THE IMPACT OF DRIVING STYLES ON FUEL CONSUMPTION: A DATA-WAREHOUSE-AND-DATA-MINING-BASED DISCOVERY PROCESS

SEMINAR REPORT

submitted by

Varun Dineshan(Reg. No – 13020029)

2013-2017 Batch

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
VISWAJYOTHI COLLEGE OF ENGINEERING & TECHNOLOGY, VAZHAKULAM- MUVATTUPUZHA

MAHATMA GANDHI UNIVERSITY, KOTTAYAM
NOVEMBER 2016

THE IMPACT OF DRIVING STYLES ON FUEL CONSUMPTION: DATA-WAREHOUSE-AND-DATA-MINING-BASED DISCOVERY PROCESS

A SEMINAR REPORT

submitted by

Varun Dineshan(Reg. No – 13020029)

Under the guidance of

Mrs. Cinita Mary Mathew

Asst. Professor, CSE Dept.

In partial fulfillment for the award of the Degree of

Bachelor of Technology in

Computer Science & Engineering

of

Mahatma Gandhi University – Kottayam

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
VISWAJYOTHI COLLEGE OF ENGINEERING & TECHNOLOGY,
VAZHAKULAM- MUVATTUPUZHA
November 2016

VISWAJYOTHI COLLEGE OF ENGINEERING & TECHNOLOGY, VAZHAKULAM- MUVATTUPUZHA

Department of Computer Science & Engineering

BONAFIDE CERTIFICATE

This is to certify that the seminar report entitled **THE IMPACT OF DRIVING STYLES ON FUEL CONSUMPTION** is a bonafide record of the seminar presented by **Varun Dineshan(13020029)** in partial fulfillment of the requirements for the award of the **Degree of Bachelor of Technology** in **Computer Science and Engineering** of Mahatma Gandhi University, Kottayam.

Date:…………..

Place:……………

Mrs. Cinita Mary Mathew

Asst. professor, CSE Department

Mrs. Dona Jose(Seminar Coordinator)

Assistant Professor- CSED-VJCET.

Dr. K.N. Ramachandran Nair

Professor & Head- CSED-VJCET

Internal Examiner

External Examiner

ACKNOWLEDGEMENT

At the very outset, I would like to give the first honor to the **Almighty** for giving me the strength, courage, knowledge and blessing to complete my seminar. It is my privilege to render my heartfelt thanks to our beloved manager, **Msgr. Dr. George Oliapuram**, for his support. I express my gratitude to **Prof. Josephkunju Paul C,** Principal, VJCET, for his guidance and support for the completion of this interesting work. I would like to thank Prof. **Dr. K. N. Ramachandran Nair,** the Head of the Department of Computer Science & Engineering, for giving me worthy suggestions and his constant encouragement, and guidance throughout the progress of the seminar. I would like to express my gratitude to **Mrs. Dona Jose,** Assistant Professor, our seminar coordinator, **Mrs. Cinita Mary Mathew,** my guide, for their personal commitment, morale boosting, technical guidance and patience for the completion of my seminar. In particular, I would like thank all the other faculties of VJCET and all other people who have helped me in many ways for successful completion of this seminar.

Varun Dineshan

ABSTRACT

Eco driving discusses the results of applied research on the eco-driving domain based on a huge data set produced from a fleet of Lisbon's public transportation buses for a three-year period. This data set is based on events automatically extracted from the control area network bus and enriched with GPS coordinates, weather conditions, and road information. Applying online analytical processing (OLAP) and knowledge discovery (KD) techniques to deal with the high volume of this data set and to determine the major factors that influence the average fuel consumption, and then classify the drivers involved according to their driving efficiency. Consequently, identify the most appropriate driving practices and styles. The findings show that introducing simple practices, such as optimal clutch, engine rotation, and engine running in idle, can reduce fuel consumption on average from 3 to 5l/100 km, meaning a saving of 30 l per bus on one day. These findings have been strongly considered in the drivers' training sessions

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

OLAP	- online analytical processing
KD	- knowledge discovery
IF	-Impact factor
AFC	-Average fuel consumption

CHAPTER 1

INTRODUCTION

Evaluating driver performance and promoting energy efficient driving has received scarce attention from the research community. This is due to the difficulty of objectively evaluating human driver performance. The driver controls the speed, acceleration, braking, engine rotation speed, the gear engaged and the position of the vehicle on the street in an environment characterized by certain traffic conditions, itinerary, load, etc. Different driving styles result in different fuel consumption levels, thus related to driving efficiency. Different external conditions result in different levels of consumption. For example, levels of fuel consumption in a public bus are strictly linked to the number of stops made per itinerary unit. The number of stops is a parameter that is not controllable by the driver, but on traffic, the bus route, or the number of passengers. One of the most efficient approaches to evaluate driver performance is to register a set of events(parameters)read from the CAN bus, which stores messages from all driving events on anon-board recorder, from where data is retrieved and stored in a database for subsequent analysis. Due to analytical needs and the huge data generated from this approach, this is a field where the application of online analytical processing (OLAP) and knowledge discovery (KD) techniques is needed. KD in databases has been attracting a significant amount of interest from both research and industry. There is an increasing need for approaches and tools to assist humans in extracting useful information from the fast growing volumes of digital data. KD techniques are being applied with success into big data scenarios and in different application demands Eco driving discusses the results of an applied research on the eco-driving domain that used the data automatically produced from the Lisbon’s fleet of public transport buses, for a 3-year period, and involving 1,041 drivers with 745 different buses and 73 carrier bus routes.applying OLAP and KD techniques to deal with the high volume of data and extract the associated knowledge

CHAPTER 2

RELATED WORKS

Driving style is seen as the attitude, orientation and way of thinking for daily driving, and is usually captured by questionnaires and surveys. Recent works use a virtual driving simulator to collect realistic driving data from human drivers and to model human driving actions, or to classify driving actions into styles by combining objective rank methods. However, this driver behavior can be analyzed through real driving parameters obtained from vehicle interfaces such as the CAN bus or driving data obtained from mobile device sensors. The authors and achievements are given below:

Author (year) [ref]	Description	Achievements claimed
Rolim et at. (2014) [16]	Monitor driver behavior (20 drivers) with lessons towards fuel savings	Around 4,8% reduction in fuel consumption
Barth (2009) [17]	Investigates the impact of providing real time eco-driving advice to the drivers based on real-time traffic speed, density and flow	Reduction in fuel consumption of 10-20% can be achieved without a significant increase in travel time
Boriboonsomsin (2010) [20]	Investigate how real-time feedback affects driving behavior	20 sample drivers show a reduction of 6% in fuel consumption on city streets and 1% on highways
Wahlberg (2007) [21]	Monitored fuel consumption reduction in buses during the 12 months after training	2% reduction in fuel consumption
Zarkadoula et al. (2007) [19]	Training monitoring driving experience for a period of two months	Mentioned fuel savings on buses of 4.35%
Ecomove (2013) [22]	FP7 European project, for testing and evaluating a series of "green" technologies and applications that aim to reduce fuel consumption	Reduced fuel consumption and CO2 emissions in road transport by 20%
Raghu K. Ganti, (2010) [23]	GreenGPS gives drivers the most fuel efficient route for their vehicle as opposed to the shortest or fastest route.	10% reduction in fuel consumption
Bart Beusen (2009) [18]	Eco-driving, on 10 drivers, based on the evaluation of individual driving style analysis for 10 months (real data)	Initial (first four months) the average fuel consumption after the course fell by 5.8% and then became stable. Some tended to fall back into their original driving habits

Table 2.1 Eco driving related projects

The availability of such data offers new opportunities due to the interpretation of raw data from real human drivers. Also, it is used in the application of intelligent algorithms to identify changes in driver behavior.

The main problem concerning the collection of data from CAN buses is that handling a huge data set of events is not an easy task and requires proper techniques and tools. Within the context of energy efficiency, much has been achieved at the level of performance of engines and vehicles, obtaining substantial improvement and energy savings. However, little attention has been focused on driving quality and on methods for continually promoting energy-efficient driving. This is due to the difficulty of objectively evaluating driver's performance.

Recommendation	Description	Explanation
Engine rounds per minute (RPM).	Drive in the highest possible gear at the lowest possible RPM.	Fuel consumption is lower at low RPM due to friction.
Maintain a steady speed	Avoid constantly braking and accelerating	Fuel is primarily consumed when accelerating
Eliminate idling	It is more fuel efficient to switch off the engine than leave the engine running	An average modern vehicle uses around 0.9-1.3 liters per hour (l=h) during idling [25]
Braking	Slow down by using the engine brake or the neutral gear instead of the actual brakes.	Modern vehicles use no fuel when using the engine brake, i.e. the vehicle is in gear and the accelerator is released
Acceleration	High accelerations consume much more fuel	Recommendation of acceleration in low gears with the RPM below 2,000 for diesel and with the throttle at half position
Speed	Drive at or below the speed limit	Fuel consumption increases at higher speeds.
Weight and air resistance	Minimize extra weight and air resistance	Both increase the load on the engine, thereby increasing fuel consumption.
Approach curves	Performed at the correct speed and in the highest possible gear	This reduces the need to accelerate after the curve
Tire pressure	Incorrect tyre pressure increases the rolling resistance	
Vehicle consuming accessories.	Air-conditioning, heating and other accessories that consume fuel	

Table 2.2 Recommendations for Eco-Driving

This also relates the previous work in eco-driving, which has primarily taken the form of advice to drivers based on the several studies performed. In an attempt to reduce fuel consumption and CO_2 emissions, the analysis of drivers' behavior has been applied and several companies have long before recognized the value of training their drivers to that purpose. As identified in Table 2.1, numerous studies have started to address these issues by monitoring drivers' behavior before and after they were given eco-driving sessions and presenting drivers with driver-specific information and the best practices. Additionally, a list of practical recommendations and advice for eco-driving that may reduce fuel consumption. Moreover, the identification of driving actions that influence consumption is yet more important in the cases of bus fleets due to high consumption and the number of hours that drivers perform on a daily basis.

CHAPTER 3

SYSTEM DESIGN

3.1 INTRODUCTION

The eco driving system is developed to reduce the fuel consumption to a greater extend. This system consist of four stages and at each stage different parameters that effects the fuel consumption and different impact factors are considered and found out.

3.1.1 Data Collection

Eco driving technique is implemented using four different process , data collection process, the data preprocessing process The OLAP process, data mining process which are designed to find the impact factors and reduce the fuel consumption. The main techniques employed in the data collection are:

This data set has the following details:

1. Around 1,500 different professional company drivers were considered, with an average age of 39, 10 years of driving experience.

2.These drivers have an average driving time of 6,000 hours, distributed throughout 700 days of driving, on average.

3.This data acquisition process was performed by the XtraN commercial product by Tecmic.

4.This product allows us to measure 44 event parameters, such as engine, ignition on/off, acceleration, braking and clutch use, fuel consumption, engine RPM's and date.

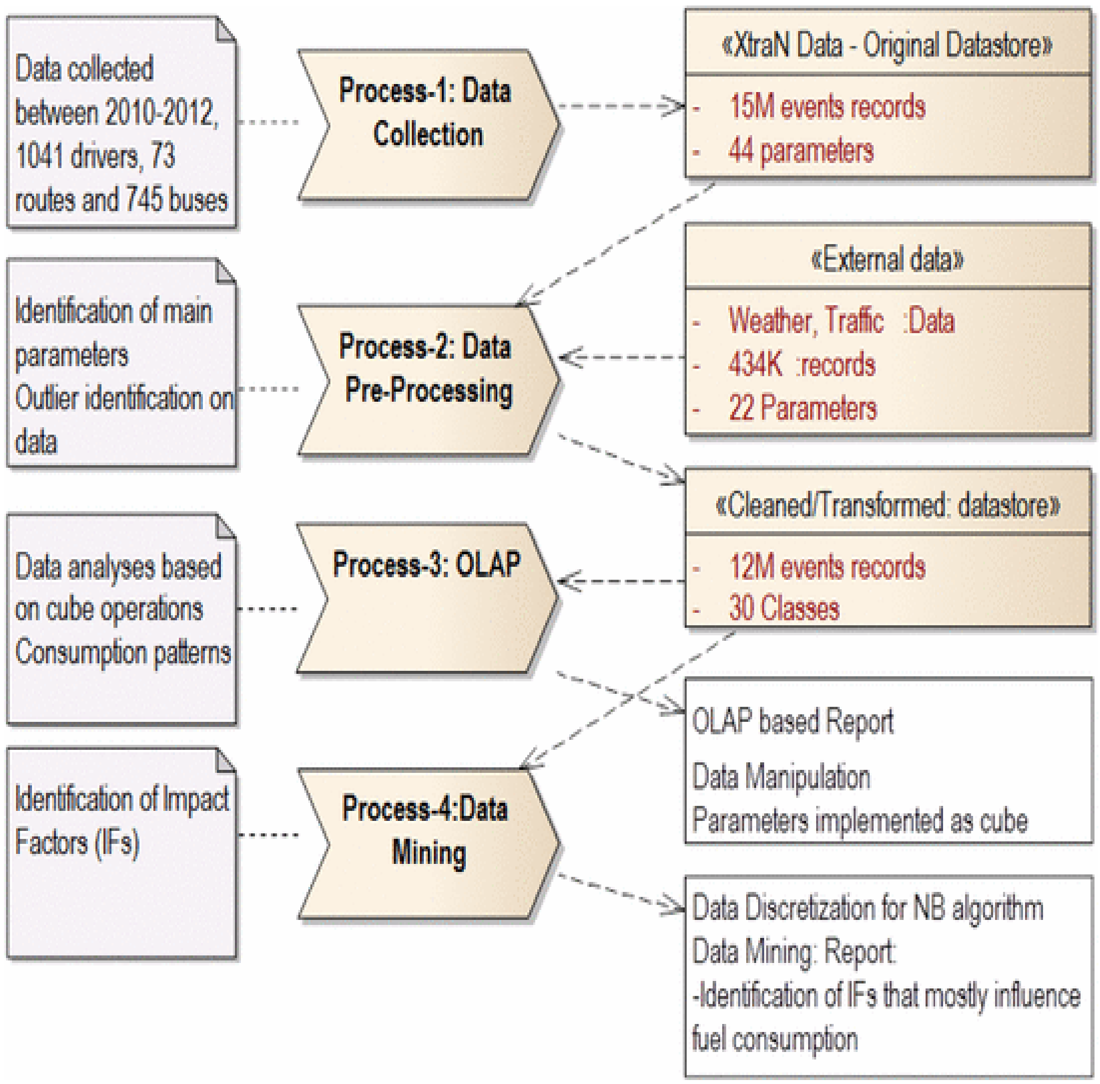

Fig. 3.1. Work methodology

3.1.2 Data Preprocessing

This process is supported by an SQL SSIS package to load a star-shaped data repository. Event records were analyzed to identify error measurements and inconsistent data based on outlier identification. By using the SQL tool and the Remove Outliers wizard, can either display a graph, a line or a bar chart, The developed Data Warehouse follows the “star shape” design principle proposed. The analysis was conducted per Driver, Route, Bus Engine, Date and Time. Historical meteorological data was used as a characteristic of the day as well as a single meteorological station. The Weather Underground site was designated to represent the geographical area covered. The 66 parameters collected were grouped by their origin into pre-defined classes. The data collected from CAN buses was reduced and transformed into 12 pre-defined classes. This process was performed by Tecmic experts, due to their deep knowledge of the data acquisition process. The information regarding engine RPMs was divided into three classes (green, yellow and red) and based on the engine type.

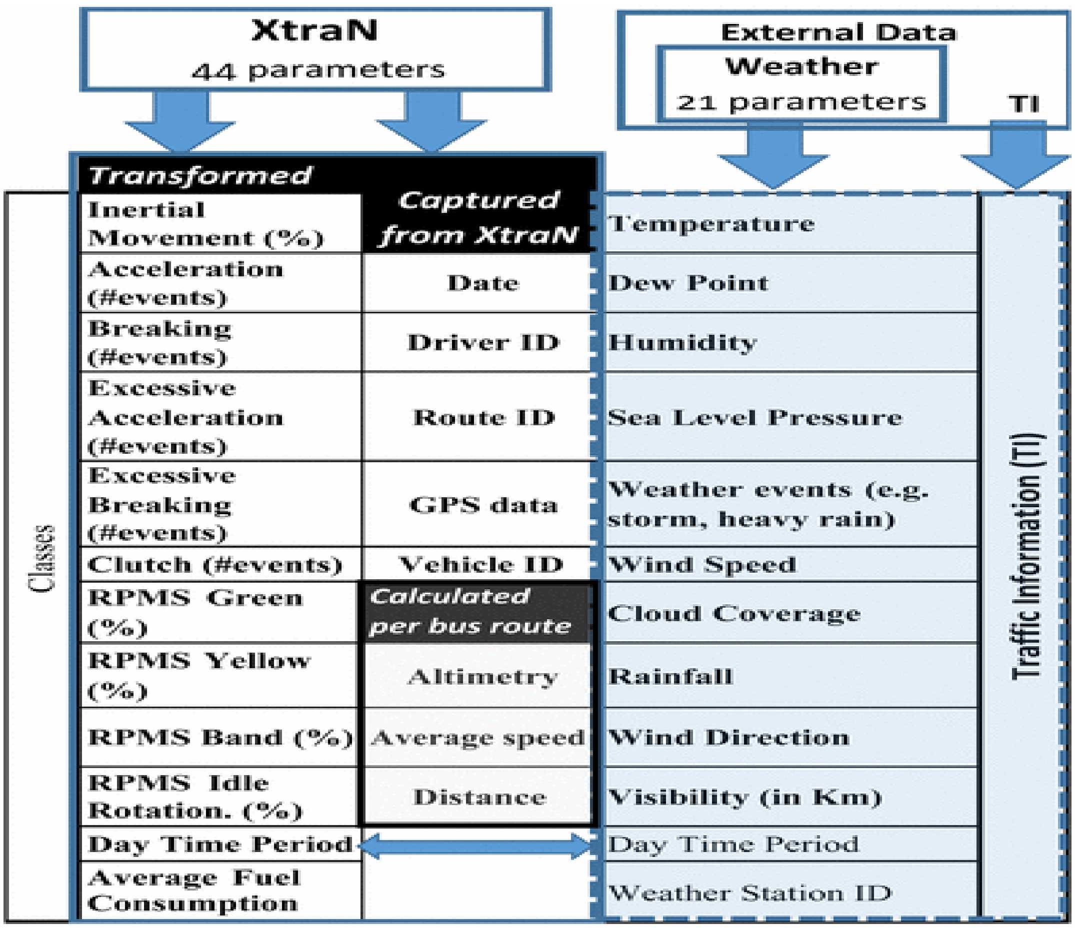

Fig. 3.2 Data transformation: the major parameters involved

The acceleration and braking events also allowed the measurement of the intensity (based on two pre-defined angular information concerning the acceleration and braking on the vehicle's pedal), thus representing excessive braking and acceleration events when the angular information of braking and acceleration was above a pre-defined threshold. From these operational variables, some ratios were derived as cubic calculated members to obtain the ratio of the sums, instead of the common trap of obtaining the sum of ratios, as data is being rolled up, down or sliced. This data set was also enriched with more data, directly captured from XtraN, namely, date (week, semester and year), route identification, driver identification, vehicle identification and GPS data. From all this data,calculated, per bus route, the following parameters: average speed, distance and accumulated altimetry. Additional external meteorological data and traffic data, (consisting of 22 variables), were transformed into pre-defined classes. This generated data set was huge, so intensive human analysis was complex, therefore followed the “Online Analytical Processing (OLAP)” approach, which is named after a set of principles proposed by Codd

3.1.3 The Olap Process

OLAP tools are designed to simplify and support interactive data analysis, but the goal of the KD approach is to automate this process, as much as possible, towards the identification of patterns. Pattern identification is based on fitting existing data to a model or commonly make any high-level description of a set of data. The KD process comprises many activities, namely data simplification from the outliers' identification, pre-processing, the search for patterns, knowledge identification, and refinement. All these activities should be repeated in several iterations. This means that KD is ahead of what is currently supported by most standard database systems.

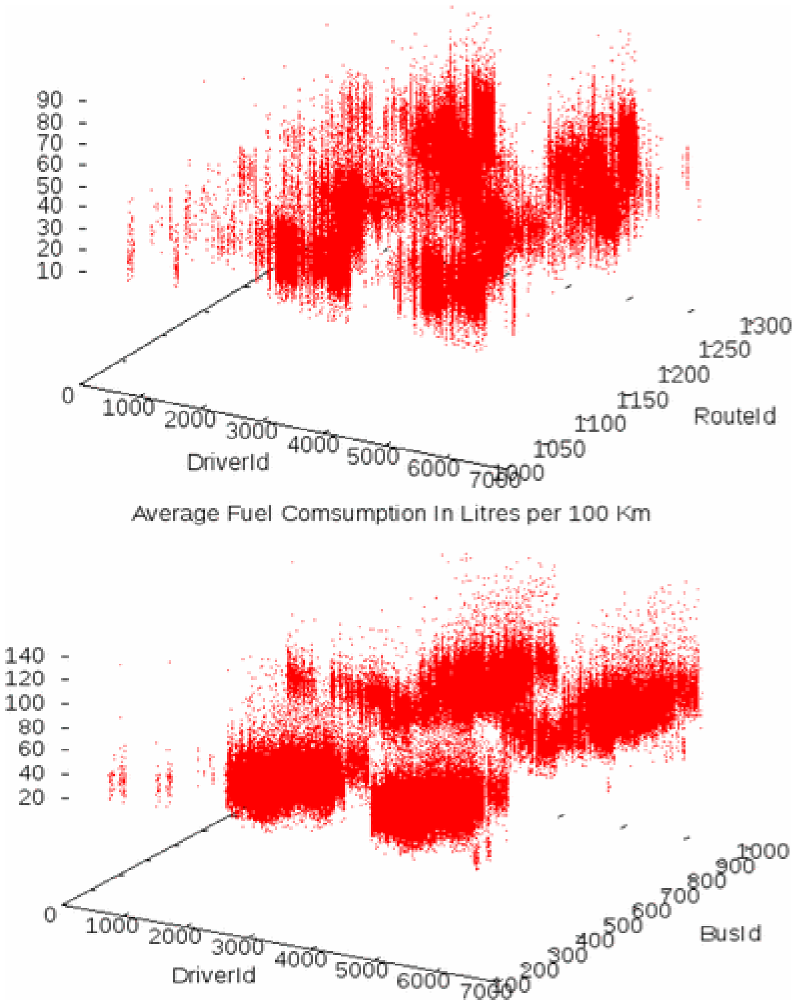

Fig. 3.3 Average fuel consumption (liters per 100 km) per Driver and Route (top, identified as (a)) and per Driver and Bus (bottom, identified as (b)). The vertical axis represents the consumption in liters per 100 km of each event.

The platform used in this research has several OLAP features available to help analyze multidimensional data interactively, from multiple perspectives. In general, OLAP involves three analytical operations: (i) consolidation or roll-up; (ii) drill-down; and (iii) slicing and dicing.

Slicing and dicing with the multidimensional cube (as a result of Process-2), supports the most common questions, such as:What are the most efficient vehicles operating on a given route This can be verified by the combination of both graphics represented in Fig 3.3 For example, the average consumption of bus 886 on route 1021 is 37.2 liters per 100 km whereas it is 54.2 liters per 100 km on route 1024. This is one of the biggest and most difficult buses to operate in an inner city route with narrow streets. Fig.3.2 shows the consumption based on drivers and route taken and Fig.3.1 b shows the consumption based on drivers and buses driven. As suggested in Fig.it is necessary to check the diversity of the consumption in drivers using the same route or bus. For example, on the route identified as 1024, the average consumption goes from 37.2 l/100 km (liters per 100 km) to 61.8 l/100 km amongst different drivers. This OLAP analysis allows multiple perspectives in multidimensional data arrays, And can aggregate data in one or more dimensions interactively. An example of this approach is the identification of above-average consumption from major drivers on an inner city route. The comparison with weather variables makes it possible to verify that this high consumption was on rainy days, especially with heavy rain. From this route, it was possible to identify problems with storm water runoff, for example, flood situations. Fig.3.3 shows a diversity of consumption levels that have an impact on drivers, routes and buses, but are not able to identify their cause.

3.1.4 The Data Mining Process

The major objective with the data mining process was the analysis of fuel consumption per driver. The approach was complex due to the 30 dimensions involved. In the literature, can see several data mining approaches and algorithms, but because provided data, used the Naive Bayes (NB) approach supported by the Microsoft platform (SQL Server 2008R2). This approach has a better performance with discrete data. For this reason performed a discretization process based on the following:

1)Heuristics were applied based on pre-defined criteria. One example is the time variable, for example, 8.35am or 9.45am has no particular meaning. When it was needed to extract knowledge, it is better to associate 8.35am to a discrete period of time, e.g. the “morning rush hour” or 10.45am to “day time”. So the time associated with the class (day time period) is divided into five subclasses: week (periods from 5am to 8am, 10am to 5pm and 8pm to 11pm); morning rush hour (8am to 10am); afternoon rush hour (5pm to 8pm); weekend and bank holidays, night period (11pm until 5am). Another example is the data variable that can be classified in winter or summer period, weekday or weekend, and also, the school and holiday period.

2) Equal area clustering based on the method of equal areas The main idea is to divide the data population in subclasses with approximately the same number of events. An example of this is the average fuel consumption by imposing it onto different type in the training data set, with the input of five subclasses. This discretization process is based on an interactive maximized expectation algorithm, in order to divide training data into groups of similar population size. The output of this process applied to is shown in Fig. 3.2 with the creation of five subclasses. This technique was selected due to the presence of pronounced peaks, and because this method selects ranges of buckets to contain equal quantities of cases. A result of this process is the subclass division in Fig. 3.2, where the disparity of consumption events from 20 to 110 liters per 100 km was divided into five subclasses.

3)The division of data population based on percentage. And decided to divide each class into five subclasses, namely: on average, below average, above average, extreme above average and extreme below average. For example, the acceleration events (class), captured from XtraN data, may have events ranging from 100 to 3500 (these numbers mean the number of times in 100 km that the accelerator pedal was used). The first subclass, ranges from 50 to 510, so the upper limit corresponds to 15% of 3400 (maximum value less minimum value). The second subclass, goes from 510 to 1190 (so the upper limit corresponds to 35% of 3400). The third subclass, goes from 1190 to 2210 (so the upper limit corresponds to 65% of 3400). The fourth subclass, goes from 2210 to 2890 (so the upper limit corresponds to 85% of 3400). And the last subclass, goes from 2890 to 3400.

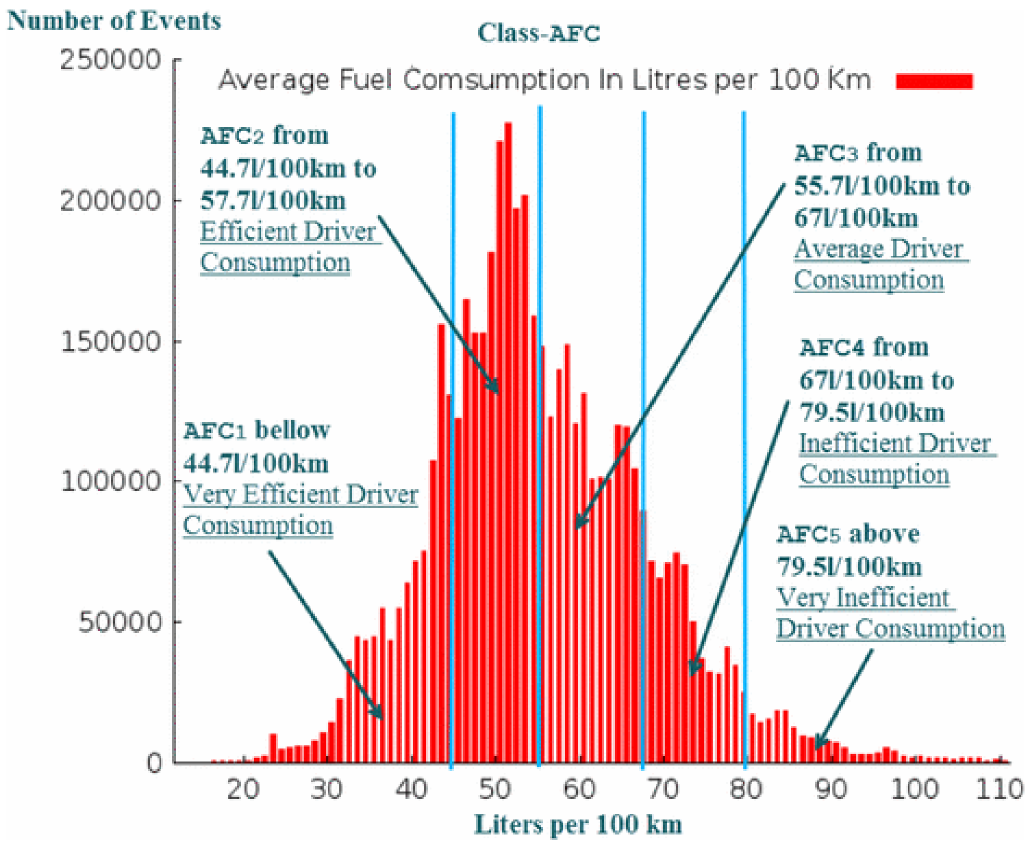

Fig 3.4 subclasses created from the discretization process.

Fig 3.4 shows the result of this discretization process for the main event classes. Since it is not possible to show the representation of all data, it highlight a few major findings: the number of excessive braking events per 100 km occurs more than excessive acceleration (approximately twice as much) but the number of events tend to occur more than events per 100 km. These events—acceleration, braking and also clutching—are related to traffic conditions and weather conditions (especially heavy rain). Traffic situations, on average, also increase these events.

After the discretization process and taking into account the data available, use Naive Bayes' algorithm to identify the main IFs that may have a major influence on fuel consumption. This process is performed based on the estimation of probabilities, which means the probable fuel class determination based on measurements performed using the Bayes theorem, where represents the five subclasses and the number of event classes:

a diversity of probabilities can be calculated using the 3-year data set. Table 3.1 shows the highest values for these probabilities, for two lower fuel consumption subclasses. This process allows the identification of the IFs that most influence the reduction of fuel consumption, namely: (1) lowest clutch use; (2) maximize the time using inertial movement; (3) maximize the time of rotation in idle; (4) minimize the time of excessive engine rotation, avoid yellow and red band (this is already avoided); (5) avoid daytime periods of traffic.

Event Classes		**Subclasses**				
	Abr	**1**	**2**	**3**	**4**	**5**
Inertial Movement (%)	IM	<3	3 to 5	5-8	8-11.5	>11.5
Acceleration (#events)	Ac	<510	510-1,190	1,190-2,210	2,210-2,890	>2,890
Braking (#events)	Br	<500	500-1,155	1,155-2,145	2,145-2,805	>2,805
Excessive Acceleration (#events)	EAc	<38	38-88	88-163	163-213	>213
Excessive Braking (#events)	EBr	<50	50-116	116-217	217-283	>283
Clutch (#events)	Cl	<276	277-604	604-976	977-1,223	> 1,223
RPMS Green band (%)	RG	<4	4-15	15-50	50-80	>80
RPMS Yellow band (%)	RY	<0.5	0.5-2	2-4.5	4.5-9	>9
RPMS Red Band (%)	RR	<0.1	0.1-0.2	0.2-0.5	0.5-1	>1
RPMS Idle Rotation. (%)	RI	<24.2	24.2-31.1	31.1-36.8	36.8-43.2	>43.2
Day Time Period	DTP	Weekday	Morning rush hour	Afternoon rush hour	Weekend and Holidays	Night

Table 3.1 Discretization Intervals of the Main Event Classes

NB Probabilities	Sub-Classes (X)									
	Cl_1	RY_1	Cl_2	IM_5	RI_4	RI_5	DP_5	DP_4	RY_2	IM_4
$P(X\|AFC_1)$	24.5%	2.1%	15.1%	10.5%	8%	7%	3.6%	3.3%	2.5%	0.1%
$P(X\|AFC_2)$	12%	24%	5%	2%	12%	11%	2%	1.5%	10%	9%

Table 3.2 Top NB Probability Subclasses for AFC1 and AFC2

Weather parameters were also considered and have concluded that extreme weather has an impact on fuel consumption. This effect does not appear as the major consumption parameter listed in Table 3.2, due to good weather conditions in Lisbon. Examples of this can be seen on temperature and visibility distance distribution events in Fig. 3.3. Extreme weather events are very few, but that affects driver and engine behavior. The main parameters on extreme weather, are rain (mainly heavy rain that appears in 40 days of the 3-year period, representing less than 0.1%), high temperatures (above 30 °C) and some fog in weather events. These events represent less than 1% of weather events and are mainly the high temperature during some of the days in the summer period. Found a correlation between hot days, with the class engine rotation in idle thus increasing times, because the drivers do not tend to turn off the engine between route carriers, mainly due to the use of air conditioning.

Only taking weather variables into consideration, the NB probabilities are represented in Table 3.3 only taking into account the effect of weather and distributed by their impact on the five *AFC's*. The weather data was captured on an hourly basis on most of the available parameters. The majority of these parameters were divided into three subclasses, for example, the temperature was divided into for temperatures below 4 °C,

		All	AFC_1	AFC_2	AFC_3	AFC_4	AFC_5
Probability	$P(T_2\|AFC_x)$	70%	90%	22%	79%	42%	1%
	$P(R_4\|AFC_x)$	19%	1%	42%	12%	44%	1%
	$P(R_5\|AFC_x)$	5%	0%	1%	7%	1%	39%
	$P(Fog\|AFC_x)$	2%	1%	9%	0%	1%	1%
	$P(T_3\|AFC_x)$	1%	5%	2%	1%	1%	43%
	Other Weather classes	3%	3%	24%	2%	6%	16%

Table 3.3 NB Probabilities for all Consumption Classes Taking Into Account Only Weather Parameters

for the temperatures in the range of 4 °C to 30 °C and for temperatures above 30 °C. The same for points of pressure, humidity or dew. The rain was divided into four subclasses: no rain, drizzle (drops with diameters of less than 0.02 inches falling closely together), light rain (less than 0.1 inches in an hour) moderate rain (more than 0.1 and less than 0.3 inches in an hour) and heavy rain when it's above 0.3 inches in an hour. Visibility was mainly satisfactory during the 3-year data:only had bad visibility for 20 days, totaling 80 hours. This visibility parameter is divided into two subclasses: fog and good visibility. The the (heavy rain) and (hot temperature) being the most influential. The rain and temperature subclasses are also the main influential weather parameters. From the data available, there was also a correlation between events and traffic. In a day's events, traffic news increases and also its severity. There is also a correlation between the events with traffic information and the class (the time percentage of an engine in idle increases).

The majority of events above 40% were traffic situations and on the left side of Fig. ,some routes show a higher percentage of ri, which means a route with more traffic and could also imply that the buses were overcrowded with passengers (longer times at bus stops also increased this time percentage). If the traffic is reduced, the increased time on the class could mean longer times at bus stations, in the loading and unloading process, due to the high flow of passengers.

Actual Class	Predicted Class				
	AFC_1	AFC_2	AFC_3	AFC_4	AFC_5
AFC_1	90%	9%	1%		
AFC_2	9%	75%	9%	6%	1%
AFC_3	1%	9%	70%	13%	7%
AFC_4		6%	13%	64%	17%
AFC_5		1%	7%	17%	75%

Table 3.4 Confusion Matrix With the Precision Measurement

NB also allows the prediction mode that is used for the estimation of driving actions in fuel consumption.Used 70% of the data for training (around 8 M event records) and the others to evaluate the results. shows the confusion matrix of the results in terms of precision, with the predicted class in columns. Precision is lower at high subclasses and most of this wrong classification is due to measurements near the subclass' limits. has a 90% correct prediction and 9% of errors are due to values near the class limits. This precision worsens in high subclasses mainly because 92.3% of the driver population belong to the first three subclasses and middle subclasses have more neighbors. Since the main goal of this research was the identification of IFs,Did not explore this prediction aspect.

CHAPTER 4

PERFORMANCE EVALUATION

4.1 INTRODUCTION

Taking into account the impact of the driving style of bus fuel consumption, the data from this study shows a huge disparity among drivers' consumption as seen in Table 4.1 shows the average annual consumption among all driving events and the results show a continuous improvement. These improvements were due to the introduction of this study as a monitoring activity with the identification of IFs.

AFC 2010	AFC 2011	AFC 2012	AFC 2013	AFC 2014
56.6	**54.4**	**52.6**	**50.1**	**47.2**
Note: Data from 2013 was not shown in our study; Data from 2014 only included data from the first semester.				

Table 4.1: Average Fuel Consumption Of Different Classes

Considering the AFC1 subclass, the IFs on fuel efficiency for the driver's case are shown in Table 3.2 and Table 4.5, where the top-10 IFs are highlighted:

1. IF-1: Use of clutch events;
2. IF-2: Observation of optimal engine rotation;
3. IF-3: Minimum engine idling;
4. IF-4: Maximize the inertial movement;
5. IF-5: Day period associated with traffic;
6. IF-6: Use of excessive acceleration events;
7. IF-7: Use of excessive braking events;
8. IF-8: Use of acceleration events;
9. IF-9: Use of braking events;
10. IF-10: Severe weather conditions.

That means that more efficient drivers tend to use the clutch less times, thus enjoying more inertia, and the percentage using the engine to idle and less time in the range “Yellow and Green” rotations.

Sub-Classes (X)									
Cl_1	RY_1	Cl_2	IM_5	RI_4	RI_5	DP_5	DP_4	RY_2	IM_4
50%	22%	-40%	13%	-10%	-7%	5%	4%	-6%	-4%

Table 4.2 IF Probabilities to Change From AFC1 to AFC2

4.2 EVALUATION

It is similarly trivial to determine which driver styles to promote in order to maximize fuel efficiency. For example, to educate a driver to move from the into the class: emphasis should be placed on the importance of optimal clutch, engine rotation and avoiding engine running in idle. The most important IF is the clutch use per 100 km, the driver uses the clutch, on average, 276 less times in 100 km (belonging to subclass) and the drivers' use of the clutch parameter is in the interval of 276 to 604 times (belonging to subclass). This also depends on the route driven, but in the same route it is possible to see considerable differences among the drivers.

This research also permits the identification of the main IFs dependency on a scale from the strongest to the weakest. Traffic (day time periods) is also an important IF, identified as rush hours in the morning and afternoon period. Traffic is also related to weather conditions—mainly rain and bad visibility events. and fog event sub classes are present in more than 70% of traffic events. Severe weather conditions show an impact on , but due to a reduced number of these events in the Lisbon area, the overall view of this effect is neglected. However, Simulating the effect of a higher percentage of these events (more than five times) and so the effect of (heavy) rain appears in the top-5 IFs.

	All	AFC_1	AFC_2	AFC_3	AFC_4	AFC_5
Population	1,041	368	394	202	43	34

Table 4.3 Distribution of Driver Population per Classes

It shows the number of drivers in each class (the average consumption during the 3-year period), but only 1041 drivers were considered, corresponding to a complete 2-years of data registers of day time drivers that performed all 73 routes. Since each driver has different consumption patterns, the class in which he belongs is determined by the class in which he has more events.

As a relevant result of research, Identifies the following lessons learned and recommendations regarding eco-driving practices:

1. **Use the clutch moderately** (IF-1): This practice is related with the change of speed (braking and acceleration) only available in manual transmission engines (MTE). Most eco-driving studies are based on an automatic transmission engine and do not take this MTE parameter into account. This class of events averaged in the range of 300 to 650 events per 100 km, with the maximum value of 1840 events per 100 km.
2. **Keep a stable engine round per minute** (IF-2): Added few red class events (drivers did not force the engine, perhaps due to their knowledge of this monitoring process), but the yellow band class was detected in all drivers, on an average of 1% to 2% of the time and from an NB probability
3. **Maintain a steady speed** (IF-3): This is related to inertial movement. This can also be related with and that accounted for separately. This class of events averaged in the range of 4% to 8% of time, with a maximum value of 14%.
4. **Eliminate idling** (IF-4): Original data showed this was the main IF because drivers did not turn the engine off between driving periods (mainly for air-

conditioning use). The idling time is directly related to the traffic as well as the number of passengers at a bus stop. This class of events averaged in the range of 25% to 35% of time, with a maximum value of 80%.

5. **Minimize the use of braking** (is the IF-7 and is the IF-9):Defined two classes of breaking:when the intensity of braking is above a certain value. This appears, on average, for all drivers in 120 to 160 events per 100 km with a maximum of 333 per 100 km. Br events, on average, range from 750 to 1250 events per 100 km with a maximum of 3300 per 100 km.
6. **Minimize the use of acceleration** (is the IF-6 and the IF-8):Defined two classes of acceleration: when the intensity of braking is above a certain value. The appears, on average, for all drivers from 50 to 100 events per 100 km (less than) with a maximum of 250 events per 100 km. events, on average, range from 1100 to 1900 events per 100 km with a maximum value of 3400 events per 100 km (more).
7. **Other conditions and practices that might well influence the eco-driving style**: Severe weather conditions show an evidence on the impact of fuel consumption, mainly regarding temperature, which is related to the use of air-conditioning and heating. Rain also has an important impact due to its direct correlation with traffic. In addition, the weight of a vehicle, air resistance, approach to curves, tire pressure or the vehicle's deterioration status directly influences the level of fuel consumption.

		AFC_1	AFC_2	AFC_3	AFC_4	AFC_5
IF-1: Clutch Class #events per 100km	Cl_1 <276	87%	47%	30%	29%	20%
	Cl_2 [276-604]	12%	46%	50%	32%	30%
	Cl_3 [604-976]	1%	5%	15%	17%	29%
	Cl_4 [976-1223]	0%	2%	4%	13%	11%
	Cl_5 >1223	0%	1%	2%	9%	10%
IF-2: RPMS Yellow Band in time %	RY_1 <0.5%	77%	50%	30%	20%	5%
	RY_2 [0.5-2]%	17%	34%	35%	25%	43%
	RY_3 [2-4.5]%	1%	9%	13%	11%	16%
	RY_4 [4.5-9]%	5%	5%	12%	27%	16%
	RY_5 >9%	1%	2%	10%	17%	20%
IF-3: RPMS Idle Rotation in time %	RI_1 <24.2%	10%	15%	18%	44%	70%
	RI_2 [24.2-31.1]%	15%	27%	22%	28%	17%
	RI_3 [31.1-36.8]%	24%	23%	29%	14%	9%
	RI_4 [36.8-43.2]%	28%	22%	21%	9%	3%
	RI_5 >43.2%	23%	13%	10%	5%	1%
IF-4: Inerial Movement in time %	IM_1 <3 %	1%	4%	11%	20%	42%
	IM_2 [3-5]%	9%	13%	38%	42%	32%
	IM_3 [5-8]%	37%	30%	20%	22%	19%
	IM_4 [8-11,5]%	27%	29%	18%	11%	5%
	IM_5 >11,5%	26%	24%	13%	5%	2%
IF-5: Day Time Period Class	DTP_1	21%	31%	36%	23%	19%
	DTP_2	10%	14%	16%	25%	29%
	DTP_3	9%	14%	17%	26%	29%
	DTP_4	31%	20%	15%	13%	12%
	DTP_5	29%	21%	16%	13%	11%

Table 4.4 Normalized Probability Values of the Five Top IFs Divided by the Five Pre-Defined Classes

Table 4.4 shows the probability distribution of the top-5 IFs for all the subclasses. For example, for the subclass it is important for the IF-1 (Clutch) to be in subclass one, so drivers should avoid the use of unnecessary clutching, thus enjoying more inertia (IF-4), the percentage of the engine use to idle (IF-3) and less time in the range “Yellow” rotations (IF-2).

CHAPTER 5

CONCLUSTION

Considering road conditions, the number of passengers on the bus, and especially traffic can influence fuel consumption. Overall, findings show that adopting appropriate driving styles can reduce fuel consumption on an average between 3 to 5 liters per 100 km. This can save 15 to 30 liters per bus in just a working day. Taking into account fuel prices, these savings represent 20€ to 40€ a day, per bus. Considering the working days in a year and with around 1,500 involved drivers, this may impact significant savings that can go up to 1.5M€ per year. Another important output of the current study is the possibility to collect real-time data. Based on the events collected in real-time and available in a central database, it is possible to apply this methodology (based on the discussed processes) to dynamically analyze and identify major driving parameters that have an impact on the reduction of fuel consumption. However, that approach should be slightly different because it should allow the monitoring of a huge set of driving parameters and simultaneously identify the most influential ones. As future work Different impact factors can be considered The driver's state of mind could also be another important parameter to be considered in this process.

REFERENCES

[1] M. Chan, A. Herrera, and B. André, “Detection of changes in driving behavior using unsupervised learning,” Proc. IEEE Int. Conf. Humans, Inf. Technol., 1994, vol. 2, pp. 1979–1982.

[2] U. Reiter, “Modeling the driving behavior influenced by information technologies,” in Proc. Int. Symp. Highway Capacity Level Service, 1991, pp. 309–320, (Ed. Brannolte).

[3] M. Natale, H. Zeng, P. Giusto, and A. Ghosal, Understanding and Using the Controller Area Network Communication Protocol :Theory and Practice. New York, NY, USA: Springer-Verlag, Jan. 2012.

[4] J. Gray, A. Bosworth, A. Layman, and H. Priahesh, “Data cube: A relational aggregation operator generalizing group-by, cross-tab, and subtotals” in Proc. 12th IEEE Int. Conf. Data Eng., 1995, pp. 152–159.

[5] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Advances in Knowledge Discovery and Data Mining. Menlo Park, CA, USA: AAAI, 1996.

[6] G. Mariscal, O. Marbán, and C. Fernández, “A survey of data mining and knowledge discovery process models and methodologies,” Knowl. Eng. Rev., vol. 25, no. 2, pp. 137–166, 2010.

[7] J. Almeida and J. Ferreira, “BUS public transportation system fuel efficiency patterns,” in Proc. 2nd IMLCS, Kuala Lumpur, Malaysia, 2013, pp. 4–8.