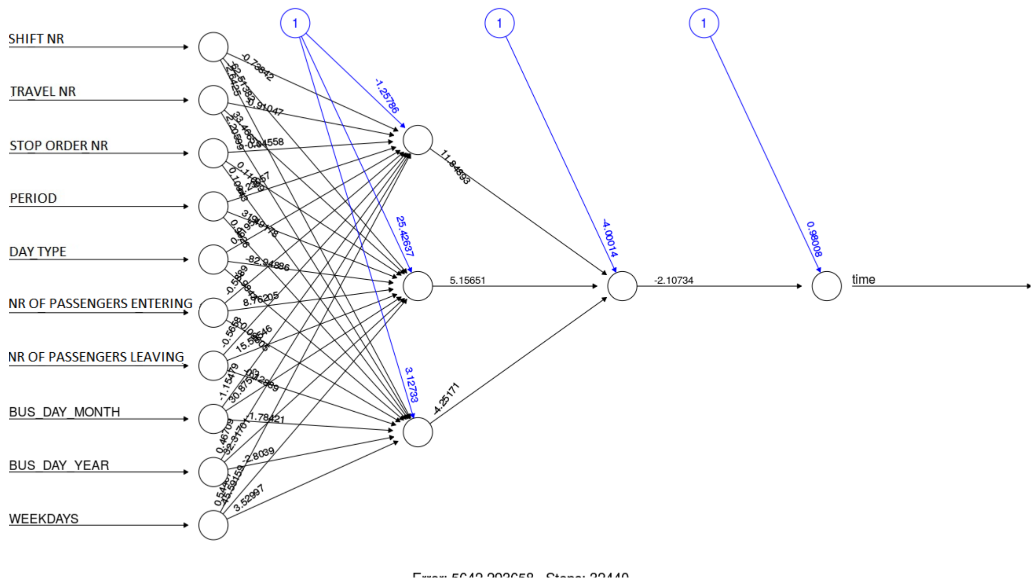


Figure 4.10: Graphical representation of the ANN model for bus line 600

These graphs 4.10 and 4.11 show the relation between the input data and the outcome. In both figures, we can see the black lines which show the connections between each layer and (the blue lines) represent the bias added in each step and connection.

Afterwards, we applied the ANNs algorithm and used RMSE metrics for predicted and real values it was possible to notice an error decrease of 40% comparing both datasets for bus line 600 and 44% for bus line 805.

In section 4.9, we evaluate and compare the results of the proposed models. Some improvements are perceptible when we analyse bus travel time prediction from AVL data in comparison to other models. We also evaluate the results from the combination of both datasets in the same predictive model applied for AVL data.

4.9 Presenting the Results

The statistical metric used to evaluate the predictive performance of the different methods tested using the two different sets of variables was the Root Mean Squared Error (RMSE). The RMSE measures is the performance of the models if they are close related to the different forms and even distributions of these models (Yu et al., 2017). We can evaluate the models after predicting the values using the test data. The results of the RMSE measure applied in the datasets with RFs, SVR and ANNs algorithm are shown in table 4.3:

The calculated measure on both datasets using RMSE consisted in knowing how much the noise is in the data frame after using the ML models. These RMSE values shown in table 4.3 parse the summarizing of points that are far from the regression line with RFs. It is also shown a decrease of values, which indicates that residuals of RMSE get lower by the comparison to the

Table 4.3: RMSE measures for bus travel time prediction using predictive models

Datasets	RFs		SVR		ANNs	
	L600	L805	L600	L805	L600	L805
AVL data	8504.82	12327.53	15320.09	17364.67	7.729	10.803
AVL and Smart Card data	7122.03	11346.68	16731.59	19482.72	4.611	6.092

prediction of the algorithms (RFs and ANNs), considering the observed values between the AVL data, AVL and SC data together.

As we can see the comparison to RMSE for datasets using RFs and ANNs algorithms, where there is an improvement in metrics when aggregating data with AVL and smart card. With the RMSE metric we try to compare the results of SVR algorithm with other proposed algorithms, so that it can also confirm whether the performance of SVR algorithm is better than other algorithms.

Evaluating the results shown in table 4.3, we also notice a decrease comparing AVL dataset and AVL plus smart card dataset for the bus lines. It means that the comparison of RMSE with ANNs for bus line 600, AVL data has a decrease. ANN obtained a RMSE of 7.729 and 4.611 respectively for the AVL dataset and the AVL plus SC dataset. For bus line 805, the RMSE of AVL data is 10.803 and for AVL and smart card data together is 6.092. We also observe an error decrease model for the datasets comparison, with advantage for the AVL alone, as shown in table 4.3.

The RMSE metric of SVR algorithm had an increase of 9% for bus line 600 when comparing both datasets. The RMSE metrics of SVR algorithm also had an increase of 12% for the bus line 805.

Chapter 5

Conclusion and Future Works

We undertook this dissertation as part of an analysis to compare and evaluate bus travel time prediction using machine learning algorithms for two bus lines in Porto. As an option to apply machine learning models in the data set, which contains information about each bus line along its route and these data are separated into two different data frames.

Throughout this study, we did some assumptions by considering a scenario where buses had the same condition to do their routes. The comparison between the two datasets was motivated by extracting forecast values of both datasets and comparing the results of predictive models using the RMSE metric. After that, we joined AVL and smart card data for each bus line. The aggregation of data was formed by the second analysis where the amount of data for AVL and smart card could be a positive factor or not. By the fact we proposed a machine learning approach, so the predictive models have their particular modelling process and the way they treat data.

Our algorithms were built to fulfil the requirements of the ML models to tend the proposed problem. The three implemented predictive algorithms (Random Forests, Support Vector Regression and Artificial Neural Networks) were a source of statistical resources to infer our analysis during the process of implementation, development and presentation of results.

Such information can come from these databases that we chose to check if joining data from each database, could return some insights about travel time prediction for public transportation. With a rich database containing bus data along its route, many other analyses appeared in mind to reinforce the use of ML models in order to get some detailed analysis or different approaches that we have addressed on this dissertation. However, this will be explained as an issue for future works in sections 5.2.

5.1 Satisfaction of Objectives

Nowadays, data is provided to experts in large quantity, and these data also have much to explain to us about the behaviour on the system, an organization or even more about people. In the research area of bus travel time prediction, we addressed the issue about how much information is raised from a company's bus lines comparing the dataset from each line, taking into account the AVL data in the first time, in addition to aggregating data we also compare the two datasets (AVL and smart card) together.

In order to reach our goals, we have done many tests using different machine learnings under the same experimental setup. As re-sampling method sliding window was used. Moreover, variable importance was used for variable selection.

In terms of algorithm implementation, to get satisfactory results, we have verified how important sliding windows and imputation of importance variable are because with these methods, we were able to input only the necessary data to build a more concise predictive algorithm. The sliding windows method provided us with the most recent data usage for training and test data samples through the ML algorithm. The positive response for this implementation was the obtained result that support the implementation process and provided realistic results in our approach.

The potential of the applied algorithms and the aforementioned implementation conducted us to an understanding that we assessed the datasets individually in order to see some statistical metrics, although these same metrics were reviewed in the implementation of the second dataset which also have more data than the first one.

We also follow the results brought from each proposed algorithm and predictive model. The two algorithms (Random Forests and Artificial Neural Networks) were the ones that we found best results comparing the datasets. For these two algorithms, we had a decrease of the RMSE for AVL data in comparison with the AVL and smart card data together, while for Support Vector Regression had an increase of RMSE when comparing the first and second datasets of each bus line.

5.2 Future Works

For the next works regarding bus travel time prediction or even to the framework studied on this dissertation, we have some recommendations for further research or another academic work, which are:

- Corresponding to the bus travel time, another analysis might be done with the same database we used in this study, implementing a different approach, which consists in relating a bus line data and generates information for other bus lines that share a similar route segment, so with predicted values it is possible to inform passengers about the bus real time at a specific bus stop and about a bus delay.
- Developing a framework also for the database used in this work, that can analyse bus travel time predictions considering external traffic data and weather condition data.
- Another work can be done by using different machine learning models to apply them for other bus lines data, considering another approach which might be to develop strategies for transport service reliability and travel behaviour.

Bibliography

Algueró, P. S. (2013). Using Smart Card Technologies to Measure Public Transport Performance: Data Capture and Analysis. Technical Report Insdustrial Engineering.

Altinkaya, M. and M. Zontul (2013). Urban Bus Arrival Time Prediction: A Review of Computational Models. *International Journal of Recent Technology and Engineering 2*(4), 2277–3878.

Bagchi, M. and P. White (2004). What role for smart-card data from bus systems? *Proceedings of the ICE Municipal Engineer 157*(1), 39 –46.

Basak, D., S. Pal, and D. C. Patranabis (2007). Support vector regression. *Neural Information Processing-Letters and Reviews 11*(10), 203–224.

Bates, J., J. Polak, P. Jones, and A. Cook (2001). The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review 37*(2), 191–229.

Bates, J. W. (1986). Definition of practices for bus transit on-time performance: Preliminary study. *Transportation Research Circular* (300).

Bin, Y., Y. Zhongzhen, and Y. Baozhen (2006). Bus arrival time prediction using support vector machines. *Journal of Intelligent Transportation Systems 10*(4), 151–158.

Bleik, S. (2015). Permutation feature importance. cortana intelligence and machine learning. https://blogs.technet.microsoft.com/machinelearning/2015/04/14/permutation-feature-importance/. Accessed: 2017-06-30.

Bowman, L. A. and M. A. Turnquist (1981). Service frequency, schedule reliability and passenger wait times at transit stops. *Transportation Research Part A: General 15*(6), 465–471.

Braverman, V., R. Ostrovsky, and C. Zaniolo (2012). Optimal sampling from sliding windows vladimir. *Journal of Computer and System Sciences* (1666), 260–272.

Breiman, L. (2001). Random forests. *Machine Learning 45*(1), 5–32.

Brownlee, J. (2013). Machine learning processhow to prepare data for machine learning. http://machinelearningmastery.com/howto-prepare-data-for-machine-learning/, note = Accessed: 2017-06-30.

Brownstone, D. and K. A. Small (2005). Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice 39*(4), 279–293.

Carrel, A., A. Halvorsen, and J. Walker (2013). Passengers' perception of and behavioral adaptation to unreliability in public transportation. *Transportation Research Record: Journal of the Transportation Research Board* (2351), 153–162.

Ceder, A. (2007). Public transit planning and operation: Theory. *Modeling and practice. 0xford: Elsevier*.

Cham, L. C. (2006). *Understanding bus service reliability : a practical framework using AVL/APC data.* Ph. D. thesis, Massachusetts Institute of Technology.

Chang, H., D. Park, S. Lee, H. Lee, and S. Baek (2010). Dynamic multi-interval bus travel time prediction using bus transit data. *Transportmetrica 6*(1), 19–38.

Chien, S. I.-J., Y. Ding, and C. Wei (2002). Dynamic bus arrival time prediction with artificial neural networks. *Journal of Transportation Engineering 128*(5), 429–438.

Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician 35*(1), 54–54.

Cover, T. and P. Hart (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory 13*(1), 21–27.

Dahl, E., A. A. Sjåfjell, and S. Skogen (2014). On Implementations of Bus Travel Time Prediction Utilizing Methods in Artificial Intelligence. (June), 1–32.

Delgado, F., J. C. Munoz, and R. Giesen (2012). How much can holding and/or limiting boarding improve transit performance? *Transportation Research Part B: Methodological 46*(9), 1202–1217.

Drucker, H., C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik (1997). Support vector regression machines. In *Advances in neural information processing systems*, pp. 155–161.

Elhenawy, M., H. Chen, and H. Rakha (2014). Random forest travel time prediction algorithm using spatiotemporal speed measurements. *21st World Congress on Intelligent Transport Systems, ITSWC 2014: Reinventing Transportation in Our Connected World* (July 2016).

Ferreira, M., R. Fernandes, H. Conceiçao, P. Gomes, P. M. d'Orey, L. Moreira-Matias, J. Gama, F. Lima, and L. Damas (2012). Vehicular sensing: Emergence of a massive urban scanner. In *International Conference on Sensor Systems and Software*, pp. 1–14. Springer.

Ferreira, P. M. and A. E. Ruano (2009). Online Sliding-Window Methods for Process Model Adaptation. *Ieee Transactions On Instrumentation And Measurement 58*(9), 3012–3020.

Friedman, J. H. and W. Stuetzle (1981). Projection pursuit regression. *Journal of the American statistical Association 76*(376), 817–823.

Furth, P. G. (2000). *Data analysis for bus planning and monitoring*. Number 34. Transportation Research Board.

Furth, P. G., B. Hemily, T. Muller, and J. G. Strathman (2003). Uses of archived avl-apc data to improve transit performance and management: Review and potential. *TCRP Web Document 23*.

Gayah, V. V. and J. S. Wood (2016). Estimating Uncertainty of Bus Arrival Times and Passenger Occupancies Estimating Uncertainty of Bus Arrival Times and Passenger Occupancies.

Giraud-Carrier, C. (2000). A note on the utility of incremental learning. *Ai Communications 13*(4), 215–223.

Golob, T. F., E. T. Canty, R. L. Gustafson, and J. E. Vitt (1972). An analysis of consumer preferences for a public transportation system. *Transportation Research 6*(1), 81–102.

Huang, G.-B., Q.-Y. Zhu, and C. K. Siew (2006). Real-time learning capability of neural networks. *IEEE Trans. Neural Networks 17*(4), 863–878.

Jeong, R. and R. Rilett (2004). Bus arrival time prediction using artificial neural network model. In *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, pp. 988–993. IEEE.

Jin, Y. and B. Sendhoff (2008). Pareto-based multiobjective machine learning: An overview and case studies. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 38*(3), 397–415.

Jou, S. (2017). Machine Learning :.

Kashid, S. and S. Kumar (2012). Applications of artificial neural network to sheet metal work-a review. *American Journal of intelligent systems 2*(7), 168–176.

Kimpel, T., J. G. Strathman, and S. Callas (2004). Improving scheduling through performance monitoring using avl/apc data. *Submitted to University of Wisconsin-Milwaukee as a Local Innovations in Transit project report under the Great Cities University Consortium*.

Kimpel, T. J. (2001). *Time Point-Level Analysis of Transit Service Reliability and Passenger Demand*. Ph. D. thesis.

Kumar, V., B. A. Kumar, L. Vanajakshi, and S. C. Subramanian (2014). Comparison of Model Based and Machine Learning Approaches for Bus Arrival Time Prediction. (3946).

Levinson, H. S. (1991). *Supervision strategies for improved reliability of bus routes*. Number 15.

Li, F., Y. Yu, H. Lin, and W. Min (2011). Public bus arrival time prediction based on traffic information management system. *Proceedings of 2011 IEEE International Conference on Service Operations, Logistics and Informatics*, 336–341.

Lin, W.-H. and J. Zeng (1999). Experimental study of real-time bus arrival time prediction with gps data. *Transportation Research Record: Journal of the Transportation Research Board* (1666), 101–109.

Ma, Z.-L., L. Ferreira, M. Mesbah, and A. T. Hojati (2014). Modelling Bus Travel Time Reliability Using Supply and Demand Data From Automatic Vehicle and Smart Card Systems.

Mendes-Moreira, J. P. C. L. (2008). *Travel time prediction for the planning of mass transit companies: a machine learning approach*. Ph. D. thesis.

Mitchell, T. M. (1997). *Machine Learning*.

Moreira-Matias, L. (2014). *On Improving Operational Planning and Control in Public Transportation Networks using Streaming Data: A Machine Learning Approach*. Ph. D. thesis.

Moreira-Matias, L., O. Cats, J. Gama, J. Mendes-Moreira, and J. F. De Sousa (2016). An online learning approach to eliminate Bus Bunching in real-time. *Applied Soft Computing Journal 47*, 460–482.

Moreira-Matias, L., J. Mendes-Moreira, J. F. De Sousa, and J. Gama (2015). Improving Mass Transit Operations by Using AVL-Based Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems 16*(4), 1636–1653.

Müller, K.-R., A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik (1997). Predicting time series with support vector machines. In *International Conference on Artificial Neural Networks*, pp. 999–1004. Springer.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications 9*(1), 141–142.

Oshyani, M. F. and O. Cats (2014). Real-time bus departure time predictions: Vehicle trajectory and countdown display analysis. *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, 2556–2561.

Padmanaban, R. P. S., L. Vanajakshi, and S. C. Subramanian (2009). Estimation of bus travel time incorporating dwell time for APTS applications. *IEEE Intelligent Vehicles Symposium, Proceedings*, 955–959.

Park, D. and L. R. Rilett (1999). Forecasting freeway link travel times with a multilayer feedforward neural network. *Computer-Aided Civil and Infrastructure Engineering 14*(5), 357–367.

Pelletier, M.-P., M. Trépanier, and C. Morency (2009). Smart Card Data in Public Transit Planning: A Review. *Transportation Research Part C: Emerging Technologies*.

Prashker, J. N. (1979). Direct analysis of the perceived importance of attributes of reliability of travel modes in urban travel. *Transportation 8*(4), 329–346.

Recknagel, F. (2001). Applications of machine learning to ecological modelling. *Ecological Modelling 146*(1), 303–310.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review 65*(6), 386.

Russell, S. and P. Norvig (2013). *Artificial Intelligence A Modern Approach*.

Shalaby, A. and A. Farhan (2004). Prediction model of bus arrival and departure times using AVL and APC data. *Journal of Public Transportation 7*, 41–61.

Strathman, J. G., K. J. Dueker, T. Kimpel, R. Gerhart, K. Turner, P. Taylor, S. Callas, D. Griffin, and J. Hopper (1998). Automated Bus Dispatching, Operations Control and Service Reliability: Analysis of Tri-Met Baseline Service Date.

Taylor, M. A. (2010). Intelligent Transport Systems. *European Commission. Handbook of Transport Systems and Traffic Control*, 461–475.

Tirachini, A., D. A. Hensher, and J. M. Rose (2013). Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand. *Transportation Research Part A: Policy and Practice 53*, 36–52.

Turner, S. M., W. L. Eisele, R. J. Benz, and D. J. Holdener (1998). Travel time data collection handbook. Technical report.

Turnquist, M. A. (1981). Strategies for improving reliability of bus transit service.

U.S. Department of Transportation, T. (2007). Its deployment statistics – transit buses with automatic vehicle location (avl) and computer aided dispatch (cad). Technical report, Intelligent Transportation Systems.

Van Lint, J. and S. Hoogendoorn (2008). Reistijdvoorspellingen en reistijdbetrouwbaarheid. *MN 2*(3).

Verikas, A., E. Vaiciukynas, A. Gelzinis, J. Parker, and M. C. Olsson (2016). Electromyographic patterns during golf swing: Activation sequence profiling and prediction of shot effectiveness. *Sensors 16*(4), 592.

Viviano, D. (2016). *Bus Travel Time Analysis using Real-Time Data*. Ph. D. thesis.

Wang, J.-n., X.-m. Chen, and S.-x. Guo (2009). Bus travel time prediction model with ν - support vector regression. *2009 12th International IEEE Conference on Intelligent Transportation Systems*, 1–6.

Wu, C.-H., J.-M. Ho, and D.-T. Lee (2004). Travel-time prediction with support vector regression. *IEEE transactions on intelligent transportation systems 5*(4), 276–281.

Yetiskul, E. and M. Senbil (2012). Public bus transit travel-time variability in ankara (turkey). *Transport Policy 23*, 50–59.

Yu, H., Z. Wu, D. Chen, and X. Ma (2016). Probabilistic Prediction of Bus Headway Using Relevance Vector Machine Regression. pp. 1–10.

Yu, Z., J. Wood, and V. V. Gayah (2017). Using survival models to estimate bus travel times and associated. *Transportation Research Part C: Emerging Technologies 74*, 366–382.

Zeng, X. and Y. Zhang (2013). Development of recurrent neural network considering temporal-spatial input dynamics for freeway travel time modeling. *Computer-Aided Civil and Infrastructure Engineering 28*(5), 359–371.

Zhang, Y. and A. Haghani (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies 58*, 308–324.

Zhao, J., M. Dessouky, and S. Bukkapatnam (2006). Optimal Slack Time for Schedule-Based Transit Operations. *Transportation Science 40*(4), 529–539.

Zhou, C., P. Dai, and Z. Zhang (2016). Passenger demand prediction on bus services. *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, 1430–1435.

Zou, Y., X. Zhu, Y. Zhang, and X. Zeng (2014). A space–time diurnal method for short-term freeway travel time prediction. *Transportation Research Part C: Emerging Technologies 43*, 33–49.