SAPIENZA
UNIVERSITÀ DI ROMA

Identifying Eco-driving Behavior and Estimating CO2 Emissions with Machine Learning

Facoltà di ingegneria dell'informazione, informatica e statistica

Corso di Laurea Magistrale in Data Science

Candidate

Malick ALexandre Ngorovitch Sarr

ID number 1788832

Thesis Advisors

Prof. Ioannis Chatzigiannakis

Co-Advisor

Prod Aris Anagnostopoulos

Academic Year 2018/2019

Thesis defended on 31 October 2019
in front of a Board of Examiners composed by:

prof. Tardella Luca (chairman)
prof. Brutti Pierpaolo
prof. Chatzigiannakis Ioannis
prof. Cuomo Francesca
prof. Galasso Fabio
prof. Liseo Brunero
prof. Maggino Filomena

Identifying Eco-driving Behavior and Estimating CO2 Emissions with Machine Learning

Master thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Version: October 21, 2019

Author's email: ngorovitch@gmail.com

Dedicated to
All the driver's out there trying to save fuel

Abstract

Transport is responsible for nearly 30% of the EU's total CO2 emissions, of which 72% comes from road transportation. The European Union, for the past few years, has set a goal of reducing emissions from transport by 60% by 2050 in comparison to the levels in the previous century. One way to reduce CO2 emission implemented by the EU the past few years was to promote the idea of Eco-driving to the general population. In this paper, with the data and experiments provided by GAMECAR, we will analyze the effects of Eco-driving on car fuel consumption, identify 10 Eco-driving behaviors before estimating the impact of those driving behaviors on CO2 emission with machine learning.

Acknowledgments

I would first like to thank my thesis advisor Prof. Ioannis Chatzigiannakis of the Department of Computer Engineering, Automatic and Management at the University of Rome "La Sapienza". Prof. Chatzigiannakis was always available whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my work but steered me in the right direction whenever he thought I needed it.

I would also like to thank the experts in the project GAMECAR who made the dataset available for my research Without their passionate participation and work, the test done in this paper could not have been successfully conducted.

I would also like to acknowledge Prof. Aris Anagnostopoulos of the Department of Computer Engineering, Automatic and Management at the University of Rome "La Sapienza" and I am gratefully indebted to him for his very valuable comments on this thesis.

Finally, I must express my very profound gratitude to my parents, my family and friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Contents

Introduction

Transport is a vital part of our modern economies as the quality of most of our life depends on a modern, accessible and efficient transport system. The development of the transport industry comes with costly environmental repercussions. Indeed, it contributes heavily to air pollution, noise but most importantly to climate change. According to the European Environment Agency, the transport industry is the biggest source of carbon emission. It contributes to 27% of EU's total carbon footprint of with 70% are issued from cars [6]. Ergo, the European Union put a plan focus on decarbonizing transport with a goal set towards 'net-zero' greenhouse gas across the euro Zone by 2050 [6]. It entails that the greenhouse gas emissions are expected to be reduced by 24% by 2020 and by 32% by 2030 [23].

One solution implemented in order to reach that goal is the introduction and promotion of eco-driving. Eco-driving could be defined as an attempt to change people's driving behavior through so-called advises such as shifting gears sooner, driving at high gear with low speed, as smooth driving by anticipating traffic, avoiding sharp acceleration or deceleration among other [8]. Those advices, through the use of efficient and safe driving, theoretically lead to the reduction of greenhouse gas emission,fuel consumption and of the number of road accidents [51].

The main objective of Eco-driving, compared to the basic method is to reduce fuel consumption while not affecting the average vehicle speed. The intuition behind developing such driving techniques comes from the mathematical relationship that exists between the gas emission/fuel consumption and the machinery within the car meaning that the more intensively we drive the more fuel we use [26]. It has as an effect the reduction of the average fuel consumption without any repercussions on the average speed of the vehicle. Consequently, there should not be any decrease in the quality of the driver's usual drive, but eco-driving should have a significant contribution to environmental protection through the reduction of harmful emissions but as well towards the driver's yearly bill for fuel as a result [26].

The effectiveness of eco-driving depends on the ways experiments were designed, on the type of evaluation methods used and on the external factors affecting the drivers[60]. The fuel reductions prior and post eco-driving instructions in several experiments shows a decrease in fuel consumption varying from 5% to 25% and from 10% to -05% depending on the type of road the vehicles are using such as highways, urban road, off-roads etc [59].

The objective of this paper can be separated into two main tasks:

1. The ability to discover eco-driving behavior periods within a drive. Meaning that using a route, we observe periods in which drivers followed or not eco-

driving advices

2. The ability of machine learning models to predict CO2 emission based on the above (eco-driving advises)

In summary, this paper aims to put in place a method that will enable the identification of driving events and provide a way to quantitatively derive eco-driving advises. The basic application of such advises can help drivers better comprehend their driving behavior. On top of that, we will see based on those driving behaviors, if we are able to derive the total CO2 emission at a particular time within a drive.

The organization of this paper would be as follow. The next chapter will introduce the background research that led to the development of the above model. Then, we will discuss the data collection and the methods applied to the dataset. In the following chapter the result will be presented before ending with a discussion and direction of future works.

Chapter 1

Literature review

With the goal of decreasing general CO2 emissions levels for a "greener" earth, we observed a need for sustainable worldwide road transport. Consequently, there has been an increased effort by most EU member states for the decrease of CO2 emissions and the promotion of energy efficiency in all sectors. It resulted in the initiation of various projects to reach that goal, one of which was the promotion of eco-driving. Eco-driving, as said before, are techniques that involve the use of safe and smart driving style which leads to the reduction in fuel consumption, in the number of road accidents, in greenhouse gas emission or in engine noise. In this section, we will be exploring the type of work/research that has been done relating to eco-driving. First, we will see how eco-driving has been implemented, then we will mention the important components of eco-driving, before noting some research issues and challenges we observed regarding eco-driving. This should give us a solid base which we used to make our research.

1.1 Implementation of eco-driving

The application of eco-driving techniques can be divided in multiple ways. Here, we will first be exploring how eco-driving was applied to operation level, choices of vehicle and choice of route, then we will see what studies have been done for training the drivers to be eco-drivers.

1.1.1 Operation Level

The operation level relates to the behavior of the driver during a drive [21]. There has been a lot of advancements in that aspect that somewhat aimed at reminding the drivers to drive sustainability. Ergo, a lot of focus has been put into implementing devices that would provide real-time or trip-summarized eco-driving advices during or after trips. It would allow drivers to adjust their driving behavior for the maximization of fuel economy. An example of such a project is from the project GamECAR whose data will be used for analysis. This EU funded project aimed at motivating the promotion of eco-driving through a multiplayer gaming platform giving the drivers the opportunity to set missions and invite others to participate collaboratively or competitively.[10] [56] Through the use of visualization techniques such as Augmented reality, they managed to monitor the eco-driving score evolution while providing the drivers with a personalized plan for improvement [55] [47].

Other applications about eco-driving have been developed and you can see the illustration of few on them in figure 1.1. For example, we have a few android applications that could display second-to-seconds eco-score ranging from 'bad' to 'very good'. The previous result was based on three fuzzified variables which are velocity, road slope and power consumption [29]. Going a step further there is a smartphone application called DrivingCoach that could detect the users driving behavior and pattern and then suggest new behaviors to reduce your fuel consumption in real-time [9]. In here again 8 variables have been used in a fuzzy way to provide feedback on the user's driving.

Another implementation of operation level apparatuses is a device that allowed the introduction of eco-driving feedback which provided real-time trip emission information [17]. It basically used a color scheme to show real-time feedback of both the fuel consumption and CO2 emission. This information displayed on a dashboard showed features like average speed, travel time, fuel consumption levels along with the C02 emission. The results of such devices on a small sample of drivers within a city showed an average improvement of fuel economy by 6% [21] [26].

1.1.2 Vehicle Choice

The vehicle choice relates to the type of vehicle you are driving and the emission levels produced by the energy released by its motor and the equipment present within the vehicle. So before buying a car, knowing if a particular car is friendly, plays a major role in the driver's potential consumption levels. A lot of policies and regulations have been implemented by the EU over the past two decades to provide a more eco-friendly automobile market. The definition of eco-cars is:

- Vehicles with alternative fuels such as ethanol with energy consumption:
  - 9.7 m3 CNG/100 km
  - 9.2 l/100km
  - 37 kWh electric energy/100 km
- Vehicle with conventional fuels (hybrids included) with CO2 emission less than:
  - 120 g/km
  - PM $<$ 5 mg/km for diesel-powered vehicles

[54]

A lot of governmental policies have been used pushing consumers towards buying a more eco-friendly car. For example in the EU, if the average CO2 emission of a car exceeds the maximum C02 threshold in a year, the manufacturer has to pay a fee called excess emission premium for every car registered in the market [1]. That fee varies from :

- €5 for the first g/km of exceeding

(a) CarbonDiem (b) Green Meter

(c) Fuel Good (d) Eco Drive Speedometer

Figure 1.1. Sample of ecodrive mobile applications

- €15 for the second g/km
- €25 for the third g/km
- €95 for each subsequent g/km.

[1]

In Japan, the central government imposed tax incentives called the Green Tax Scheme for buying low emission and fuel-efficient cars which are identified as such by a certification process[54]. In Germany, measures to promote electric mobility have been put in place. Such measures include a tax break for zero-emission vehicles as well as policy measures such as special lanes or parking spots for zero-emitting cars. In Sweden, an incentive scheme that depends on several performance parameters has been applied. As a result, a person may get up to 1100€ for the registration of an eco-car meeting certain environmental criteria.

Besides the benefits various government started to implement when buying a new car, other factors should be taken into consideration concerning the vehicle choice. Things like vehicle weight, tire pressure, aerodynamic drag, and vehicle maintenance should affect the driver's decision into acquiring a particular car. Indeed vehicle weight should be minimized as the lighter the car is, the less energy is required to move the car which translates into lower fuel consumption. The same principle goes about the general maintenance of the car and its components may reduce fuel consumption. On top of everything, minimizing aerodynamic drag, like closing fully the windows or avoiding to put external cargo, will decrease the resistance the air has on the car ergo decreasing the amount of energy required during the drive [32].

1.1.3 Route Choice

Route planning relates to planning what road to take before the drive taking into account things like traffic congestion, time of the day, etc. [21]. Map applications such as Fueoogle were developed to address the route choice. It uses a participatory sensing service that maps vehicular fuel consumption to enable drivers to figure out the most efficient route for their vehicles between two endpoints [49].

With more recent implications, the term eco-route has been put in place which purpose is to recommend the driver's cost-effective routes in real-time. It is based on the hypothesis that one can decrease time traveled in order to maximize lower fuel consumption [35]. The obtention of those eco-routes has been done through various techniques mainly using graph theory principles where the nodes represent junctions, the edges are roads and the costs are estimated by the energy/fuel needed in order to travel between two connected nodes [35].

Eco-routing navigation systems have now been developed in order to full-fill that purposed [16]. They are composed of the connection to a Dynamic Roadway Network database, which is a roadway network integrating real-time and historical information about traffic from various data sources with the help of an embedded data fusion algorithm. Those are paired with an emission parameter set which is defined as a compilation of emission factors for different types of vehicles, road characteristics, and conditions of traffic. The third component would be the routing engine which is the optimal shortest path algorithms used for route calculation paired finally with a user interface that receives and display route information to the users.

1.1.4 Training

The goal of an eco-driving training program is to train experienced or new drivers to become sustainable drivers. In order to do so various strategies and studies have been put in place in order to test the efficiency of the training programs. Fig **??**

summarizes a series of published eco-driving training programs and their effect on fuel consumption before or after straining. Generally, we can see that those training programs help reduce fuel consumption by 2-15%, which depends on the type of program [32]. This variation is mainly due to the fact that each training program is different and varies in terms of strategies, driving conditions, country, etc. [30]

The programs are usually divided into 3 steps which are theoretical training, practical training and the combination of both. For instance, a study shows the effect of simple advice and eco-driving training and how it changes driving behavior. As a result, it showed that the fuel consumption decrease by about 12% when given simple eco-driving advice which was higher than the full eco-driving training which experiences a decrease of 11% [22]. Another study compared the effectiveness of online learning and hard-copies, of the later and 2 hours of driving lessons, of the online learning and 1.5 hours of driving, of the combination all the previously cited and finally on online learning and half a day workshop [33]. The result showed that all the above 5 actions had effects on fuel economy but no significant difference between them [33].

Another paper compared the results of two eco-driving strategies that are an in-car feedback system and a feedback system with a personal trainer. The results concluded that both strategies resulted in a fuel saving of about 6.8 % but no difference was observed between the two strategies [53]. Controversially, another study reported that solely theoretical training did not have any effect on either the short term nor the long-term driving behavior, it stated that practical training is absolutely necessary [50].

In general, the results obtained by most studies are usually recorded straight after a set experiments which usually shows a positive outlook. However the long term studies show that the eco-driving skills tend to disappear over time. It is mainly due to the fact that it is generally very hard for someone to change their driving habits, especially if one has been driving a particular way for years [32] [2] [12].

Additionally other factors influence fuel consumption, ergo the training result may be affected by it. We will go through those factors later in the paper. A survey showed that eco-driving practical learning were more effective than high level of theoretical learning or motivation [39]. Additionally, eco-driving training was more efficient in city rather than in highway and received better result for manual transmission rather than automatic transmission cars [12].

On average, most studies were made for existing experienced drivers with very few considering new learning. The ECOWILL European project was such a study that focused its effort on new drivers for a period of 3 years in 13 different European countries [27]. That training program provided eco-driving seminars for both experienced an new comers. The new drivers were taught standard driving techniques that include the golden and silver rules of eco-driving. The ECOWILL project aimed at educating 10 million learner and beginner drivers with the principles of eco-driving and the long lasting benefits sustainable driving has. By the end of the project, the European Commission made eco-driving a mandatory component of the driver's test in all states in the European Union[27]. As a result, studies have shown that eco-driving, for learners has been instated and shaped through their education whereas for the experienced drivers, their understanding was broader which included strategic to tactical decisions [**?**].

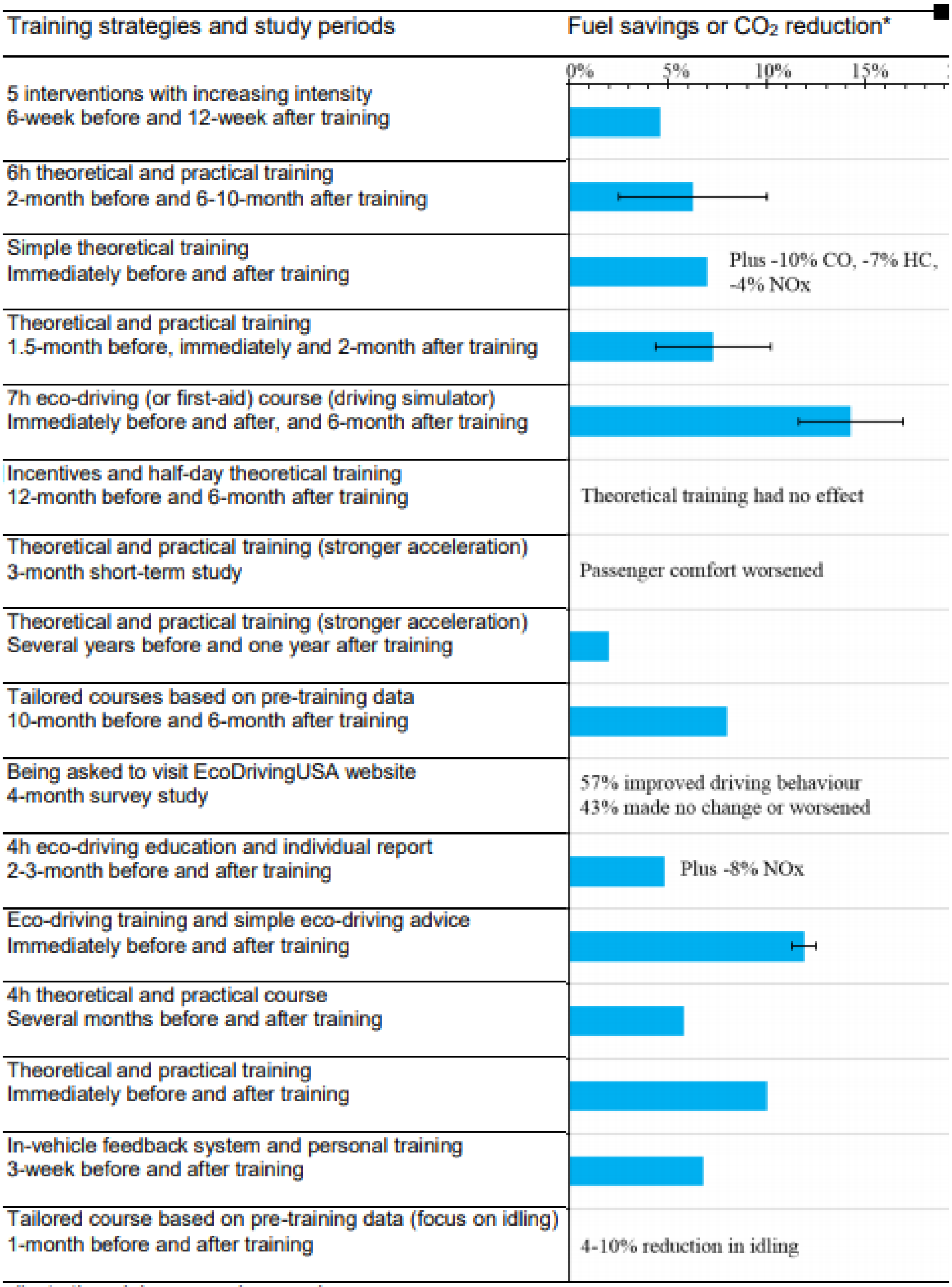


Figure 1.2. Efficiency of various training program on fuel reduction [32]

1.2 Eco-Driving Standards

Eco-driving is made up of a fair amount of components that are used to establish rules regarding driving style. The project ECOWILL is a EU wide project whose main goal was to widen the implementation of eco-driving techniques for both learner drivers and licensed drivers established the so-called golden and silver rules of driving. They are considered as smart and safe driving techniques which can lead to significant fuel savings [27]. In this section, we will explore what are the golden and silver rules of driving and why implementing them may save fuel consumption according to various research.

1.2.1 Golden Rules of Eco-driving

The Golden rules of eco-driving are the 5 most important eco-driving tips in other to decrease fuel consumption [27]. They are generally considered key and are considered as follow [27]

1. Greater anticipation
   The idea behind anticipating traffic is based on the idea of constantly (as much as possible) keeping a smooth driving style. Ergo anticipating potential events by looking and paying attention to the traffic stream will decrease the amount of forceful breaking and avoiding unnecessary acceleration [52].

2. Maintaining a steady RPM at Low Speed
   Diesel engines reach usually their optimal level of efficiency at lower engine speeds, so a maximum engine speed of 2000 rpm for up-shifting is recommended. For a petrol car, it is usually at around 2500 rpm [37].

3. Shifting up early
   A study even showed that the fuel consumption at the same average driving speed can be increased up to 20% only because of the variation in the way of gear shifting. Figure **??** shows a study made by Volkswagen where generally driving with higher gear is more fuel-efficient [14].

4. checking tire pressures frequently
   It is advised to keep tire pressure at the recommended manufacturer's levels because low tire pressure reduces gas mileage. One study even showed that keeping optimal tire pressure can reduce 40% of vehicle emissions [43].

5. Removing all unneeded ancillary loads
   It is generally known that as mass increases more energy will be required for acceleration and driving at a constant speed. Indeed studies have shown that for every 100 kg of extra load, fuel consumption increases up to 7% with an average estimate in the order of 4%. Additionally, having boxed roofs or towing a trailer can increase fuel consumption by 20% to 37.2% respectively because of additional conditions such as increased air drag and rolling resistance[61].

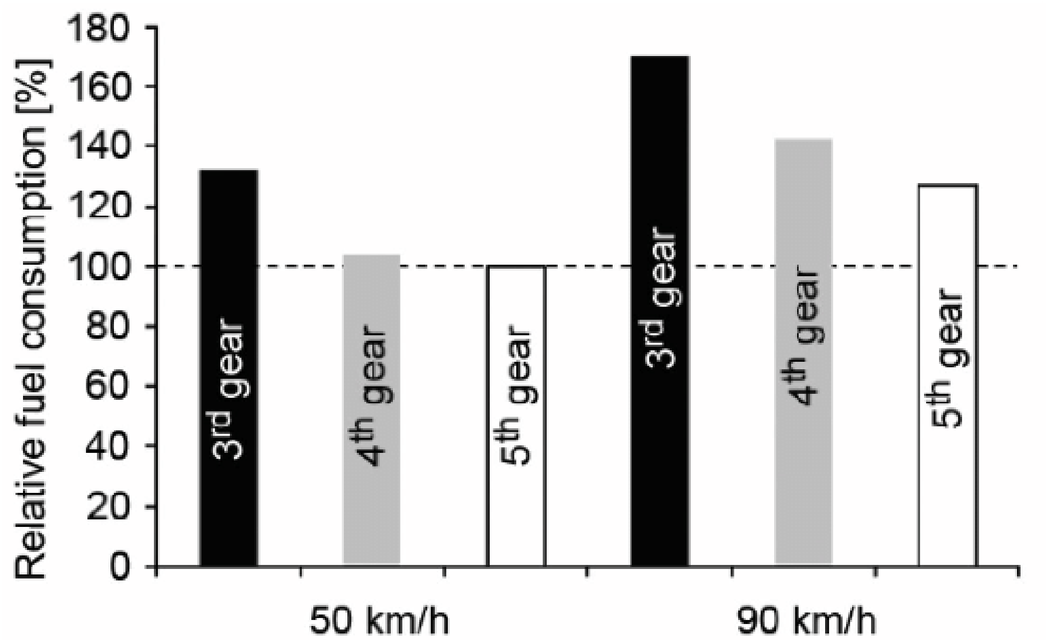


Figure 1.3. Volkswagen's Study on Fuel Consumption level at 50km/h and 90km/h at different gears

1.2.2 Silver Rules of Eco-driving

The silver rules of eco-driving are important eco-driving tips that could be followed for a lower fuel consumption

1. Choosing low emission car
   Since 2003, the technologies within the vehicles such as the powertrain, fuels, transmission or the emission control devices have been significantly upgraded. Low emission vehicles complying with current emission regulations have been built with advanced emission control technologies such as diesel particulate filters, selective catalytic reduction, three-way catalytic converters and contain sophisticated electronic systems to monitor engine operations. [21]

2. Avoiding short car trips
   More than half of all car trips are less than 5 km long. If we consider the individual driving patterns and constraints, in principle, walking, cycling or even taking public transportation could technically substitute up to 41% of short car trips a study has shown. This short car trip reduction could lead to a decrease of 5% from car travel [44].

3. Start engine when you are actually about to leave
   Even though motor vehicles perceived as using a trivial amount of fuel compared to a moving vehicle, we do not need to be running the engine a long time before we leave for a trip. Indeed, a study showed that vehicle may use up to 100 ml of fuel every 10 mins while idling [36].

4. Turn off the engine if stationary for a long period of time
Similarly to the previous silver rule, and the engine simply running is still using fuel thus turning it off when stationary for a long period of time is advisable.

5. Use low friction oil and low energy tires
The idea behind low friction oil is that it reduces the friction between surfaces in mutual contact. Using low friction oil has shown to reduce fuel consumption by 3-5% a study showed [7].

6. Close windows when driving fast
Driving with an open window increases aerodynamic drag, which as a result increase power consumption and fuel consumption. Studies have shown that it is more economical to drive under closed windows with air conditioner rather than with open windows [27] [62] [61].

7. Check car regularly
checking the state of the various equipment within a car and keeping them at an optimum level of performance can potentially decrease your general levels of fuel consumption up to 5% [62].

8. Consider alternative way of transport such as cycling, walking. public transportation etc. [27]

1.3 Components of Eco-driving

When we think about the components of Eco-driving, we are referring to the main driving behaviors, during or before a journey, used in order to decrease fuel consumption and emissions. In this section, we are going to explore some literature results that are related to the vehicle speed, the acceleration and deceleration, the vehicle accessories, the type and choice of route, the type of fuel, the effect of weight and other general characteristics that may affect the fuel consumption.

1.3.1 Vehicle Speed

One of the golden rules of eco-driving was to maintain somewhat a constant speed by applying smooth driving techniques. However, this is theoretically impossible to do in real life due to the number of unpredicted events happening while driving such as the traffic signals, the speed limits, the level of fatigue the driver's suffering, the slope of the road and all the like. Ergo, it is often recommended to drive below the speed limit [32]. Indeed, many studies have showed that based on the slope of the road [25] [32], the signal on the roads [31] and the congestion levels on the road [58], that keeping your speed below the limit would allow the drivers to drive more smoothly which will have significant impact on your fuel consumption.

Cruise control, whenever possible and available, is also recommended to be used. In this context, we cannot establish a proper basis to which speed should cruise control be set regarding fuel economy since each Internal Combustion Engine has its optimal speed for fuel economy. Thus, fuel consumption varies depending on the cruising speed. This means that the fuel consumption driving at cruising speed is

U-shaped; fuel consumption first decreases with the engine speed through reduced heat losses, reaches the optimal point then increases because of the increased friction loss [**?**].

Studies have shown that on average, the optimal cruising speed range between 50-90 km/h. It is suggested that fuel consumption increased significantly with speeds over 90 km/h, meaning that a car would use up to 25% more fuel when cruising at 120 km/h than at 90km/h [5].

In terms of policies, since driving a bit below speed limits helps save fuel, the European Environment agency did estimate that reducing the speed limits within a motorway from 120 km/h to 110 km/h, may help reduce significantly fuel level consumption by 12% for diesel cars and by 18% for gasoline car, assuming 100% compliance to speed limit and smooth driving style. On top of that, we will observe a decrease in other pollutant gases such as NOX and PM emissions [32] [6].

However, achieving the above is not an easy task as we do not live in a perfect world. The improvement of people understanding about the benefits and cost of driving generally at a suitable speed, through training programs mixed with tighter road enforcement, should help reach the above-mentioned goals.

1.3.2 Vehicle Idling

The golden and silver rules of eco-driving generally state that idling should be minimized. Indeed, as previously described, a typical car uses up to 100 ml of fuel per 10 mins of idling [36]. In the US, it is estimated that about 11.2 billion liters of fuel were wasted annually by car in idling state. Another way of looking at it is that removing all idling states would be approximately the same as taking off 2.5 million cars off the roads [48].

Various techniques are used to decrease the idling period. Those techniques may or may not entirely be controlled by the driver. First, as previously said in the silver rules of eco-driving, being idle in congestion could be avoided through proper route planning, anticipation, and smooth driving [27]. But first, through an efficient training program, people's understanding of how and what idling is must change. For example, modern cars do not need to warm up the engine. The optimal temperature will be reached quicker by simply driving the car and that applies for cold climate. Most manufacturers suggest driving off gently for about 30s to warm up the engine [20].

In addition to that, most modern cars do not get damaged by being turned off and on. A 20-second idling has more fuel consumption than a stop - turn off the engine - restart uses [20]. However, surveys have shown at least 80% of the respondents thought that staying idle for more than 30s was better than turning off the engine and more than half of the respondent believed that letting the engine running for about 2 minutes before a start (longer during cold periods) was a good habit [38]. As a result, a lot of fuel is being wasted just by outdated knowledge about cars. This is toppled by the fact that a recent survey showed that most people knew about eco-driving and had positive views about it, but their knowledge about specific fuel economy behavior was low [38].

1.3.3 vehicle Acceleration/Deceleration

The acceleration refers to the increase in vehicle speed. The golden rules of eco-driving advise avoiding sharp acceleration or deceleration which means a general reduction in aggressive driving style. A study in the US showed that aggressive driving could increase fuel consumption by about 15 - 25% in highways and for about 10-40% in a stop-go situation [54]. It is suggested that drivers should do their very best to avoid the aggressive type of driving by keeping a proper distance between the various cars, by driving smoothly and by anticipating the flow and the road ahead[32].

Controversially, other studies have suggested that driving more aggressively to a particular optimal speed might be more fuel-efficient. For example, in Sweden, an eco-driving program advised buses do accelerate strongly and earlier, at the expense of passenger comfort [2].

The effect of anticipating traffic properly allows us to change gear more efficiently and thus avoiding any unnecessary acceleration or braking. Driving aggressively produces a lot of additional pollutants such as CO2, HC NOx, etc. [28] [21] [57] A study showed that by reducing your speed in highways, one can potentially save the same amount of fuel as reducing acceleration during a full drive session [41]. Yet, on an individual basis, it was suggested that drivers with aggressive tendencies reduce the sharpness of their acceleration whereas drivers with less aggressive tendencies reduce their speed in highways [41].

Using low engine speed and a moderate throttle position has been a good acceleration strategy [54]. Indeed a study recommended using smooth acceleration to reach high gears and the target cruising speed through the use of the throttle position not going above 45 degrees out of 90 [13]. A similar principle applies to the deceleration. Indeed, applying smooth deceleration by pressing the engine break without changing gears if possible was recommended.

1.3.4 Vehicle Mass

Vehicle mass is another important component of eco-driving. Indeed the heavier the vehicle the more energy is required to move it making fuel consumption increase as mass increase. As previously said in the Golden rules, every 100 kg of added weight increases the fuel consumption by 7 %. Besides adding extra load on the rooftop or in the trunk, another interesting factor adding some variation on the vehicle weight is the number of passengers within the car. The mean passenger weight is about 75 kg, and the amount of equipment within the trunk depends on the purpose of a trip which can range from a few kgs to 100 kg [61].

The occupancy rate is the number of occupants within a car, driver included [4]. Since we have an average occupancy rate of about 1.6 in Europe adding a mass of 45 kg and if we add another 10 kgs for the equipment present within the car, we end up having a 50-60 kg of extra load. And if we take the about ratio of 7% increase for every 100 kg, we end up having a 4% increase in fuel consumption.

Table **??** shows the increase in fuel consumption based on additional mass from a few studies. We can observe a trend in which there is an increase in fuel consumption as more mass is added to the car. However, the average occupancy rate in Europe has decreased from 1.75 in the 80s to around 1.6 now [4]. The trailers and the luggage on the road contribute as well to the increase in mass and fuel consumption as they mess with the Aerodynamic properties of the car.

Aerodynamics is defined as the design and shape of the car and its projected frontal area. A car can pierce through the air and to minimize the amount of resistance the air is putting on a car while driving. Any modification to that frontal shape will increase the aerodynamic resistance, hence increasing the fuel consumption. For example, studies have shown that fuel consumption increases for about 5% when adding boxes on a roof, 5.1% for open windows when driving at about 130km.h, 2% due to the effect of side-winds [61].

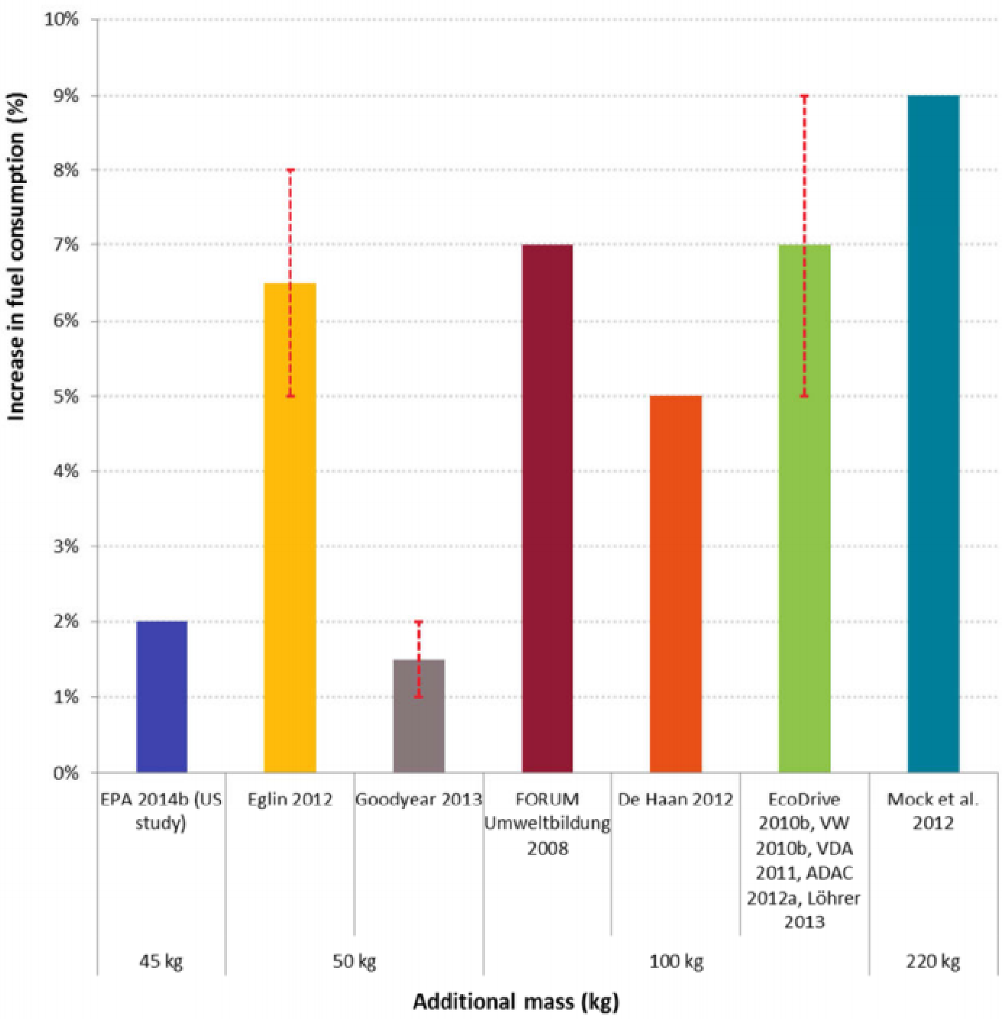


Figure 1.4. Increases in fuel consumption due to additional mass [61]

1.3.5 Type and Choice of Route

The type of road is another factor contributing to increasing or decreasing fuel consumption. Indeed different types of road surfaces and the quality of that surface can influence fuel consumption by 2%. Additionally, the slope of the road is known to affect fuel consumption up to 18% for slopes higher than 2% [61]. Even though fuel consumption is reduced during downhill driving, the consequences of driving the same road uphill and downhill have on average a negative result on fuel consumption.

Looking at the geomorphological properties of the road, we realize that things like the slope, the shape, and the altitude have the most impact on fuel consumption [61]. Indeed at high altitude, there is a lower aid density which decreases the aerodynamic resistance. Yet, in a mountainous road, we will observe more sharp corners or turn and a higher road slope. Riding up a hill, meaning with a road with a positive slope will increase the fuel consumption as more power will be needed to move the vehicle [16]. The opposite is true.

Table **??** shows the results of the above, through and experiment. 2 runs have been done on 4 different types of roads; one route went Uphill then downhill and the others went up and down followed with a flat flow. Looking at the figure we realize that there is an increase of fuel consumption by about 15-20% one the route that was going uphill then downhill, on a road averaging a slope of 4-6% [16]. Another road aspect to take into account is the quality, the texture, the roughness and materials the road was made up, all have some kind of effect on how the car behaves ergo on fuel consumption [61].

Additionally, road traffic hurts fuel consumption. Indeed if we take into account the number of start, stop, traffic signs, and all the stress of driving in a city, we could experience an up to 50% increase in fuel consumption [26]. The research presented in Table 1.1 shows a few papers that did experiments on that.

1.3.6 Type of Fuel

Fuel types are highly regulated through some corresponding standards. It is usually made of different types of hydrocarbon or organic compounds. The composition of the fuel differs based on the geographical regions, the availability of a certain blend or the time of the year [34]. Over the past decade, we have seen the rise of the so-called biofuel which can act as a healthier alternative to the standard fuel than are being used. Indeed, using biofuel could significantly decrease the driver's carbon footprint, however, during actual travel, drivers may experience an increase in volumetric fuel consumption. Indeed, literature has shown that the increase in volumetric fuel consumption (l/km) from different types of biofuels range from 2% [11].

The level of CO_2 emission and even fuel consumption are established during the certification test of a car using a specific fuel. However, the fact that there exist different blend of fuel may affect the engine in different ways, possibly affecting the consumption. For instance, Biodiesel has as benefit a reduced life-cycle GHG intensity, yet the drivers may experience an increase in the volumetric fuel consumption, meaning that the engine was a bit less efficient [61].

Studies have shown that the average CO2 emission for any type of trip or

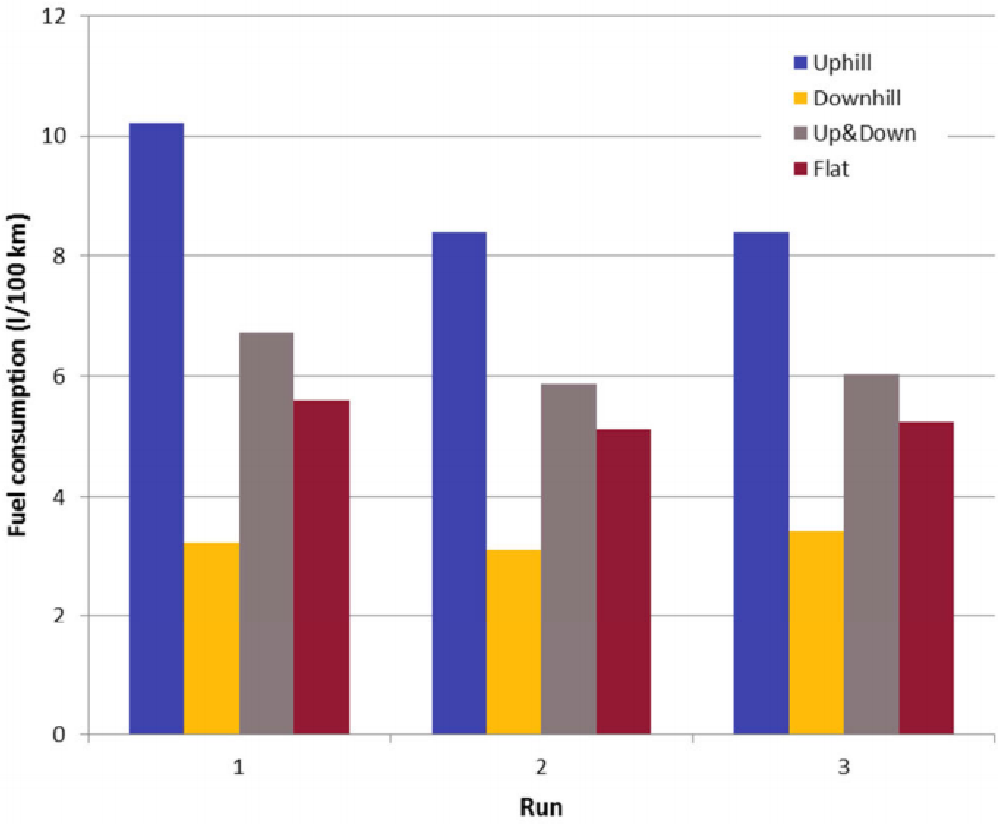


Figure 1.5. Fuel Consumption according to different type of road slopes [61] [16]

vehicle is about 158 g/km. As shown in the table, with normalized data, obviously traveling between cities is the activity consuming the most fuel, followed by the urban environment than long and short distance travel. This trend does not change based on the type of fuel used [61].

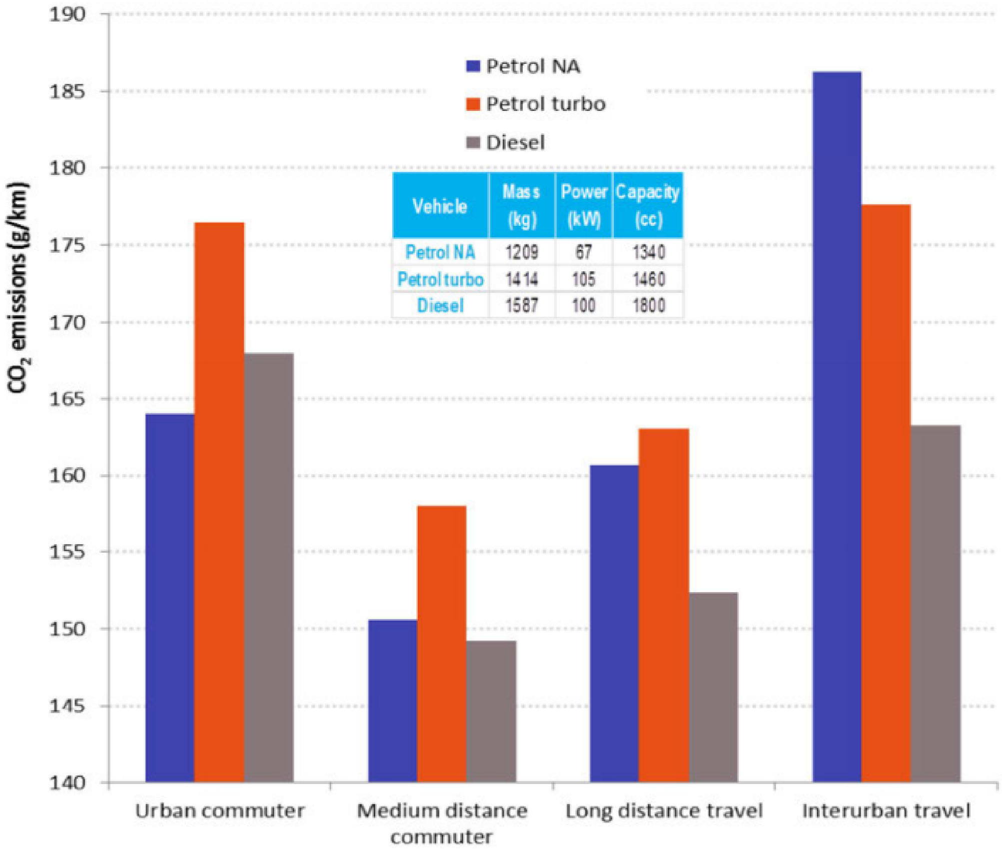


Figure 1.6. Fuel Consumption according to different type of fuels [61] [16]

1.4 Research Issues/Challenges

The application of the methods of eco-driving is mainly done into two-phase: the first one is through the gain of the theoretical knowledge of how, when and why to apply those advices and the second one is through practical on-road training. Over the years, many experiments have been tried and interestingly, they have all observed a decrease in fuel consumption.

The results of the experiments run in the past in table 1.1 show the effect obtained on fuel consumption through various methodologies, year and road type. We observe a general decrease in fuel consumption indicating a positive result. Indeed, all of those experiments show that there is a great advantage in implementing eco-driving and integrating eco-driving as a solution to decrease fuel consumption.

Table 1.1. A summary previous eco-driving experiments and results on fuel consumption [59].

Year of Publication	Road Type	Methodology	Fuel Consumption (fc)
1999	Mixed type	Results compared after instructions	−10.90%
2003	Mixed type	86 drivers; results compared after instructions	−8% with fc monitoring and −1.2% without
2007	15 km route	3 bus drivers	10–15%
2008	Mixed type	300 drivers; results compared after training	−25% short term, −10% long term
2011	City and highway	20 drivers; results compared after 2 weeks, receiving instant feedback	City −6%, Highway −1%
2012	70 km Mixed type	20 drivers; results compared before and after training	−11.3%
2013	Mixed type	After training; results compared after one month	−1.7 kg CO2 emissions per day
2013	16 km urban road	54 drivers; results compared after 6 weeks	−6.80%
2015	Mixed type	116 drivers; pre-test, 30 min training, re-test	Less than −10%
2015	Mixed type	91 logistic drivers	No effect

However, we do not yet fully understand nor able to predict the effect of eco-driving in the general population. Indeed, most of the experiments were set under some very specific setting with driving performances varying based on the drivers, the road type among other factors. For example, experiments in 2011 presented in table 1.1 states that eco-driving performances are more effective in cities rather than in highways [15], but almost half of the drivers already practiced eco-driving [59], meaning we cannot derive the true effect of the eco-driving performance as we have lost the baseline on that experiment. The following experiment in table 1.1 shows an experiment made in 2012 in a 70km mixed type of road. The result of that experiment shows a general decrease in fuel consumption by 11.3% [22], yet the road type and the type of vehicle used were different meaning that their results cannot be compared.

Another problem with some research is that most of the experiment setting was done over short distances with controlled or reduced traffic which consequently reduced the number of external elements that may affect the driver's drives. This means that certain of the result presented in some research ignores the influences that real traffic has on the drivers mindset[59]. Ergo, even tho there is a general decrease in fuel consumption under a very specific setting, those results may not be valid under real-life situations as they do not take into account the driver's stress endured under real-life conditions.

Chapter 2

Methods

In this chapter, we will be exploring how the data was obtained, then we will elaborate on the preprocessing and feature engineering applied to the data before going through the method used to predict the CO2 levels.

2.1 Data

2.1.1 Experiment Setting

The initial dataset was obtained through the same type of experimentation in 3 different test sites (CTAG, IFSTTAR and Leeds). The experiment followed the following settings:

1. Each test site recruited 24 participants
2. Participants divided into 4 group of 6
3. Each participant drove 6 times
   (a) Fist two drives done without any advices
   (b) last four drives done with eco-driving advices

Those drives were saved into CSV files that were corresponding to one trace. Those files were saved with the following nomenclatures

$$X - Gi - Pj - Dk - Ln \tag{2.1}$$

where X is the test site letter (C for CTAG, I for IFSTTAR and L for Leeds), i is the group number that ranges from 1 to 4, P is the Participant and j is the participant number that ranges from 1 to 6, D is the test drive and k is the test drive number that ranges from 1 to 6 and finally l is the logging number. For example, I-G1-P4-D2-L1 will be the trace saved on the spark server corresponding to a drive made in the IFSTTAR location for participant number 4 from group number 1 driving his/her fourth drive.

2.1.2 Data Collection

The data collection had two main sources of input. One set was derived from the vehicle itself and the other one derived from a sensor attached to the driver.

The vehicle data contained information on the CAN bus data that was provided by the OBD-II adapter of the vehicle. That data included:

- Throttle position
- Vehicle Speed
- Intake Air Temperature
- Engine revolutions per minute (RPM)
- Short term fuel trims
- Fuel type, consumption and tank levels
- engine coolant and oil temperature

Besides the vehicular data, we will be monitoring the driver as well to somewhat get the stress levels the driver is under at a given point in time. To do so we are going to be using a Fitbit or the Spire which will provide information about the heart rate, the respiration rate, and the muscle activities.

2.1.3 Equipment

Since we needed two sets of recording, one that takes information about the car itself and the other one takes information about the driver's state. Here are the various software and hardware setup.

1. Software
   (a) Sparks mobile app for data logging
   (b) CTAG app for observer protocol
2. Hardware
   (a) Heartbeat sensor (Xiaomi Mi Band)
   (b) OBD II Bluetooth reader
   (c) Android nomadic device and accessories

2.1.4 Test Route

Three different test sites were used, one in England in Leeds, one in France in Versailles, and the last one in Spain in CTAG in this section we will give more details about the various routes used.

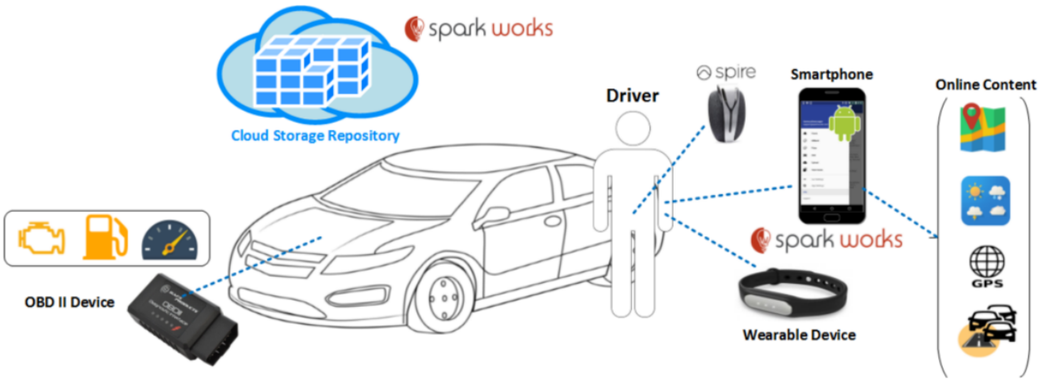


Figure 2.1. System Architecture for data collection [45]

Leeds University route

The Leeds university route was the test drive made in England. The route is about 20km long with very little traffic. It takes approximately 35 min to fully complete it. The vehicle speed in this route ranges from 50 km/h to 100 km/h on the rural and secondary roads respectively. The route is originally selected due to the high pedestrian activity and sections of free-flowing activities. There were various types of slopes as well as a lot of traffic lights and roundabouts. The route can be visualized in figure 2.2

IFSTTAR Route

The IFSTTAR route is the test site used in France. It is about 16km long and takes between 30 to 40 mins to complete. The driving is quite dense in some sections. This route has been selected because it represents different types of driving conditions that include city driving, highway driving, and interurban hilly road. The detailed route can be seen in figure 2.3 with colors representing the intensity of the traffic.

CTAG Route

The CTAG is the test site used in Spain. It lasts about 40 mins and is 15 km long. Like the IFSTTAR route, it is composed of various section that includes city driving, highway driving and interurban hilly road

2.1.5 Vehicles

The IFSTARR test site failed to report the type of car used. The Leeds university test site used a Hyundai i30 for all the trips and all the participants. The CTAG test site used two different vehicles, a Citroen Xsara Picasso and a Renault Scenic.

The main characteristic for the Citroen are:

1. Power: 90 HP
2. Type of gearbox: Manual

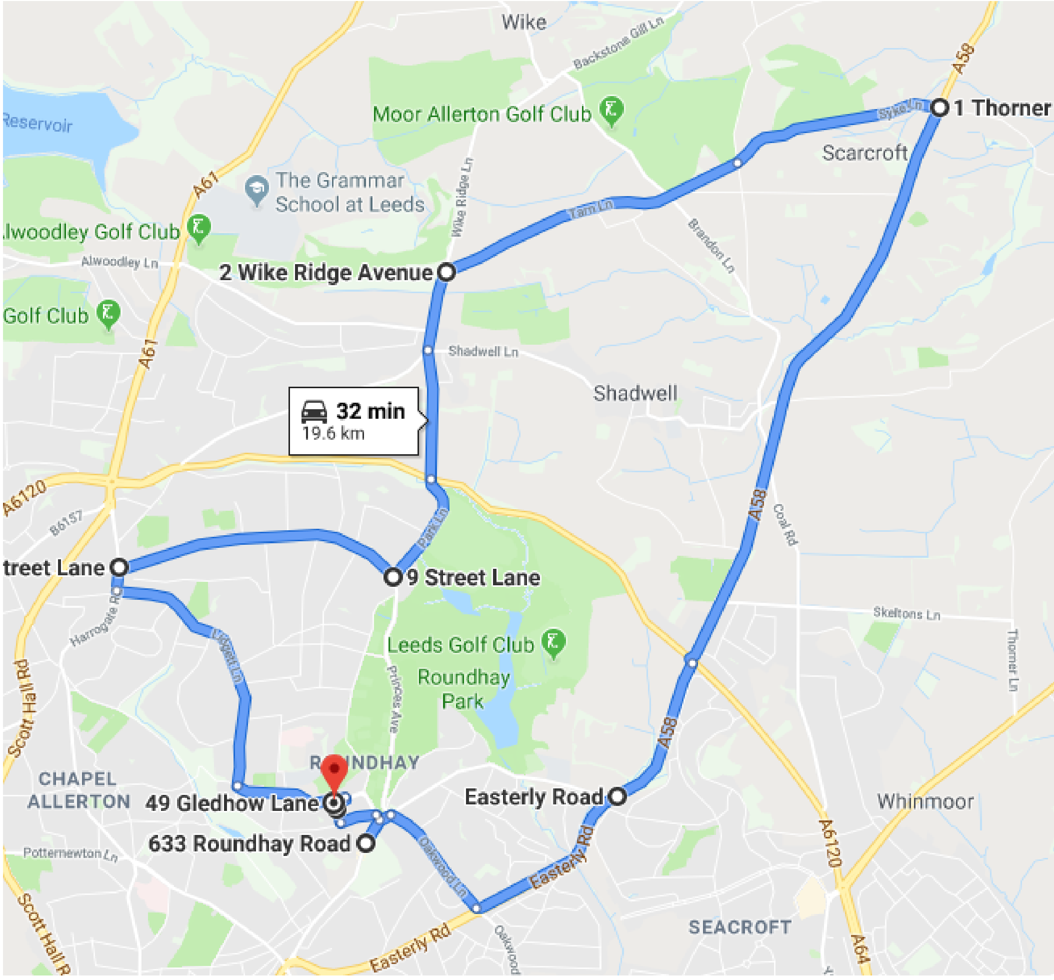


Figure 2.2. Leeds test route

3. Gearbox (number of gears): 5
4. Maximum speed (km/h): 175
5. Acceleration 0-100 km/h (s): 14.5
6. Acceleration 0-1000 m (s):
7. Urban consumption (l/100 km): 7.0
8. Extra-urban consumption (l/100 km): 4.6
9. Average consumption (l / 100 km): 5.5
10. CO2 emissions (g /km): –

The main characteristic for the Renault are:

1. Power: 130 HP

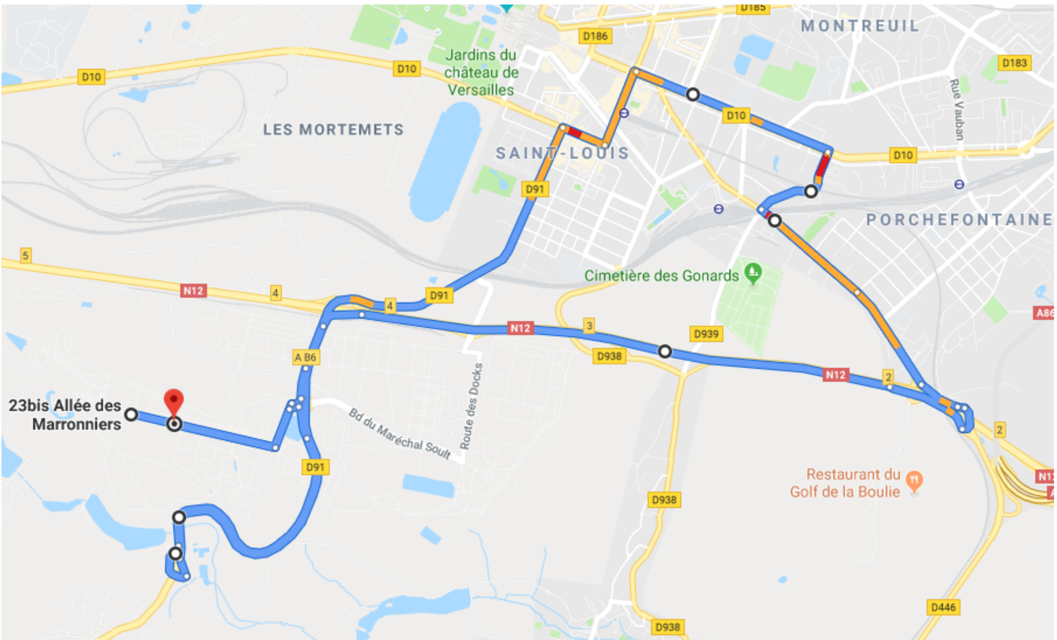


Figure 2.3. IFSTTAR (France) test route

2. Type of gearbox: Manual
3. Gearbox (number of gears): 6
4. Maximum speed (km/h): 192
5. Acceleration 0-100 km/h (s): 9.6
6. Acceleration 0-1000 m (s) 31.2
7. Urban consumption (l/100 km): 7.3
8. Extra-urban consumption (l/100 km): 5.2
9. Average consumption (l / 100 km): 6.0
10. CO2 emissions (g /km): 159

2.1.6 Test Drivers

The participating driver's age varies from 23 years to 59 years old across all test sites. The UK test site drivers are slightly older than the French and Spanish ones, but all three test sites driver's ages are relatively normally distributed.

The driving experience was another factor taken into consideration. Indeed the annual millage on average for all the drivers within the test site was 8291 for the UK, 10635 in France and 17 587 in Spain. Important to note that the test site in Spain had the younger test drivers.

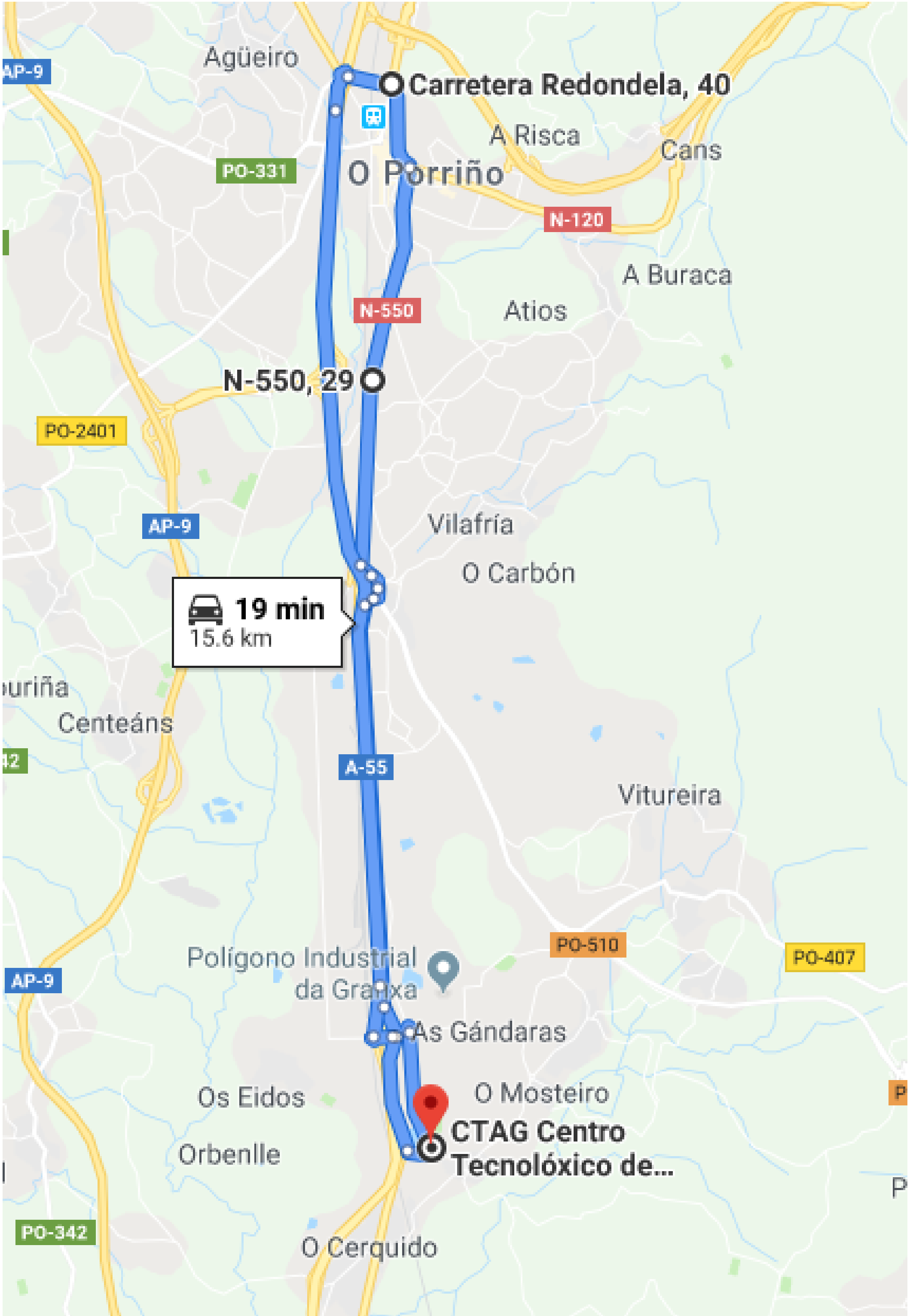


Figure 2.4. CTAG (Spain) test route

2.2 Data Pre-processing

During the data pre-processing step, a file was created containing general information about the various trips. Table 2.1 represent the information file allowing us to navigate through the various trips and trip information. That table contains the file name with the above nomenclature giving us the kind of trip it was, the location of the file corresponding to the physical location where the trip was made, the duration of the trip corresponding to how long the trip was, the name of the non-empty (active) and empty variables.

Table 2.1. Sample of the information table

File	Location	Duration	Non-Empty Variable	Empty Variable
./data/I-G4-P4-D5-L1.csv	IFSTTAR - Satory, 12, Allée des Marronniers, S...	00:33:28	['throttle$_p$osition','longitude','short$_t$erm...	['fuel$_t$ype','fuel$_l$evel','fuel$_c$onsumption
./data/I-G4-P4-D6-L1.csv	IFSTTAR - Satory, 12, Allée des Marronniers, S...	00:33:13	['throttle$_p$osition','longitude','short$_t$erm...	['fuel$_t$ype','fuel$_l$evel','fuel$_c$onsumption
./data/I-G4-P5-D1-L1.csv	Circuit de Versailles-Satory, Satory, Versaill...	00:28:45	['throttle$_p$osition','longitude','short$_t$erm...	['fuel$_t$ype','fuel$_l$evel','fuel$_c$onsumption
./data/I-G4-P5-D2-L1.csv	IFSTTAR - Satory, 12, Allée des Marronniers, S...	00:26:58	['throttle$_p$osition','longitude','short$_t$erm...	['fuel$_t$ype','fuel$_l$evel','fuel$_c$onsumption
./data/I-G4-P5-D3-L1.csv	IFSTTAR - Satory, 12, Allée des Marronniers, S...	00:29:29	['throttle$_p$osition','longitude','short$_t$erm...	['fuel$_t$ype','fuel$_l$evel','fuel$_c$onsumption

At the end of the experiment, we were supposed to have about 432 trips coming from all test sites. Upon receiving the data, 10 trips were removed from the IFSTTAR test site, 120 from the Leeds test site, and we had to ignore all trips from the CTAG test site. From the test site from IFSTTAR, trips were removed due to empty content or because the variable "Throttle Position" was constant during all recordings. For the CTAG test site, a similar reason why all trips were ignored and not received in the place.

The trip data are made of several features coming from the car's sensors and other connected components (android phones and bracelets). The prepossessing steps consisted of 2 major part which is the data integrity and the data frequency.

2.2.1 Data Integrity

The data integrity consists of checking the so-called quality of the data by checking the various inconsistencies present within a trace. Indeed, we will verify the presence of outliers and checking the validity of the data within a trip.

Before considering a file as valid and adding its information content to the analysis apparatus. We do an integrity check that consists of the following steps:

1. Check for the duration of trip: Most of the trips are about 40-mins long on average, an hour-long at most. Ergo any trips that are below 20 mins or above an hour-long will not be considered as they may have been tempered or will not reflect the true drive according to the circuits.

2. Check empty or constant columns: Most of the traces were filled with empty variables. Indeed, the equipment failed to capture the readings from all the sensor or the readings were constant. For example, the CTAG throttle position or the fuel tank level were had constant values. Those type of trips were removed from the final analysis

3. Imputation missing value with low ranked SVT imputation or filled based on the previous value

4. Remove the Outliers within the variables.

SVT Imputation Procedure

Singular Value Thresholding [46] is a technique that has been used in Exact Matrix Completion. The matrix completion allow us to solve nuclear norm minimization problem and by extension problem in the form of

$$\begin{aligned} &\text{minimize} && \| X \\ &\text{subject to} && A(x) = b \end{aligned} \tag{2.2}$$

where A is a linear operator on the space of $n_1 \times n_2$ matrices and $b \in R^m$. This algorithm is well suited for problem of large size problem in which solution has low rank. This above algorithm will then be sketched in the matrix completion setting. Let P_Ω be the orthogonal projector of the amount of matrices outside Ω in a way that (i, j)th component of $P_\Omega(X) = X_{ij}$ if $(i, j) in \Omega$ and 0 otherwise [19]. The problem will then be expressed as

$$\begin{aligned} &\text{minimize} && \| X \| \\ &\text{subject to} && P_\Omega(X) = P_\Omega(M) \end{aligned} \tag{2.3}$$

[19] where of optimization variable $X in R^{n_1 X n_2}$. We will then fix $\tau > 0$ and a sequence $\sigma_{k k>=1}$ as scalar step sizes. Ergo, we start from $Y^0 = 0 in R^{n_1 X n_2}$, the SVT defines

$$\begin{cases} X^k = shrink(Y^{k-1}, \tau) \\ Y^k = Y^{k-1} + \sigma_k P_\Omega(M - X^k) \end{cases} \quad [19] \tag{2.4}$$

until a stopping criterion is reached. The shrink function in corresponds to a nonlinear function that applies a so called soft threshold rule of τ on the input matrix singular values. The larger τ is, the sequence X^k will converge to a solution that minimise 2.3. So at every step of the algorithm, we compute at most one singular value decomposition and perform basic matrix operation[19].

2.2.2 Data Frequency

The other trip prepossessing is the downsampling of the various values. The reason why we need to do that is since the different sensor readings had different timestamp values attached to them. Indeed, they were emitting in a range of milliseconds to a few seconds depending on the type of sensor which created a lot of empty values.

To deal with that problem, we decided to downsample the various readings from milliseconds to seconds. The new readings would be equal to the average reading of the millisecond's values within a particular second.

Doing this allowed us to synchronize the various feature within our dataset. SVT was then used to impute the missing values.

2.3 Feature engineering and driving behavior discovery

After doing the prepossessing and the data exploration (the result will be shown in the next section), we went on an added few features. The first important feature added is the CO2 emission rate along with 9 driving behavior events.

2.3.1 CO2 Emission

Since the dataset did not contain any CO2 readings; we needed to figure out a way to estimate it. Thankfully, there exists a relationship between the speed and the acceleration, along with the power demand for the car that gives us an estimation of the CO2 emission level using a linear model. Previous models have been developed in the past and they have all shown a correlation between the vehicle speed, the acceleration, and the power demand [3].

The multiple linear regression model is represented generally with the following equation:

$$Y_i = \beta_0 + \beta_1 X_1 + ... + \beta k X k + \epsilon i \tag{2.5}$$

where Y_i is the target variable, X_k is the predictor variable with the range $k = 0, 1, 2, 3...p - 1$ where p is the number of parameter; ϵ_i is the random error in Y for observation i

In our setting the CO2 emission level is going to be used as a Target variable, with as predictor variables features representing the vehicle speed, the acceleration, the power demand and the at the intensity of traffic.

The details about the predictors are as follow

- Vehicle speed in kilometers per hour (km/h)
- Acceleration in km/h per second
- Deceleration in km/h/s
- Power demand which corresponds to the product of vehicle speed and acceleration in km/h
- Intensity of traffic a boolean with 1 representing intense traffic and 0 representing free-flowing traffic mostly

[3]

the CO2 emissions are going to be represented in g/s. Since we did not have the acceleration in km/h but g acceleration. We then need to convert this g acceleration to km/h/s. The conversion is done with the formula below.

$$1 \text{ acceleration of gravity} = 0.00980665 \text{ kilometer/second}^2 \tag{2.6}$$

Once the conversion is done we end up having a pretty good estimate CO2 emission estimate on both arterial or highway roads. We would then scale the CO2 emission to g/km to be able to compare it to the EU regulation standards [3].

2.3.2 Driving Behavior Events

The next step in our feature engineering was to identify the various driving behavior events. The 10 events that will be identified are the periods of sharp acceleration, sharp deceleration, long acceleration, long idle, constant speed driving, low-speed driving, high-speed cruise, moderate start, start and stop periods and finally the moderate break [21].

These above driving events are the behaviors that have a significant impact on fuel consumption and CO2 emission if we compare them to the golden rules. Identifying the above events was done according to the following rules.

1. Sharp Acceleration: ACC $> 4km/h/s$. If the speed change between the second t and t+1 is above 5km/h/s. the we eperienced a period of sharp acceleration and aggressive driving as a standard car should not accelerate this fast.

2. Sharp Deceleration: DES $> 4km/h/s$. Same principle as the previous event

3. Long Acceleration: $T_{\text{ACC}>5/s}$A period of long acceleration is defined as an increasing acceleration for more than 5 seconds

4. Long Idling: $T_{\text{dec}} or \text{Speed}_0 > 30s$ a period of long idling is definied as a period with a speed of 0 and no rpm input that last more than 30 seconds

5. Constant Speed Drive: $T_{\text{dec}} or \text{Speed}_0 > 60s$ a period of Constant Speed drive is definied as a period with no speed change and no rpm input that last more than 30 seconds

6. low speed driving : It is a period in which the mean speed for an interval of 60 second is below 20 km/h

$$\text{mean}(s(t), s(t-1)..., s(t-59)) \geq 20\text{km/h} \tag{2.7}$$

7. high speed cruise: This event basically identify period in which the drivers drives fast, and is defined as having the properties below:

   (a) Average speed that is higher than 50 km/h

   $$\text{mean}(s(t), s(t-1)..., s(t-4)) \geq 50\text{km/h} \tag{2.8}$$

   (b) the Speed Standard deviation fpr speed is no more than 1.5 km/h

   $$\text{std}(s(t), s(t-1)..., s(t-4)) \leq 1.5\text{km/h} \tag{2.9}$$

   (c) the Speed variation is below 2km/h

   $$\text{abs}(s(t) - s(t-4)) \leq 2\text{km/h} \tag{2.10}$$

   (d) the acceleration is no more than 2 km/h

   $$\max(a(t), a(t-1)..., a(t-4)) \leq 2\text{km/h} \tag{2.11}$$

8. moderate start, A period of moderate start is when the vehicle accelerate from an idling position not in a sharp way. it will have one of the following property.

   (a) the vehicle speed $\in [10, 20]$ km/h/s

   $$10\text{km/h} \leq s(t) - s(t-4) \leq 20\text{km/h} \tag{2.12}$$

(b) the maximum acceleration speed is below 5 km/h/s

$$\max(a(t), a(t-1)..., a(t-4) \leq 5\text{km/h/s} \tag{2.13}$$

9. start and stop periods: This is an event in which the vehicle may be stopped in traffic. It is defined as a period in which the vehicle is idle within 3 sec after starting from an idle position

10. moderate break: like a moderate start, it is when the vehicle decelerate not in a sharp way. it will have the following properties

    (a) deceleration $\in [15, 25]$ km/h/s

$$-25\text{km/h} \leq s(t) - s(t-4) \leq -15\text{km/h} \tag{2.14}$$

    (b) the minimum acceleration speed is below -5 km/h/s

$$\min(a(t), a(t-1)..., a(t-4) \geq -5\text{km/h/s} \tag{2.15}$$

The above is saved in a fuzzy system with 9 additional variables. Each variable will have 1 when an event is active and 0 when it is not.

2.4 Methods for Predicting the CO2 Consumption

The other aspect of the investigation was the ability of the driving behavior event in predicting the CO2 levels. The goal of this section is to better understand the relationships and the impact of the features on the CO2 emissions.

2.4.1 Feature Selection

Here, we are going to be selecting the most important features to make a prediction. This main goal of this phase is to pick relevant features that will enable us to better understand what and to what extent does a variable affect CO2 consumption.

Our final data set was made of 18 features of which the CO2 will be our target variable we are going to try to predict, and we will be removing the vehicle speed. The idea behind removing the vehicle speed is that we should theoretically always be able to predict very accurately the CO2 consumption with the vehicle speed as a feature. Doing so will decrease our understanding of the impact of the other variable on the CO2 emissions.

As a result, we will be left with 14 variables namely, the engine rpm, the throttle position, the short term fuel trim, the acceleration, the heartrate and the 10 driving behaviors identified in the previous step. Those 10 variables of 375838 records representing a recording of driving per second are going to be our final dataset.

2.4.2 Model Selection and Evaluation

Our first phase will be to attempt to identify a model or family of models that would work the best with the dataset in our disposition. At the end of this phase, we will pick a family of models that would be the most appropriate to make a prediction and extract the most relevant information to understand the impact of the variables on fuel consumption

In this first phase, we will be running the dataset through a pipeline with a regressor at the end of the pipeline. The pipeline is shown in fig 2.5 is made up of multiple steps. First, we will divide the dataset into two sections representing the numerical and the categorical variables. To the numerical variables, we will be imputing missing values, if any, before applying a scaling whereas for the categorical data only imputing will be applied. Then both numerical and categorical data will be combined to a preprocessed dataset to which we will apply our regressor. The neural network will need to be reshaped to a proper input format before being used.

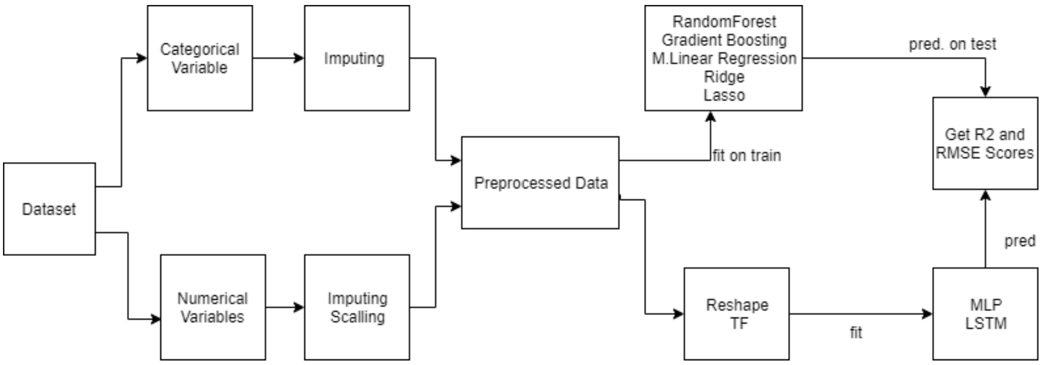


Figure 2.5. Model Selection Pipeline

The families of models to be used are the linear models along with some regularization algorithms, the tree family models and the neural networks. Within the linear models we will be running a multiple linear regression, then both a Lasso model (using an L1 type of regularization) and a Ridge regression model (using an L2 type of regularization) They are defined as

- multiple linear regression: It works similarly with the simple linear regression but the difference is that instead of using 1 variable to predict the outcome of a target variable, the MLR will use 2 or more variables; in our case, it will use 15. The output will be represented as a mathematical function defined as $y = \beta_0 + B_1.x_1 + ... + \beta_n x_n$ where y is our target variable β represent the coefficients and x x_n our independent variables[40].

- Ridge Regression: it is a sort of extension for the MLR. It is a regularized linear regression used like the Lasso to avoid overfitting the model. Indeed, model evaluation is made of a loss function that should be minimized. The ridge regression is defined as a regularization term that somewhat punishes the coefficient with high values to simplify a model as much as possible, adding a constraint for the maximum minimization of coefficients. Mathematically, it is expressed as $L = \sum(\hat{Y}_i - Y_i)^2 + \lambda \sum \beta^2$ where $\sum(\hat{Y}_i - Y_i)^2$ is the sum of distance between each prediction and the ground truth and $\lambda \sum \beta^2$ is our regularization term that sum over the squared value of the coefficient β multiplies it is with another parameter λ [40].

- Lasso: the lasso is another regularization algorithm. The main difference from the ridge regression is that the regularization term is not the sum of squares but the sum of the absolute values. The loss of the lasso is mathematically defined as $L = \sum(\hat{Y_i} - Y_i)^2 + \lambda \sum |\beta|$ [40].

The next model family will be the tree-based family in which we will be testing the decision trees, the random forest regressor along with the gradient boosting algorithm. They are defined as follow:

- Decision tree: it is an algorithm that forms a tree structure that breaks down a dataset into a smaller and smaller subset, while at the same time, tunes an associated decision tree developed incrementally. A node represents a single input variable and a split point for that variable. The tree is finishing with terminal or leaf nodes that represent the variable y used to make a prediction. The creation of the binary tree involves dividing up the input space through a process called recursive binary splitting, in which all the values are lined up and different splits are tried and tested using a cost function [42].

- Random Forest: Random Forest is a bagging type of algorithm which uses as learning techniques the decision tree. As previously said, in a decision tree, the input is fed on top and is traverses the tree as the data get divided into smaller and smaller sets. The random forest takes that principle and combines the trees to form an ensemble. Consequently, the trees itself are week learners and the random forest is a strong learner [42]. The mathematical definition of a Random Forest will be provided in the next section.

- Gradient Boosting: The gradient boosting is an ensemble type of model which makes a prediction based on a group of weak learners. The gradient boosting involves 3 basic elements which are a loss function which we optimize, a week learner such as a decision tree and an additive model. We will be using the root mean square error as the loss function. The idea behind using decision trees is that they output real values for splits which can themselves be added together, which allows subsequent models output to be added and corrected. They are constructed in a greedy way that chooses the best spit point based on scores like Gini to minimize the loss function. Finally, the additive model adds the tree one at the time and the already existing trees in the model are not changed. Gradient descent procedure is used for the minimization of the loss[42].

The final model family that will be testest will be the neural networks. Namely:

- Multi-Layer Perceptron: It is the classical type of Neural network. They are made of one or more layers of neurons. Data is put through an input layer made of one or more hidden later which gives the level of abstraction through a series of activation functions. Gradient descent is then used to adjust the weight and minimize the loss function. The result is obtained by the last layer called the output layer.

- LSTM: An LSTM is a recurrent Neural Network which we train using back-propagation through time to overcome the vanishing gradient problem. As opposed to an MLP, and LSTM is made of memory blocks that are connected to layers. Each of those blocks contains gates which are responsible for the management of the block's state and output. Each gate within that unit uses a sigmoid activation function that controls the activation of that particular gate ergo changing, or not, the state and adding the information flowing through the conditional unit. Within this memory unit, we have the input gate, the output gate and the forget gate. The intuition behind the use of the LSTM here it to maybe predict the CO2 level using the previous driving pattern.

Upon passing the model through the pipeline, available in figure. We will then evaluate the models using two scoring measures; the Root Mean Square Error and the R^2 score. The RMSE provides an idea about the magnitude of the error. Taking the square root of the mean squared error converts the unit back to the original output variable. The R^2 score indicates how well the prediction fits the actual value or, in other words, the goodness of fit (coefficient of determination). The scoring values will be between 0 and 1 with the former representing no fit and the later a perfect fit [18].

2.4.3 Hyperparameter Tuning

The idea behind hyperparameter tuning is basically to adjust the settings of an algorithm to improve its performance. The standard procedure for optimizing the hyperparameters is called cross-validation(CV). The most common method of cross-validation is K-fold CV, where we splot our training set into K number of subsets which are called Folds. Having that, we iteratively fit the model K times each time training and evaluation on the Kth fold. In the end, we take the mean performance on each fold to have a final validation metric for the model. Consequently, in the K-Fold CV, we will be performing many iterations each time using a different model setting and saving the model settings that work the best for our final predictions.

The hyperparameter tuning phase will be done in two steps. In the first one, we will tune the hyperparameter using the Random Search Cross-Validation. Indeed, in here, we will be evaluating a wide range of hyperparameter values. We will first create a wide grid of hyperparameter we would like to test, and perform K-Fold CV by hyperparameter values at random. Once we are done training we will evaluate the best hyperparameters deduced from CV and check if there were any improvements. If they were improvements we will move to the next step, if there wasn't any improvement, we will keep the default values.

The next step will be a grid search with cross-validation. Indeed performing the random search allowed us to narrow the range for each hyperparameter. We can use a full grid search by specifying all combination settings we which to try. We then evaluate the model and save the best parameters. If there is an improvement we save that model, if not we revert to the original hyperparameter from the previous step.

2.4.4 Model Results

In order to do that we will be using, after performing the step 1 test, the random forest and the gradient boosting to do so. The random forest will be a bagging ensemble technique where as the gradient boost, as the name implies, will be a boosting technique.

The final data is combined and separated in to a training set and testing test. Since we are working with a timeseries, the train-test dataset have not been shuffled because it is inappropriate but they were split respecting the temporal order of values that were observed. Multiple splits have been used to train-test separating the two with a ratio of 50-50, 70-30 and 90-10. The result of those analysis will be displayed in the next function. The Root Mean Square Error has been used as a scoring metric.

As the name said the random forest is a tree based ensemble algorithm with each tree relating to a collection of random variable. Given a p-dimensional random vector $X = (X_1.....X_p)^T$ representing an input or predictor, in our case here would be the driving behavior events; and a random rariable Y representing the response, in our case, that would be the levels of CO2 emissions, we will assume an unknown joint distribution $P_{XY}(X, Y)$. The aim of the algorithm is to find a prediction function $f(X)$ that would be predicting Y. It is obtained by a the loss function $L(Y, f(X))$ and minimized the expected values of the loss.

$$E_{XY}(L(Y, f(X))) \tag{2.16}$$

let $D = \{(x_1, y_1), ...(x_N, y_N)\}$ be our training data xith $X_i = (X_{i,1}..., X_{1,P})^Y$. So for $j = 0$ to J:

1. Extract the bootstrap sample D_j of size N from D
2. With the sample D_j as the training data, we fit a tree using binary recursive partitioning
   (a) Partition Starting with all observation in a single node
   (b) Until stoping criterion met, we do the following recursively:
      i. Select predictors p at Random from the P available predictors
      ii. On the m predictors from step i, find the best binary split among all binary splits
      iii. Split the note onto two descendant nodes using the previous split

In our case, the prediction will be made with $f'(x) = \sum_{1...J} h'_j(X)$ where $h'_j(X)$ is the prediction of the response variable at x using the jth tree [24].

Regarding the Gradient boost, the prediction we will make will be under the format $f(x) = w^T\phi(x)$ where $\phi(x) = [k(X, \mu_1), ..., k(x, \mu_N)]$ where μ_k represent our training data. The concept of boosting is genrally associated with a greedy algorithm that fis adaptative basis function generated by an algorithm called a weak/base learner. Boosting works by applying week learners by sequence to a weighted version of the data in which more weigh is assigned to sets that were misclassified in earlier rounds [42].

The problem that the Gradient Boosting will try to solve, like the random forest, is the minimization of a loss function, which we will represent with:

$$f' = argminL(f) \tag{2.17}$$

where $f = (f(x1)...f(X_n)$ represents the parameters. The algorithm will be preformed in the following steps:

1. Initialization of $f_0(x) = argmin_\gamma \sum_{i=0}^{N} L(y_i, \phi(x_i; \gamma)$
2. With a step size of m going to M, we will then perform the following
   (a) Find the residual gradient which is $r_{im} = -[\frac{(y_i, f(x_i))}{}(x_i)]_{f(xi)=f_m-1(x_i)}$
   (b) Using the weak learner to find γ_m that will minimize $\sum_{i=1}^{N}(r_{im} - \phi(x_i; \gamma_m))^2$
   (c) Finally, we update $f_m(x) = f_{m-1}(X) + v\phi(x; \gamma_m)$
3. Upon the completion of the above steps, we retrieve our $f(x) = f_M(X)$

Both the Random forest and the gradient boost are the 2 algorithms that will pass through parameter tuning in the next session. The results of the gradient boost and the random forest are presented in the following section and in the annexes.

Chapter 3

Results

In this section, we will be presenting the various results. We will begin with the data exploration followed by the results for the driving events test and finish with the CO2 estimation tests.

3.1 Data Exploration

After data pre-processing, our final dataset was made of 105 hours worth of driving time. Looking at the correlation between variables in a single trip, we realize that there is a limited amount of variables sharing a strong positive correlation. We can observe the details of the correlations in figure **??**

To start of, we realize that engine rpm and vehicle speed have a high correlation. This is what we expect as in other to gain more speed, more rpms need to be put in. We have, to a less extend, some kind of relation with the throttle position and the engine rpm and the vehicle speed too. We observe a high positive correlation between the free flow and the current-flow speed.

Since we did not have anything we did not already know about from the correlation figure. We decided to investigate in detail the main CO2 consumption feature driving variable. We established those variables by running a feature importance test in predicting the co2 emission.

Figure 3.5 shows the distribution of those features. After looking at the graphs and describing the feature behavior here are our conclusions:

- Vehicle speed: The average vehicle speed of all the experiment combined is about 30 km/h with a standard deviation of 20 km/h. The median of all the drives are about 27km/h, with the minimum speed of naturally 0 and the maximum speed of 128 km/h. The speed reading is consistent with the type of drive as it was a city drive with high traffic intensity with a portion on a highway.

- Engine rpm: Most drives had an average engine rpm of about 1734 rpm with a standard deviation of +/- 512 rpm. the median and even the 3rd quantile was at 2002 rpm. This makes sense since a lot of people are taught to shift up at 2000 rpm. As we drive, drivers tend to reach that 2000 mark and stick to it. We had a min rpm value per second at 0 and a maximum value at 4883 rpm.

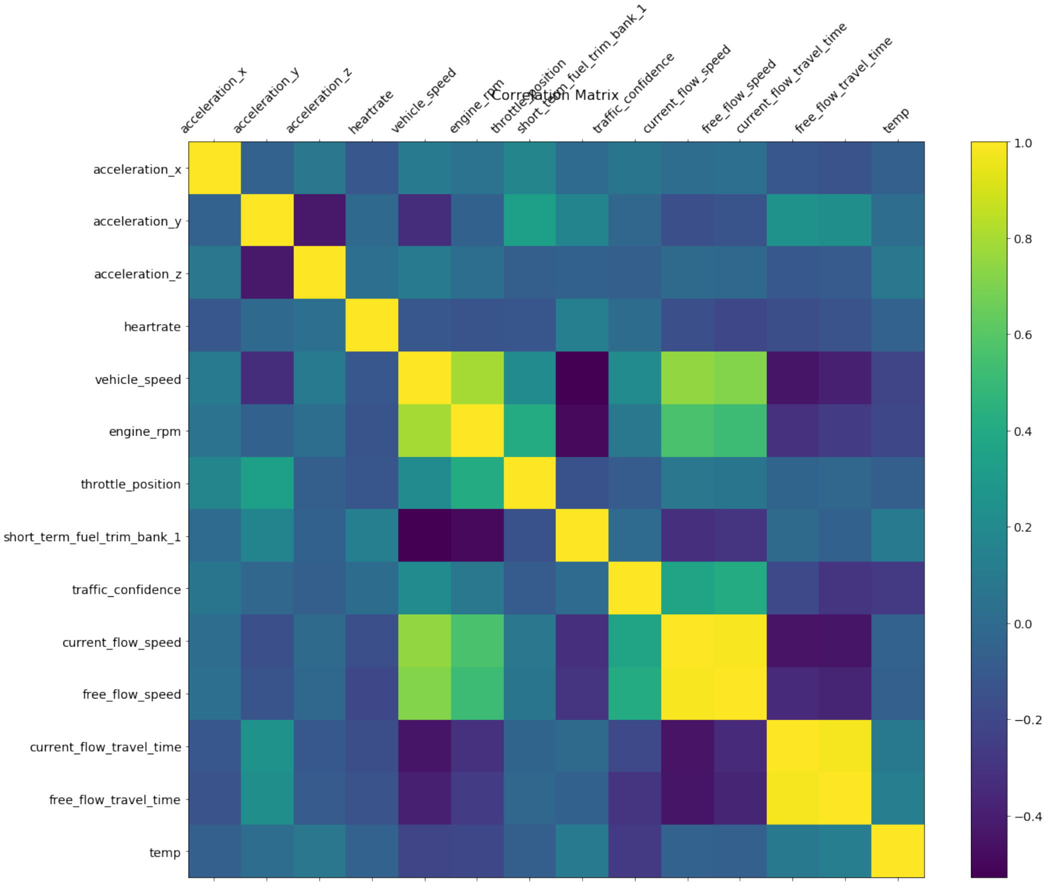


Figure 3.1. Correlation of variable within a trip

- Throttle position: The angle of the throttle used most often by most drivers is 15°, with the average position being at 19.25° with a standard deviation of about 12°. It had most of its values centered around the first 20°, and a maximum angle of 85°. Once again nothing too surprising here, most people drive in a city using an throttle position angle of no more than 20°, where as we use a higher throttle angle when accelerating quickly when reaching a highway.

- CO2 emission: After converting and scaling the co2 consumption to g/km we have an average emission of 87 g/km with a standard deviation of 16. The minimum emission levels were at 62 g/km with a maximum value of 269. Most drivers drive at around 85 with most of the values centered between 62 and 94 g/km.

- heartrate: The average heart rate across all experiment is about 83 bpm with a standard deviation of about 29. The minimum heart rate encountered is at 30 and the maximum is at 240 bpm. Most of the values are centered around 71 to 91. Given that a well rested heart rate for a human is between 60-100 bpm, everything seem normal. We have few spikes in the heartrate which might be due to the stress caused while driving.

Now that we have looked at the general distribution for those different variables, now we are going to explore the behavior of those variables during the different phases of our driving experiment. Indeed, the behavior of those variables are going to be plotted based on the baseline drive, the 2nd session of driving and the last session of driving. We display the result of those experiments in figure.

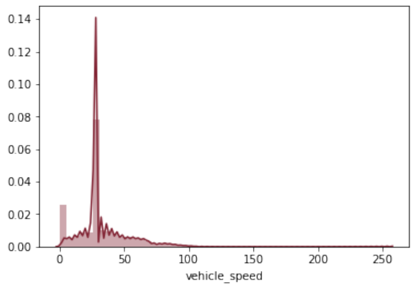
Looking at the vehicle speed, we realize that most of the values are centered around a speed of 30km/h, where as the drives for both the baseline and the last session was more spread about. Additionally, both the base line and the last session observed longer period of stops. This may be due to the fact that, during drive 2 the driver did their very best to somewhat converve their constant speed.

Regarding the engine RPM, we can distinguish a high concentration of values at around 2000 rpm, specially for session 2. session 3 and the baseline had their rpm values a bit more spread out. We can notice a higher concentration of engine rpm values at around 900 rpm for both baseline and session 3.

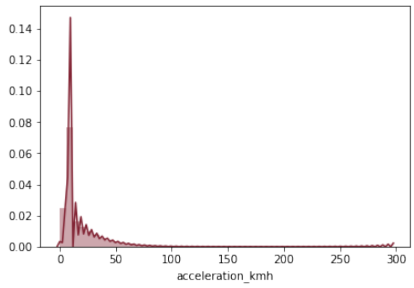
The throttle position share almost the same shape between the 3 session. The baseline and the session 3 drive share approximativelly the same shape. Whereas the drive in section 2, we have again appoximatelly the same shape but with a concentration abount 10-15. The two other phase had a concentration between 15 and 30. Additionnally, session 2 shows higher conventration of values at higher throttle position degree. This might be the effect of driving ecologicaly.

3.2 Driving Events

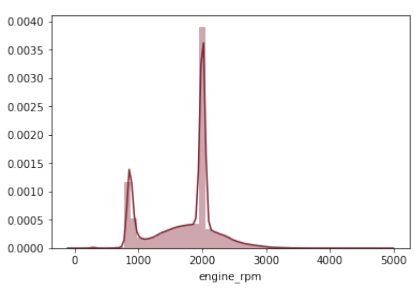
In this section, we will showcase the result obtained for the driving event identification. L



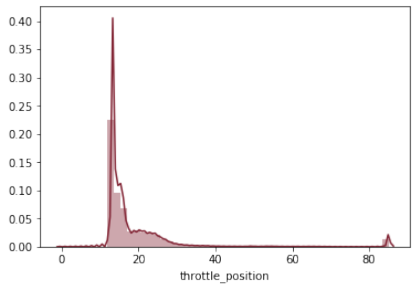
(a) Vehicle Speed



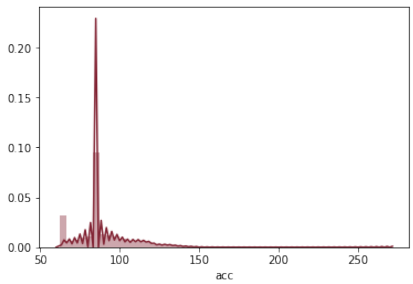
(b) Acceleration



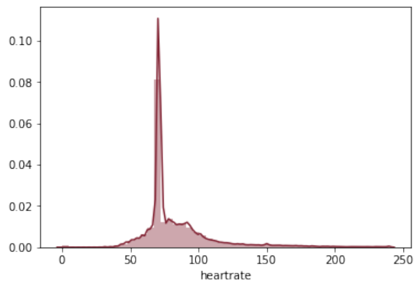
(c) Engine RPM



(d) Throttle Position

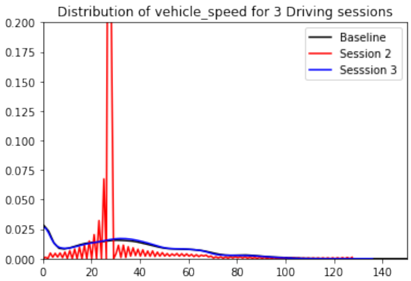


(e) CO2 Emission

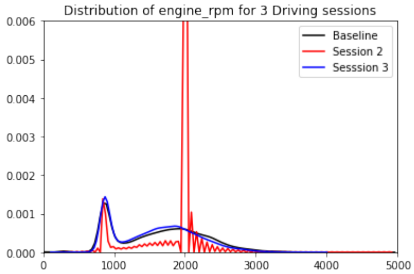


(f) Heartrate

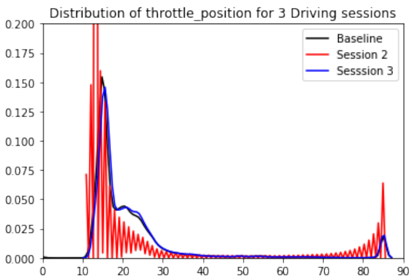
Figure 3.2. Histogram showing the distribution of various features



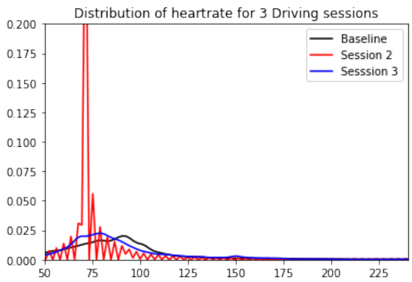
(a) Vehicle Speed



(b) rpm



(c) throttle position



(d) Heartrate

Figure 3.3. Histogram showing the distribution of various features

3.2.1 Driving Behavior Identification

ooking at the driving behaviors results individually.A period of any of the events described below is defined as the time in second between the start of the sharp acceleration event and the end of it. We have noticed:

- Sharp Acceleration: The golden rules of eco driving advises us to reduce the amount of sharp acceleration there are. In our experiment, out of 375838 seconds of driving, 3915 seconds have been spent on shart accelerations. If we divide those based on section, we realized that there was 992 periods of sharp acceleration during the baseline drive, 889 during the first eco-driving session and 882 during the second period of eco-drive. The base line drive spent a total of 1250 seconds in sharp acceleration. Sharp acceleration causes and excess in RPMs which highly affect CO2 emmissions

- Long acceleration: We observed more long acceleration during the eco-drive sessions rather than the baseline drive. Long acceleration is advised for smooth type of driving. We observed 3811 of such periods during the baseline drive, 3909 during the first eco-drive and 4345 during the second eco-drive.

- Long idling: Those are the periods in which the car is in full stop. Like expected we found more of those periods in the eco-drive session. We found 67 idle period during our baseline drive, 82 during our first eco-drive and 85 during our second eco drive. If we look at where the idling happens. We notice that it is more prominent in the city portion of the test drives rather than in the sub-urban or the highway.

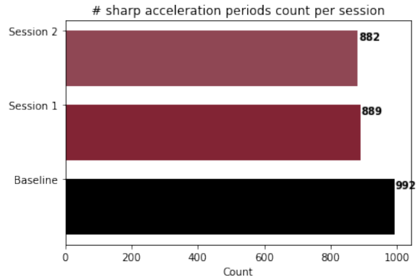
- Low speed drive: The period of low speed drives are relatively consistant with all the drives. We observed around 260 for both baseline drive and the first ecodrive and around 300 for the last eco-drive.
- High speed cruise. Regarding the high speed drives, we realized that drivers tend to drive much faster during the baseline drives than during the two eco-driving sections. Indeed we observed 273 periods of high speed drives.
- moderate start: Interestingly, we realised that moderate starts were approximativelly the same across all drives. We observed 2055 of such periods duing the first eco-drive, 2249 during our baseline drive and 2275 during our second ecodrive.
- Start/stop: We realized that the start stop sessions happen more often during the baseline drive rather than the eco-drive with periods being 21, 18 and 13 respectivelly. This may be due to the fact that either car used to completely stop during the baseline drive more often, or there was more traffic on the road during the baseline drive
- moderate brake. Like the moderate start, the values were quite similar between the baseline drive and the eco drive which counted 1236 and 1219 of such periods respectively. The second eco-drive however recorded 1341 moderate break which is relatively more than the two previous ones.

Table 3.1 shows the percentage time each driving event were spent on during the combined drives, the baseline drive, the first and second eco-drive. The driving events do overlap sometimes so the percentages shown in table 3.1 do not reflect the proportion of time an event has happen compared to the other event, but rather how much time was spent on an event independently to the others.

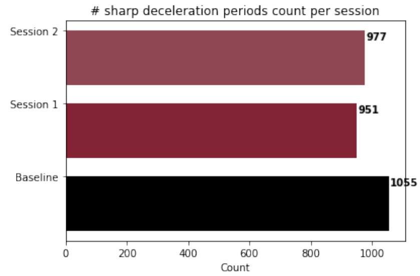
Looking at the table results, We did not witness too many sharp accelerations or decelerations, they were usually seen 1-2% of the session you are in. It goes the same for the start-stop event as it was observed sometimes less than 1% of the time. This suggest that the experiment drivers experienced very minimal situation of heavy traffic. This means than most of the drive within the experiment were quite fluid. Additionnally we observe more sharp acceleration and deceleration during the baseline drive compared to the two other driving session. This means that without prior advice, drivers tend to drive a bit more aggressively.

The moderate start and moderate stop appeared globally and respectively 5.2% and 2% of the time. Once again we can make the assumption that the car spent more time driving rather than being stuck in traffic. The baseline drive has spent more time using moderate start and break techniques rather than during the eco-driving session.

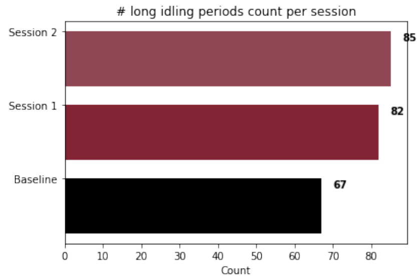
Drivers spent globally more time during long acceleration and long idling, In this paper idling refers to the drive at constant speed. This suggest that the car was spending most time driving to gain speed and keep the speed at a constant level. The Baseline and the eco-drive session 2 spent more time idling and accelerating and



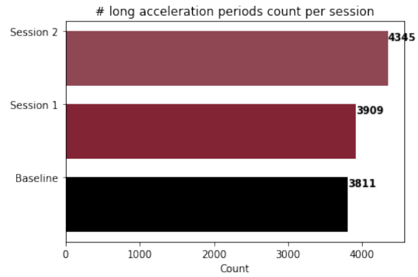
(a) sharp acceleration



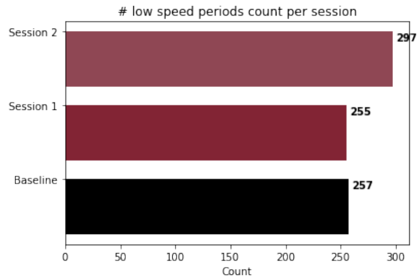
(b) sharp deceleration



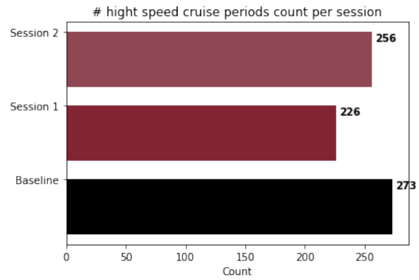
(c) long idling



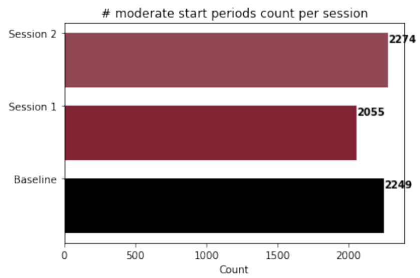
(d) long acceleration



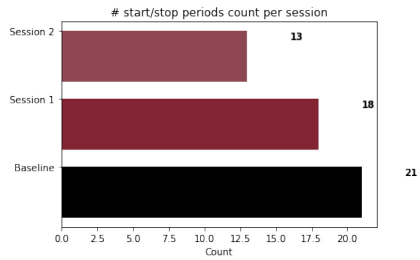
(e) slow drive



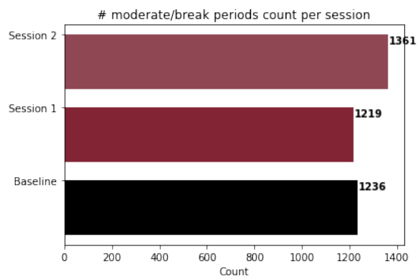
(f) high speed



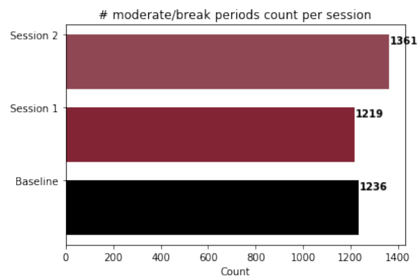
(g) moderate start



(h) start/stop



(i) moderate break



(j) Constant Drive

Figure 3.4. Histogram showing the distribution of various features

the eco-drive at session 1 spent more time accelerating and idling. This behavior may suggest that the baseline and session 2 drive were accelerating and keeping driving at a constant speed (idle) where as the session 1 drive was focused on accelerating continuously when the vehicle speed goes below a certain speed.

Drivers from the baseline drive tend to drive faster than the drivers for the other eco-driving session. We can observe that, globally, 28% of the time the drivers drove at low speed as opposed to 22% of the time spent driving at high speed. It's interesting to note that the drivers during the first eco-drive did not drive much at high speed spending only 5% of their respective time driving at high speed.

as seen

Table 3.1. % time spent on each driving event during different driving session

Driving Behavior	Global (%)	Baseline (%)	Session 1 (%)	Session 2(%)
Sharp Acceleration	1	1.8	1	1.3
Sharp Deceleration	1.5	2.6	1	2.1
Long Acceleration	63	45	79	44
Long Idling	14.79	21.5	8.6	23.4
Constant Speed	41	43.12	17.6	49
Low Speed Drive	28.2	43	17	41.4
High Speed Drive	22	14.8	5	13.5
Moderate Start	5.2	8.3	3	7.8
Start Stop Drive	1	1	1	1
Moderate break	2	4.3	1	4.32

3.2.2 Driving Behavior Effect on CO2 Emissions

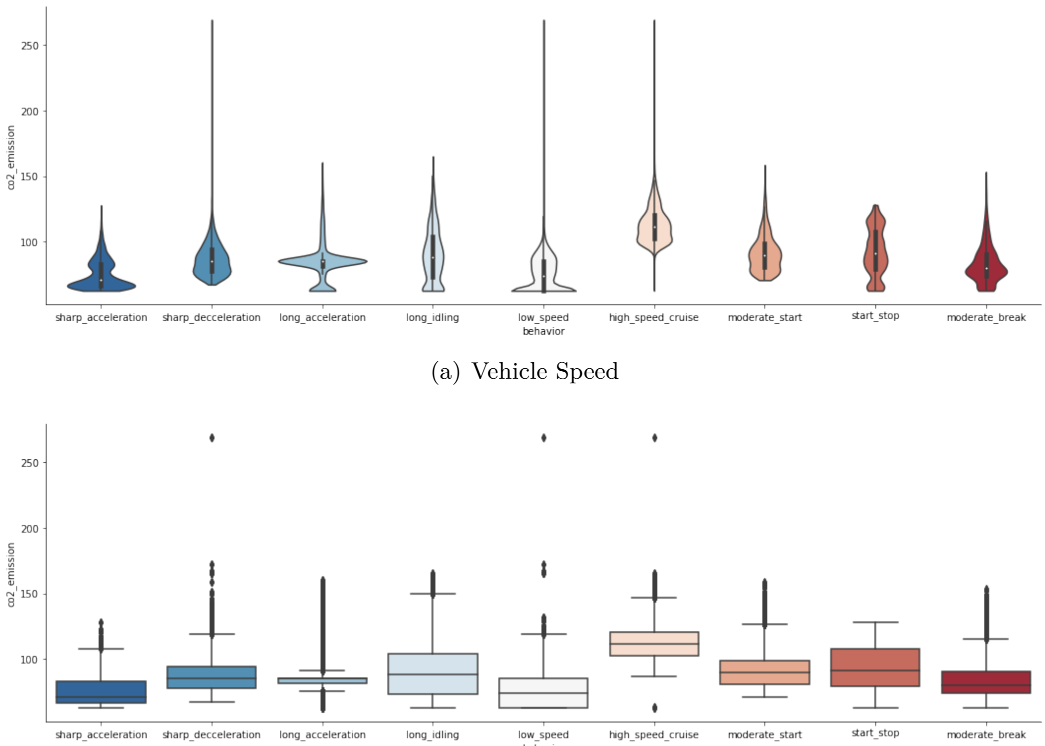
Now that we have established that, let's go ahead and check how those drives are affecting the co2 consumption on a seconds basis. We have plotted the various driving events against their respective emission rate per second.

First, one thing that we realize is that the both the sharp acceleration and deceleration are not the biggest contributors on the fuel emission level. Looking at the box plot 3.5(b) we can see that the bulk of the values are between 60-100 which is below the EU standards. We observe as well outliers that show a high level of co2 emission but those were insignificant. From that observation, we can somehow make the assumption that sharp acceleration and deceleration do not contribute directly to co2 emission, as opposed to any other driving event. However, eco-driving rule suggest to avoid this. It may not be due to its direct effect on the driving but rather due to the wear and tear that sharp acceleration or deceleration put on the engine and on the tires.

The long acceleration and the low speed drive were the behaviors that saved the most fuel. Indeed, looking at both the violin plot 3.5(a) and boxplot 3.5(b), we can see that they are lower that the other variables. Indeed, the bulk of the values for the long acceleration values are between 80 to 100 g/km with very little values going outside that range. And for the low speed drive almost all the values are below 100 which is the EU standard. Long acceleration are somewhat related to slow driving as the drivers is gaining speed in a smooth and somewhat constant way. Doing so allowed to have very narrowed emissions within the recommended range. On the other hand, it was expected that low speed drive generate the lowest level of emission as the slower we drive the less co2 we emit.

Surprisingly the long idling, moderate start and break had higher value than we originally hypothesised. The moderate break is way lower than the sharp deceleration with the bulk of its values being below 100 and having less outliers. The moderate start values though are a bit higher and more spread out than the sharp acceleration. Once again, as previously said, sharp acceleration seem to strain more the engine and tire than a moderate start would but has less direct effect on the instantaneous co2 emission rate. The long idling had tho the largest variance of all box plots but most of its values are below 100.

The highest level of CO2 emission in general was contributed by the high speed drive. Indeed, once again it make sense as the faster we drive the more CO2 is emitted. And this tendency is reflected by our data.



(a) Vehicle Speed

(b) Acceleration

Figure 3.5. Violin and box plot of the driving behavior compared to emission levels

Now that we have already established a general outlook for the effect of driving behavior on the CO2 consumption. Now we can go further and check the same effects on the different driving sessions.

Regarding the baseline drive, we can observe the that are more values above 160 g/km compared to the other two eco-driving sessions. However we observe in general the same tendencies that the general overview regarding the effect of driving behavior on CO2 emission. We can observe that driving at high speed has in general the highest Co2 emission rate as all most all the values are above 100 g/km. Similarly low speed driving has the lowest Co2 emission rate. In addition to that, compared to the other driving sessions, we realized that they were more values exceding the 160 mark.

Regarding the fist session on driving, the effect of high speed and low speed drives on CO2 consumption is even more visible here. Indeed we can really differeciate graphically that high speed drive hold the highest values where as the low speed drive bare the lowest values. The long idling possesses the highest level of variability. This is the same behavior that we have noticed during the second eco-driving session. The maing difference between the two eco-driving session is in the variance of the ling acceleration. Indeed, the noticed that in the first eco-drive session the bulk of the long acceleration values were centered at 80 where as in the second drive, and to some extend the baseline, the values were more spread out ranging from 80 to 100 for the bulk of it, but extends all the way to 150 g/km.

To conclude, wether its generally or separated in the experiment session, we can see that the high speed drive is the behavior that is emitting the most co2. The lowest co2 emission were recorded during slow drives and long acceleration. Moderate breaks emit less that sharp deceleration whereas moderate start emits more than a sharp acceleration. Long period of idling or driving at the constant speed has generally the highest level of variance (spread of the data).

3.2.3 Driving Behavior Effect on Heartrate

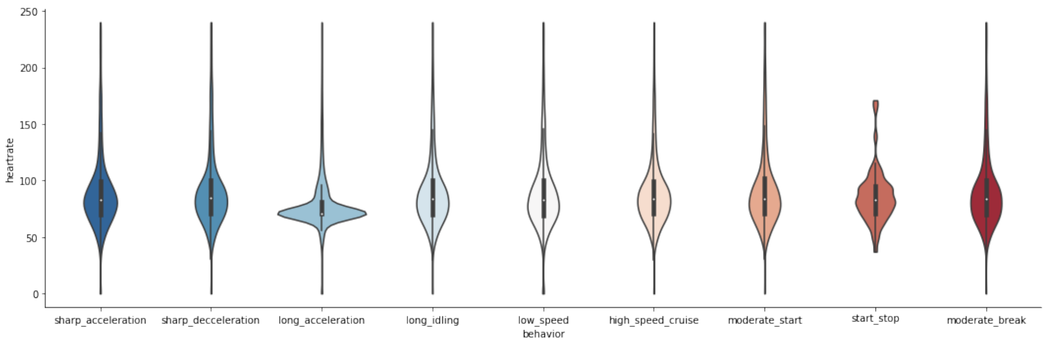
Looking at the effect of driving behavior on the heart rate at figure 3.6, we can observe that the bulk of the values are around 90 bmp which is absolutely normal for a healthy individual.

Nonetheless, there are few observations to note. First people see, to be calmer during long acceleration and during start and stop evends as the variance of those evens is a bit narrower with the majority of the values between 60-80 bpm.Drivers had a pretty similar agitation level during the rest of the driving events.

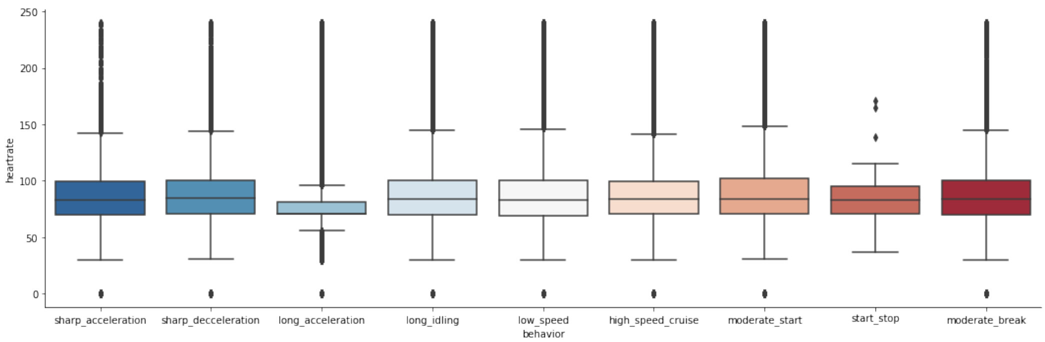
Since no interesting observation could have been deduced from the combined dataset, we will see the effect on the different driving sessions. That as well seem to be counter productive since no clear difference could have been found. The only observation to be made is that the variance is more narrow during the start and stop wheres the range of values is more spread. The long acceleration for the two eco driving session seem to be more narrow as well which shows a more calm driver.

We can conclude from those observations that drivers tend to be more calm during start and stop event, long acceleration and a bit during high speed drive. This might be due to the fact that during start and stop drivers are usually stuck in traffic, ergo not much things are happening that could increase the drivers stress level. I is the same principle that goes for the high speed cruise and the long acceleration, even those two contained a lot of period of stress.

Since those figures did not tell much about how the heart rate affects the data, we decided to filter down those results. Indeed, the more stress we have should theoretically translate to more aggressive driving behavior. This is shown by an



(a) Violin Driving Behavior/Heart rate



(b) Box Driving Behavior/Heart rate

Figure 3.6. Violin and box plot of the driving behavior compared to heartrate

abnormal heart rate per second. An abnormal heart rate is any rate at resting state that is the equivalent of an intensive exercise. In other words, any heart rate that is above 150 bpm shows a situation of stress.

We went ahead and recorded the data that match that criteria and counted in seconds the amount of time that that happens. This resulted in some interesting events, for starter, it seems like the event in which stress is less likely to happen is during start and stop. I make sense since most of the time that that start-stop happens is when stuck in traffic, no many stressful events during that period. Compared to the other 9 events a high level of heart rate during a start-stop situation is less likely to happen. We have noticed as well a low amount of stressful events during both the sharp acceleration and deceleration but as well during the moderate start and break.

Looking at the events that may have affected the heart rate the most, we notice that the low speed, constant speed, and long acceleration have the highest value count. Indeed, those events are most seen in a city portion of the test drive. This makes sense since more traffic means more road awareness from the drive. In addition to that, in cities lines are not separated which creates a lot of situations that increase the driver's awareness and therefore their stress level. Much fewer stress situations were recorded during the high-speed cruise, which is related to driving on the highway. This is because not as many road events happen on the highway as opposed to the city.

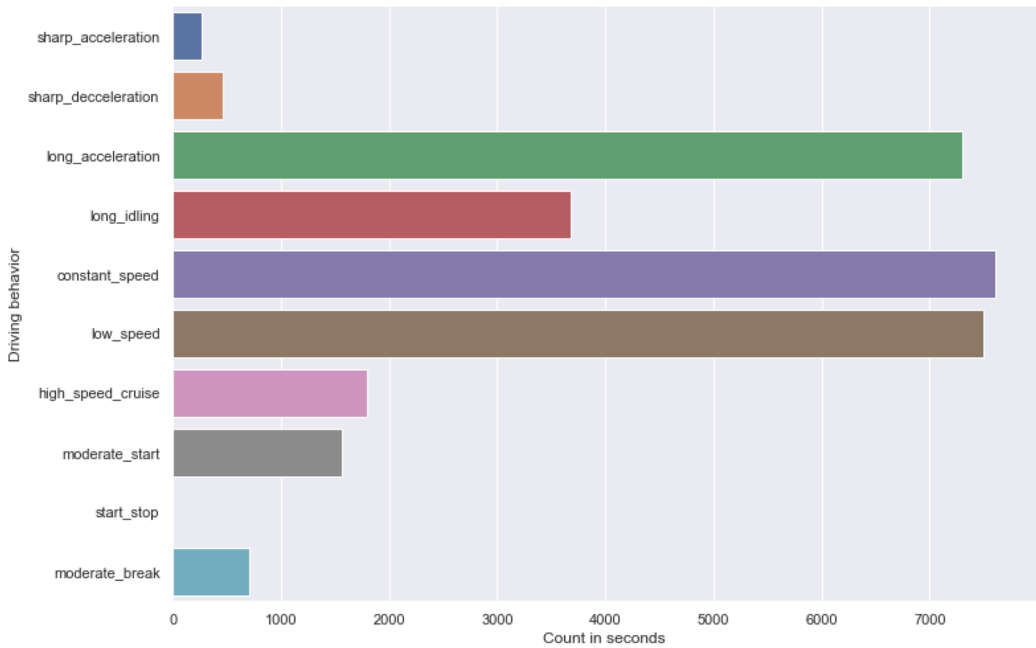


Figure 3.7. Count in seconds for increased heartrate per event

We counted over 7000 sec of high heart rate between 24 drivers. This shows that during a drive, drivers will experience some sort of stress. This stress can be associated with the fact that they are driving a new car or they are driving in various parts of the city. First, the first observation here is that an abnormal heart rate happened in all the parts of the circuit, whether the drives happened within the city, in between the cities or on the highway. It does not matter where you drive an expected event that may happen at any portion of the road which is why it is advised to pay attention when driving.

We narrowed down the areas that contained the highest density in terms of increased heart rate, represented in figure **??**. The blue markers represent events in which there were heartrates above 150 bpm. The first event we notice is a high concentration of the abnormal heart rate activity at the beginning of the circuit, even though a normal rate of traffic was found at the car was driving at a normal pace between 30-40 km/h. This stress might be caused by the fact that the drivers were getting accustomed to driving with a new or unknown car. This initial stress seems to go away after the first kilometer of driving.

The next portion of the road takes us through an intercity road taking us to the city. One thing that we realize in this portion is a small decrease in the amount of increased heart rate. We observe those increase only during a period of turn or close to an exit road. This decrease in heart rate may be due to the decrease in events that happen when driving in a suburban road.

The nex portion leads us to the city portion of the road. The urban area is the portion of the circuit in which the test drivers experienced the most stress. Indeed we have observed a high density of increased heart rate in areas having a high level of traffic, according to Google Map Typical Traffic for the test period. This increased level of heart rate might be due to the increased awareness when driving to the city as the number of obstacles, cars, traffic signs, and unexpected events tend to increase within the city.

The last portion of the road takes us to a highway and back into the intercity/suburban roads. In that portion, we saw again a decrease in the increased heart rate. The part in the highway that contained the highest density of abnormal heart rate values was close to a highway exit or entrance. This may be due to an increased awareness of the drivers to new cars entering the highway. The other increases that are not close to an exit/entrance may be related to car surpassing another or being surpassed.

To conclude, driving under stress as said before leads to unsustainable driving, and even though during an average drive, the heart rate seems to be the same; there exist portion of the road in which we experience short periods of high stress. Even though this unsustainable driving due to stress is minimal during an average drive, and does not have a high direct impact on fuel consumption, as opposed to speeding, for example, it may lead to the general wear and tear to the car whether it is on the engine, tires, etc. This general wear and tear might down the line, reduce the driving efficiency of the car. An interesting observation from this section is that drivers tend to be more stressed in the city rather than the highways or suburban streets.

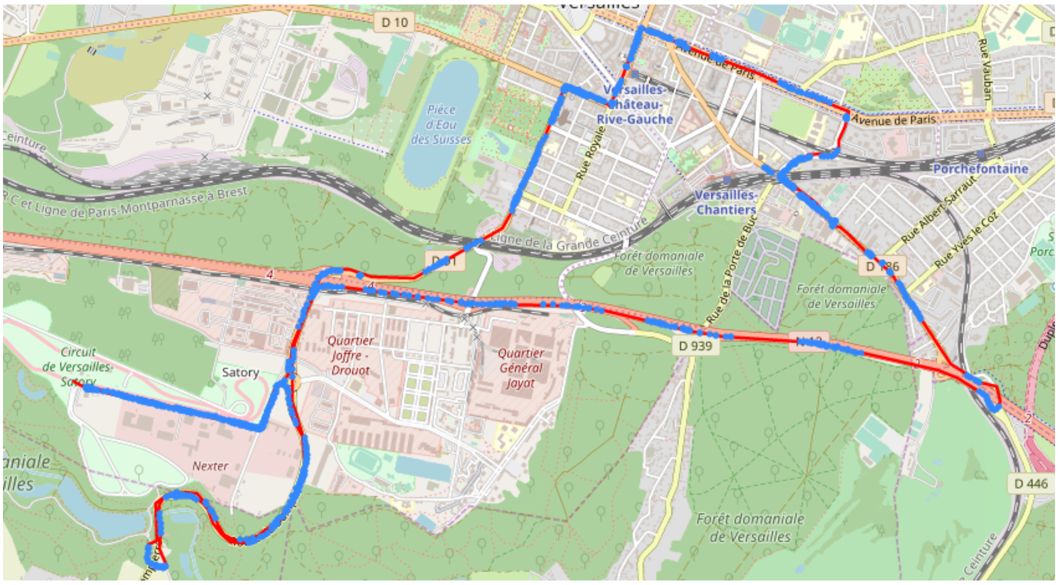


Figure 3.8. Area with increased heartrate

3.3 Machine Learning Model Evaluation

In this part, we will be displaying the results of the machine learning evaluation for the prediction of the CO2 emissions. The goal of this section is to better understand the relationships and the impact of the features on the CO2 emissions.

The final dataset was made of 9 variables that corresponds to the behavior identified.

3.3.1 Features Selection

We have already established in section 1 that the faster we drive the more CO2 consumption there is and consequently the more fuel we consume. As a result, the vehicle speed is the more important feature in estimating the CO2 levels. Indeed, combined with the other variables present in our dataset, we managed to get an R2 score of about .99 for the decision tree, random forest, gradient boost, and MLP regressor. This means that having the vehicle speed will allow us to predict to C02 consumption of the user almost all the time.

However, for this paper, we want to know the influence of the other variables on speed and how well do they predict fuel consumption. Our final features will be made of the 9 driving events and the throttle position, heart rate, engine rpm, and short term fuel trim bank.

3.3.2 Model Selection & Evaluation

Few models were run through a pipeline where and we recorded the model results based on the RMSE and the R2 score. The final model that we will pick, is the model with the best score. Table 3.2 represents the results for the different models.

Those models used the default hyper-parameters values for our first run of the model. All those models were trained on a ratio of 70/30 train and test set.

Table 3.2 shows the result of our first model run with different types of models. The scoring is represented by the RMSE and the R2 score. Looking at the range of values, we can notice that the ensemble/tree-based methods are the ones that are performing the best. We can see that the gradient boosting and the random forest have the highest score in terms of predictive accuracy. They are closely followed by the Neural Network Methods and the linear models. Regarding the scores themselves, linear models seem to be underperforming if we compare them to the other model families. Indeed, they had an R2 score of around 77 (with the lasso being around 70) but a very average RMSE of above 4.5. Again those results are not bad, but the other families of models have done a bit better, both the tree and the neural network. The tree-based models (Random Forest and Gradient Boosting) are the ones that were able to better predict the CO2 emissions with R2 scores of .86 and .87 and RMSE of 2.1 and 2. This tells us that those models fitted almost perfectly and were generally of by 2 points, which is better than the rest of the other families. The Neural Networks perform a bit less good than tree-based families with R2 scores of .84 and .83 and RMSE of 2.2.

as seen

Table 3.2. Scores for the model used to predict CO2 emission rate

Model	R2	RMSE
Multiple Linear Regression	0.7789	4.5478
Ridge	0.7789	4.5478
Lasso	.7057	5.24762
Decision Tree Regression	0.7497	4.8860
Random Forest Regression	0.8656	2.2039
Gradiant Boosting Regression	0.8706	2.1638
Multi Layer Perceptron	0.8432	2.787
LSTM	0.8539	2.6967

Consequently, we will be using the tree-based algorithm to attempt to improve the results. In addition to that, the tree-based methods have a feature importance feature which will allow us to understand which variable contributed the most/least to the prediction, ergo telling us the impact of each variable on the fuel consumption.

3.3.3 Hyperparmeter tuning

The hyperparameter tuning is going to be done in two phases as said in the previous chapter, first, we are going to be performing random search cross-validation by setting a hyperparameter grid and then we are going to be using the result from the random search to perform a Grid Search with cross-validation.

Our initial random search will have the folling grid.

- 'bootstrap': [True, False],
- 'max depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
- 'max features': ['auto', 'sqrt'],
- 'min samples leaf': [1, 6, 12,
- 'min samples split': [2, 10, 20],
- 'n estimators': [10, 20, 50, 100, 200, 400, 600]

The random search training took approximately 2h 55 min to train 100 iterations cross-validated 3 times. This improved our model by 0.58 % (0.8657) from the base model. The best parameters found by our random search are as followed.

- 'bootstrap': True,
- 'max depth': 80
- 'max features': 3
- 'min samples leaf': 6,
- 'min samples split': 10,
- 'n estimators': 10

Finally, we are going to create our final grid for the grid search, we will pick the values close to our best estimator from the previous check. The grid search took 2h 20 min to train all possible combinations. Consequently here is the grid used for the grid-search with the best parameter highlight in bold.

- 'bootstrap': True,
- 'max depth': [80, 100, **150**, 200]
- 'max features': [2,**3**, 4]
- 'min samples leaf': [**4**, 6, 8],
- 'min samples split': [8, **10**, 12]
- 'n estimators': [10, **15**, 20]

Using the above allowed us to improve the result of our model by 1.58 %. This makes our random forest model as optimized as it possibly can be and we are ready to move to the next step which is the extraction of the various feature importance.

3.3.4 Model Result

After hyperparameter tuning, we train the model on our dataset using the best parameters found on the previous step.

Table 3.3 shows the result of the prediction of the random forest model on the different ratios of train test. The best performing on has a ratio of 70/30 with a score of 0.8766 which is around the best score it can get, which is acceptable considering that there was not the speed variable.

Using both the random forest and the gradient boosting (see the result in Annex), we were able to have a good fit with an R2 score close to 1. It is not perfect but it gives us a good prediction of the emission rate which may be off by 2-5 g/km which is acceptable. The result both with the gradient boosting tends to plateau at an R2 score of 0.88, which is a good indicator of how well the model is predicting the CO2 rate.

The RMSE for most models was quite low in general, we can see that any model with an R2 score above .80 will have an RMSE below 2.5 which is good and show acceptable predictions. Most of the value of the CO2 emissions are between 60 to 120 g/km, ergo being off by 2 points is acceptable. Anything model with an RMSE score above 3 shows an average or poor model performance.

Taking a second look at the results show on table 3.3, we realize that the ratio 50/50 held the worse result, with an R2 score of 0.4271 which shows a very poor fit and an RMSE of 5.5 which shows an even larger variability to the regression line. This model cannot be used further as it cannot predict very well the CO2 emission making extracting feature importance from that model irrelevant. The 90/10 we have a pretty good R2 score but a very average RMSE for this data. This may be a sign of overfitting. In general, the 70/30 ratio works the best, as we have both a good R2 and RMSE score.

Looking at the Gradient boost results presented in Appendix B.1, we can see a similar trend as the one for the random forest> Indeed, we observe pretty low results for the 50/50 train test ratio with an R2 score of around .08 and an RMSE of 12. This model completely failed to learn the CO2 emission rate from the data. The 90/10 train test ratio did a bit better but the high RMSE of 6.2 shows again an overfitting trend. The 70/30 train/test ratio, however, performed the best.

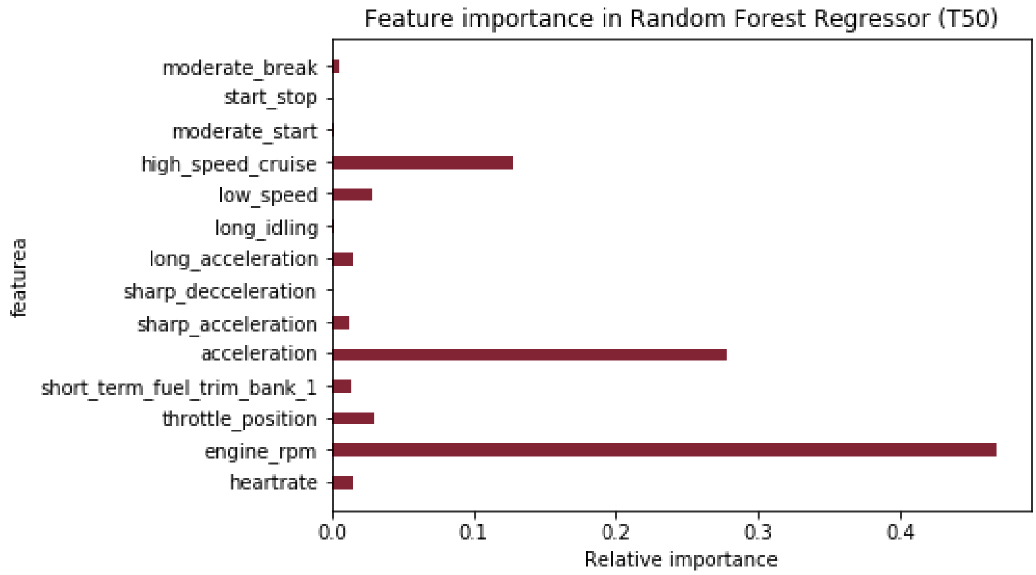
Table 3.3. Scores for the Random Forest model used on various train/test ratio

train/test Ratio	R2	RMSE
50/50	0.4271	5.504
70/30	0.8766	2.016
90/10	0.8658	3.4984

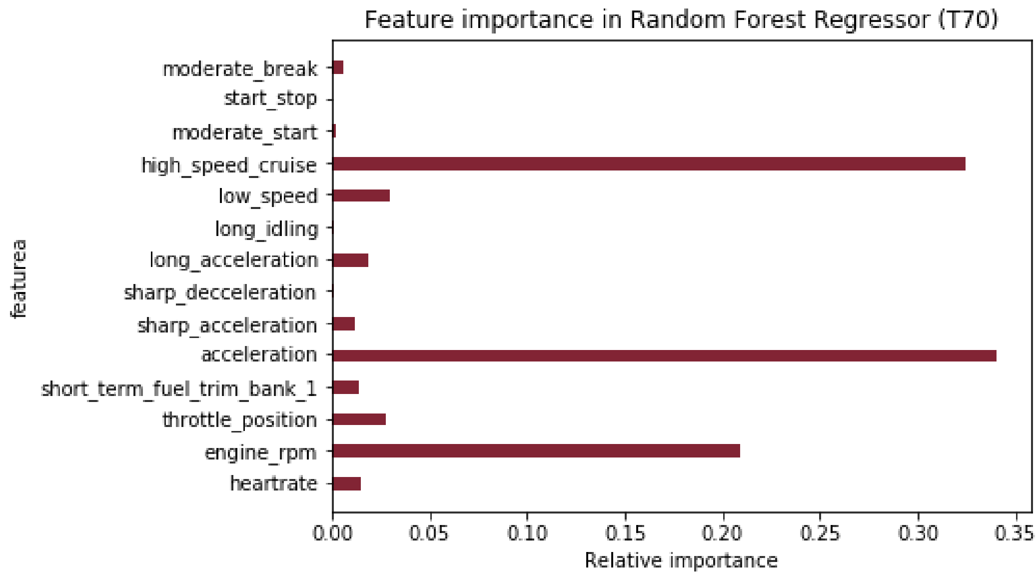
Fig 3.9 shoes the most important features that have been used to make the prediction. We ran the random forest model at different training/ testing ratios and in the first training/test ratio of 50/50 we notice that the engine rpm was the most important feature followed by the acceleration and the driving event of a high-speed cruise. If we look in detail about the driving events identified, we noticed that the high-speed cruise, low-speed drives, long acceleration, and sharp acceleration helped to some extent make a prediction; whereas the rest of the driving events were irrelevant in predicting the CO2 emission rate during a drive. Knowing the relevance of driving behavior may reflect which of the driving behavior directly or indirectly affects the C02 emission rate.

Looking at the features regarding the driving, we realize that, in general, the acceleration was the most important feature, followed by the engine rpm and the throttle position. We can observe that any feature that helps the car gain momentum tends to have higher importance. Another interesting aspect is that the heart rate plays a small role in predicting the C02 emissions but only for the random forest prediction. The results for the gradient boosting (See annex) however do not take into account the heartrate as a contributing variable for the prediction.

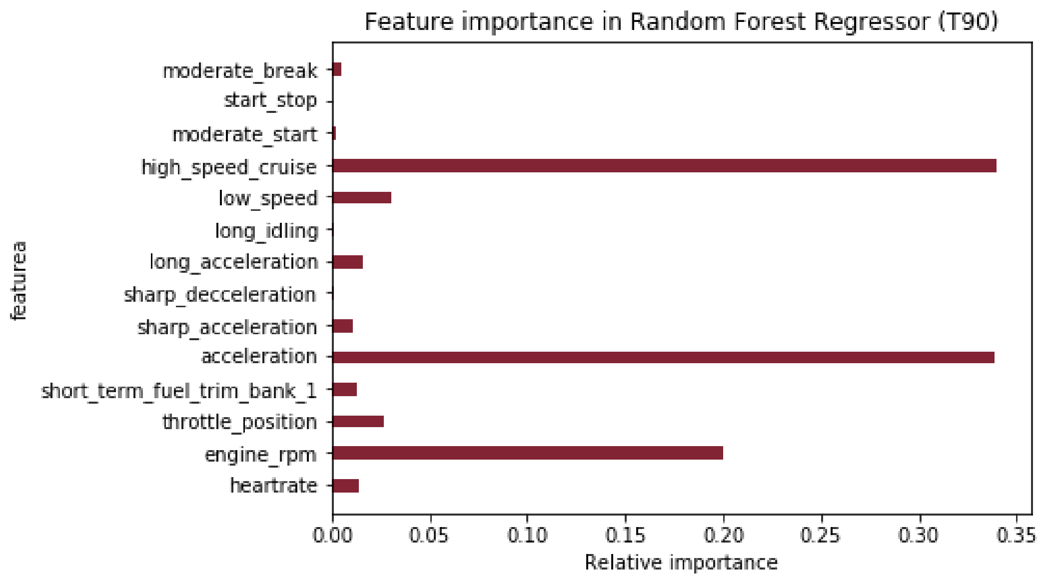
Generally, both the random forest and the gradient boosting take the same variable as best suited for prediction. Indeed we observe in both cases, trained and tested with a different ratio that, from most important to least important, the acceleration, high-speed cruise, and engine rpm are the top 3 (besides the vehicle speed) variable to prediction CO2 emission. The common part aspect about those variables is the fact that they are all somewhat an indicator of how fast a car is driving, and the faster the car drive, as seen in the literature review, the more CO2 is emitted. Ergo any feature that helps know approximatively the speed of the car, will be considered as the more relevant feature to predict the level of C02 emissions.



(a) train/test ration 50/50



(b) train/test ratio 70/30



(c) train/test ratio 90/10

Chapter 4

Discussions

In this section we will discuss first the result of the experiment from the setting to the machine learning model, then finally we will talk about how we could apply what we have learned in a real-life situation.

Looking at the literature review, we realize that a lot of work has been done in terms of eco-driving and its effect on fuel consumption and CO2 emission. Even though all results seem to indicate that the more eco-friendly we drive the less fuel we consume saving anything between 2 - 20 %. In other to achieve those results and make those conclusions, experiments like the one described in the methods sections have been set and studied. However, those experiments did have their flaws which we described in the research issues section in the methods. Similarly, our experiment had its flaws which were mainly related on one hand to the obtention of the data and on the other hand on the experiment setting in itself.

Acquiring different types of data proved to be quite an issue, especially from different test sites. Indeed, the data was originally made from the combination of 3 main blocks. The first one was extracted from the OBD II reader, the second one from the android device and the recording sensor (Fitbit) attached to the driver and the last one from the data engineered by the team. The two major problems were found in the data itself. The first one is the fact that the sensors were emitting at a different rate, some emitted ever 5 sec, and others at a rate a bit less than a second. This left us with a data set made of a lot of empty value. We fixed the problem by resampling and using Singular Value Thresholding imputation methods (described in Section 2) to replace those mission values. The problem in doing so is the fact that we may lose some information or to the least simplify which may temper the originality of the value.

The other issue found within the data itself is related to the standardization of the data itself. Indeed, as described in the methods, the data was organized in traces/trips represented in CSV files. However, depending on the test site, a lot of those trips were lost because they did not pass the original quality check. Consequently, we had to delete a lot of traces which decrease the amount of information we had. Some of those traces were part of the experiment, which consequently tempered the quality of the results. In addition to that, we did not have the same features in all the traces. Indeed some traces contained certain features, others did but they were empty and that played a major role in terms of feature selection. For example, the gear was present in some trace and absent in other, which made it very hard to combine the different traces into one suitable for learning.

Another issue about the data itself is the lack of features that would indicate the level of fuel consumption directly taken from the OBD II sensor. Indeed, in some traces the fuel consumption and fuel tank level are present but most of them are empty. We needed to estimate the level of CO2 using a function from the Oak Ridge Laboratory present in the literature review. This is somewhat important because it turns our result to the estimation of CO2 consumption and not actual consumption.

The second flaw was related to the experiment itself. Indeed, we had various types of drivers of different age, driving a different car. Regarding the drivers, the experiment took the recording of a baseline drive, followed by 4 drives that are done in an eco-driving way. Since we are somewhat checking the effect of eco-driving on fuel consumption and the act of driving is considered a habit, drivers should not be knowingly driving in an eco-friendly way. That means that the drivers should have had learned how to drive before 2013 as any driving training done in the EU past that date, teaches how to drive in an eco-friendly way. In addition to the experience of the driver, no information was given about the state of the car itself. Indeed, we have seen that a not well-maintained car can lead to a higher level of CO2 emissions.

We have seen in the literature review that some external factors such as the weather, the traffic signs, the slope of the road, the additional load of the vehicle or the number of occupants within the vehicle affect the fuel consumption and the CO2 emission. The weather affects both the driver's psychological state and the mechanics within the car (air conditioner usage, general engine wear, and tear). The traffics sign and intensity affect the amount of time one needs to stop, start, accelerate and all the like which encourages aggressive behavior. The steeper the slope the more fuel the car consumes. And extra load or passenger increases the car weight which consequently increases the amount of energy required to move a car. Little to no information was given about those factors. And if those vary between drive subsequently between different drives, it may affect the accuracy of our results.

Looking at the driving events, we were able to identify 10 distinctive driving events from our data. The thresholds used to identify that behavior was based on literature and common sense. Those parameters for the threshold may not be optimal, however, the robustness of the driving behavior identification system should be further tested on the field in a supervised manner to be more certain of how well it works. Those type of testing goes unfortunately beyond the scope of this paper.

Taking a deep dive into the driving behaviors, we tend to see that globally long accelerations are the more often state and very little time is dedicated to start and stop. That means although google showed red for the typical traffic within some section of the drive, generally, the drives seem to be quite fluid. The interesting part of the experiment itself is the fact that the circuit used different types of environments from urban, semi-urban and highway which brought a diverse range of environments each challenging drivers to different eco-driving rules which somehow simulate real-life situations.

Having that in mind we tend to observe that drivers know how to make eco-driving adjustments when instructed to. Indeed, from the various baseline drive, we saw that the drivers managed to correct successfully various non-eco drives. For example, looking a the number of periods for sharp acceleration/deceleration, which is an eco-driving rule advises to avoid, we can see an adjustment made by drivers that show a decrease of those events from the baseline drives to the 2 eco-drive sessions. This information can prove useful in terms of the development of systems or trainers that could help monitor and advise drivers in the correction of their driving behavior.

Looking at the effect of driving events on fuel consumption, we tend to observe that the higher the speed leads to higher energy demands which lead to more fuel use and more CO2 emitted. The CO2 estimation tends to show that the high-speed drive consumes more fuel than the rest of the events whereas the low-speed drive tends to consume less. But we cannot conclusively say that one contributes more than the other since we do not have an actual reading of the CO2 emitted by the car, but rather an estimation. A further test needs to be done by comparing these results with traces containing actual readings to conclusively say that an event leads to more CO2 emissions.

The driving effects on the heart rate is a bit interesting on a global level. Indeed we tend to observe that most events have an average of 90 bpm which shows that the driver was in general calm. But looking at it closely we can see that drivers tend to stress more in the city portion of the drives rather than the highway or the suburban portion. We also noticed that the stress levels were more prominent during the long acceleration, long idling and long constant drive, all typical behaviors from drives done within a city.

Another way to measure and understand the impact of driving behavior on CO2 emissions was through machine learning. Our model tries to predict the CO2 emissions base on driving events and vehicle features. Generally, those systems did very well in predicting the CO2 emissions based on the R2/goodness of fit score and the RMSE showing quite good values. They were trained with different models but the tree-based model seems to be the more adequate ones for predicting CO2 emissions. One advantage of using the tree-based algorithm is the fact that we could derive the feature importance. Since the model predicted the values very well, we could deduce that those values play a more important role. Interestingly, the result shows that any feature that helps put speed to the car helps the most in predicting the CO2 levels.

If we have to talk about the possible future application of this work, having a model that could learn your driving behavior and predict what you will be emitting could be beneficial in putting in place supports that could help drivers monitor their driving behavior. Indeed, over the years, we have seen more and more integrated computers within the car dashboard that display various information about the car. Knowing that it is possible to identify driving events and that through this experiment, drivers can adjust their driving behavior based on advices given before the drive, there could be systems that monitor your driving behavior and help advise their user in driving in a more eco-friendly manner. Through the use of the data already picked by the sensor within a car, and the use of an even detection model such as the one presented in this paper, we could be able to build a useful system that could monitor the various variables and driving behavior and advise the user on what to adjust while driving.

Chapter 5

conclusion

Over the past decade, a lot of effort was put by the European Union for a greener environment. The transport sector was not spared with a lot of policies enforcing a more sustainable approach. We have seen that throughout his paper and among other related papers, that eco-driving will save anything between 2% to 20% of your annual fuel bill.

A lot of research has been done to understand the effect of, and advices on the concept of eco-drive. We have seen in Section 1 the way eco-driving evolved overtime. It started with an instauration of rules or golden/silver rules that were tested by various researchers to decrease fuel consumption due to driving. Those established rules state that in summary, vehicle mass, speed, gear, road slope, number of passengers, etc. have a role in the increase of CO2 emissions.

Based on the literature above, we investigated if we could identify driving behaviors that are related, positively or negatively to the increase of CO2 emissions. This is our we managed to find 10 different driving behaviors events which were saved in a fuzzified format. From those identified driving behaviors, we have observed how they were distributed and their direct effect on CO2 emissions. We have realized that high-speed drives are the one that contributes the most the increased CO2 emission rate. This is due to the U-shaped emission curve that emits more CO2 during low and high speeds. Drivers should do their best to drive at the car's optimal speed. This optimal speed, unfortunately, varies from car to car.

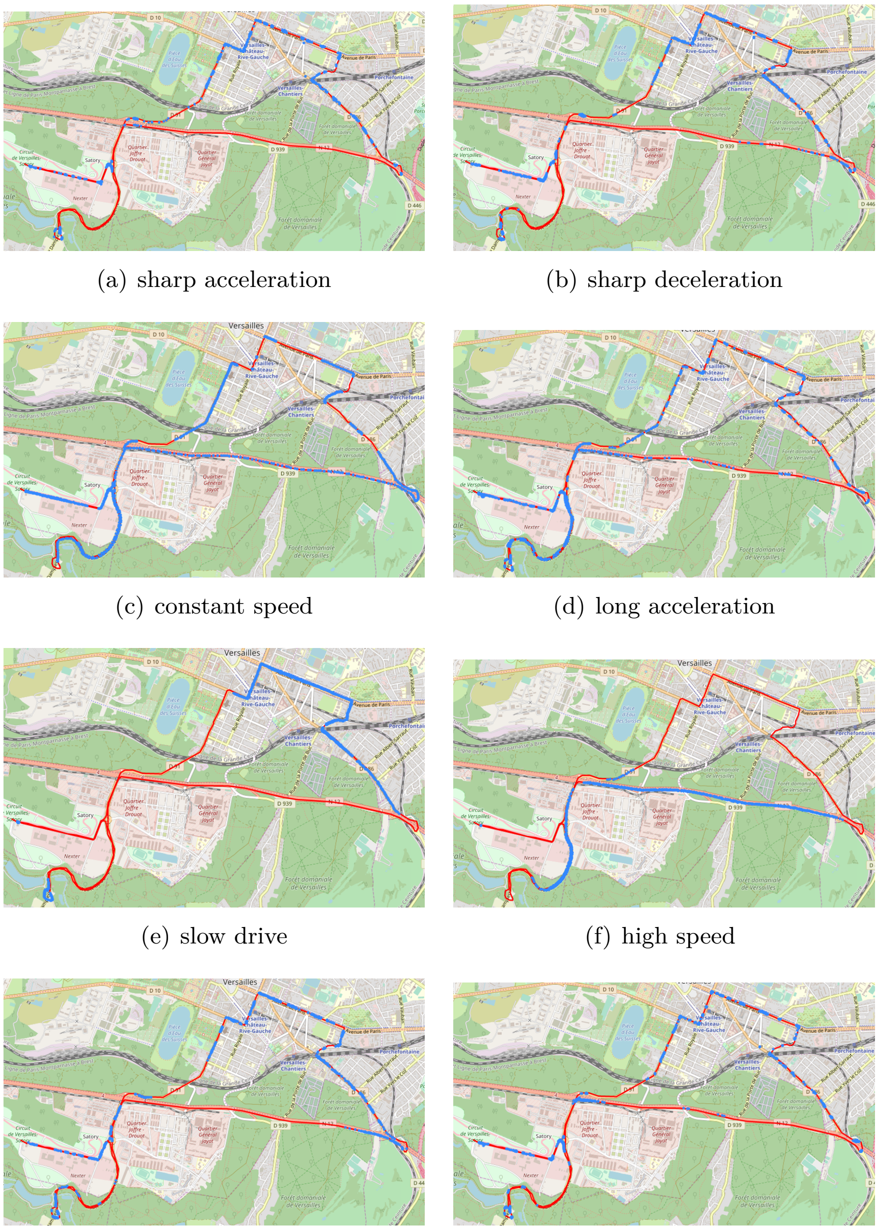
Upon identifying the behavior, we then observed the effect they had on the heart rate. We realized the average heart rate does not change over time and that situation of stress happens in every single event. However, plotting the excess stress revealed that there is more stress (higher rate of abnormal bpm) in city roads rather than in highways and suburban roads.

Finally, we wanted to see what are the most important driving behavior to predict the CO2 emission rate and how well does a model predicts those. Hence, we ran our data set through a pipeline that would impute the data, pre-process it, do hyperparameter tuning with a random search, followed by a grid search before running a regressor on it. After trying several regressors we realized that the gradient boosting and the random forest work better in predicting the C02 rate. Indeed, we have managed to create a model with a pretty decent accuracy from which we derived that high-speed drive and engine rpm are the most important features for prediction.

Lastly, there are so many applications and benefits in implementing an eco-driving habit. With the increase of technology and the EU 2050 no emission goal, we are positive to see and advancement in eco-driving behavior. The ecological benefit along with the financial saving contributes a lot for the EU to push that habit more and more into the driving plans in the Eurozone. Our first phase will be to attempt to identify a model or family of models that would work the best with the dataset in our disposition. At the end of this phase, we will pick a family of models that would be the most appropriate to make a prediction and extract the most relevant information to understand the impact of the variables on fuel consumption

Appendix A

Drive Behavior



(a) sharp acceleration

(b) sharp deceleration

(c) constant speed

(d) long acceleration

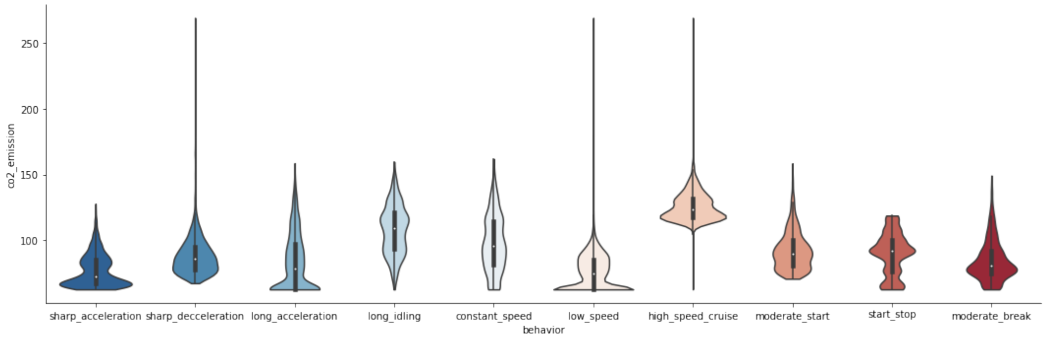
(e) slow drive

(f) high speed

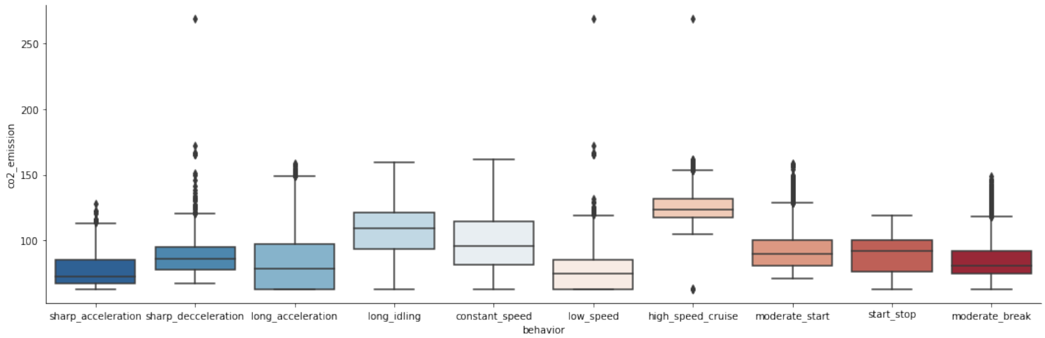
(g) moderate start

(h) moderate break

Figure A.1. Most common location of driving behaviors during drives.

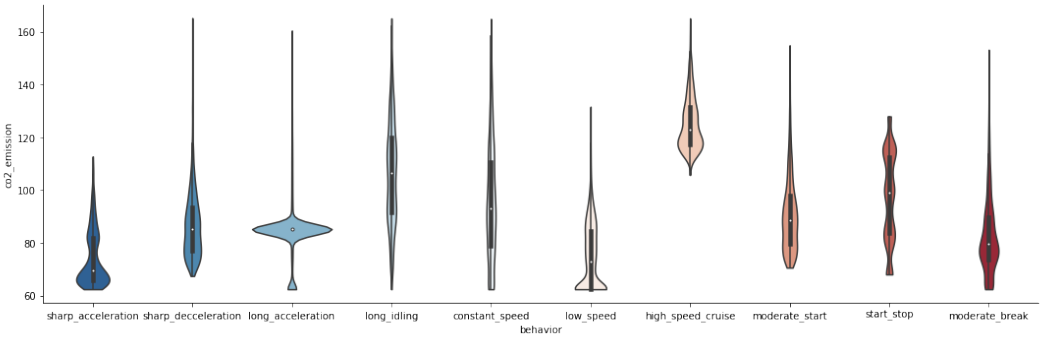


(a) Violin Driving Behavior/CO2 emission

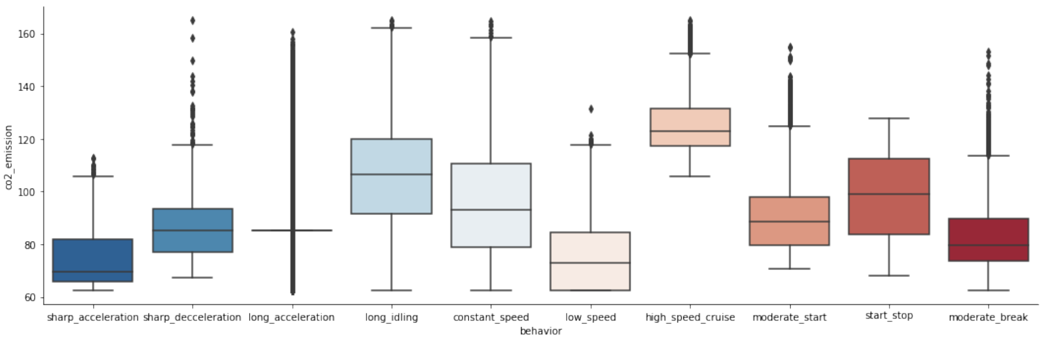


(b) Box Driving Behavior/CO2 emission

Figure A.2. Distribution of the driving behavior compared to CO2 for the Baseline Drive

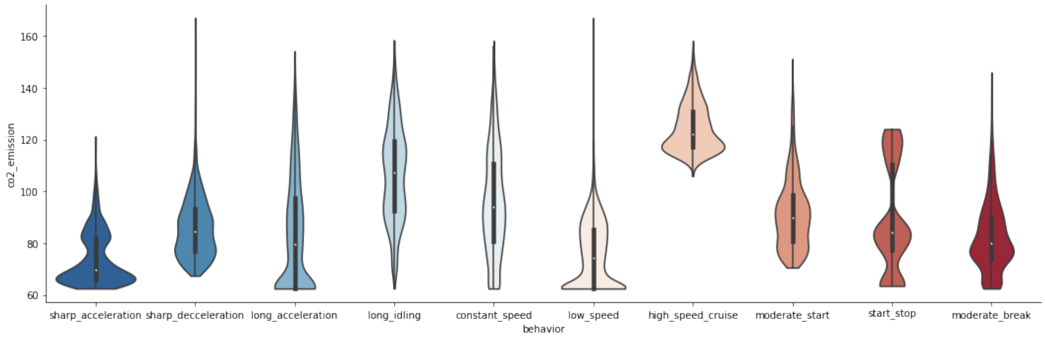


(a) Violin Driving Behavior/CO2 emission

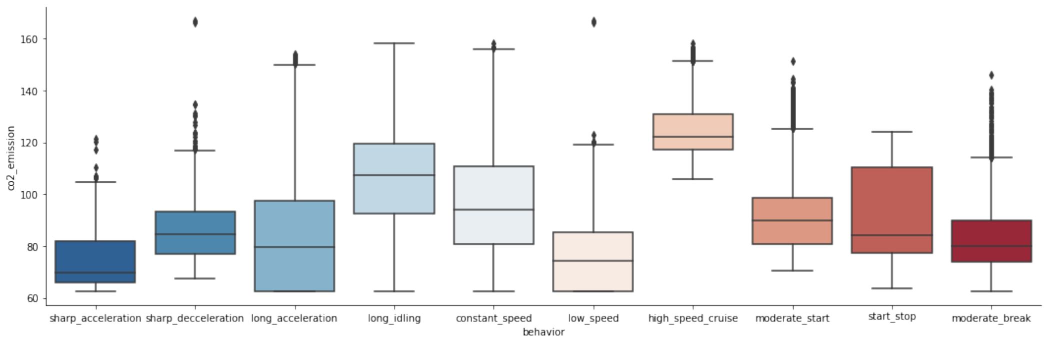


(b) Box Driving Behavior/CO2 emission

Figure A.3. Distribution of the driving behavior compared to CO2 for eco-drive session 1



(a) Violin Driving Behavior/Co2 emission



(b) Box Driving Behavior/CO2 emission

Figure A.4. Distribution of the driving behavior compared to CO2 compared for eco-drive session 2

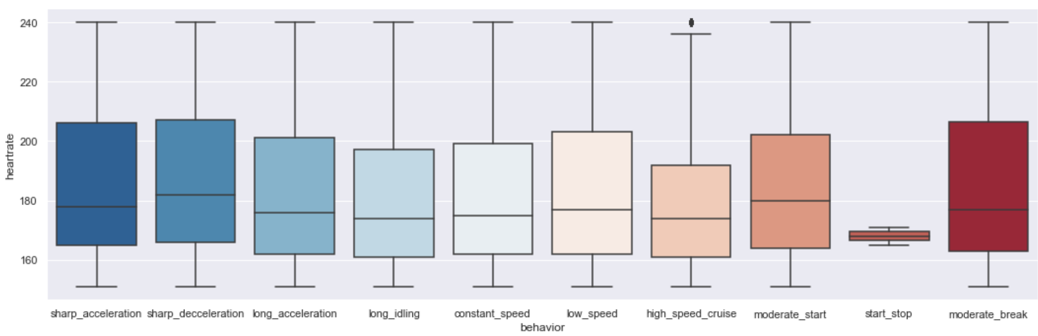


Figure A.5. Distribution of heartrate above 150 by driving behavior

Appendix B

Gradient Boost Results

Hyperparameter Tuning: Random Search initial grid:

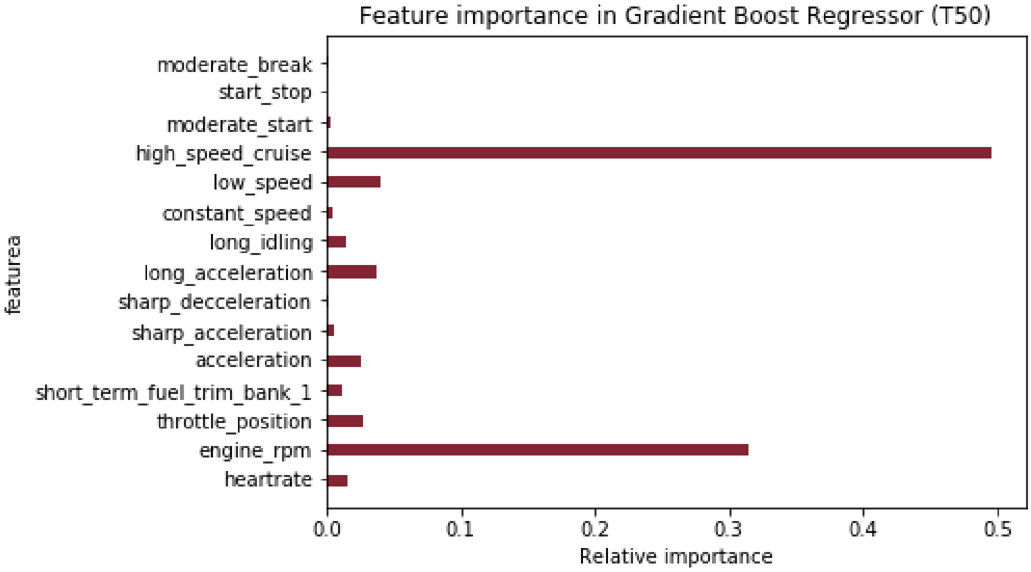
- 'alpha': [0.1, 0.5, 0.9, 1],
- 'max depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
- 'min samples leaf': [1, 6, 8, 12]
- 'min samples split': [2, 4, 6, 10, 20],
- 'n estimators': [10, 20, 50, 100, 200, 400, 600]

Hyperparameter Tuning: Random Search best hyperparameters

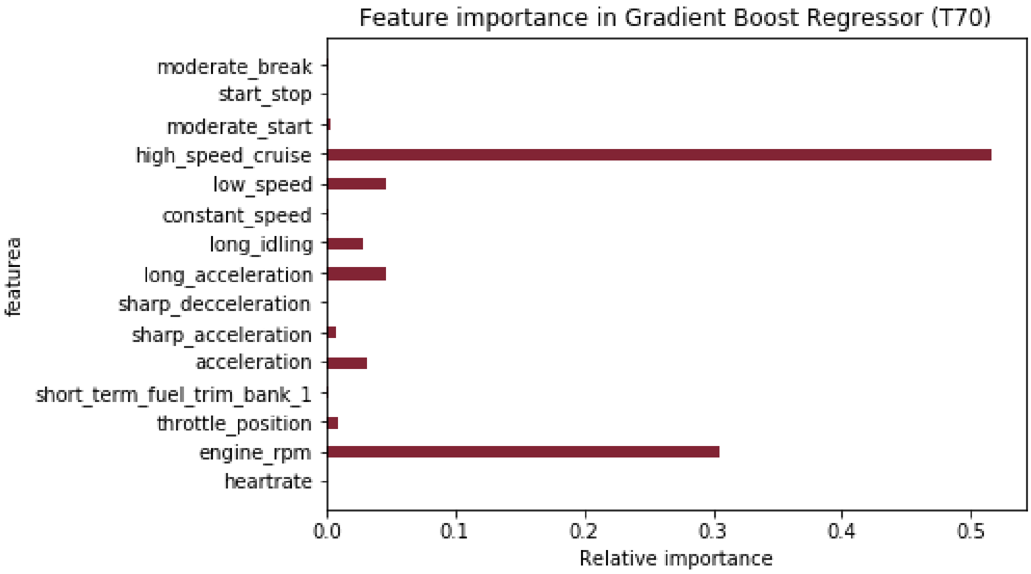
- 'alpha': 0.9,,
- 'max depth': 100,
- 'min samples leaf': 6
- 'min samples split': 4
- 'n estimators': 50

Hyperparameter Tuning: Grid Search Results

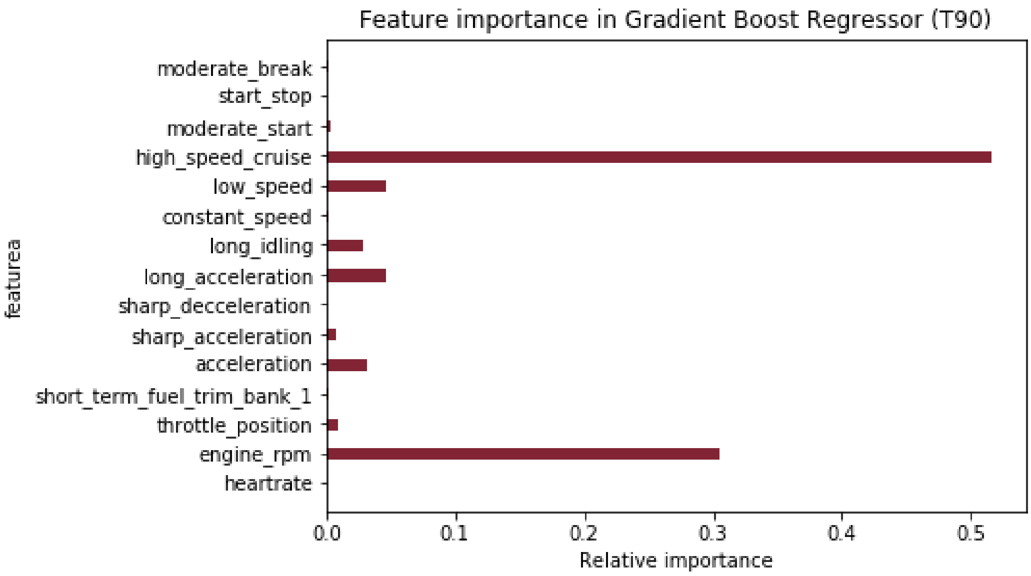
- 'alpha': [0.8, **0.9**, 1],
- 'max depth': [100, 120, **150**],
- 'min samples leaf': [4, **6**, 8]
- 'min samples split': [**4**, 6 8],
- 'n estimators': [10, 20, 50, **100**]



(a) train/test ration 50/50



(b) train/test ratio 70/30



(c) train/test ratio 90/10

Table B.1. Scores for the Gradient Boost model used on various train/test ratio

train/test Ratio	R2	RMSE
50/50	0.08	12.22
70/30	0.8866	2.016
90/10	0.8658	4.4984

Bibliography

[1] Reducing co2 emissions from passenger cars. `https://ec.europa.eu/clima/policies/transport/vehicles/cars_en#tab-0-0`. Accessed: 2019-09-02.

[2] af Wåhlberg, A. Short-term effects of training in economical driving: Passenger comfort and driver acceleration behavior. *International Journal of Industrial Ergonomics - INT J IND ERGONOMIC*, **36** (2006), 151. `doi:10.1016/j.ergon.2005.10.001`.

[3] Afotey, B., Sattler, M., Mattingly, S., and Chen, V. Statistical model for estimating carbon dioxide emissions from a light-duty gasoline vehicle. *Journal of Environmental Protection*, **04** (2013), 8. `doi:10.4236/jep.2013.48A1002`.

[4] Agency, E. E. Occupancy rates (2016).

[5] Agency, E. E. Do lower speed limits on motorways reduce fuel consumption and pollutant emissions? (2019).

[6] Agency, E. E. Laying the foundations for greener transport (2019).

[7] Ahmed Ali, M., Fuming, P., Younus, H., Abdelkareem, M., Essa, F., Elagouz, A., and Xianjun, H. Fuel economy in gasoline engines using al2o3/tio2 nanomaterials as nanolubricant additives. *Applied Energy*, **211** (2018), 461. `doi:10.1016/j.apenergy.2017.11.013`.

[8] Andry Rakotonirain, G. S.-P. P. D., Narelle Haworth. Research issues in eco-driving (2011).

[9] Araújo, R., Igreja, , de Castro, R., and Araújo, R. E. Driving coach: A smartphone application to evaluate driving efficient patterns. In *2012 IEEE Intelligent Vehicles Symposium*, pp. 1005–1010 (2012). `doi:10.1109/IVS.2012.6232304`.

[10] Aris S. Lalos, K. M., Stavros Nousias. Gamecar: Gamifying self-management of eco-driving. ercim news (2018).

[11] Arnault, N. and Monsallier, G. Diesel fuel filter designs for cold weather. *SAE Technical Papers*, **2014** (2014). `doi:10.4271/2014-01-2711`.

[12] Barla, P., Gilbert-Gonthier, M., López Castro, M., and Miranda-Moreno, L. Eco-driving training and fuel consumption: Impact, heterogeneity

and sustainability. *Energy Economics*, **62** (2017). doi:10.1016/j.eneco.2016.12.018.

[13] Birrell, S., Taylor, J., McGordon, A., Son, J., and Jennings, P. Analysis of three independent real-world driving studies: A data driven and expert analysis approach to determining parameters affecting fuel economy. *Transportation Research Part D: Transport and Environment*, **33** (2014), 74–86. doi:10.1016/j.trd.2014.08.021.

[14] Blagojevic, I., Vorotović, G., Stamenković, D., Petrovic, N., and Rakićević, B. The effects of gear shift indicator usage on fuel efficiency of a motor vehicle. *Thermal Science*, **21** (2017), 707. doi:10.2298/TSCI160806233B.

[15] Boriboonsomsin, K., Barth, M., and Vu, A. Evaluation of driving behavior and attitude towards eco-driving: A southern california case study (2011).

[16] Boriboonsomsin, K., Barth, M. J., Zhu, W., and Vu, A. Eco-routing navigation system based on multisource historical and real-time traffic information. *IEEE Transactions on Intelligent Transportation Systems*, **13** (2012), 1694. doi:10.1109/TITS.2012.2204051.

[17] Boriboonsomsin, V. A. . B. M., K. Eco-driving: Pilot evaluation of driving behavior changes among u.s. drivers. (2010).

[18] Brownlee, J. *Machine Learning Mastery with Python*. 1st edn. (2014).

[19] Cai, J.-F., Candès, E., and Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, **20** (2010), 1956. doi:10.1137/080738970.

[20] Carrico, A. R., Padgett, P., Vandenbergh, M. P., Gilligan, J., and Wallston, K. A. Costly myths: An analysis of idling beliefs and behavior in personal motor vehicles. *Energy Policy*, **37** (2009), 2881 . Available from: http://www.sciencedirect.com/science/article/pii/S0301421509001633, doi:https://doi.org/10.1016/j.enpol.2009.03.031.

[21] chen chen, Y. Y. Y. Z. J. R., Xiaohua Zhao and Liu, X. Driver's eco-driving behavior evaluation modeling based on driving events (2018).

[22] Cindie Andrieu, G. S. Comparing effects of eco-driving training and simple advices on driving behavior (2012).

[23] Commission, E. A policy framework for climate and energy in the period from 2020 to 2030 (2014).

[24] Cutler, A., Cutler, D., and Stevens, J. *Random Forests*, vol. 45, pp. 157–176 (2011). doi:10.1007/978-1-4419-9326-7_5.

[25] D'Amato, A., Donatantonio, F., Arsie, I., and Pianese, C. Development of a cruise controller based on current road load information with integrated

control of variable velocity set-point and gear shifting (2017). `doi:10.4271/2017-01-0089`.

[26] Danijela Baric, M. P., Goran Zovak. Effects of eco-drive education on the reduction of fuel consumption and co2 emissions (2013).

[27] ECOWILL. Ecodriving short-duration training for licensed drivers and integration into driving education for learner drivers; experiences and results from the ecowill project (2013).

[28] El-Shawarby, I., Ahn, K., and Rakha, H. Comparative field evaluation of vehicle cruise speed and acceleration level impacts on hot stabilized emissions. *Transportation Research Part D: Transport and Environment*, **10** (2005), 13. `doi:10.1016/j.trd.2004.09.002`.

[29] Frank, R., Castignani, G., Schmitz, R., and Engel, T. A novel eco-driving application to reduce energy consumption of electric vehicles. In *2013 International Conference on Connected Vehicles and Expo (ICCVE)*, pp. 283–288 (2013). `doi:10.1109/ICCVE.2013.6799807`.

[30] Gardelis, K., Lalos, A., and Moustakas, K. Development of an eco-driving simulation training system with natural and haptic interaction in virtual reality environments (2018). `doi:10.5220/0006619500940101`.

[31] He, X. S., Liu, H., and Liu, X. Optimal vehciel speed trajectory on a signalized arterial with consideration of queue. *Transportation Research Part C Emerging Technologies*, **61** (2015), 106. `doi:10.1016/j.trc.2015.11.001`.

[32] Huang, Y. Eco-driving technology for logistics transport fleet to reduce fuel consumption and emissions. (2018), 27.

[33] Jeffreys, I., Graves, G., and Roth, M. Evaluation of eco-driving training for vehicle fuel use and emission reduction: A case study in australia. *Transportation Research Part D: Transport and Environment*, (2016). `doi:10.1016/j.trd.2015.12.017`.

[34] Keemei, L. Effect of biodiesel fuels on diesel engine emissions keemei.l (ma13144). (2016). `doi:10.13140/RG.2.1.4568.4082`.

[35] Kubicka, M., Klusacek, J., Sciarretta, A., Arben, C., Mounier, H., Thibault, L., and Niculescu, S. Performance of current eco-routing methods. pp. 472–477 (2016). `doi:10.1109/IVS.2016.7535428`.

[36] Kumar, P., Singh, A., Sharma, N., and Chalumuri, R. S. Evaluation of idling fuel consumption of vehicles across different cities (2015).

[37] M. de Goede, R. V. S. B., J. Hogema. Effects of a gear-shift indicator and a fuel economy meter on fuel consumption (2010).

[38] Maamria, D., Gillet, K., Colin, G., Chamaillard, Y., and Nouillant, C. Which methodology is more appropriate to solve eco-driving optimal control

problem for conventional vehicles? pp. 1262–1267 (2016). doi:10.1109/CCA.2016.7587980.

[39] Mansfield, L., Guros, F., Truxillo, D., and MacArthur, J. Individual and contextual variables enhance transfer for a workplace eco-driving intervention. *Transportation Research Part F: Traffic Psychology and Behaviour*, **37** (2016), 138. doi:10.1016/j.trf.2015.12.001.

[40] Marsland, S. *Machine Learning: An Algorithmic Perspective, Second Edition*. Chapman & Hall/CRC, 2nd edn. (2014). ISBN 1466583282, 9781466583283.

[41] Michelle, I. The effects of driving style and vehicle performance on the real-world fuel consumption of u.s. light-duty vehicles. (2010).

[42] Murphy, K. P. *Machine Learning: A Probabilistic Perspective*. The MIT Press (2012). ISBN 0262018020, 9780262018029.

[43] Muslim, N., Keyvanfar, A., Shafaghat, A., Mu'azu, M., and Khorami, M. Green driver: Travel behaviors revisited on fuel saving and less emission. *Sustainability*, **10** (2018), 325. doi:10.3390/su10020325.

[44] Neves, A. and Brand, C. Assessing the potential for carbon emissions savings from replacing short car trips with walking and cycling using a mixed gps-travel diary approach. *Transportation Research Part A: Policy and Practice*, **123** (2019), 130 . Walking and Cycling for better Transport, Health and the Environment. Available from: http://www.sciencedirect.com/science/article/pii/S0965856417316117, doi:https://doi.org/10.1016/j.tra.2018.08.022.

[45] Nousias, S., Tselios, C., Bitzas, D., Amaxilatis, D., Montesa, J., Lalos, A., Moustakas, K., and Chatzigiannakis, I. Exploiting gamification to improve eco-driving behaviour: The gamecar approach. *Electronic Notes in Theoretical Computer Science*, **343** (2019), 103. doi:10.1016/j.entcs.2019.04.013.

[46] Nousias, S., Tselios, C., Bitzas, D., Lalos, A., Moustakas, K., and Chatzigiannakis, I. Uncertainty management for wearable iot wristband sensors using laplacian-based matrix completion (2018). doi:10.1109/CAMAD.2018.8515001.

[47] Nousias, S., et al. Managing nonuniformities and uncertainties in vehicle-oriented sensor data over next generation networks (2018). doi:10.1109/PERCOMW.2018.8480342.

[48] of Energy, U. D. Idling reduction for personal vehicles (2015).

[49] Pham, N., Ganti, R., Nangia, S., Pongthawornkamol, T., Abdelzaher, T., Heo, J., Khan, M., and Ahmadi, H. Fueoogle: A participatory sensing fuel-efficient maps application. (2009).

[50] Schall, D., Wolf, M., and Mohnen, A. Do effects of theoretical training and rewards for energy-efficient behavior persist over time and interact? a natural field experiment on eco-driving in a company fleet:. *Energy Policy*, **97** (2016), 291. doi:10.1016/j.enpol.2016.07.008.

[51] Sinisa Husnjak, T. B., Ivan Forenbacher. Evaluation of eco-driving using smart mobile devices (2015).

[52] Smit, R., Rose, G., and Symmons, M. Assessing the impacts of ecodriving on fuel consumption and emissions for the australian situation (2010).

[53] Strömberg, H. and Karlsson, M. Comparative effects of eco-driving initiatives aimed at urban bus drivers – results from a field trial. *Transportation Research Part D: Transport and Environment*, **22** (2013), 28–33. doi:10.1016/j.trd.2013.02.011.

[54] UNECE. Environmentally friendly vehicles and the world forum for the harmonization of vehicle regulations. (2012).

[55] University of Patras, B. M. S. F. p. p. d. l. i. i. y. d. t. e. l. i. d. a. d. G. S. W. L. F. I. o. S., University of Leeds, Technology for Transport, D., and Networks, K. S. s. Gameecar (2018).

[56] Vlachos, E., Lalos, A., Berberidis, K., and Tselios, C. Autonomous driving in 5g: Mitigating interference in ofdm-based vehicular communications. pp. 1–6 (2017). doi:10.1109/CAMAD.2017.8031619.

[57] Wang, A., Ge, Y., Tan, J., Fu, M., Shah, A., Ding, Y., Zhao, H., and Liang, B. On-road pollutant emission and fuel consumption characteristics of buses in beijing. *Journal of environmental sciences (China)*, **23** (2011), 419. doi:10.1016/S1001-0742(10)60426-3.

[58] Wang, M., Daamen, W., Hoogendoorn, S., and Arem, B. Potential impacts of ecological adaptive cruise control systems on traffic and environment. *IET Intelligent Transport Systems*, **8** (2014), 77. doi:10.1049/iet-its.2012.0069.

[59] Wang, Y. and Boggio-Marzet, A. Evaluation of eco-driving training for fuel efficiency and emissions reduction according to road type (2018).

[60] Yuhan Huanga, J. L. N. C. E. F. C., Elvin C.Y.Ng and Hong, G. Eco-driving technology for sustainable road transport: A review (2018).

[61] Zacharof, N., Fontaras, G., Ciuffo, B., Tsiakmakis, S., Anagnostopoulos, K., Marotta, A., and Pavlovic, J. Review of in use factors affecting the fuel consumption and co2 emissions of passenger cars (2016). doi:10.2790/140640.

[62] Zhang, Y., Li, J., and Zhang, Z. Analysis on car aerodynamic drag for open windows. **31** (2010), 651. doi:10.3969/j.issn.1671-7775.2010.06.007.