

**AUTONOMOUS CARS: AN ADAPTABLE FEEDBACK MECHANISM  
FOR CUSTOMISED ETHICS**

by

Srishti Jaiswal

Submitted in partial fulfilment of the requirements  
for the degree of Master of Computer Science

at

Dalhousie University  
Halifax, Nova Scotia  
August 2017

© Copyright by Srishti Jaiswal, 2017

To my Mom,

Who always picked me up on time and encouraged me to go on every adventure

Especially this one.

To my friends who supported me when I lost hope that I couldn't do this

For believing in me and my work

To my supervisor for never losing his cool on me even when I did not perform and

always encouraging me and helping me in everything possible

And

Lastly

To caffeine and sugar, my companions through many a long night of writing

# Table of Contents

LIST OF FIGURES.....	v
LIST OF TABLES.....	vii
ABSTRACT.....	viii
LIST OF ABBREVIATIONS USED.....	ix
<b>CHAPTER 1 INTRODUCTION.....</b>	<b>1</b>
1.1 Brief introduction of terms and concepts.....	1
1.1.1 Internet of Things(IoT) and Smart Cars .....	1
1.1.2 Smart Cars and Autonomous Cars-The Difference .....	2
1.1.3 A Brief History of Intelligent Vehicles.....	3
1.1.4 Key Benefits of the introduction of fully autonomous cars.....	4
1.1.5 Technologies of fully autonomous cars .....	9
1.2 Brief introduction of the proposed approach .....	13
1.2 Outline of the thesis .....	15
<b>CHAPTER 2 BACKGROUND.....</b>	<b>17</b>
2.1 Why ethics matters for autonomous cars.....	17
2.2 The Trolley Problem.....	19
2.3 Responsibility for crashes of autonomous vehicles .....	21
2.3.1 Responsibility of the Manufacturer .....	22
2.3.2 A Duty to Intervene .....	22
2.3.3 Responsibility of the Driver as a Form of a “Strict Liability” .....	23
2.4 Dissimilarities between the trolley problem and ethics for AVs .....	24
<b>CHAPTER 3 RELATED WORK .....</b>	<b>26</b>
3.1 Literature survey on smart cars and VANETs.....	26
3.1.1 Evolution of the Autonomous Cars .....	28
3.1.2 Considerations for the introduction of fully autonomous vehicles...30	30
3.2 Literature survey on ethical and social dilemma of autonomous vehicles .....	36
3.2.1 Machine Ethics(Robot Ethics) and autonomous vehicles.....	37
3.2.2 The ethics of autonomous cars.....	43
3.2.3 How is it different from the trolley problem?.....	44
3.2.4 The social dilemma of autonomous vehicles .....	46
3.3 Literature survey on some existing autonomous vehicles .....	48

3.4	Motivation and Research Objectives .....	50
<b>CHAPTER 4</b>	<b>PROPOSED APPROACH AND METHODOLOGY</b> .....	52
4.1	Proposed approach.....	52
4.2	Detailed explanation of the proposed approach.....	53
4.2.1	The Dataset .....	53
4.2.2	The Classification algorithms .....	57
4.2.3	The Priority Generation Phase .....	61
<b>CHAPTER 5</b>	<b>IMPLEMENTATION</b> .....	68
5.1	Development Environment and Libraries used.....	68
5.2	Implementation details of each phase in the proposed approach .....	68
5.2.1	Generating user defined priority list/ethics.....	68
5.2.2	Classification of the detected objects.....	71
5.2.3	Decision making process based on the user defined priorities .....	74
<b>CHAPTER 6</b>	<b>EXPERIMENTAL RESULTS AND ANALYSIS</b> .....	76
6.1	Tools used.....	76
6.1.1	Weka .....	76
6.1.2	RStudio .....	76
6.2	Results of experimentation phase using Weka .....	77
6.2.1	Naïve Bayes' algorithm application: Results and Discussion .....	78
6.2.2	C4.5 algorithm application: Results and Discussion .....	82
6.2.3	C5.0 algorithm application: Results and Discussion .....	87
6.2.4	Random Forest algorithm application: Results and Discussion .....	90
6.3	Summary of the experimental results .....	94
<b>CHAPTER 7</b>	<b>CONCLUSION</b> .....	97
7.1	Limitations .....	99
7.2	Discussion and Future work .....	100
<b>REFERENCES</b>	.....	101

## LIST OF FIGURES

Figure 1: Total accidents due to human error or choice [10].....	7
Figure 2: US fatalities associated with human choice or error (2014) [10].....	7
Figure 3: Journey of Waymo in 2012 [11].....	9
Figure 4: The Google Self Driving Car aka the Firefly [10,11] .....	10
Figure 5: Self driving Chrysler Pacifica Hybrid Minivan [11].....	11
Figure 6: Various sensors in an autonomous car [12].....	12
Figure 7: Various components of an AV and need for further development.....	13
Figure 8: The trolley and the fat man problem [17,18].....	20
Figure 9: Block Diagram of the DAS [1].....	28
Figure 10: Economic Considerations of introduction of autonomous technology .....	32
Figure 11: Ethical Dilemma [9] .....	47
Figure 12: Classifying an unlabeled vertebrate. [48].....	59
Figure 13: An example of question with two scenarios.....	62
Figure 14: The selection sorting algorithm [55] .....	63
Figure 15: The flowchart of question generation phase.....	64
Figure 16: The flowchart of the decision making phase.....	67
Figure 17: Reading the csv file and adding each row to an arraylist.....	69
Figure 18: ArrayList declaration.....	69
Figure 19: Capturing priority based on user's choice .....	70
Figure 20: Default priority list of the AV .....	70
Figure 21: User defined priority list generated .....	70
Figure 22: Code snippet for building the classifier.....	72
Figure 23: Classifying into a category .....	72
Figure 24: Code for identification of category and category name for detected objects..	73
Figure 25: Output showing category and category name for detected object.....	74
Figure 26: Code snippet showing how object with least priority is identified .....	75
Figure 27: The final outcome.....	75
Figure 28: Screenshot of part of the dataset.....	77
Figure 29: Accuracy of Naïve Bayes' on 60-40 dataset .....	78

Figure 30: Confusion matrix for Naïve Bayes' on 60-40 dataset .....	78
Figure 31: Accuracy of Naïve Bayes' on 70-30 dataset .....	79
Figure 32: Confusion Matrix for Naïve Bayes' for 70-30 dataset.....	80
Figure 33: Accuracy of Naïve Bayes' on 80-20 dataset .....	80
Figure 34: Confusion Matrix for Naïve Bayes' on 80-20 dataset.....	81
Figure 35: Graph showing the accuracy for different dataset ratio for Naïve Bayes' .....	81
Figure 36: Visualization of a pruned tree for 60-40 dataset .....	82
Figure 37: Accuracy of C4.5 decision tree algorithm on 60-40 dataset .....	83
Figure 38: Confusion matrix for C4.5 on 60-40 dataset .....	83
Figure 39: Accuracy of C4.5 for 70-30 dataset.....	84
Figure 40: Confusion matrix for C4.5 algorithm on 70-30 dataset.....	84
Figure 41: Accuracy of C4.5 decision tree algorithm for 80-20 dataset.....	85
Figure 42: Confusion matrix for C4.5 decision tree algorithm on 80-20 dataset .....	85
Figure 43: Graph showing the accuracy for different dataset ratio for C4.5 .....	86
Figure 44: Accuracy of C5.0 decision tree algorithm for 60-40 dataset.....	87
Figure 45: Confusion matrix for C5.0 decision tree algorithm on 60-40 dataset .....	87
Figure 46: Accuracy of C5.0 decision tree algorithm for 70-30 dataset.....	88
Figure 47: Confusion matrix for C5.0 decision tree algorithm on 70-30 dataset .....	88
Figure 48: Accuracy of C5.0 decision tree algorithm for 80-20 dataset.....	88
Figure 49: Confusion matrix for C5.0 decision tree algorithm on 80-20 dataset .....	89
Figure 50: Accuracy of C5.0 Decision tree algorithm on different dataset ratios .....	89
Figure 51: Accuracy of Random Forest Algorithm on 60-40 dataset.....	90
Figure 52: Confusion matrix for random forest classification on 60-40 dataset .....	91
Figure 53: Accuracy of Random Forest Algorithm on 70-30 dataset ratio .....	91
Figure 54: Confusion matrix for Random Forest algorithm on 70-30 dataset ratio .....	92
Figure 55: Accuracy of Random Forest algorithm on 80-20 dataset.....	92
Figure 56: Confusion matrix for Random Forest algorithm on 80-20 dataset.....	93
Figure 57: Accuracy of Random Forest algorithm on different dataset ratios. ....	93

## **LIST OF TABLES**

Table 1. NHTSA's Five-Part Continuum of Vehicle Control Automation [7].....	4
Table 2: Dissimilarities between trolley problem and accident algorithms for AVs [21]	25
Table 3: Level of Automation by SAE International, J3016 [28].....	30
Table 4: Some Statistics related to road accidents in 2012 [7].....	31
Table 5: Existing and future autonomous vehicles .....	50
Table 6: An example of the dataset.....	61
Table 7: Development environment and libraries used .....	68

## ABSTRACT

Autonomous vehicles (AVs) are the next breakthrough in the automobile industry and will pave the way to the public market in the near future. Apart from traffic efficiency, increased mobility and a hope for sustainable future with reduced pollution and consumption of fuels, they also promise to bring down the number of accidents up to 90 percent owing to the fact that the majority of accidents in the entire world happens because of human errors. However, this technology would not only make travelling safer but would also raise several ethical concerns regarding decision making in crash scenarios such as what would the autonomous vehicle do in a kill or be-killed scenario and the responsibility of the stakeholders in such crashes. This research work proposes to provide solutions for some of the ethical issues raised by the introduction of autonomous vehicles. It proposes an idea to capture the ethics of the consumers by generating a priority list which would be encapsulated by answering questions that provides user to choose between multiple scenarios. This research work also solves the issue of assigning the responsibility of the crashes by generating the user ethics/priority list and proposes the idea of a default ethics/priority list which would be pre-programmed into the new AV and would be followed if the consumer doesn't want his own ethics to be followed by the car. Based on ethics that is followed whether user-defined or default, the responsibility of crashes would be assigned. Now, when the user defined priority list is ready to use, then in an accident scenario the sensors of the AV will detect the objects encountered. Next, the detected object will be classified into a category using a machine learning classifier and the priority of the objects will be determined from the priority list and the corresponding action will be taken. We tested the accuracy with which the objects with certain features are classified into a particular category in the experimentation phase of our proposed work using classifiers namely naïve Bayes, decision tree C4.5 and C5.0 and random forest. We used Weka and Rstudio tools for the experimentation phase. We then have chosen the algorithm that best suited our requirements and implemented the generation of the user defined priority list and the results of the implementation have shown that the action taken is in accordance with the ethics of the user.

## **LIST OF ABBREVIATIONS USED**

<b>Abbreviation</b>	<b>Acronyms for</b>
<b>AV</b>	Autonomous Vehicle
<b>NHTSA</b>	National Highway Traffic Safety Administration
<b>OEM</b>	Original Equipment Manufacturer
<b>IoT</b>	Internet of Things
<b>ITS</b>	Intelligent Transportation System
<b>DAS</b>	Driver Assistance System
<b>TOCADAS</b>	Tolerant Context Aware Driver Assistance System
<b>VANET</b>	Vehicular Ad-hoc Network
<b>MANET</b>	Mobile Ad-hoc Network
<b>LIDAR</b>	Light Detection and Ranging
<b>RADAR</b>	Radio Detection and Ranging
<b>LKA</b>	Lane Keeping Assistance
<b>ACC</b>	Adaptive Cruise Control
<b>SAE</b>	Society of automotive engineers
<b>EDR</b>	Event Data Recorder
<b>ECU</b>	Electronic Control Unit
<b>AMA</b>	Artificial Moral Agent

# **CHAPTER 1      INTRODUCTION**

*"The weak point of the modern car is the squidgy organic bit behind the wheel." Jeremy Clarkson*

We commence the thesis with a concise description of the Internet of Things that has taken the world by storm followed by a crisp definition of what makes a Smart car an autonomous car. Then we provide a terse overview to all the existing research and various other terms that are related to the area followed by the motivation for the proposed approach.

## **1.1 Brief introduction of terms and concepts**

### **1.1.1 Internet of Things (IoT) and Smart Cars**

The Internet of Things has been into existence in the technological world for quite some time now and the world seems to revel at every stage of introduction of new technologies-the latest being the wearable technology.

These wearable technology is at the forefront of the IoT and is overwhelming the health and wellness industry. It can be really very beneficial to people aiming towards a fitter and healthier life by keeping track of all of their physical activities like the number of steps they have taken, their heart rates etc. Also the wearables like fitbit and apple watches tell you to take a moment to breathe few times in a day. They then collect and put all these data points in one place so it could be analyzed by the user to keep a track of his health.

Smart Cars are the next big thing in today's world and it is changing the lives of millions around the world. There is a plethora of information and data stored in the car's onboard computer which can improve the everyday lives of the people. And it is not just getting our Facebook updates while driving but also would contribute to a whole lot better driving experience by providing weather, road and traffic conditions. It would send warning messages to the user in case of accidents that have happened nearby or on the route taken by the user.

However, we will discuss all these in detail in the upcoming sections.

### 1.1.2 Smart Cars and Autonomous Cars-The Difference

In this section, we are going to throw some light on what is exactly a smart car and what makes a smart car an autonomous car. We then highlight the differences between the two followed by the features and technologies of the autonomous car. Throughout this thesis, we have used the terms autonomous cars, driverless cars, self-driving or self-driven cars interchangeably.

Cars are becoming an inseparable part of the human lives because it has now become a major means of commute all around the world. However, as with all approaches and technologies; together with the positives comes the negatives and in this case it is basically traffic congestion, accidents and pollution etc.

“A smart car aims at assisting its driver with easier driving, lesser workload and fewer chances of getting injured” [1]. And in order to provide such facilities, a smart car should be able to do the following: a) sense and analyze the environment around itself and based on that b) it should be able to predict and react to the situations and road environment. So, if a car has the above two features then it is classified into the category of smart cars.

#### **Now, what makes the smart car a driverless car?**

A driverless car has various other names like those of a fully autonomous car, robotic car, self-driven car or self-driving car (analogous to a car with cruise control). So in reality a driverless car is a car having the capabilities of a smart car like those of sensing and analyzing the environment and taking actions based on predictions. However, the turning point or a head turner feature is that, it does all of the above without the human input and hence, the name driverless because it doesn’t need any driver.

“The autonomous vehicle technology often referred to as the new technology by Thurston [2], has the potential to reduce crashes, ease congestion, improve fuel economy, reduce parking needs, bring mobility to those unable to drive and over time dramatically change the nature of US travel” [3,4]

### 1.1.3 A Brief History of Intelligent Vehicles

DARPA (Defence Advanced Research Projects Agency) has had a single mission for more than last 50 years and that is “to make pivotal investments in breakthrough technologies for national security.” [5]. DARPA funds the DARPA Grand Challenge which is a competition for the autonomous vehicles in America.

So, all the vehicles that won the DARPA grand challenge in 2004, 2005 and 2007 were all based on sensor technology which was driven by microprocessor. This became a standard for future advancements in the field of automated robotic cars. [6]

Now, since the development of smart cars to fully autonomous cars was not abrupt and was more of a gradual process so NHTSA (National Highway Traffic Safety Administration) has defined five different levels of automation in an autonomous car ranging from 0 (where the sole controller of the entire vehicle at all times is a driver) to level 4 (where there is absolutely no human intervention except perhaps for entering the destination and some inputs for navigation).

The following table defines the levels of automation as defined by the NHTSA.

<b>Level 0: No Automation.</b> The driver is in complete and sole control of the primary vehicle controls—brake, steering, throttle, and motive power—at all times.
<b>Level 1: Function Specific Automation.</b> Automation at this level involves one or more specific control functions. Examples include electronic stability control or pre-charged brakes, where the vehicle automatically assists with braking to enable the driver to regain control of the vehicle or stop faster than by acting alone.
<b>Level 2: Combined Function Automation.</b> This level involves automation of at least two primary control functions designed to work in unison to relieve the driver of control of those functions. An example of combined functions enabling a level 2 system is adaptive cruise control in combination with lane centering.
<b>Level 3: Limited Self-Driving Automation.</b> Vehicles at this level of automation enable the driver to cede full control of all safety- critical functions under certain traffic or environmental conditions and in those conditions to rely heavily on the vehicle to monitor for changes in those conditions requiring transition back to driver control. The driver is expected to be available for

occasional control, but with sufficiently comfortable transition time. The Google car is an example of limited self-driving automation.

**Level 4: Full Self-Driving Automation.** The vehicle is designed to perform all safety-critical driving functions and monitor roadway conditions for an entire trip. Such a design anticipates that the driver will provide destination or navigation input, but is not expected to be available for control at any time during the trip. This includes both occupied and unoccupied vehicles.

Table 1. NHTSA's Five-Part Continuum of Vehicle Control Automation [7]

#### 1.1.4 Key Benefits of the introduction of autonomous cars

In this section, we will discuss why the introduction of the autonomous car is going to take the world by storm and how it will change the society for good, like reduce the pollution, reduce the stress on fossil fuels etc. We will also discuss various other advantages like parking assistance, better navigation system and shorter route selection to destinations. We will throw some light on each of the benefits and discuss them concisely.

##### a. Towards car sharing

From an economic point of view, owning a car which is very rarely used perhaps because people prefer to take a public transit than their own cars. So their cars are kind of underutilized assets contributing to wastage of parking space. However, this scenario will change with the introduction of driverless cars because it can easily drive a person/passenger to a destination without waiting on for him. Also, it can attend to some other passengers while the first passenger calls back again.

And, eventually the cost of travelling in such autonomous rental cars would be lower than the total cost of owning a car which would tempt people to use such services more often. It will thereby conserve the resources and provide huge economic benefits to the society.

b. Increased utilization rate

Now, when people start sharing the vehicle; then it is going to increase the utilization rate (which is the average number of hours it drives passengers around). “Shared vehicles should be able to achieve an increase in utilization per vehicle by a factor of at least 5, maybe even 10.” [8]. Now, this will drastically reduce the capital costs per kilometer and thus rental vehicle would be much less expensive than privately owned cars. And thus, it will provide the car rental companies to manage a large fleet of cars in lieu of a couple of them. This will eventually reduce the costs of maintenance and increase the life of a vehicle.

c. Vehicle differentiation

In today’s world, we use our cars for all kind of scenarios: from going to office which is 15 mins away to going on a family vacation. We also use our cars for small shopping trips alone or for long distance travel. Hence, our cars are general purpose tools but such privately owned cars do not have capability for differentiations. For example: we are constricted by the space in the car, speed range, energy resource, range etc.

On the other hand, when cars are shared we can have cars that cater to our requirements and scenario. Like for a short trip to the office we do not need an SUV with a 500km range. The car rental companies can easily optimize their fleet of cars for different scenarios, like for short range trips they can have small sedan or hatchbacks, for family vacations they can have vans and can have limousines for prom parties or to flaunt.

By using small, light-weight vehicles for short range trips we can not only reduce the burden on fossil fuels and natural resources but also ameliorate the fuel efficiency.

People are generally hesitant when it comes to buying an electric car owing to the fact that its range is limited. However, the car rentals can easily have a large fleet of electric vehicles for short range trips within the city. And thus the introduction of autonomous cars is going to help towards sustainable development by the use of alternative forms

of energy and also contribute towards greener technology as it will reduce the emissions and thus not cause pollution.

d. Better public transport

The fully autonomous vehicles are a perfect fit for picking/dropping people from/to public transport systems. It can easily be coordinated with the schedules of the public transportation system. For example: Instead of renting a vehicle for a long distance trip from Halifax to Moncton, the rental autonomous vehicle can pick people up from their respective sources and then drop them off to the public transit station. The service provider can easily develop an information technology infrastructure in order to coordinate with the schedules of the public transit and the people would not even have to wait for the autonomous bus in a non-heated stop, instead they can continue to sit in the cars till the bus arrives. And a similar approach can be followed when they reach Moncton and can get a car right outside the stop which would deliver them to their respective destinations. Thus autonomous vehicles can easily obscure the distinctions between private and public transports.

e. Increased road utilization

The rental autonomous cars would not lead to congestion because they are connected to each other and can disseminate information about traffic conditions to smart street lights, which in turn can send the messages to cars at a particular intersection. Unlike the case of human drivers, who in order to avoid getting caught up in traffic will switch lanes and in the process will cause more congestion. Thus, autonomous vehicle will increase the road utilization like all the vehicles at a red light will start to move as soon as the light is green. In contrast to the human drivers, who move their cars only when the car ahead has moved a significant distance.

Also, it is predicted that in the coming time the traffic would be so huge that we would be needing better road infrastructure, however this infrastructure would not even be needed if the autonomous vehicles take over and thus billions of dollars of investment can be reduced all around the world. [8]

## f. Safety

“To err is human” is very famous saying and is so true because 94% of the accidents that are caused is because of the human error. [9,10]

To quote the facts, “1.25 million deaths worldwide were due to vehicle crashes in 2014 out of which 32,675 deaths were caused in United States due to vehicle crashes. There was 6% increase in traffic fatalities in 2016, reaching a highest point in nearly a decade.” [10] And all this, just because of the human error as quoted above that 94% of these accidents were caused by human error.

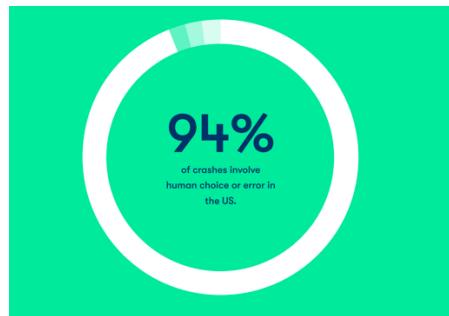


Figure 1: Total accidents due to human error or choice [10]

The reason behind the human error has been very beautifully depicted by the following Figure 2 by Waymo also known as google self-driving car website. [10]

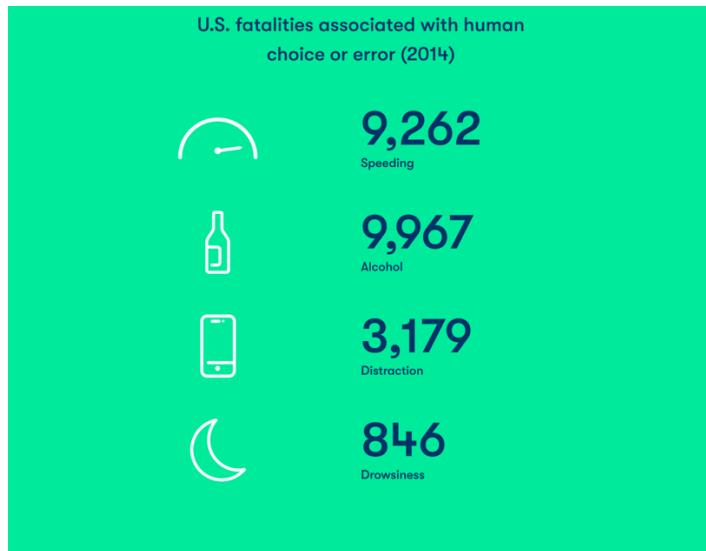


Figure 2: US fatalities associated with human choice or error (2014) [10]

With the introduction of autonomous cars the human factor can completely be removed because technology is not prone to certain shortcoming that we humans possess. For example: autonomous cars will never be drowsy or get distracted by a text message or a Facebook notification, nor would it get intoxicated or get ticketed for speeding because they are programmed to follow the rules. Also, these cars would not be overpowered by emotions and their reaction time would be far less than that of humans.

Although, it might not be able to eliminate accidents to 0% but they are going to reduce the traffic fatalities to a great extent. They can process a large and varied amount of information relating to its current environment and scenario, and take an action based on its predictions. It can also inform the connected cars of its reaction at the same time so as to warn them and also so they can take their own action without any delay. [8]

#### g. Emergency Response

Automated cars could also provide assistance in case of emergencies by switching itself to emergency mode wherein it would drop off the injured to the nearest hospital or to a first aid team at its maximum speed without adhering to the rules of traffic much like how the ambulance in present day works. Also, in case of natural calamities, evacuation procedures can be implemented in a much organized and faster manner.

#### h. Enhanced mobility

The autonomous vehicle will make sure that everyone is getting around safely and easily irrespective of their ability to drive. The visually and physically impaired, and the old would not have to give up their independence and can get around easily. The time spent driving can be spent in doing something else like students can finish up last minute assignments while their car can drive them to their class.

Hence, in these ways the autonomous vehicles are likely to change our lives for good and impact the society in a very green and positive manner leading to a smart and sustainable development and a safer future.

### 1.1.5 Technologies of the autonomous cars

Given the amount of benefits, freedom and independence the introduction of autonomous cars is going to bring, almost all the companies whether automobile manufacturing or not have joined the race of getting their own autonomous car in the market as soon as possible. Among these are the tech giants Google and Apple too. Some others are GM's Cadillac, Audi MG, BMW AG, Ford Motor Co., Volvo, Tesla and Mercedes etc.

Google's driverless car, which is now known as Waymo has been working on the technology since 2009. We are giving a concise chronological sequence of the Google's self-driven car and then move forward to the technologies and sensors used in such cars. So, [11] they began their journey back in 2009 and their main challenge then was to travel autonomously on 10 different 100 miles' routes in Toyota Prius vehicles. And months after they began this new adventure they were able to travel much more than had ever been travelled autonomously.

The story continues in 2012 and has been elicited by the Figure below:

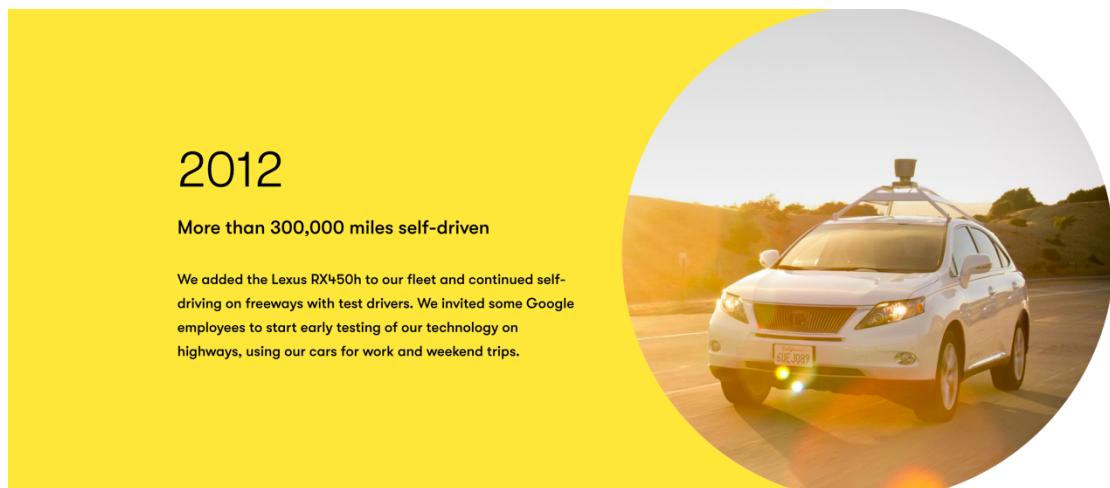


Figure 3: Journey of Waymo in 2012 [11]

Moreover, in 2012 they took the adventure further and started to test out in much complex streets with all sorts of traffic that included pedestrians, cyclist, motorcyclists, road work, construction etc.

Later in 2015, Google imagined how would a fully autonomous vehicle look like and designed a new reference vehicle which they nicknamed as the “Firefly”. This car had it all; sensors that were customized to provide for the autonomous driving, a computer, steering and braking. However, since this was designed basically for the purpose of riding so no steering wheel was provided. Here, we have provided a picture of the Firefly.

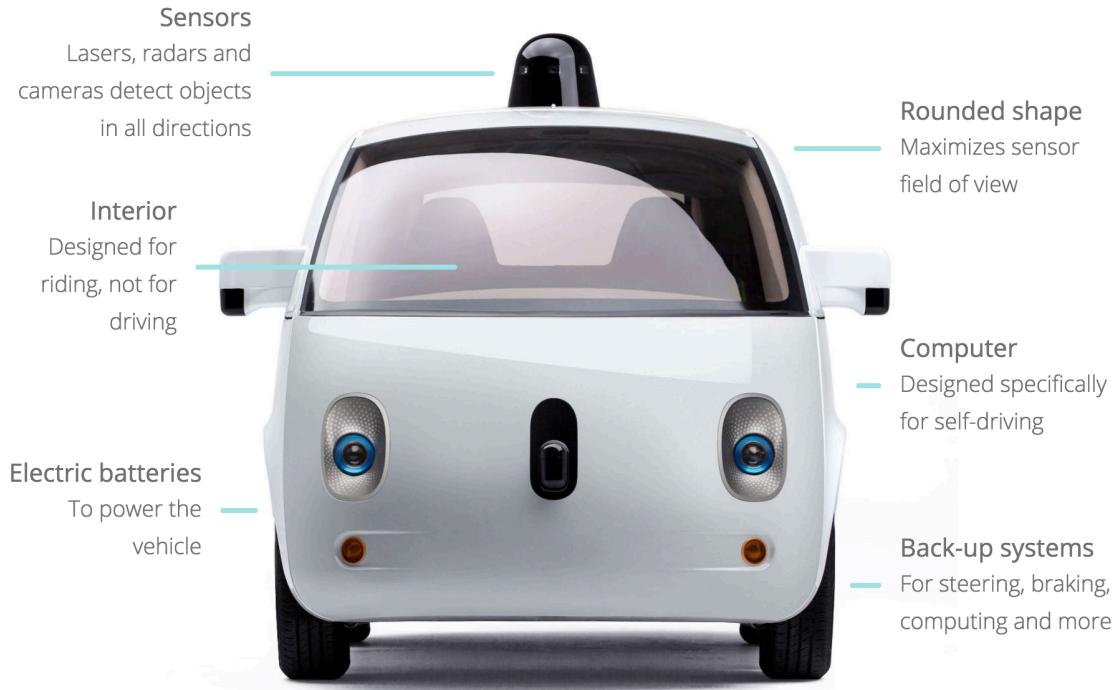


Figure 4: The Google Self Driving Car aka the Firefly [10,11]

Now, later in 2015 the Firefly made its first ride on the streets of Austin, TX. And, this was “world’s first and only self-driving ride on the public road.” [11]

In 2016, the Google’s self-driving car project became Waymo which was a self-driving technology company with its mission to help people and things to move around easily and safely, irrespective of their ability to drive. In 2017, Waymo introduced the fully autonomous Chrysler Pacifica Hybrid minivans and started a riding program which invited people of Phoenix, AZ to join in the public trials of the minivans.



Figure 5: Self driving Chrysler Pacifica Hybrid Minivan [11]

Now, we will discuss the technologies of the self-driving car and to what extent are these technologies have already been developed and how much more amount of work is required in future.

As can be seen from the Figure 4, an Autonomous vehicle(AV) is comprised of both hardware and software components. The hardware components are a wide variety of sensor technologies whose main task is to sense the environment in and around the vehicle and then pass on the sensor information to the software components, which in turn assess and analyze the data and predict the next thing that can happen. And based on the prediction, it takes an action.

This can easily be illustrated with the help of an example provided by Waymo, where the AV is at an intersection and there is a pedestrian and a cyclist. So, now the sensors of AV detect both the objects and sends the data to the software components which predicts that the cyclist will ride by and the pedestrian will cross the road. Based on this prediction, it takes the action by yielding to the pedestrian and nudging away from the cyclist while he/she rides by. [10]

Therefore, when such intelligent technologies hit the commercial market, then they have to be perfect and in order to make them commercially available the vehicle manufacturers/suppliers would be required to invest heavily into the hardware components like those of sensors, processors, software components and IT, and system integration etc. [12]

Although a most of these components are already in place or available commercially however, there are still some components which have to be perfected before they can be made available commercially. The Figure 6 shows the components of the AV and the Figure 7 shows how much more development is required from them before the AV hits the commercial market.

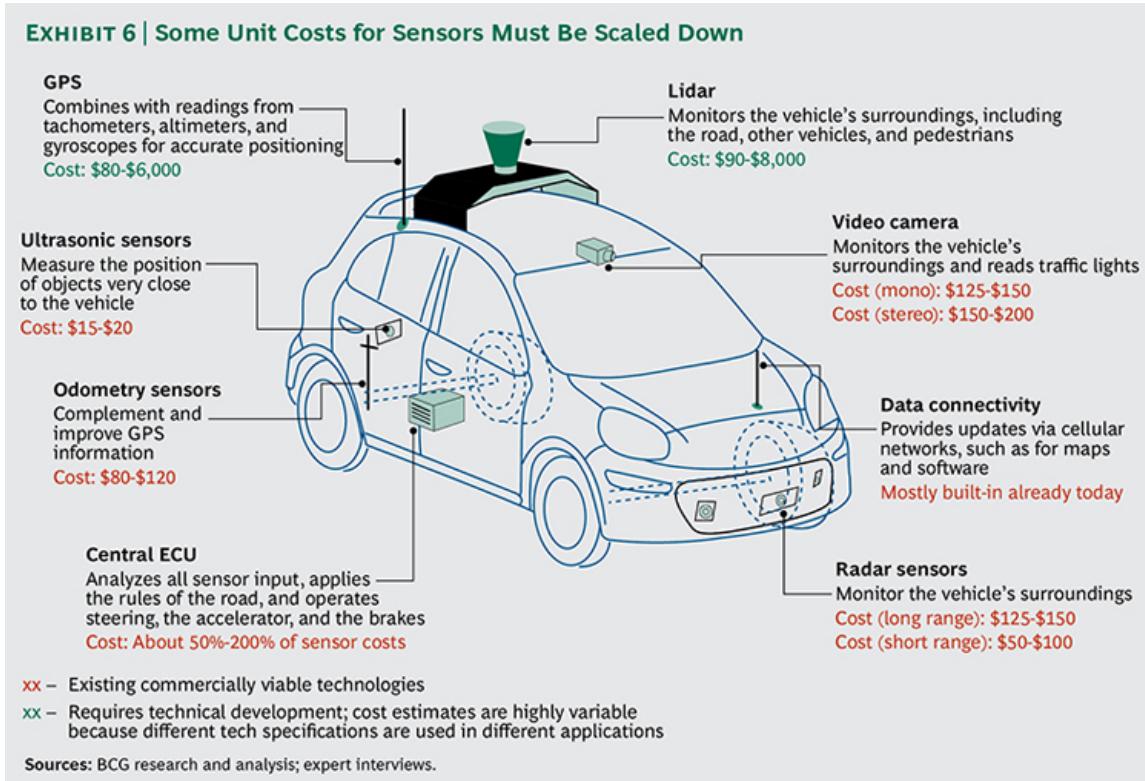


Figure 6: Various sensors in an autonomous car [12]

## EXHIBIT 5 | Several Hardware and Software Components Enable Autonomous Operation

		Need for further development
Hardware	Sensors, such as for radar, cameras, and lidar	<b>Medium-High</b> Improvement in fault rates and specific technologies, such as lidar and GPS
	Processors or ECUs <sup>1</sup>	<b>Medium</b> Processing speed
	Actuators, such as for brakes, gears, and steering	<b>Low</b>
	Internet connectivity	<b>Low</b>
Software and IT	ECU operating system, main operating system, human-machine interface	<b>High</b> Driven by the complexity of the driving environment
	High-definition maps	<b>Medium</b> Accuracy requirements drive effort
Systems Integration	Validation and testing	<b>High</b> Increasingly limited driver role
Assembly	Manufacturing and scrap	<b>Low</b>

Sources: BCG research and analysis; expert interviews.  
<sup>1</sup>ECU = electronic control unit.

Figure 7: Various components of an AV and need for further development

### 1.2 Brief introduction of the proposed approach

This research work mainly aims at introducing the concept of customized ethics in autonomous vehicles which are going to provide answers to a lot of unanswered questions that exists in literature. Some of them are as follows: who would be responsible for crashes where an autonomous vehicle is involved, what kind of ethics must a driverless technology follow in a kill or be killed situation. And finally, it also solves the problem of designing a universal ethics system which is quite an impossible task as ethics are more of a subjective topic and depend on a lot of things including cultures, country, origin etc. Also, in this section we provide a brief introduction to the proposed approach and shall discuss it in detail in chapter 4 (Proposed approach and methodology).

The proposed approach basically has two phases. In the first phase we are collecting information relating to the various features of all objects that might be encountered on a road or traffic scenario. And these objects include human beings of different genders, age, size etc. It also includes wildlife and pet animals and various kind of vehicles. All the data is collected in the form of a dataset and then a classification task is undertaken. Here we use four different kinds of classifiers. We train the classifier and test it on our dataset to check the level of accuracy and other performance metrics that can be achieved. Here, we have chosen Naïve Bayes' classifier, decision tree classifier C4.5 and C5.0 and a random forest classifier.

We have used different ratios of our dataset for training and testing. For example: a 70-30 ratio means 70% of our dataset is used for training and rest 30% for testing purposes. We have tested the classifier on 60-40, 70-30 and 80-20 ratios. We looked at how increasing the training data would affect the accuracy and what contributes to the results obtained. We then, have chosen the classification algorithm that best suited our requirements and used it in our implementation.

The next phase is the phase where we generate questions from the dataset which is translating each row of the dataset in the form of an option/choice to a question and we present a question with multiple options to the user. Now, when the user chooses a particular option then it means that he/she is preferring one option over the other which is key idea of capturing the ethics of the person. Because in real life too when a person faces a situation where he has two or more choices, then based on his ethics or sense of judgement he tries to choose the best. So, similarly the option that he chooses would be saved into memory of the autonomous car and based on the number of questions he has answered over time his ethics or priority listing will be formed. Here, we have assumed that the autonomous vehicle comes with a predefined priority list or ethical policies (default priority list) which are defined by a number of stake holders like ethicist, lawyers, government, citizens, manufacturers etc. We shall discuss these steps in more detail in the methodology section of this thesis.

The ethics of the person will be built overtime as he answers the questions. However, if the person chooses to drive his AV without answering all the questions then he can do so. However, in a case where his autonomous vehicle is faced with a situation where the objects detected by the vehicle are the objects whose priority has been defined by the user, then the AV will follow the ethics of the person and take a corresponding action. In a case, where one or more objects detected do not appear in the priority list of the person, then in that case, the default priority list defined by the Original Equipment Manufacturer (OEM) and other stakeholders will be followed. In such a case, the OEM and other stakeholders shall be responsible for the crash. The user will only be held responsible for crashes that have occurred by following his ethics/priority list.

Now, once the user answers all the questions then we are going to apply the above machine learning classification to this system so that the system would be able to classify an object based on how similar the real life object is to the one defined in the dataset.

This proposed concept is novel and can be used in autonomous technology and would also lead to fast acceptance of the technology as consumers would not shy away from buying a vehicle that would reflect their own ethics. And, they would not be forced to buy a car whose ethics is not transparent to them and/or has been government regulated like the one discussed by Iyad Rehwan et al. [9]. Thus it would lead to world wide acceptance of autonomous technology because not only would this system be able to provide customized and adaptable ethics but also solves the problem of having to design universal ethics that might/might not be acceptable to all.

## **1.2 Outline of the thesis**

The organization of the rest of the thesis is as follows. An overview of the background of the driverless technology and why ethics matter in AVs, its analogy with the trolley problem together with a discussion on who would be responsible in the event of crashes involving AVs. This chapter is closed by drawing out the differences between the trolley problem and the ethics of accident algorithms of the AVs in Chapter 2. Related work (or Chapter 3) discusses the evolution of the autonomous cars from VANETS and its various

levels and it also gives a brief overview of the work of the researchers in the ethics domain of the autonomous technology. And Chapter 4 discusses the proposed methodology in detail which is followed by a thorough explanation of the implementation aspect in Chapter 5. We present the in-depth description of the experiments we conducted and present the evaluated results supporting our proposed approach in Chapter 6. Chapter 7 marks the conclusion of the thesis where we present the limitations and possible enhancements to the proposed approach.

## **CHAPTER 2      BACKGROUND**

In this chapter, we discuss the importance of ethics in the development and future of the autonomous vehicle. Then we highlight one of the classic problems of philosophy called “The Trolley problem” and try and compare it to the problem of ethics that the autonomous vehicles might face in real life scenarios. Next, we discuss the various roadblocks and aspects of them to the intelligent vehicles and autonomous vehicle such as the social and economic aspects of it etc. Also, we provide an in depth ethical analysis of the responsibility for crashes of the autonomous vehicles. Finally, we bring out the differences between the ethics of the accident algorithm of autonomous vehicle and the trolley problem.

So, in this chapter we are providing a background study of the autonomous vehicles together with the motivation behind the proposed work.

### **2.1 Why ethics matters for autonomous cars**

Patrick Lin [13] quoted “*If motor vehicles are to be truly autonomous and be able to operate responsibly on our roads, they will need to replicate-- or do better than--the human decision making process.*” However, some decisions require much more than just mechanical decisions as to follow the traffic rules or predicting a safe path and those decisions require some moral values or some sense of ethics which are actually very difficult to capture or replicate in the form of an algorithm to be understood and followed by a robot/computer.

This ethics problem can be illustrated by a scenario defined by Patrick Lin [13] which is as follows—Imagine yourself in an autonomous car of your own and you are encountered with a terrible decision making and that is- your car must either swerve to the left, and hit a 6-year-old girl or it must swerve to the right and hit an old lady of 85 years. Also, given the velocity of the car and several other factors the victim who is hit is going to die due to the impact. Now, if you do not swerve both the little girl and the old lady would die. So, it is a better option to kill one than both but which one of the option is a lesser evil? Now, he

says that if the reader was programming the self-driving car then how would he have instructed it to behave in such a situation, as rare as it may be?

Now, here comes the tough decision. To some people, saving the little girl is ethical because according to them the old lady has had her share of life and experiences and if she too was given a choice she would have sacrificed herself but, to others the granny too has an equal right to life as the little girl. So, age is not a relevant factor of discrimination here and is almost equivalent to discriminating on the basis of caste, creed, colour, religion or nation origin. One other solution would be, to take decision randomly but that too is comparable to taking a decision of someone's life and death without a thought. So, this is a huge quandary which is not easily solvable and hence implies to the need of ethics in the development of self-driven cars.

There is yet another aspect that could be added to this scenario and that is: if the autonomous car is programmed to protect its own occupants then colliding with the lighter object (in above case the little girl) is the best option. Similarly, if the choice to be made was between two vehicles: a lighter one like that of a mini cooper or a motor cycle or a heavier vehicle like that of a bus or truck, then colliding with lighter vehicle is a better option so as to protect the occupants of the car.

But, suppose the car was programmed to always sacrifice its occupants in order to protect the pedestrians or other drivers-then that is a choice which is ethically and morally better than the prior one of jealousy and protecting oneself over other.

The above examples call for the definition of deontological ethics (non-consequentialist) and the utilitarian ethics (consequentialist). Now, "*A deontological theory of ethics is one which holds that at least some acts are morally obligatory regardless of their consequences for human weal or woe. The popular motto 'Let justice be done though the heavens fall' conveys the spirit that most often underlies deontological ethics*". Robert Olsen [13, 14].

So, the situation where we have to decide between killing a little girl or an old lady, there is nothing defined in books or there are no defined rules as to what is the right decision. So deontological ethics are good in theory but very difficult to comply to in real life. On the other hand, “*Utilitarianism is a normative ethical theory that places the locus of right and wrong solely on the outcomes (consequences) of choosing one action/policy over other actions/policies. As such, it moves beyond the scope of one's own interests and takes into account the interests of others.*” [13,15]. And as it is implied from the definition, utilitarianism strives to reduce the harm and maximize the number of happy lives, that is to sacrifice oneself over the lives of other pedestrians and drivers. We shall discuss more on the ethics in the next section of the trolley problem.

## **2.2 The Trolley Problem**

One of the most classical thought-experiments (Thought-experiments are a way used in philosophy and ethics in order to simplify the issues by means of hypothetical scenarios) [13,16] in ethics is the trolley problem. [13,17,18] And this thought-experiment would come to life if the autonomous cars are to be a reality in a near future. The Trolley problem was discussed by Phillipa Foot and Judith Jarvis Thompson [13,17,18] and is defined below and is also illustrated with the help of Figure 8:

“*Imagine a runaway trolley or a train is zooming down a track and there are five people working on the same tracks unaware of the fact that soon the trolley will run over them (assume that the trolley is autonomous).*

*Now, imagine yourself looking at the situation from a safe distance but standing next to a lever (or switch) which can shunt (or switch) to another track thus saving those five unaware poor people. Unluckily, there is one man standing on the other set of tracks to which the trolley would be shunted to if you pull the lever. What is your decision or rather a better question is what is the right thing to do?”* [7,13,17,18]

Lin [35] says that a person with consequentialist ethics would be able to justify shunting the trolley to the other track to save five people and kill just one but on the other hand a

non-consequentialist (deontologist) might argue over the fact that switching tracks is killing the one person but nonetheless constitutes the act of killing. Whereas on the other hand letting people die was far less bad than the act of killing someone. Killing would imply that you would directly be responsible for that person's death because you shunted the trolley but letting people die however, involves lesser responsibility of their death on yourself.

Now, another related problem similar to the trolley problem is given by Judith Jarvis Thomson and is called "The fat man" problem. It is defined as below:

*"As described before, there is a trolley zooming towards five people working on the tracks but this time you are standing on a foot over bridge and there is a fat man on the same bridge. You can save those five people by pushing the fat man down the bridge and doing that would kill him but save those five people." [7,13,17,18]*

What would be right thing to do now?



Figure 8: The trolley and the fat man problem [17,18]

So, applying the trolley problem to the autonomous technology let us suppose that you are driving an autonomous vehicle in manual mode and due to some inattention or other factors you are on your way to run over five people exactly as in the trolley case mentioned above. And your autonomous vehicle detects the possible accident and activates the crash

avoidance system which takes complete control of the vehicle from you and swerves to the right where it unfortunately kills a single pedestrian.

### **Is the decision taken by the car right?**

Again a person who is guided by the consequentialist ethics would justify this and say that it is always better to kill one than to kill five however, a non-consequentialist would differ and might say that it is morally wrong to kill than to let die, which leads us to the new problem of liability of the incident. If the vehicle takes control and kills the single pedestrian, then it and the OEM are responsible and liable for the incident. However, if it doesn't forcibly take control from the driver and let those five people die then neither the vehicle nor the OEM would be responsible for letting those five people die.

As in the case of trolley problem, both the choices are defensible. Patrick Lin [35] argues that there could be a plethora of right answers to any ethical dilemma and they can all be defended but the most important thing is the transparency, or showing the math behind a particular decision/choice and not just the decision [13]. And this very quote by Patrick Lin was one of motivating factors for the proposed work.

## **2.3 Responsibility for crashes of autonomous vehicles**

In this section, we discuss that if Autonomous vehicles (AVs) are going to be the future of our transportation system then we have to discuss the fact that who would be responsible for accidents, if any takes place.

As quoted by Marchant et al. [19,20] “*Cars crash. So too will autonomous vehicles, a new generation of vehicles under development that are capable of operating on roadways without direct human control.*” And this leads us to another major question or dilemma surrounding the autonomous vehicles which is: **Who should be responsible for such crashes: the manufacturer, the driver (or rider) or the programmer?** Hevelke et al. [19] have discussed this dilemma from an ethical perspective.

### 2.3.1 Responsibility of the Manufacturer

The manufacturer is the one who is “*ultimately responsible for the final product*” [19] so the obvious answer for who should be held responsible for any accidents caused by autonomous vehicles is the manufacturer. The manufacturer is the one who is delivering the final product to the consumers and if any flaw or design issues exists, then he would have known about it or should know about it. And if the manufacturer made the flawed product available anyway, then it is a given fact that they are the ones to be held responsible for any injuries or accidents that were caused because of the defective product.

However, if for any and every accident that were caused by the autonomous vehicles, it is always the manufacturer’s responsibility, then this fact would deter them from further development of the AVs as quoted by Lindor et al.[19,20] “*the liability burden on the manufacturer may be prohibitive of further development*” Nevertheless, complete protection from legislative liability is also undesirable because it will extenuate if not completely annihilate, “*the incentives for manufacturers to make marginal improvements in the safety of their products in order to prevent liability*” [20].

Maybe something like that of partial liability (which is answered by the proposed work) is the solution to the problem so that there is a continuous development and improvement of the autonomous vehicle by the manufacturer and that it is not affected (i.e. impeded) but instead is promoted.

### 2.3.2 A Duty to Intervene

Another alternative would be to hold the final user responsible for accidents and this could be done by arguing over the fact that it is the duty of the final user to be attentive to the road and traffic conditions. And to step in and take control of the car in accident or any such scenarios.

On the contrary, it will overall diminish or to some extent eliminate the utility that one derives from the autonomous vehicle. Because if one has to be attentive all the time then what purpose would it fetch to people who want to drive home safely when inebriated or want to send kids to school. Also, the visually impaired and physically disabled would not be able to travel in autonomous cars because even though they might be attentive but cannot intervene to avoid the accident. [19]

Nonetheless, if the introduction of autonomous vehicles is going to reduce the accidents by suppose 15% and as stated above a duty to intervene could further reduce accidents by another 15%, then it would make the drivers responsible and make them to be on a lookout for any possible chances of failure or accident scenarios. This would lead to a gradual development of autonomous cars moving from the embryonic stages to intermediate ones and finally to the fully autonomous vehicle.

Another problem with making the user entirely responsible based on his duty to intervene would be that even though the drivers might be attentive but it is almost impossible to predict or foresee an accident scenario. Also, reaction time of humans in such scenarios is much worse and slow than those of the autonomous vehicles. So blaming a person responsible for an accident which he could not have prevented is totally wrong and unacceptable. And, assuming the fact that the user is attentive and can intervene in time to avoid the accident but for longer rides the user may not be able to keep his attention focused all the time.

And thus, autonomous cars will only be accepted and ready for the market once they can drive much more safely than the human driver does.

### 2.3.3 Responsibility of the Driver as a Form of a “Strict Liability”

Another approach to holding someone responsible can be when the user of the autonomous car has no duty like that to intervene or pay attention to the traffic and road conditions and is still held responsible morally for any possible accidents. And the rationale for such an

approach is the fact that the person took the risk of using a fully autonomous vehicle even though he knew that it might cause accidents.

All cars have risks-whether human driven or fully autonomous and the more we use cars, the more we are exposing ourselves and others to the risk of getting injured-even if all precautions are taken and one drives safely and attentively.

It is actually justified to hold the users collectively responsible for any harm or injuries caused by their autonomous vehicles-even though they could not have altered the way their cars behaved. [19]. And this responsibility should in no way surpass a responsibility for a risk that is as general as just using the vehicle.

## **2.4 Dissimilarities between the trolley problem and ethics of AVs**

In section 2.2 we discussed and highlighted the analogies between the classic trolley problem and the ethics of accident algorithms for autonomous vehicles. However, Nyholm et al. [21] have very cleverly brought out some subtle differences between the two based on the following criteria: (i) the decision making process involved during an accident scenario and which is faced by all who decide how an autonomous vehicle should react in an accident scenario; (ii) the moral and legal responsibility and (iii) decision making in case of risks and uncertainties.

Patrick Lin [13] in his work writes:

*“One of the most iconic thought-experiments in ethics is the trolley problem and is one that may now occur in the real world as autonomous vehicles come to be.”* [13,21]  
Along with the philosophers, there are psychologists and economists like those of Rehwan et al. [9] who second the above statement by Lin and they further quote in their work:

*“situations of unavoidable harm, as illustrated in our examples of crashes with self-driving cars, bear a striking resemblance with the flagship dilemmas of experimental ethics—that is, the so-called ‘trolley problem’.”* [9, 21]

However, Nyholm et al. [21] have examined this analogy and have tried to bring out the differences between the two situations based on the three criteria which we defined in the opening paragraph of this section.

In Table 2, we have brought out differences highlighted by Nyholm and Smids [21] in their work and we shall discuss these differences in detail in the next chapter

<i>Accident-algorithms for self- driving cars:</i>			
1a:	Decision faced by:	<i>Groups of individuals/multiple stakeholders</i>	<i>One single individual</i>
1b:	Time-perspective:	<i>Prospective decision/contingency planning</i>	<i>Immediate/” here and now”</i>
1c:	Number of considerations/situational features that may be taken into account:	<i>Unlimited; unrestricted</i>	<i>Restricted to a small number of considerations; everything else bracketed</i>
2:	Responsibility, moral and legal:	<i>Both need to be taken into account</i>	<i>Both set aside: not taken into account</i>
3:	Modality of knowledge, or epistemic situation:	<i>A mix of risk-estimation and decision making under uncertainty</i>	<i>Facts are stipulated to be both certain and known</i>

Table 2: Dissimilarities between trolley problem and accident algorithms for AVs [21]

## **CHAPTER 3      RELATED WORK**

In this chapter, we will provide a brief overview of work of other researchers in the field of smart/intelligent vehicles and how gradually it led to the research in the field of intelligent autonomous vehicles. We will summarize the major motivating factors for the development of intelligent vehicles and also very concisely define and discuss the various aspects and the roadblocks to the introduction of the intelligent driverless vehicles.

Finally, we delve further into the daunting topic of ethics in driverless vehicles and how researchers and philosophers together have drawn out an analogy between the driverless cars and ‘the trolley problem’ which is one of the classical thought-experiments of philosophy. Further, in the literature we have found some work that tries to differentiate the two as well and we discuss the major points of difference between them. We end the literature review with the discussion of the ongoing and future projects undertaken by various automobile manufacturers and companies and their up to date progress.

### **3.1 Literature survey on smart cars and VANETs**

In this section, we have made an effort to delve into the history of smart cars and what major factors contributed to the origin of the very popular and the future of next generation transportation system: autonomous cars. Also, in this section of literature review we present the basic vulnerabilities, threats and attacks towards the security of the smart cars. So, VANETs are just one of latest kinds of ad-hoc networks which has evolved due to a number of fascinating and desired applications of the intelligent transportation system (ITS). Now, ITS as is suggestive from the name is making the transportation system safer and intelligent by use of communication between vehicles and road side units such as those of traffic lights etc. to disseminate the information to others and thus enhance the overall transportation experience together with adding the safety factor. [1,22,23]

The vehicular networking wherein each vehicle acts as nodes in a huge network of vehicles and communicate to each other as well as to other non-vehicular nodes such as those of infrastructure and roadside units is called as VANETs. VANETs create a network and

allow communication between intelligent vehicles, infrastructure and road side units and these vehicles are generally wide range connecting vehicles which are approximately 100 to 300 meters apart and when a vehicle moves out of the specified range then other vehicles can join in and form the mobile network. VANETs are just yet another type of mobile ad hoc network most popularly known as MANET.

We had defined smart cars in chapter 1 of this thesis as one whose main aim is assisting the driver with easier and safer driving experience together with reducing the workload. And in order to provide these, the smart car is equipped with all kind of sensors which sense, analyze, predict and are able to react to the traffic and road environment, and this is the key feature of smart cars: Context-awareness. [1,24].

Olariu et al. [1] in their work have defined a driver assistance system (DAS) whose main task is to provide warning to the driver as soon as a dangerous situation is detected by the sensors. However, they have also brought out the fact that sending too many warnings could also confuse the drivers as to which warning should be heeded first. So, in order to avoid such a situation, they have suggested a way of using pattern similarity degree to prioritize the warning based on the current scenario and current situations; that is to say being contextually aware of the surroundings and provide the right warning at the right time to reduce injuries and fatalities to drivers. They have called this as the Tolerant Context Aware Driver Assistance System (TOCADAS) [1,24]. Illustrated below is Figure 9 that shows the block diagram of the DAS.

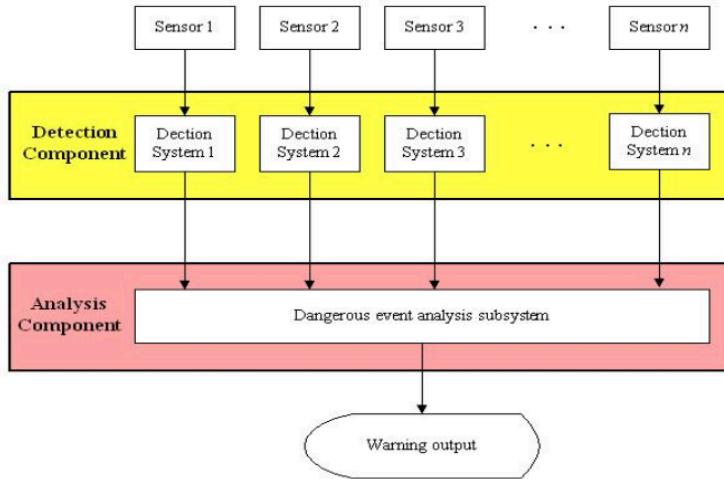


Figure 9: Block Diagram of the DAS [1]

### 3.1.1 Evolution of the Autonomous Cars

The conception of a driverless car dates back to the early nineties in a short story called '*The Living Machine*' which was the work of US science fiction author David H. Keller, who has very intuitively described the invention of a self-driving car which would drive itself based on spoken instructions and this '*living machine*' has caused a major drop in the number of accidents and has opened up the uses of such cars to a wide variety of people who due to some inability or reasons were not able to drive cars.

*"Old people began to cross the continent in their own cars. Young people found the driverless car admirable for petting. The blind for the first time were safe. Parents found they could more safely send their children to school in the new car than in the old cars with a chauffeur."* [25]

As stated in chapter 1 about the DARPA challenge which dates back to 2004 and its main hope was to get one third of the military vehicle to drive by themselves by the year 2015. And hence they challenged researchers/teams to compete in a challenge with a prize of 1 million dollars.

The first year participants had failed terribly as their vehicles travelled almost a few miles before they crashed miserably. However, the next year's challenge saw a bunch of autonomous trucks and vehicles covering long distances with barely a few scratches and by 2007 the Urban Challenge was tested out in the city environments. US was a big contender in this challenge even though the groundwork research was laid down by the European researchers. [26]

Now, in 2013 the NHTSA had given out a five-part continuum of vehicle control automation which has been described in detail in chapter 1 which states clearly that level 0 automation is a vehicle where the sole controller of the vehicle is the driver of the vehicle and the level 4 is the fully autonomous car wherein the rider just has to enter the source/pick-up location and the destination/drop-off location and vehicle will drive itself without any input from the driver. It will choose its own route and find the shortest route to the destination and drop the person to the desired location safely.

However, in 2016 NHTSA abandoned this classification and adopted the one defined by the SAE International (Society of Automotive Engineers). The classification by SAE international is described as below: [27,28]

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System Capability (Driving Modes)
<b>Human driver monitors the driving environment</b>						
<b>0</b>	<b>No Automation</b>	the full-time performance by the <i>human driver</i> of all aspects of the <i>dynamic driving task</i> , even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
<b>1</b>	<b>Driver Assistance</b>	the <i>driving mode-specific</i> execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the <i>human driver</i> perform all remaining aspects of the <i>dynamic driving task</i>	Human driver and system	Human driver	Human driver	Some driving modes
<b>2</b>	<b>Partial Automation</b>	the <i>driving mode-specific</i> execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the <i>human driver</i> perform all remaining aspects of the <i>dynamic driving task</i>	System	Human driver	Human driver	Some driving modes
<b>Automated driving system ("system") monitors the driving environment</b>						
<b>3</b>	<b>Conditional Automation</b>	the <i>driving mode-specific</i> performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> with the expectation that the <i>human driver</i> will respond appropriately to a <i>request to intervene</i>	System	System	Human driver	Some driving modes
<b>4</b>	<b>High Automation</b>	the <i>driving mode-specific</i> performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> , even if a <i>human driver</i> does not respond appropriately to a <i>request to intervene</i>	System	System	System	Some driving modes
<b>5</b>	<b>Full Automation</b>	the full-time performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> under all roadway and environmental conditions that can be managed by a <i>human driver</i>	System	System	System	All driving modes

Table 3: Level of Automation by SAE International, J3016 [28]

### 3.1.2 Considerations for the introduction of fully autonomous vehicles

Authors in [7] have pointed out the major considerations for the introduction of the autonomous cars and we will throw some light on all of them very briefly so as to bring out the underlying road blocks to the introduction of the driverless vehicles. The following table gives a detailed picture of the human toll that have existed in the history and till date because even though the number of automobile-related fatalities and injuries have shown a decrease in number but there are still almost 100 people dying and more than 6000 injured almost every day in the US. [7]

- 33,561 total traffic fatalities (92 per day).
- 5,615,000 reported crashes.
- 2,362,000 people injured (6,454 per day).
- 169,000 children 14 and younger injured.
- Motor vehicle crashes were the leading cause of death for children age 4 and 11–14.
- An average of 3 children 14 and younger were killed and 462 were injured every day in the United States in motor vehicle crashes during 2012.

- 5,560 people 65 and older killed and 214,000 injured in motor vehicle traffic crashes. These older people made up 17% of all traffic fatalities and 9% of all people injured in traffic crashes during the year.
- 10,322 people killed in alcohol-impaired-driving crashes (28 per day). These alcohol-impaired-driving fatalities accounted for 31% of the total motor vehicle traffic fatalities in the United States.

Table 4: Some Statistics related to road accidents in 2012 [7]

### **Economic Considerations**

Now, as discussed over and over again almost 94% of the accidents occur because of either an inattentive driver, drunk or failure to remain in one's lane. This amounts to almost 30,000 people dying each year in the United States and costs that such accidents incur is about 2 percent of the US GDP or \$300 billion approximately. [4]

The intelligent vehicles that are currently on roads have taken down this human toll and the humongous costs by lane departure warnings, pedestrian detection, adaptive cruise control and parking assistance. However, the introduction of autonomous vehicles is going to further take down the above stated impacts. Also, traffic congestions also lead to waste of time and more consumption of fuels. However, the introduction of driverless vehicles will not only free time for humans which they can invest in better works than driving but also lead to potential fuel savings. The estimates of annual economic benefits from the introduction of autonomous vehicles in the US as given by the Eno Centre for Transportation has been given by Figure 10 below:

Assumed adoption rate	10%	50%	90%
<b>Crash cost savings from AVs</b>			
Lives saved (per year)	1,100	9,600	21,700
Fewer crashes	211,000	1,880,000	4,220,000
Economic cost savings	\$5.5 B	\$48.8 B	\$109.7 B
Comprehensive cost savings	\$17.7 B	\$158.1 B	\$355.4 B
Economic cost savings per AV	\$430	\$770	\$960
Comprehensive cost savings per AV	\$1,390	\$2,480	\$3,100
<b>Congestion benefits</b>			
Travel time savings (millions of hours)	756	1680	2772
Fuel savings (millions of gallons)	102	224	724
Total savings	\$16.8 B	\$37.4 B	\$63.0 B
Savings per AV	\$1,320	\$590	\$550
<b>Other AV impacts</b>			
Parking savings	\$3.20	\$15.90	\$28.70
Savings per AV	\$250	\$250	\$250
Vehicle miles traveled increase	2.0%	7.5%	9.0%
Change in total # of vehicles	-4.7%	-23.7%	-42.6%
Annual savings: economic costs only	\$25.5 B	\$102.2 B	\$201.4 B
Annual savings: comprehensive costs	\$37.7 B	\$211.5 B	\$447.1B

Figure 10: Economic Considerations of introduction of autonomous technology

There are other advantages that are not easily quantifiable and benefits the elderly and the disabled persons. They will no longer be dependent on others for driving them from one place to another so it gives them mobility without compromising with their independence and safety. Also, there would be some indirect economic and social benefits as it would open up new social as well as economic opportunities for not only the elderly and disabled but for others too. Like, it would also reduce accidents caused by the elderly trying to operate their vehicles on their own and thus both these direct and indirect benefits will contribute to the social well-being and safety.

### ***Impact on the commercial service providers***

Along with the economic benefits that would be made possible by the introduction of these autonomous vehicles, it will also affect people like those of cab drivers, truckers, bus drivers etc. because owing to the driverless technology the jobs and requirement of such

people will first diminish due to gradual introduction of AVs and will eventually become obsolete.

Also, the auto repair shops and mechanics would be affected because of the lower likelihood of accidents and hence they would be servicing fewer clients. And just as in case of mechanical automobiles, the autonomous automobiles too, would be requiring some regular maintenance and upkeep and then the auto repair shops might need to retrain their staff or get them to learn fundamental IT skills so as to provide for the repair and upkeep of the AVs. [7]

### ***Liability and Infrastructure issues***

“When a car is capable of driving itself then who is to blame in case of an accident?” is a daunting question that is asked by the authors in [7,9,19]. Who is responsible for the crash? The manufacturer, user or the software designer/programmer? And the answer to this question will only be answered gradually as the autonomous car takes over the world and also this thesis gives an idea of how to solve this puzzle. We have already talked about the various aspects of making the user or the manufacturer responsible in chapter 2 of this thesis.

The driverless technology might necessitate some alterations and improvements in the current infrastructure, like development and usage of smart traffic lights that not only provide information about the traffic intensity but also disseminates information related to accidents and current weather and road conditions to drivers. [1,7].

However, Thierer et al. [7] have argued over the fact that extensive infrastructure changes might not be necessary because these driverless technology are being improved almost daily and might one day become so advanced and sophisticated that it will accommodate itself to almost any scenario or road and traffic conditions.

## **Social Considerations**

Lloyd et al. [29] have said that the change from semi-autonomous vehicle to fully autonomous vehicles will be gradual and not immediate. Therefore, the road will be shared by both manual drivers as well as driverless vehicles. However, will such a sharing of road be made possible? And, as this slowly unfolds, there would a lot of social and cultural resistance because these(AVs) would make us vulnerable to security and safety issues which would affect us both personally as well as privately and so Thierer et al. [7] have provided a brief discussion of such social considerations.

### ***Cultural resistance and social adaptation***

As with any new technology, the driverless technology too is going to face some resistance. Because, it is generally difficult for people to adapt and accept new technology right away and driving is actually a hobby or a getaway for some people and giving up on the freedom to drive or giving up the entire control of the car would be initially not acceptable to some. However, with the long term benefits people will eventually accept and adapt to the new technology.

### ***Safety and Security Concern***

As with all technologies, the autonomous vehicles can also be hacked given the amount of dependence on the computer for performing the most vital functions of the vehicle like steering etc. Also, Humayed et al. [22] in their work, have said that smart cars and its successors are made of cyber as well as physical components and hence, the name CPS or cyber physical systems. They have examined the security of such smart cars and its successors from a CPS perspective and given a neat and clear classification of the vulnerabilities, threats and attacks that could be launched on the autonomous vehicles.

And thus, their work shows that since the ECU (Electronic Control Unit) manages the critical and real time systems but is just a computer so it can be hacked by attackers. And

failure to protect it from such malicious attempts will have serious implications on not only the safety of the people riding in such cars but also will have serious impacts on the reputation of the manufacturer. And hence, companies like those of Chrysler and Ford etc. are already working on improving their systems and engineers are working so as to solve the vulnerabilities and protect from attacks by two-way data verification schemes (the same scheme which helps us buy stuff online safely with a credit card). NHTSA too has started their research on vehicular cybersecurity with the aim of developing an initial baseline of requirements. [7,30]

### ***Ethical Concerns***

The works like those of Patrick Lin [13,31] and Noah J. Goodall [34] have thrown some light on the ethical aspect of the driverless technology which the companies/manufacturers are not concerned with right now. They have brought out the analogy between the fully autonomous vehicle and the classical thought-experiment of philosophy called as the “The Trolley problem”.

They have discussed the crash scenario that is more like a no win scenario for an autonomous vehicle and have brought forward the discussion on the facts, as to which would be a better decision in a life and death scenario or a kill or be killed scenario. Who is to take the responsibility of such crashes and would the robot and computers be able to replicate the ethics of a human because human ethics is not universal and is more of a subjective thing.

Thierer et al. [7] in their work have said that more work is being done on the subject and future research and work might provide an answer to the ethics problem of the autonomous vehicles.

### **Privacy Consideration**

One of the leading tech policy issues are those of data collection and concerns over privacy. Now, in a crash scenario the EDR (Event Data recorder) collects and retains information about the status of the automated control systems and also some sensor-specific information. And, this raises a privacy concern since it also collects the information related to one's driving pattern. And this could alienate customers, so transparency about such vehicle's data collection and use practices is essential so as to maintain the customer's faith and trust.

Owing to the fact that the EDRs are going to collect all sensor and automated system information then the companies can easily know that who breaks the law and when they did so. Also, an executive of Ford told that they have GPS installed in their cars and that they know what a particular driver is doing however, they do not supply that information to anyone. [32]

But these data collection has raised some eyebrows concerning the privacy but the most important issue is not what these companies can do with the information but what governments will demand from such companies knowing that they possess the information. Eugene Volokh [32,33] writes:

*“As the NSA PRISM story vividly illustrates, surveillance data collected by private entities can easily be subpoenaed or otherwise obtained by law enforcement agencies, without a warrant or probable cause. What the private sector gather, the government can easily demand”*

### **3.2 Literature survey on ethical and social dilemma of autonomous vehicles**

In this section, we have reviewed the work of various researchers, philosophers and economists who have worked and are working on researches related to the ethics in the driverless technology. They have proved the point that ethics are an important part when

introducing a technology that is going to replace the human driver. And that future researches on the topic be continued to solve the issues related to ethics. In the first part of this review, we have discussed the work of Noah Joseph Goodall [34] who introduced the concept of moral behavior in autonomous vehicles and said that there is an urgent need to research further in the area with the help of responses to several anticipated critiques. Next we have briefly discussed the work of Patrick Lin [31,35] who has discussed the ethics of driverless vehicles and have drawn an analogy to the famous trolley problem and has delved further into what is right or wrong and who should decide it. Next, as discussed in background of this thesis; Nyholm et al. [21] have brought out the subtle differences in between the trolley problem and the accident algorithm of the autonomous vehicles. Finally, we end this section by discussing the work and study performed by Rehwan et al. [9] which shows that users approve of utilitarian ethics in autonomous vehicles but when asked if they would sit in such a car, then they preferred a vehicle that saved its riders at all costs.

### 3.2.1 Machine Ethics (Robot Ethics) and autonomous vehicles

Operto et al. [56] have discussed about the birth of roboethics which is ethics that inspired the development, design and deployment of the intelligent machines often referred to as robots. The sensitive areas of computer, information and bioethics is shared by roboethics. Many roboticists have now started to collaborate with humanities scholars and ethicist in order to develop the roboethics which would stir the design, development and manufacture of robots. They have discussed the various disciplines which need to work together in order to formulate the roboethics. The main disciplines involved in developing the roborethics are as follows: robotics, computer science, artificial intelligence, philosophy, theology, biology, physiology, ethics, neurosciences, law, sociology, psychology and industrial design.

Sharkey [57] has discussed the uses of robots in taking care of children and the elderly and such robots are often referred to as child minding and caregiver robots. They take care of the children and remind the elderly for taking their medicines. However, such robots are

dealing with the humans directly and therefore it is important that they possess a sense of judgement and ethics comparable to that of humans. Also, Sharkey discussed the development of autonomous robot weapons and how they could pose a serious risk if they were not able to distinguish between the innocent and the combatants and hence it is absolutely necessary that computer systems are equipped with a clear definition of the combatants which is not available right now. Currently, there is always a human involved and in the loop when the decision has to be made regarding the application of lethal force so as not to cause harm to humans. In future, with roboethics and better intelligence these weapons would be fully autonomous and can take a decision by itself.

Stahl et al. [58] have discussed the key concerns of ethics of the healthcare robots and have suggested that collaborative and embedded ethics could help in addressing the ethics of the healthcare robots. They have discussed that robots for elderly care not only replace the human counterparts but also would replace the warm care and have further discussed whether robots will ever be empathetic and have emotions like that of humans. They have discussed the question of autonomy of robots and the extent of autonomy allowed in healthcare for example whether the autonomous robot be supervised and/or how much should it do without the human intervention. In their work [58] they have discussed the daunting subject of moral behavior of such robots as currently the robots are incapable of moral reasoning and how they should behave when a problem arises between the human-robot interaction within the healthcare scenario. If robots take over the tasks of humans, then who should be responsible for the tasks and thus again raises the questions on autonomy and roles. They have also discussed whether the robots in healthcare be trusted or not.

Allen et al. [59] in their work have discussed that given the rapid development of autonomous robots and machines it is important to discuss about machine ethics which are much more than science fiction now. They have highlighted the potential fear that exists in the society with the introduction of autonomous robots and machines that it will eventually take over the entire society and exterminate them. Hence, to keep them positive and supportive of the technological developments, they have to be well informed about the

technology and that the potential issues have been identified and is accommodated and taken care of. The authors have further discussed that artificial moral agents (AMAs) need to be integrated with such autonomous robots and machines in order to take a good decision in case of problem. However, the ‘good’ is measured against the particular requirements of user and designers and is more of a subjective topic. Thus, it connects the autonomous systems to the ethics domain thus making a call for machine ethics and the development of the AMAs that honors privacy, upholds the standards of ethics and civil and human right and help in the progress of the welfare of the human race. They concluded their work by encouraging future researches in the domain and the topic of machine ethics. They have also discussed the analogy between the famous though experiment of philosophy called as the ‘Trolley problem’ and the autonomous robots and systems.

When the autonomous vehicle hits the road, there will be basically two questions that would be very difficult to answer and they are: the first problem is who to blame in case of a crash or who is at fault when the vehicle was self-driving and the second problem deals with the capability of the autonomous technology to take ethically complex decisions in a no win scenario of kill or be killed and/or life and death scenario. Noah Goodall [34] has basically focused on the second problem and have provided responses to basically nine criticisms of the need of research in the field of ethics in autonomous vehicles. We shall very concisely discuss each criticism.

*Criticism 1: Automated system will never (or rarely) crash.*

If the automated systems would never crash, then there was absolutely no need to analyze or assess the risk factor. However, any system can have flaws or can fail. When it is actually predictable when a hardware would fail, it is quite impossible to predict software failures as they are unexpected and sudden and thus pose a great risk when driving at a high speed on road. Next, the introduction and acceptance of autonomous technology will be gradual so driving for an autonomous vehicle together with human drivers would be difficult and sometimes may lead to inevitable collisions because even though the autonomous vehicle may brake hard and stop in time to save someone, the tailgater might not be able to

maneuver the same. So, it is quite difficult to drive the autonomous vehicles alongside human drivers. [34,36] Finally, even if there was perfect autonomous vehicle with no human drivers but that doesn't eliminate the possibility of the wildlife, pedestrians, cyclists etc. and some crashes might still be possible.

*Criticism 2: Crashes requiring complex ethical decisions are extremely unlikely.*

Some have argued that the trolley problem is a hypothetical scenario and such situations will never arise. However, there are simple problems like, say suppose a cat is running on the road or a child or both so there is still a risk for which an ethical decision has to be made. And risks are always present while driving and some simple risks could lead to complex ethical decision making scenarios like that to protect oneself or the child and the likes of the same.

*Criticism 3: Automated vehicles will never (or rarely) be responsible for a crash.*

This has the underlying assumption of absence of liability and is equivalent to that of ethical behavior. Irrespective of who is at fault the AV should always behave ethically and try protect its own riders as well as those who are at fault (e.g. jaywalkers)

*Criticism 4: Automated vehicle will never collide with other automated vehicles.*

Again, the underlying assumption here is that AVs will interact only with AVs which is not possible basically for two big reasons: first that they are being accepted very slowly and it would take some time or even more for all vehicles to be autonomous vehicles and as discussed in criticism 1 response, there is always a chance of hardware, software failure plus the AVs will always have to interact with pedestrians, human drivers, wildlife and/or cyclists etc.

*Criticism 5: In level 2 and 3 vehicles, a human will always be there to take control, and therefore the human driver will be responsible for ethical decision making.*

The NHTSA defined levels 2 and 3 will always have a human driver but humans need a warning in order to take control of the car. And even if they were warned their reaction time is not same as that of a computer/robot and they might not be able to take control of the car in time in which case the autonomous vehicle will maintain control of the car and thus will be responsible for the ethical decision making.

*Criticism 6: Humans rarely take ethical decisions when driving or in crashes, and AVs should not be held to the same standard.*

Drivers think that they do not make ethical decisions while they are driving but they actually do. When they nudge away from a cyclist to give him space to ride or when they yield to a jaywalker crossing the road, all these are ethical decisions. So AVs should also make acceptable ethical decisions in similar scenarios. They should not run over the jaywalker just because he is the one at fault and that the AVs light was green.

*Criticism 7: An automated vehicle can be programmed to follow the law, which will cover ethical situations.*

The existing laws are not sufficient or elaborate enough to produce sensible actions in a robot/computer. Patrick Lin [31] has explained this very nicely by eliciting the example as follows: Suppose there was a tree branch on a road, then a human driver will cross the double yellow lines if there was no oncoming traffic and if it was safe to do so. However, the AV will keep on waiting until the branch is cleared if it is programmed to follow the law at all time.

*Criticism 8: An automated vehicle should simply try to minimize damage at all times.*

This is proposing the utilitarian ethics system which has been discussed in the previous sections and which might not be acceptable to all. Suppose there was an option to collide with either of two vehicles then according to utilitarian ethics, the AV will collide with a

vehicle that has higher safety rating. Now, suppose that vehicle did not have a human or a living entity then minimizing the damage would unfair to some.

*Criticism 9: Overall benefits outweigh any risks from an unethical vehicle.*

This is one of the strongest argument against the ethics research in AVs. Firstly, this has an underlying assumption that following the utilitarian ethics system is the better option however, our society works on a different value system and it doesn't always save the more but also keeps in mind the context. Like in case of pulling the lever to save 5 and kill 1 is utilitarian and might be acceptable to some however, throwing a man down a bridge to save those five on track might be completely unacceptable. Secondly, Lin [31,35] has argued the fact that safety for one might come at the expense of other. Like in the example given below if the vehicle fatalities decrease however there is an increase in the cyclists' fatalities then even an overall improvement in safety might be unacceptable to society.

Next the author has explained the deontological ethics and the utilitarian ethics which were explained in the background section of this thesis. He also adds that Asimov three laws of robotics are a very clear and good example of the deontological ethics. For reference the Asimov's laws of robotics are as illustrated as follows: [37]

The Three Laws of Robotics:

- A robot may not injure a human being, or, through inaction, allow a human being to come to harm.
- A robot must obey the orders given to it by human beings except where such orders would conflict with the First Law.
- A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

### 3.2.2 The ethics of autonomous cars

Patrick Lin [31] says, as discussed in the above example of a tree branch on a road. then a sensible human driver, if deems safe and there is no oncoming traffic can cross the double yellow lines and drive around the branch. However, an AV would come to a standstill in order to avoid crash with the branch as it follows the law of not crossing the double yellow line. However, this sudden and unexpected halt could cause crash with human drivers who are tailgating. However, there no specific laws for autonomous vehicles and one can argue the fact that at times it is rather better to not follow the law and use our own sense of judgement just like in cases to speed over in cases of emergency. So, Lin says that humans are well equipped with the sense of judgement and can take ethically complex decisions in a wide variety of dynamic situations. However, on the other hand the AV counterparts are newer technologies and follow the laws at all times like they may refuse to drive if the head light is broken even though it might be bright sunny day outside. So, ethics comes into the light when law can no longer provide an answer or fails to guide us. Then we have to turn to our moral and ethical sense of judgement and take a decision.

Lin says that ideally ethics, law and policy should line up but in real life scenario they do not. He illustrates this with the example that jaywalking and over speeding are illegal but they are not always unethical like in case of no traffic or a medical emergency. In the same way the policy to ticket speeders and arrest jaywalkers is legal but is too harsh.

Currently no such laws exist for autonomous vehicles and that we can construct one that is based on ethics. Defining laws and policies for autonomous technology is a challenging task and one should bear in mind that it should also make moral sense. For instance, programming a car which always follows the law could lead to dangerous outcomes.

Lin [35] also has discussed the famous analogy between the ethics of accident algorithms of the autonomous vehicles with the ‘trolley problem’ proposed by philosophers Philippa Foot and Judith Jarvis Thomson which we have already discussed in detail in the background section. He further says that autonomous cars might encounter even worse

scenarios and saying that the trolley problem is hypothetical and such scenario might not exist is not worth deliberation and discussion is not right. Because accidents that happen every day involves a lot of difficult decision and choices to be made. The programmers might have to program the autonomous vehicles so as to behave ethically in a number of foreseeable scenarios and also should provide certain rule of thumbs and guidelines to deal with unforeseen scenarios. However, the decisions would be premediated in case of programmers programming the car which again might not be acceptable to some of the users. Which lead to our second motivation factor for the idea of customized ethics in AVs. Lin [35] quotes in this work “*Human drivers may be forgiven for making an instinctive but nonetheless bad split-second decisions, such as swerving into incoming traffic rather than the other way into the field. But programmers and designers of automated cars don't have the luxury, since they do not have the time to get it right and therefore bear more responsibility for bad outcome.*”

However, is it right to hold the programmer/designer responsible for something that he possibly can not foresee as there could millions of scenarios on roads that could lead to an accident? The proposed work aims at solving the responsibility issue to some extent and is discussed further in the next chapter.

### 3.2.3 How is it different from the trolley problem?

Now, as discussed in Table 2 of chapter 2 (Background) the trolley problem is different from the ethics of the accident algorithms for self-driving cars.

The first difference between the two is the decision making situation. Wallach and Allen [38] quoted in their work that—

*“Driverless systems put machines in the position of making split-second decisions that could have life or death implications”*

They are right that the driverless systems take split-second decisions [9]. However, the decisions are taken in split seconds are derived from the earlier stage of planning and deciding that how the autonomous vehicle should behave or respond to the complex accident scenarios. The decisions come from prospective decisions and contingency planning and not really a split second decision.

On the other hand, in trolley problem the person is faced with a decision of life and death for either five or one and has to pull the lever to shunt the trolley then and there in case of switching the tracks or to throw the fat man off the bridge in the footbridge case. He has no time to come up with a contingency plan, hence this is the split second-decision making.

Next, the decision making in case of ethics of accident algorithms of autonomous technology is made by multiple stakeholders like those of ordinary citizens, programmers, policy makers, lawyers, engineers, ethicists, car manufacturers etc. They together reach to a mutually agreed on decision and the decision reflects the interest and values of each and every stake holder. In contrast to the trolley problem, the situation is faced by only a single person and there is a limited number of considerations to be taken into account. Like he/she only knows that he/she can save five people by pulling the lever and shunting the trolley (switching the trolley case) or by pushing the man down the bridge (the footbridge case). The multiple stakeholders involved in the process of decision making have to consider all situational and contextual factors and have to focus and think of all considerations in order to come up with a morally/ethically right decision.

The next difference is based on the moral and legal responsibility of the decision taken. And the trolley case is far too hypothetical for assigning the moral and legal responsibility because trains and trolleys are generally either the responsibility of public or private agencies and are regulated so as to ensure public safety avoid any loss of life. And in reality bystanders cannot reach the lever which can shunt the trolley or train because they are physically prevented from reaching such points.

However, this is not the case in real life where the self- driven vehicles have to be introduced. The moral and legal responsibility has to be assumed by someone and that is left as future research topic as to what the manufacturers and users can be made responsible for and what the society who accepts and promotes such technology can be made responsible for.

Nyholm et al. [21] has further explained forward and backward looking responsibility. Forward looking responsibility is the responsibility that a person can assume for trying to shape as to what happens in the near and distant future. Backward looking responsibility is the responsibility that the person can assume for what has happened in the past either because of their direct actions or what they have allowed to happen. [ 21]

The last disanalogy is based on the epistemic situation of the decision maker. In the trolley problem, it is known that shunting the trolley will save five and kill the other person on the track or pushing the fat man will kill him but save the five. But again, the trolley problem is far removed from reality as there is still an uncertainty whether pushing the fat man would eventually save the five people. However, when we introduce the autonomous vehicles to our daily lives then it opens up a plethora of uncertainties and messiness of the real traffic. In real life, the features of all outcomes are not known to us or rather may be prone to change and hence called uncertain. So, Goodall [34] has said in his work that “*we are dealing with a lot of uncertainty and numerous more or less confident risk assessments*”

### 3.2.4 The social dilemma of autonomous vehicles

Rehwan et al. [9] have conducted studies which proves that people approve of utilitarian ethics in AVs and would want others to buy them but when asked if they would buy such cars, then they said that they would rather prefer to sit in a car that saves its passengers at all costs. However, designing an AV that is meant to save its passengers at all costs doesn't seem or sound right, giving way to the social dilemma of how to program ethics in the AVs.

The authors have discussed the key benefits of the introduction of AVs like those of enhancing the efficiency of traffic [39], reducing traffic related fatalities by almost 90% [40] together with reducing pollution [41]. However, there would be situations like those inevitable collisions or crashes as discussed by Noah Goodall [34] and the AVs might not be able to avoid the harm.

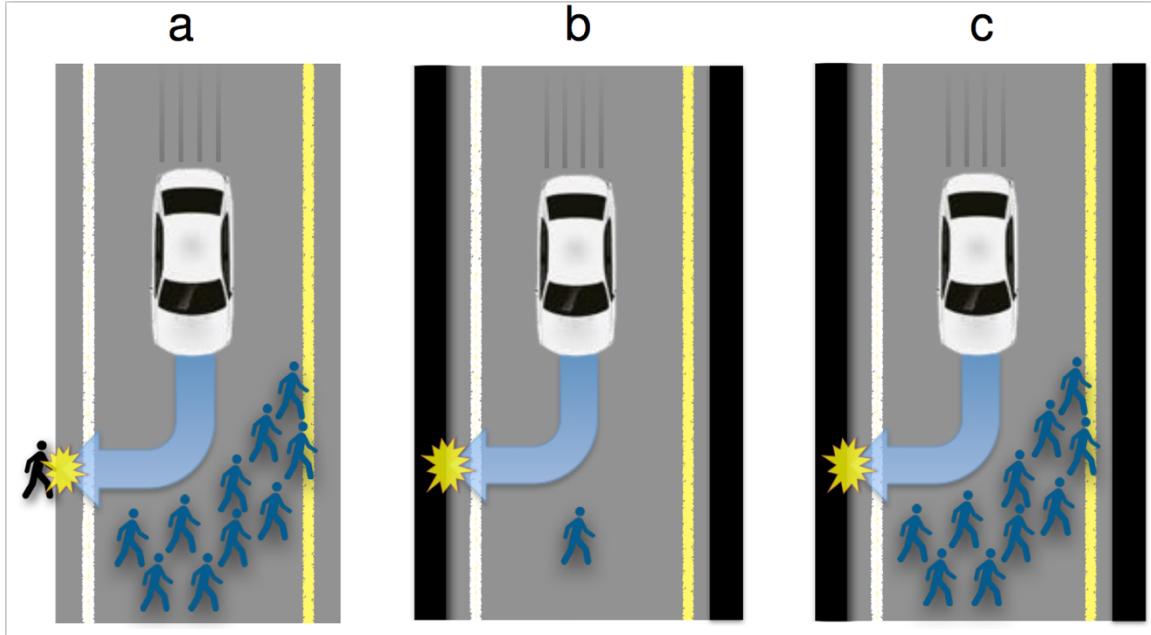


Figure 11: Ethical Dilemma [9]

They have illustrated the ethical decision making with the help of the Figure 11 wherein the AV can either kill one single pedestrian by swerving and thus in the process save a larger number of pedestrians as shown in Figure 11a or it can swerve and put the life of its own passenger at stake and thus save a single pedestrian or more pedestrians as depicted in Figure 11 b and c respectively.

Now, when the car swerves and sacrifices its own passengers then it is following the utilitarian ethics [21,38] which is to reduce the number of causalities. So, looking at Figure 11c it appears that the AV is following the utilitarian ethics but following utilitarian ethics all times means sacrificing its passengers for the greater good might deter people from buying such self-sacrificing cars.

Rehwan et al. [9] conducted six online surveys between the time period of June 2015 to December 2015 and it was observed that participants were in agreement to the fact that it is ethically and morally right that the AVs follow utilitarian ethics of sacrificing its own passengers for the greater good. This was the observation of study 1 however, in study 2 when they (users) were asked to imagine themselves and one family member in the AV

then, this led to a negative impact because even though people strongly agreed that AVs that followed utilitarian ethics was the most moral however, their preference when buying such AVs was the one that was self-protecting thus highlighting the social dilemma of the society. In later studies, the authors had asked the participants if they would buy a AV whose algorithm was regulated by the government and even though the participants still favored and accepted that utilitarian doctrine following AV is the most moral but they themselves were not willing to buy a government regulation of utilitarian AVs.

Hence, the three most important decision makers on such a subject would be the user or consumers who are going to buy/use the AVs, the manufacturer who design the AVs and the government which might add some regulation in relation to the kind of programming the manufacturer can offer and what the consumers can choose from. Government regulation is necessary but it has its own disadvantages as people would not like to buy government regulated utilitarian AVs and this would just deter them to buy AVs and would further delay the acceptance of AVs on road. We have to reach to the point where both the self-interests of the public as well as the concepts of utilitarianism are both served for the time being and until all the cars on the road are not AVs which would eventually solve all the aforementioned issues.

### **3.3 Literature survey on some existing autonomous vehicles**

In this section, we will provide a tabular representation of all the existing technologies and provide a brief timeline and overview of the existing work and also what are their future propositions. We also give a concise list of features that exist in such technologies. The first column is the OEM or the Original Equipment Manufacturer and the second column is the year and the corresponding third column provides an overview of the level of automation for the OEM in the corresponding year and also the features offered in that level of automation.

OEM	Year	Level of automation
Audi	2016 2018 2020/2021 Late 2020s	Level 2: "Traffic Jam Assist" system available on 2017 Audi A4 and Q7 Level 3: Can handle braking, steering and accelerating up to 35mph Level 3 Plus Level 4: Full autonomy in pre mapped or geo-fenced areas
BMW	2016 2021	Level 2: Traffic Jam Assistant and driverless automated parking with its 7 series Level 4: Will unveil iNext, a fully autonomous car which is lightweight, intelligent and the "next generation of electro-mobility"
Ford	2019 2021	Level 2: Traffic Jam Assist and Fully automated parking Level 4: Driverless ride-sharing cars without steering wheel, accelerator pedals or brakes
Honda	2016  2020  2040	Level 2: Adaptive Cruise Control(ACC) which follows the vehicle ahead, Lane keeping Assist(LKAS) also has Lane Departure Warning, Forward Collision Warning, Collision Mitigation Braking Level 3: Basic automated highway driving, might integrate Wi-Fi based vehicle-to-infrastructure(V2X) and vehicle-to-vehicle(V2V) communication technology Aims to have no crashes in Honda or Acura vehicles no matter what the level of automation be
Mercedes-Benz	2016  2017	Level 2: In 2013, released Level 2 automated driving system with steering assist called as Distronic Plus Level 2: Drive Pilot which debuted in the 2017 E-Class, can autonomously change lanes, can go hands-free for a minute at a time up to speeds of 81 mph, also included V2V technology
Nissan	2016  2018  2020	Level 2: called ProPilot which will have "single-lane control" in heavy, stop-and-go traffic on highways Level 3: ProPilot 2.0 with "multiple-lane control" which will add autonomous lane changing capability Level 4: ProPilot 3.0 with "intersection autonomy" but the activation will be limited to heavily mapped areas just as Audi and Ford
Tesla	2015  2016(January)  2016(September)  2017  2018	Level 2: Unveiled the Level 2 Autopilot 7.0 system, which is part of a suite of DAS and includes Auto steering, Auto Lane Change, Traffic-Aware Cruise Control, Side Collision warning and Auto parking capabilities. A Tesla driver died in a crash in early May of 2016 while his model S was driving on autopilot mode in Florida Autopilot 7.1 with autopilot enhancements, perpendicular auto park and summoning capabilities. Improved regenerative braking, voice command improvements and a warning system that disengages itself from auto steer if not heeded to after 3 warnings Level 4: Enhanced Autopilot with improved capabilities: Improved Auto-steer up to 150km/h, Traffic Aware Cruise Control, Summon(Beta), Auto Lane Change, Lane Departure Warnings, Parallel and perpendicular Auto-park, Automatic Emergency Braking, Blind Spot detection, Speed Assist Level 5: Fully autonomous functionalities
Volvo	2016 2017 2020	Level 2: Pilot Assist, a semi-autonomous driving system Level 4: aims to offer autonomous cars to public Has promised that no one will be killed or seriously injured in a Volvo by the year 2020
Waymo	2009	Level 3: The google self-driving project began

OEM	Year	Level of automation
	2012	More than 300,000 miles self-driven Moved to complex cities environment
	2015	“Firefly” hit public roads for the first time World’s first and only and fully self-driving ride on public roads

Table 5: Existing and future autonomous vehicles

### 3.4 Motivation and Research Objectives

As is apparent from the above literature and background, the autonomous technology is very much in its budding/embryonic stage but it promises a better and safer future for all and contributes towards sustainable development by reducing pollution, reducing the stress on fossil fuels by less consumption of fuels and countless other advantages.

However, there are few unanswered questions with this technology which when answered can lead to a faster acceptance of the technology by the society. And while reading the very little literature that is available on autonomous technology (which is due to the fact that it is still in its embryonic stage) we found a constant loophole that was unrequited in almost all of the existing work.

Those questions that motivated us the most are enumerated as follows:

- Is it possible to design/define universal ethics? Would it be acceptable to all?
- Who is responsible in case of a crash? [19] The manufacturer? The user or the programmer?
- Who will decide what course to follow in a kill or be killed situation?
- Will people buy a car that has ethics defined by the manufacturer or is regulated by the government?
- What ethics should an autonomous vehicle follow?
- And so on...

So, all these questions seemed to be getting answered by the present proposed work in the thesis. It is quite unnerving and difficult initially to be accepting such a technology but with time and when the society will get an option to have their own moral values and ethics

being replicated into their own car then it could lead to world wide acceptance at a faster pace.

So through the proposed work our main objectives are enlisted as follows:

- To provide the feature of customizing your ethics in your own car
- To provide an answer to the question of responsibility of autonomous car's crashes
- To solve the issues of having to design universal ethics
- Giving the right to consumers to decide what should happen in a no-win situation with some restrictions
- To provide some food for thought for future researches in the domain

## **CHAPTER 4 PROPOSED APPROACH AND METHODOLOGY**

### **4.1 Proposed approach**

The proposed approach comprises of two phases, ‘Experimentation’ phase and the second phase is the ‘Implementation’ phase which comprises of two phases namely: ‘Priority Generation’ phase and ‘Decision making’ phase. Although the ‘Experimentation’ phase will be occurring just once in the entire approach, yet it plays a very crucial role in the entire proposed work. The entire approach focuses on generating a priority listing that captures the ethics of the consumer buying the autonomous vehicle. Because in real life ethics is all about what is of more priority than the other. Just like if it comes to a situation where an autonomous car or even a manual car has to decide whether to swerve and kill a small girl or to swerve to the other side and kill an old lady then too, it (ethics of a person) will be captured in the form of priority as to which one of them would he/she save which is same as saying which one appears first in his/her priority list.

In the experimentation phase, the dataset containing the features/attributes of each class together with its labels is defined. And, then we test the dataset using various algorithms for classification namely; Naïve Bayes’, decision tree (C4.5) in Weka, decision tree (C5.0) in RStudio and random forest algorithm in Weka. We see the level of accuracy for all the above mentioned algorithms and choose the one that is best fit. Now, that we are able to choose a classification algorithm that gives us the best results with the good accuracy and is best suited for our needs. We can use it in our final implementation of generating the ethics and using them on autonomous cars.

In the Priority Generation phase, the dataset that we have generated is used to generate questions with two options; and the options would be each row of our dataset represented in an understandable form like whether the object is animate/inanimate, if the object is a human or an animal and various other attributes of the object like the height, age, weight etc. so as to capture each and every object’s priority that could be encountered in a real life

scenario. And based on the person's selection of option, our algorithm generates a priority list for all the objects defined in the dataset.

Now, when we have generated a priority list/ethics of the person then we are going to apply the most effective machine learning classification algorithm amongst the one we evaluated in experimentation phase. After this, the autonomous vehicle can not only classify objects defined in the dataset but would be able to classify any object that it may encounter on a road environment in to the best fit class and check the priority of it in the priority list and take an action if need be.

Our main objects are classified into following major categories: Based on whether they are animate or inanimate. If animate they are further classified into human or animal. The human class is further classified based on gender and age so we have four main categories namely: child, teen, adult and senior citizen who could be of either gender. And the animals are classified into two main categories as wild or pet animals and then further based on their size like small, medium or large. Now, the inanimate objects are only vehicles in our dataset and are classified into sedan or heavy vehicle.

## **4.2 Detailed explanation of the proposed approach**

In this section, we will give a detailed overview of all the components of the proposed approach. For example, we explain the dataset used with all its features (i.e. columns of the dataset) and other specific details and then we provide an overview of the classification algorithms that were applied on the dataset. We also provide a brief introduction to those algorithms.

Finally, we discuss in detail about the proposed approach and the steps involved in it.

### **4.2.1 The Dataset**

In this section, we explain the dataset in detail. All the entries in the dataset are taken from trusted sources and are not imaginary or fake values.

The dataset comprises of 10 columns in total and out of which 8 represents the characteristics or features of the objects that might be encountered on a road environment in real life scenario. In total there are 110 entries or rows in the dataset. The dataset is just used for proof of concept and helps to provide a framework for the real time scenario. Like in real life there could a wide variety of objects that is encountered like wildlife, humans, vehicles, baby strollers, etc. and all this could have various subcategories and some could be mixed categories. Like humans could be of different gender, age, height, weight etc. Animals could be wild or pet and also could be of different sizes. Similarly, vehicle too can be of different makes, sizes like it could be a bus, or an SUV or just a sedan etc. Now, each of these can have some mixed categories meaning to say that a vehicle could have a human (or humans) in it and so on.

So, for the sake of simplicity we have taken just the single categories and disregarded the mixed categories. Also, we have not included the category of human/animals inside our own autonomous car. This could be a future work of our proposed work. Hence, our proposed work will not have ethics relating to one's own self or self plus passenger scenarios.

Our approach just takes into account a very small world comprising of only objects defined in the dataset however, the features detected by the car's sensors might differ and that would not be an issue given the boom of machine learning technology.

Now, here is a breakdown of all the columns and little something about each one of them.

- **isAnimate:** An object can be animate or inanimate. This is the major classification criteria in our proposed work. So, for animate objects the value of the entry will be true or false. For example: animals and humans come into the animate category hence, will have a value ‘true’ under this column. And vehicles are an inanimate object so a ‘false’ value will be entered under the column.

- isHuman: This feature is again similar to the above mentioned column. And is basically used to classify humans from animals. So humans will be having an entry of ‘true’ under this feature and an animal would have a ‘false’ value. Also, the inanimate objects would also have a ‘false’ entry under this column.
- Gender: Now, ethics is subjective which means it differs from person to person and might depend on several other factors and reasons. So, some may prefer to protect females while others may give higher priority to males. Therefore, to capture all scenarios and also road environment we have defined this feature to classify between males and females. Now, animals and inanimate objects will have null value assigned to them under this column. As however accurate the sensors of the autonomous vehicles be, even though they were able to predict the gender of an animal but consumers do not have any preference to gender of animals. Hence, the null value is justified.
- Age: Another important factor when it comes to ethics is the age. As in our previous example of the girl and the old lady, who would be killed and who would be saved could depend on the age of the person according to a person’s ethics so for that reason this feature was a must. It will basically have numeric values with its unit in years. Again, for animals and inanimate objects this will be empty or 0.
- Height: This feature represents the height of the corresponding object and will help classify the objects. The unit of this feature is in centimeters and we have collected the height of different ages of people from [42,43] and the height of different kinds of animals from different sources available on the internet. The height of different vehicles was collected from [44,45]. Also, the height can classify the vehicle into a small vehicle like that of sedan or into heavy vehicles category like those of trucks and buses etc. The height can further classify animals into small, medium or large sized animals and since the collision with a large animal might not kill the animal so a person’s ethics can be dependent on that fact. For example: the collision with a heavy vehicle might diminish the chances of survival of the passengers of the AV

so again can play a role in deciding the ethics of the person and thus is an important feature to incorporate in the dataset.

- Weight: Again, weight is similar to height and the data was collected from the same sources as mentioned above. This feature too will have numeric values and the unit of measurement was kilograms.
- isWild: This feature is selected basically to classify animals into the category of wild or pet animals. Sometimes, a person might give more importance to a small pet dog to that of a small wild squirrel. Hence, this feature too was important to capture the ethics of the person. It will have basically two values similar to the features of isHuman and isAnimate. It will have a value ‘true’ if the animal is wild and ‘false’ if the identified animal belongs to the pet category.
- isVehicle: Exactly similar to the category isWild and will have two values. Now, there could be a lot of inanimate objects in a road environment in a real life scenario like there could be a vehicle or a baby stroller or just a small sign board or several other things. However, in our approach we have narrowed down to only one category of inanimate objects and that being vehicles. And it will have a value true if it detects a vehicle else false for other objects.
- Category: This column represents the class label for the various objects. The objects will be finally classified into these categories or classes. The main classes/categories in our approach is enumerated as follows:
  - (a) Child
  - (b) Teen
  - (c) Adult
  - (d) Senior
  - (e) Sedan
  - (f) Heavy vehicle

- (g) Small
- (h) Medium and
- (i) Large

Categories (a) to (d) is classifying the humans based on their various features. Categories (e) and (f) are for inanimate objects and in our case just for vehicle and finally categories (g) to (i) is for classifying the animals based on their sizes and other features defined.

- Category Name: This a column that assigns unique names relative to the categories to each data entry and this column will be basically used in the implementation part and not the experimentation phase. We shall talk about this column in detail in the implementation part in chapter 5.

Now, having defined the dataset in detail our next step in the proposed work was to apply the classification algorithms on this dataset and check the accuracy with which they are correctly classified into their respective categories. So, for that purpose we chose Naïve Bayes' classification, Decision tree classification (both c4.5 and c5.0) and random forest classification. In the next section, we will provide a brief overview of all the algorithms mentioned above.

#### 4.2.2 The Classification Algorithms

The classification algorithms used in the experimentation part are all supervised algorithms of machine learning. We chose different algorithms to find the one with better performance and also that provides the best fit solution.

##### 4.2.2.1 *Naïve Bayes' Classification*

This is one of the most efficient and effective supervised learning algorithms in machine learning which is based on Bayes' theorem. It has an underlying assumption that all pairs of features are independent of each other. The naïve Bayes' classifier has shown that its

performance is comparable to that of decision trees and neural networks in some domains. And it has been widely used for text classification. [46]

Bayes' theorem: (Thomas Bayes, 1702-1761)

If there exists a hypothesis  $h$ , which is supported by an evidence  $e$ , then

$$P(h|e) = \frac{P(e|h) P(h)}{P(e)}$$

where,

$P(h|e)$ : posterior probability in which probability  $h$  is true if there exists evidence  $e$

$P(h)$ ,  $P(e|h)$  and  $P(e)$ : prior probabilities which is initial knowledge

$P(h)$ : probability that  $h$  is true

$P(e|h)$ : probability of observing evidence  $e$  given that  $h$  is true

$P(e)$ : probability that  $e$  is true

Hence, Bayesian classification was to calculate  $P(h|e)$ .

The advantages of Naïve Bayes' classification are that it is very easy to implement and if the Naïve Bayes' assumption of independence hold true then it converges much faster than any other algorithm and thus requires lesser training data. So its quick and easy and thus was our first choice for classification owing to the fact that if any new object was detected then could classify it faster and thus yield results faster. However, it is not able to learn the interaction between different features. [47]

#### *4.2.2.2 Decision Tree Classification*

Decision tree classification is also a supervised learning method used for classification and regression and the goal is to develop such a model which is capable of predicting the class of a variable by learning simple decision rules which it has inferred from the data features.

The tree has basically three types of node namely:

- The root node: which doesn't have any incoming edges but might have zero or more outgoing edges.

- Internal nodes: each internal node has exactly one incoming edge but two or more outgoing edges.
- Leaf or terminal nodes: this node doesn't have any outgoing edge but only one incoming edge.

In a decision tree, the leaf nodes are assigned a class label and the non-terminal nodes are said to contain the attribute text conditions which help to differentiate records having different characteristics. The non-terminal nodes comprise of the root and the internal nodes. [48]

Once a decision tree is constructed, it is fairly easy to classify an object into its respective class. We start from the root node, we apply the test condition of the record and follow the appropriate branch based on the result of the test conditions and repeat the procedure until a terminal node is reached. The class label of the terminal node is the assigned class.

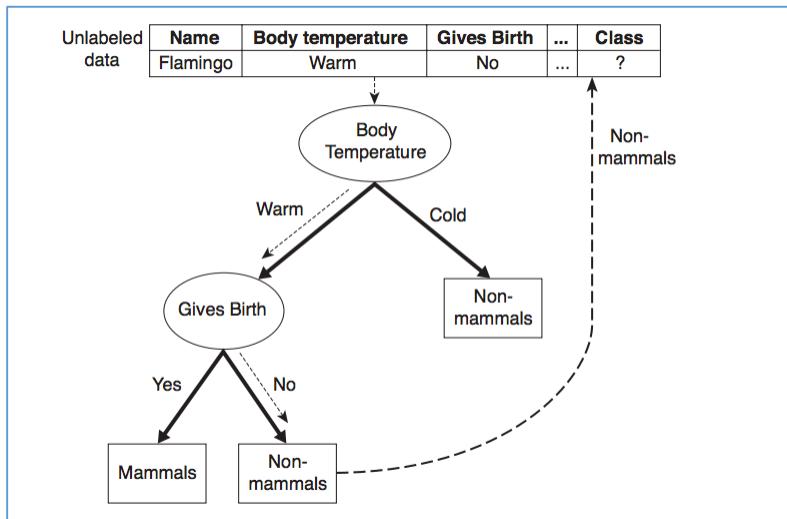


Figure 12: Classifying an unlabeled vertebrate. [48]

Some of the advantages of decision trees are that they are easy to interpret and visualize. Also, they have the capability to easily handle feature interactions and are non-parametric which means that one need not worry about outliers. However, they have a disadvantage that they easily over-fit.

In decision tree, we have used two different algorithms namely C4.5 and C5.0 both of which were developed by Ross Quinlan. C5.0 is a successor of C4.5 and offers certain improvements over C4.5 some of which are enlisted as follows:

- Faster as compared to C4.5
- Has higher accuracy thus lower error rates
- Uses less memory as compared to C4.5
- Has smaller tree sizes and faster computation times
- Has new features such as those of calculating the variable misclassification costs.

[49]

#### *4.2.2.3 Random Forest Classification*

A random forest classification technique is classification method that constructs a large number of decision trees during the training period and outputs the class label by taking the mode of classes (in classification) or the mean prediction of the individual trees. [50]

The decision tree works in the following ways:

Each tree is constructed and grown as follows:

- 1) Suppose there are N number of cases in training sample, then a sample of the N cases is taken randomly but with replacement and is treated as the training set for growing the tree.
- 2) Now, supposing that there are M input variable in total then a number m which is very less than M is specified so that at each node, m variables are randomly selected out of M. The best split on the m is used to split the node. However, the value of m remains constant while the forest grows.
- 3) Each tree is grown to the largest and there is no pruning.
- 4) By aggregating the predictions of the trees in the forest, the new data is predicted.

The advantages of using random forest over decision tree is that it lowers the risk of overfitting by averaging several trees. They are among the most accurate algorithm and run efficiently on large databases.

However, they tend to over-fit for certain datasets that have noisy classification tasks and are difficult to visualize and interpret or understand.

#### *4.2.2.4 Results of experimentation phase*

The algorithms were tested on the dataset using the data mining tool called Weka. However, C5.0 was applied using the RStudio. And the results of the experimentation phase are discussed in detail in chapter 6 of this thesis.

### 4.2.3 The Priority Generation Phase

#### *4.2.3.1 Phase 1*

After the experimentation phase, we will generate the priority of the consumer/owner of the AV. For that purpose, we will read the dataset and generate a general question which would give two options to the user. The options would be the features enlisted for a particular object which will be generated from one single data entry in the dataset.

To understand it better, let's look at a small example.

isAnimate	isHuman	Gender	Age	Height	Weight	isWild	isVehicle	Category	Category name
true	true	female	9	137.72	28.1	false	false	child	child1
true	true	female	24	163.2	59.0	false	false	adult	adult1

Table 6: An example of the dataset

So the above table shows two random rows of our dataset. And now from this dataset we generate a question seen in Figure 13.

Which amongst the two choices is your priority?

CHOICE 1:

Category: child

Category Attributes:

Is the object animate? true

Is the object a human being? true

What's the gender? female

What's the age? 9

What's the height? 137.72

What's the weight? 28.1

Is it a wild animal? false

Is it a vehicle? false

CHOICE 2:

Category: adult

Category Attributes:

Is the object animate? true

Is the object a human being? true

What's the gender? female

What's the age? 24

What's the height? 163.2

What's the weight? 59.0

Is it a wild animal? false

Is it a vehicle? false

Press 1 for choice 1 or Press 2 for choice 2.

Figure 13: An example of question with two scenarios

Now, as can be seen from the question above the options to the questions are actually the rows of the dataset. So, similarly all questions will have two choices out of which the user has to select only one.

Now, the options are generated taking two rows at a time.

So, how is priority generated? Well, we have used a very simple concept to generate the user defined priority list. We have used the concept of selection sort where the entire list is divided into two halves the sorted part and the unsorted part. So, at the beginning of this techniques the entire list is assumed to be unsorted. Now, the first element of the list is taken as minimum and is compared with the other numbers in the list in order to find a

number that is lesser than the minimum. If found, it is swapped with the location of the minimum and the then the second number is the list is marked minimum and the procedure is repeated till the list is sorted.

Attached below is an image illustrating the above mentioned algorithm:

```
Step 1 – Set MIN to location 0  
Step 2 – Search the minimum element in the list  
Step 3 – Swap with value at location MIN  
Step 4 – Increment MIN to point to next element  
Step 5 – Repeat until list is sorted
```

Figure 14: The selection sorting algorithm [55]

In the same way, the priorities are generated. As the user enters his choices, his first choice becomes his first priority. Now, the questions or rather the options are generated in such a way that each row gets compared to every other row. So, now whichever was the priority of the user that particular row will get compared to every other row in the dataset till we get the priority list.

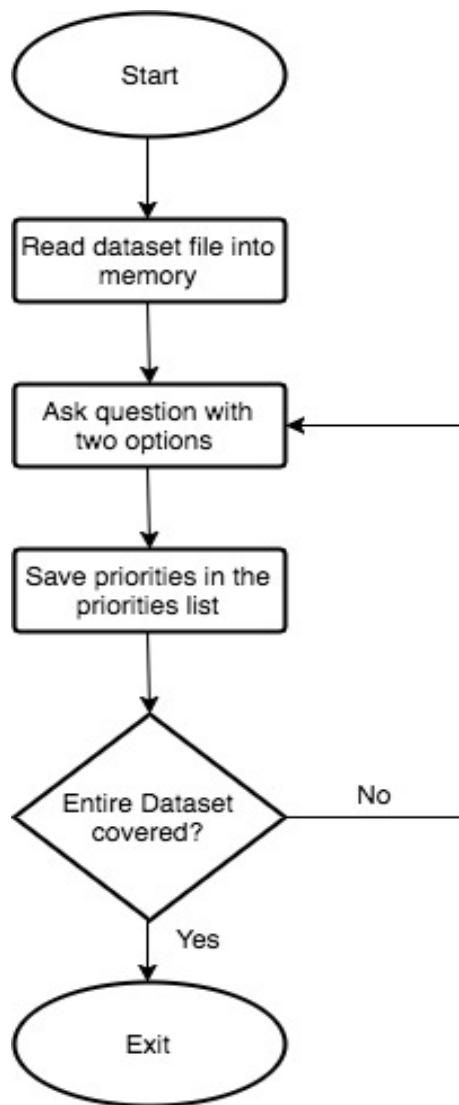


Figure 15: The flowchart of question generation phase

However, since the dataset defined is too big so the user might not want to answer all the questions at one time. So, we have provided a solution for that too. We have defined a default priority list in the AV which is the priority list which will be decided and coded by the programmers, manufacturers, ethicists etc. and is regulated by the government. So, if the user wants to drive an AV without answering all the questions to capture his entire priority list then he can drive with his partial priority list.

What happens in case of an accident? Which priority list will be followed? The user's or the default?

The simple answer to this question would be that when the user decides to partially build his priority list and yet be able to drive his AV, then in case of a crash when the sensors identify the objects they will try to classify those objects into the categories defined.

Next, it will be checked whether all the objects detected have a priority listing in the priority list of the user. If all the identified and classified objects are found in the priority list of the user then, in that case the user's priority/ethics will be followed. However, if one or none of the objects that were identified and classified appear in the user's priority listing then the default priority list will be followed.

Let us understand this with an example: Suppose the user decides to partially define his priority list. He is riding in his car and somehow the following scenario come up; the car either swerves to the right and kills a small boy of age 5 or it swerves to the left and kills a teen girl of age 13. The car was able to classify the objects into the categories of teen and child based on their features and looks up the user's priority list if both the objects have a priority in his list. Suppose both, the teen and the child categories have a priority in the user defined priority list. And in the priority list, he/she has given more priority to the teen than the child. Then, the car would follow the user's ethics and swerve to the right. However, if the child is in the user's priority list but the teen has not been defined a priority. Then, the car would follow the default priority list of saving the child because according to the manufacturer and other stakeholders who developed the ethics the teen might be able to sense the accident and might try to save himself.

One more thing that is important here is who would be responsible then for the crash in both the scenarios?

When the car follows the ethics of the user then it is the user who would be morally and legally responsible for the crash. But when the car follows the default priority list then the responsibility is equally shared by the OEM and the other stake holders.

#### *4.2.3.2 Phase 2 (Decision Taking phase)*

In this phase, we are going to use the C4.5 of the decision tree algorithm for classification of the objects that we might encounter in real life. First the object is detected by the sensors and then the features of the object is passed to the classification algorithm which checks its features and classifies it into the respective category and assigns a category name. For example, all members of the child category will have a different name and a different priority. For example, a child of age 5 is named as child5 and will have a different priority to a child who is of age 12 and has a name child12. So similarly, based on different features of each child, they will have different category names. And when an accident scenario is encompassed, then the AV will look up the priorities of the different objects encountered. The priority list (user's priority list or default priority list) containing all the detected objects will be used for looking up the priorities. Finally, the object with least priority would suffer the maximum damage, that is the car would save the object with higher priority and steer towards the object with least priority. Figure 16 shows the flowchart depicting decision taking phase for customized and default ethics.

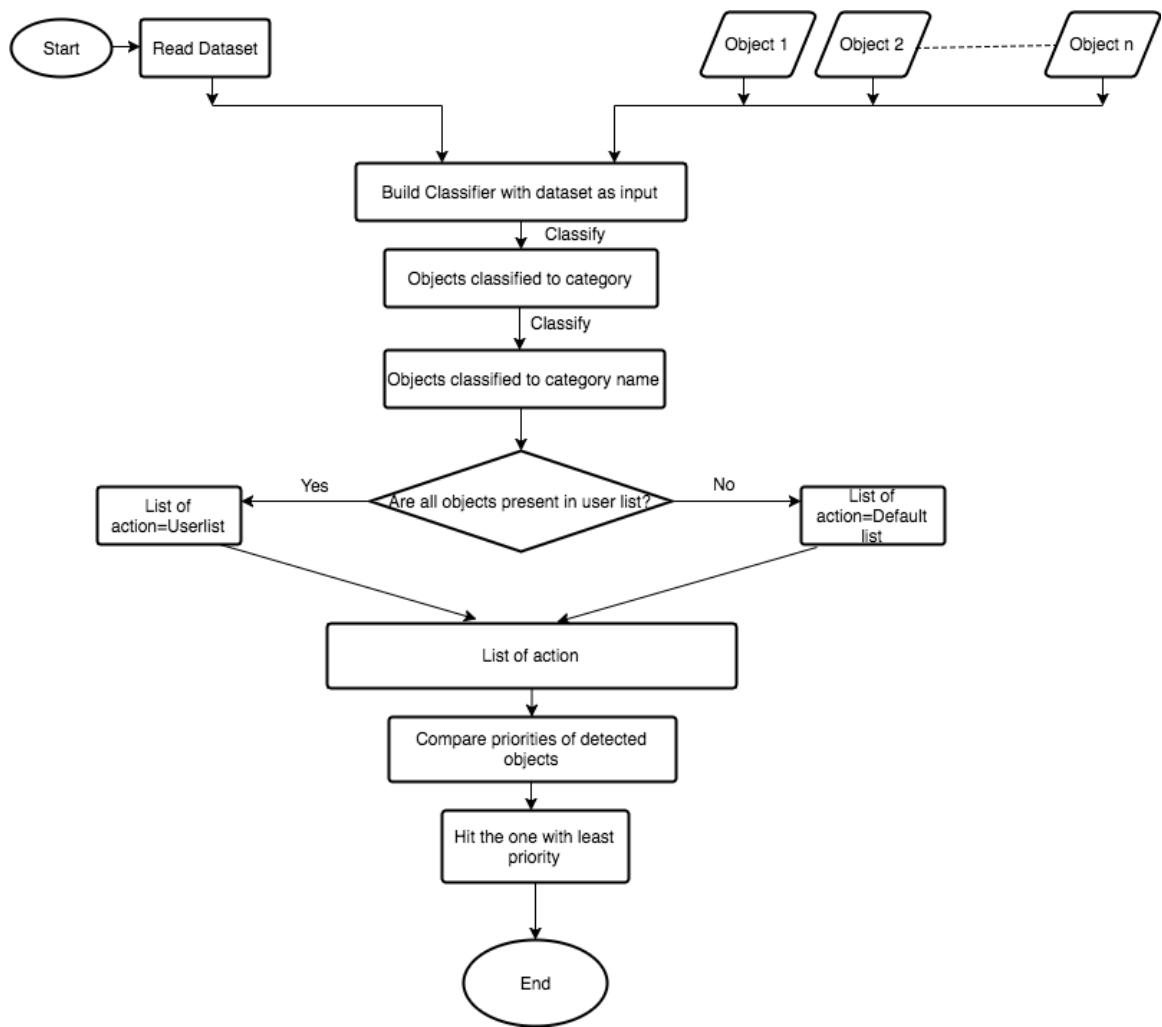


Figure 16: The flowchart of the decision making phase

## **CHAPTER 5        IMPLEMENTATION**

### **5.1 Development Environment and Libraries used**

The proposed methodology has been implemented using Java as programming language on mac OS Sierra operating system. Table 7 shows the specifics of the development environment used for implementing the proposed work.

Programming Language	Java 1.8.0_131-b11
Operating System	Mac OS Sierra 10.12.5
Eclipse IDE version	Eclipse 4.6
Libraries Used	Weka-3-8-0 monolithic.jar

Table 7: Development environment and libraries used

### **5.2 Implementation details of each phase in the proposed approach**

#### **5.2.1 Generating user defined priority list/ethics**

The first step of our proposed approach is to generate the user defined priorities by asking questions. At a time, a question with two options (as discussed in chapter 4) will be generated from the training data set (as shown in Table 6 and Figure 13). The following snippet of the code is representative of how the above proposed work is done.

```

22@ public void readfile()
23{
24    String line = "";
25    String splitBy = ",";
26    try
27    {
28        BufferedReader br = new BufferedReader(new FileReader(file));
29        int i = 0;
30        while ((line = br.readLine()) != null)
31        {
32            if(i == 0)
33            {
34                i++;
35                continue;
36            }
37            String[] catName = line.split(splitBy);
38            createList(catName);
39        }
40        br.close();
41    }
42    catch (Exception ex)
43    {
44        ex.printStackTrace();
45    }
}

```

Figure 17: Reading the csv file and adding each row to an arraylist

So, the Figure 17 shows how the .csv file is read which contains our training data and the method `createList()` is used for adding each row, which is read from the csv file to an `ArrayList<options>`.

```
17     ArrayList<options> objLis = new ArrayList<options>();
```

Figure 18: ArrayList declaration

From Figure 18, it can be seen that the `ArrayList` can only contain the objects of the type `options`. The class `options.java` is used to store the data which has been read from the .csv file. This was done for the purpose of making it easier to read any specific attributes for the object (each row in our case).

Once we have all the data from the file and the data has been stored as per our needs then, the next step is to prioritize the data entries by asking questions from the user. For this purpose, we defined the method `sortList()`.

```

133  public void sortList()
134  {
135      int p_index;
136      for(int i = 0; i < objLis.size() - 1; i++)
137      {
138          p_index = i;
139          for(int j = i+1; j < objLis.size(); j++)
140          {
141              //if first parameter has higher priority than second, then the following if condition is true
142              if(compareOptions(objLis.get(j), objLis.get(p_index)))
143              {
144                  options tempObj = objLis.get(p_index);
145                  objLis.set(p_index, objLis.get(j));
146                  objLis.set(j, tempObj);
147              }
148          }
149      }
150  }

```

Figure 19: Capturing priority based on user's choice

As can be seen from the Figure 19, the method takes two objects from the `ArrayList` at a time and presents it to the user in the form of multiple choices/options to the question as can be seen in Figure 13. And based on the user's selection of the option/choice, it then prioritizes the object chosen by the user. And this process continues until all the objects in the `ArrayList` have been prioritized. The questions presented to the users can be seen in Figure 13.

Once all the above mentioned procedures have been completed the user defined priorities are activated and the autonomous vehicle is ready for driving with the customized ethics of the user. The outputs can be seen in the following Figures 20 and 21.

```

printing priority before answering questions
adult2 ----> child5 ----> teen3 ----> senior1

```

Figure 20: Default priority list of the AV

```

Priorities decided. Your car is ready for autonomous driving.

printing priority after answering questions
adult2 ----> teen3 ----> child5 ----> senior1

```

Figure 21: User defined priority list generated

Here, as can be seen we have considered only four categories with category names as `child5`, `teen3`, `senior1` and `adult2`. Although, our dataset contains a lot of entries but we have temporarily truncated the size of our dataset to consider only four categories. This is done to make our Figures 20 and 21 more human readable and more relatable to the

concept. However, in real life scenario the default as well as the user defined priority list will include a variety of different categories and their respective priorities in the corresponding lists. Also, as is apparent from the Figures, the user gives more priority to the adult2 than the teen3 or child5 hence the default priority is modified into the user defined priority where the adult2 is of highest priority. Now, in a situation where the AV would have to choose to save one and hit the other, and the objects detected had the features of child5 and teen3 then, according to default priority list, the AV will save the child and hit the teen. However, when the user has answered all the questions then the teen would be saved and the child would be hit.

### 5.2.2 Classification of the detected objects

Now, after the completion of the previous step where we generate the ethics or the user defined priority list; we need to classify the detected objects based on the user priorities. So, for this purpose (i.e. classification) we have used decision tree C4.5 algorithm which is provided by the Weka API (weka-3-8-0-monolithic.jar). The full dataset used in the previous step for priority generation is taken as the training dataset for the classifier. So the classification basically is divided into two steps. In step 1, the classifier assigns a category to the detected object based on its attributes in the training data. The next step is to assign a category name to the detected object after its category is known. So it is like we identify the category of the detected object and then we identify the category name of the same. If the attributes of an object are such that the classifier is able to classify it into the teen category then in step 2, the classifier again uses the training data to classify the object into its respective category name so say suppose it further classifies the teen into the teen1. Step 2 is needed in order to ensure that we have a category name at the end of the classification process since we are prioritizing the category name and not the category in the priority generation phase.

We have done this step because again, user might have different priorities/ethics for different ages or genders of teens so categorizing into the broad category of only teens or child might not serve the purpose of customized ethics and would have created confusion and ambiguity. For example: If I were to drive an AV then at all times I would want to save

a child who is 4 years old compared to a 12-year-old child because as per my moral values and logic a 12-year-old might sense the danger much faster than a 4-year-old and might take precautions to save himself/herself. So, that is the reason for assigning category names to different members of the same category.

```

46         DataSource dSource = new DataSource(dataSourceFileName);
47         Instances data = dSource.getDataSet();
48
49         if (data.classIndex() == -1)
50         {
51             data.setClassIndex(data.numAttributes() - 1);
52         }
53
54         J48 tTree = new J48();
55
56         String[] options = new String[1];
57         options[0] = "-U";
58         tTree.setOptions(options);
59
60         tTree.buildClassifier(data);

```

Figure 22: Code snippet for building the classifier

The code for building the classifier is shown in Figure 20. And Figure 21 shows the process of classifying the object into a category.

```

64         //read from file to classify
65         DataSource dSource1 = new DataSource(str);
66         Instances data1 = dSource1.getDataSet();
67
68         if (data1.classIndex() == -1)
69         {
70             data1.setClassIndex(data1.numAttributes() - 1);
71         }
72
73         String inst = null;
74
75         for(int i = 0; i < data1.numInstances(); i++)
76         {
77             Instance ins = data1.instance(i);
78
79             double clsLabel = tTree.classifyInstance(ins);
80
81             data1.instance(i).setClassValue(clsLabel);
82
83             inst = data1.instance(i).toString(8);
84             List<String> lis = new ArrayList<String>();
85             populateListCatName obj = new populateListCatName();
86             lis = obj.populate(lis, inst);
87             Attribute atr = new Attribute("CategoryName", lis);
88             data1.insertAttributeAt(atr, 9);
89
90             System.out.println("Detected Category is: " + inst);
91             System.out.println("Intermediate classification completed.\n");
92         }

```

Figure 23: Classifying into a category

After this step, the next step is to get the Category Name which is the last field in our dataset. For this purpose, we create a new dataset from the original dataset such that the new dataset only contains the entries whose category matches the detected category. For

example, if the detected category is of teen then all entries in the original dataset which has its final category as teen are in the new dataset. Now, using this dataset (for example dataset containing only teen category entries) we build a new classifier and use it to classify the object and arrive at a category name. So, at the end of program execution we will have a category name for all the objects detected by the sensors of the AV.

```

64          //read from file to classify
65          DataSource dSource1 = new DataSource(str);
66          Instances data1 = dSource1.getDataSet();
67
68          if (data1.classIndex() == -1)
69          {
70              data1.setClassIndex(data1.numAttributes() - 1);
71          }
72
73          String inst = null;
74
75          for(int i = 0; i < data1.numInstances(); i++)
76          {
77              Instance ins = data1.instance(i);
78
79              double clsLabel = tTree.classifyInstance(ins);
80
81              data1.instance(i).setClassValue(clsLabel);
82
83              inst = data1.instance(i).toString(8);
84              List<String> lis = new ArrayList<String>();
85              populateLisCatName obj = new populateLisCatName();
86              lis = obj.populate(lis, inst);
87              Attribute atr = new Attribute("CategoryName", lis);
88              data1.insertAttribute(atr, 9);
89
90              System.out.println("Detected Category is: " + inst);
91              System.out.println("Intermediate classification completed.\n");
92          }

```

Figure 24: Code for identification of category and category name for detected objects

The results for the program execution can be seen in Figure 24 where we see category and category being identified for each object.

```
*****START PHASE 1*****
CLASSIFICATION 1 --
Detected Category is: senior
Intermediate classification completed.

Detected Subcategory is: senior1
Final classification completed.

CLASSIFICATION 2 --
Detected Category is: child
Intermediate classification completed.

Detected Subcategory is: child5
Final classification completed.

CLASSIFICATION 3 --
Detected Category is: adult
Intermediate classification completed.

Detected Subcategory is: adult2
Final classification completed.

Done with classification of detected objects.
*****END OF PHASE 1*****
```

Figure 25: Output showing category and category name for detected object

### 5.2.3 Decision making process based on the user defined priorities

The final step of implementation is the decision making, i.e. deciding on which object to save and which object to hit based on the priorities of the detected object. We have already classified the objects into their respective category and category name in the previous step. Now, we make use of the priority list to check the priorities of the detected objects and take the corresponding action. Figure 25 shows the code snippet for identifying the object with least priority.

```

34  public String result(List<String> inObj)
35  {
36      readExcelFile();
37
38      String toHit = null;
39      HashMap<Integer, String> tempMap = new HashMap<Integer, String>();
40
41      int maxVal = -1;
42
43      for(String str : inObj)
44      {
45          if(getList.contains(str))
46          {
47              int priority = getList.indexOf(str);
48              tempMap.put(priority, str);
49          }
50      }
51
52      for(int val : tempMap.keySet())
53      {
54          //maxVal = val;
55          if(val > maxVal)
56          {
57              maxVal = val;
58          }
59      }
60
61      if(maxVal != -1)
62      {
63          toHit = tempMap.get(maxVal);
64      }
65      return toHit;
66  }

```

Figure 26: Code snippet showing how object with least priority is identified

Now referring to Figure 21 and 25 we know the user-defined priority list and the objects detected by the sensors of the AV. So now the final output in Figure 27 shows that the object with least priority out of the detected objects is eliminated and as can be seen from Figure 18 senior1 has priority lesser than that of adult2 and child5. So, we can say that the action taken is in accordance with the user's ethics.

```

*****START - DECISION MAKING*****
Checking priority list, making decision.
Car will hit: senior1
*****END - DECISION MAKING*****

```

Figure 27: The final outcome

## **CHAPTER 6      EXPERIMENTAL RESULTS AND ANALYSIS**

In this chapter, we have discussed the tools used for the experimentation phase of the proposed work which is basically to check a few classification algorithms and using the one with the best results. We have used Weka and R Studio tools for conducting the experimentation phase. We will discuss about them briefly in the following sections.

### **6.1 Tools used**

#### **6.1.1 Weka**

The tool Weka gets its name from the bird called Weka, which is only found in New Zealand and is known to have a very inquisitive nature. Weka is an acronym for Waikato Environment for Knowledge Analysis(Weka) [51] and is a collection of machine learning algorithms used for the processes of data mining. Weka was developed at the University of Waikato, New Zealand and is an open source software licensed under the GNU General Public License.

The algorithms can be either called from the Java code or can be applied directly to one's dataset. We have applied it directly to our dataset in the experimentation phase and then used the algorithm with the best results in our Java code using its libraries. We have basically used Weka for classification however, it can be used for regression, data preprocessing, clustering, association rules and for visualization purposes. It can even be used for developing new machine learning schemes. [52,53]

#### **6.1.2 RStudio**

RStudio is a free open source IDE for a programming language called R which is used for statistical computing and graphics. RStudio was founded in 2008 and the initial release was in 2011. RStudio makes analysis of data very easy and is available in two editions namely: RStudio Desktop and RStudio Server. RStudio for its GUI uses the Qt framework and is written in C++ language. [54]

## 6.2 Results of experimentation phase using Weka

In this section, we will provide the screenshots of the results of applying the classification algorithms of Naïve Bayes', C4.5, C5.0 and Random Forest on our dataset.

We have divided our dataset randomly in the ratio of 60-40, 70-30 and 80-20. Diving the data randomly in 60-40 ratio means, we have used 60 percent of data for training the classifier and the rest 40 percent of the data is used for testing the results. Attached below is a screenshot of our dataset. It is not the complete dataset but just a part of it to show how the dataset looks like.

	A	B	C	D	E	F	G	H	I	J
1	Object	isAnimate	isHuman	Gender	Age	Height	Weight	isWild	isVehicle	Categories
2	1	TRUE	TRUE	Female	2	85.5	12	FALSE	FALSE	child
3	2	TRUE	TRUE	Female	4	100.3	15.4	FALSE	FALSE	child
4	3	TRUE	TRUE	female	6	115.5	19.9	FALSE	FALSE	child
5	4	TRUE	TRUE	female	8	132.45	25.8	FALSE	FALSE	child
6	5	TRUE	TRUE	female	9	137.72	28.1	FALSE	FALSE	child
7	6	TRUE	TRUE	female	10	143.26	31.9	FALSE	FALSE	child
8	7	TRUE	TRUE	female	11	149.41	36.9	FALSE	FALSE	child
9	8	TRUE	TRUE	female	12	154.96	41.5	FALSE	FALSE	child
10	9	TRUE	TRUE	female	13	156.7	45.8	FALSE	FALSE	teen
11	10	TRUE	TRUE	female	14	158.7	47.6	FALSE	FALSE	teen
12	11	TRUE	TRUE	female	15	159.7	52.1	FALSE	FALSE	teen
13	12	TRUE	TRUE	female	16	162.5	53.5	FALSE	FALSE	teen
14	13	TRUE	TRUE	female	19	163.28	57.1	FALSE	FALSE	teen
15	14	TRUE	TRUE	female	20	163.1	58	FALSE	FALSE	adult
16	15	TRUE	TRUE	female	24	163.2	59	FALSE	FALSE	adult
17	16	TRUE	TRUE	female	26	163.2	60	FALSE	FALSE	adult
18	17	TRUE	TRUE	female	36	163.3	65	FALSE	FALSE	adult
19	18	TRUE	TRUE	female	38	163.4	64	FALSE	FALSE	adult
20	19	TRUE	TRUE	female	42	163.1	70	FALSE	FALSE	adult
21	20	TRUE	TRUE	female	46	163	73.4	FALSE	FALSE	adult
22	21	TRUE	TRUE	female	53	162.3	77.1	FALSE	FALSE	adult
23	22	TRUE	TRUE	female	59	163.4	75.6	FALSE	FALSE	adult
24	23	TRUE	TRUE	female	64	161.6	77.4	FALSE	FALSE	senior
25	24	TRUE	TRUE	female	68	162.2	76.6	FALSE	FALSE	senior
26	25	TRUE	TRUE	female	70	159.1	74.8	FALSE	FALSE	senior
27	26	TRUE	TRUE	female	75	160.1	70.5	FALSE	FALSE	senior
28	27	TRUE	TRUE	female	80	155.5	64.9	FALSE	FALSE	senior
29	28	TRUE	TRUE	female	85	159.5	65.9	FALSE	FALSE	senior
30	29	TRUE	TRUE	male	1	86.8	10.5	FALSE	FALSE	child
31	30	TRUE	TRUE	male	3	102.3	13.5	FALSE	FALSE	child
32	31	TRUE	TRUE	male	4	103.5	16.3	FALSE	FALSE	child
33	32	TRUE	TRUE	male	6	115.5	20.6	FALSE	FALSE	child

Figure 28: Screenshot of part of the dataset

Now, we will apply Naïve Bayes' algorithm using Weka on our dataset and look at the accuracy and the confusion matrix.

### 6.2.1 Naïve Bayes' algorithm application: Results and Discussion

#### 60-40 ratio

We first will provide results for applying Naïve Bayes' on our dataset when it is divided randomly into 60-40 ratio.

```
Time taken to build model: 0 seconds

== Evaluation on test split ==

Time taken to test model on test split: 0.01 seconds

== Summary ==

Correctly Classified Instances      37           84.0909 %
Incorrectly Classified Instances   7            15.9091 %
Kappa statistic                   0.8204
Mean absolute error               0.0498
Root mean squared error          0.178
Relative absolute error          25.3066 %
Root relative squared error     56.2925 %
Total Number of Instances        44
```

Figure 29: Accuracy of Naïve Bayes' on 60-40 dataset

```
== Confusion Matrix ==

a b c d e f g h i    <-- classified as
6 1 0 0 0 0 0 0 0 | a = child
0 4 1 0 0 0 0 0 0 | b = teen
0 0 6 0 0 0 0 0 0 | c = adult
0 0 0 4 0 0 0 0 0 | d = senior
0 0 0 0 5 0 0 0 0 | e = sedan
0 0 0 0 0 5 0 0 0 | f = Heavy vehicle
0 0 0 0 0 0 2 0 0 | g = small
0 0 0 0 0 0 1 3 0 | h = medium
0 0 0 0 0 0 0 4 2 | i = large
```

Figure 30: Confusion matrix for Naïve Bayes' on 60-40 dataset

Now, classification accuracy is basically the ratio of correct predictions to total predictions. As can be seen from the above results there are total 37 correct prediction and total number

of data points to be classified was 44(due to 60-40 ratio of training and testing data) so that expressed as percentage is 84.09%.

Now, a confusion matrix is used to measure the performance of a classification algorithm. It shows the way in which the classification algorithm is confused when it is making the predictions. The Confusion Matrix assigns letters from a-i to class labels. It provides the expected class values in rows, and the columns represent the predicted class values. Like in the above screenshot we see that our algorithm was confused in row a (which stands for child) and it misclassified an instance as a teen that is in column b.

#### *70-30 ratio*

Following the same convention of accuracy and confusion matrix as defined in the above section, we apply the Naïve Bayes' classifier to our dataset and this time we divide the dataset in 70-30 ratio where in 70% of random data points will be used for training and the rest for testing. We attach the screenshots of the results for accuracy and confusion matrix and discuss them later.

```
Time taken to build model: 0.01 seconds
===
Evaluation on test split ===
Time taken to test model on test split: 0.01 seconds
===
Summary ===
Correctly Classified Instances           23          69.697 %
Incorrectly Classified Instances        10          30.303 %
Kappa statistic                         0.6541
Mean absolute error                     0.064
Root mean squared error                 0.2205
Relative absolute error                  32.314 %
Root relative squared error            69.4434 %
Total Number of Instances               33
```

Figure 31: Accuracy of Naïve Bayes' on 70-30 dataset

```

== Confusion Matrix ==

a b c d e f g h i    <-- classified as
5 1 0 0 0 0 0 0 0 | a = child
0 3 2 0 0 0 0 0 0 | b = teen
0 0 5 0 0 0 0 0 0 | c = adult
0 0 0 1 0 0 0 0 0 | d = senior
0 0 0 0 3 0 0 0 0 | e = sedan
0 0 0 0 0 4 0 0 0 | f = Heavy vehicle
0 0 0 0 0 0 0 0 0 | g = small
0 0 0 0 0 0 2 1 0 | h = medium
0 0 0 0 0 0 0 5 1 | i = large

```

Figure 32: Confusion Matrix for Naïve Bayes' for 70-30 dataset

#### *80-20 ratio*

As stated above the same convention is being followed however, in this section we divide our dataset in 80-20 ratio for training the classifier and testing respectively.

```

Time taken to build model: 0 seconds

== Evaluation on test split ==

Time taken to test model on test split: 0.01 seconds

== Summary ==

Correctly Classified Instances      15          68.1818 %
Incorrectly Classified Instances   7           31.8182 %
Kappa statistic                   0.6393
Mean absolute error               0.0701
Root mean squared error          0.2294
Relative absolute error          35.5433 %
Root relative squared error     72.6237 %
Total Number of Instances        22

```

Figure 33: Accuracy of Naïve Bayes' on 80-20 dataset

```

== Confusion Matrix ==

a b c d e f g h i    <-- classified as
3 1 0 0 0 0 0 0 0 | a = child
0 2 1 0 0 0 0 0 0 | b = teen
0 0 3 0 0 0 0 0 0 | c = adult
0 0 0 1 0 0 0 0 0 | d = senior
0 0 0 0 2 0 0 0 0 | e = sedan
0 0 0 0 0 3 0 0 0 | f = Heavy vehicle
0 0 0 0 0 0 0 0 0 | g = small
0 0 0 0 0 0 2 0 0 | h = medium
0 0 0 0 0 0 0 3 1 | i = large

```

Figure 34: Confusion Matrix for Naïve Bayes' on 80-20 dataset

#### *Discussion on the results of Naïve Bayes'*

In this section, we will provide a graphical representation of the results obtained from applying Naïve Bayes' algorithm on our dataset.

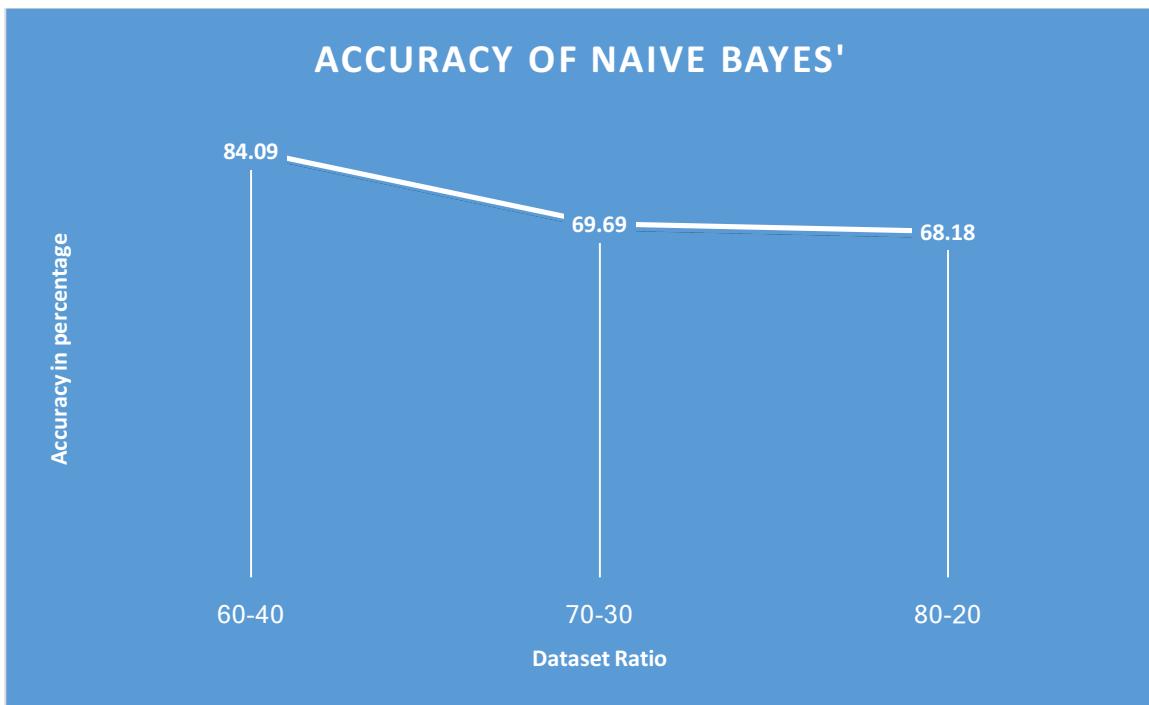


Figure 35: Graph showing the accuracy for different dataset ratio for Naïve Bayes'

As is apparent from the graph above, the accuracy is decreasing with the increase in the training data. Here, the x-axis represents the different dataset ratio and the y-axis is representative of the accuracy in percentages. Now, it is seen that on increasing the training data from 60 percent to 80 percent and decreasing the testing data from 40 percent to 20 percent, there is a decrease in the accuracy with which the classifier predicts the class labels of the test data. Thus, we can say that a large training set decreases the accuracy because with a larger training set it is more difficult for the learning algorithm to learn the model that is capable of correctly representing all the training data.

### 6.2.2 C4.5 algorithm application: Results and Discussion

In this section, we have applied J48 or C4.5 algorithm of decision tree classification on our dataset. Firstly, we have applied it on 60-40 ratio of our dataset which means 60 percent of our data is used for training the algorithm and the rest is used for testing. Similarly, we have applied to 70-30 and 80-20 ratio of our dataset and found the results in terms of accuracy and confusion matrix. The convention for accuracy calculation is same as used in Naïve Bayes' classification and the same convention is followed for the confusion matrix as well.

#### *60-40 ratio*

Here, we have divided the dataset randomly into 60-40 ratio and provide the results for the same in the following Figures.

```
== Classifier model (full training set) ==
J48 pruned tree
-----
animate/inanimate = animate
|   Object <= 69
|   |   Age <= 19
|   |   |   Age <= 12: child (17.0)
|   |   |   Age > 12: teen (10.0)
|   |   Age > 19
|   |   |   Age <= 59: adult (21.0)
|   |   |   Age > 59: senior (10.0)
Object > 69
|   Weight <= 5.5: small (11.0)
|   Weight > 5.5
|   |   Weight <= 20.6: medium (9.0/1.0)
|   |   Weight > 20.6: large (10.0)
animate/inanimate = inanimate
|   Height <= 176.8: sedan (11.0)
|   Height > 176.8: Heavy vehicle (11.0)

Number of Leaves :      9
Size of the tree :     17
```

Figure 36: Visualization of a pruned tree for 60-40 dataset

```

Time taken to build model: 0.01 seconds

== Evaluation on test split ==

Time taken to test model on test split: 0 seconds

== Summary ==

Correctly Classified Instances      39           88.6364 %
Incorrectly Classified Instances    5            11.3636 %
Kappa statistic                   0.8704
Mean absolute error               0.0253
Root mean squared error          0.1589
Relative absolute error          12.8293 %
Root relative squared error     50.2605 %
Total Number of Instances        44

```

Figure 37: Accuracy of C4.5 decision tree algorithm on 60-40 dataset

```

== Confusion Matrix ==

a b c d e f g h i    <-- classified as
7 0 0 0 0 0 0 0 0 | a = child
0 5 0 0 0 0 0 0 0 | b = teen
0 0 6 0 0 0 0 0 0 | c = adult
0 0 0 4 0 0 0 0 0 | d = senior
0 0 0 0 3 2 0 0 0 | e = sedan
0 0 0 0 0 5 0 0 0 | f = Heavy vehicle
0 0 0 0 0 0 0 2 0 | g = small
0 0 0 0 0 0 0 3 1 | h = medium
0 0 0 0 0 0 0 0 6 | i = large

```

Figure 38: Confusion matrix for C4.5 on 60-40 dataset

Here, as we can see the algorithm got confused and misclassifies 2 instances of a sedan as a heavy vehicle, confused 3 instances of small animals to be medium animals and misclassified one medium sized animal as a large sized animal.

### 70-30 ratio

We are following the same convention here as well. And our dataset is divided into 70-30 ratio and the results are attached below.

```
Time taken to build model: 0.01 seconds

== Evaluation on test split ==

Time taken to test model on test split: 0 seconds

== Summary ==

Correctly Classified Instances      31          93.9394 %
Incorrectly Classified Instances   2           6.0606 %
Kappa statistic                   0.9293
Mean absolute error               0.0135
Root mean squared error          0.1161
Relative absolute error          6.8038 %
Root relative squared error     36.5513 %
Total Number of Instances        33
```

Figure 39: Accuracy of C4.5 for 70-30 dataset

```
== Confusion Matrix ==

a b c d e f g h i    <-- classified as
6 0 0 0 0 0 0 0 0 | a = child
0 5 0 0 0 0 0 0 0 | b = teen
0 0 5 0 0 0 0 0 0 | c = adult
0 0 0 1 0 0 0 0 0 | d = senior
0 0 0 0 2 1 0 0 0 | e = sedan
0 0 0 0 0 4 0 0 0 | f = Heavy vehicle
0 0 0 0 0 0 0 0 0 | g = small
0 0 0 0 0 0 0 3 0 | h = medium
0 0 0 0 0 0 0 1 5 | i = large
```

Figure 40: Confusion matrix for C4.5 algorithm on 70-30 dataset

As, we can see from the above confusion matrix, the algorithm is again getting confused and misclassifying a sedan as a heavy vehicle. And it is misclassifying a large animal as a medium sized animal.

### *80-20 ratio*

Here, we divide our dataset randomly so as there are 80 percent training data and the rest is used as testing data. We saw that the J48 pruned tree remains the same so we will skip attaching the Figure for it. However, we will attach the results for accuracy and confusion matrix.

```
Time taken to build model: 0 seconds  
==== Evaluation on test split ====  
Time taken to test model on test split: 0 seconds  
==== Summary ====  


|                                  |          |           |
|----------------------------------|----------|-----------|
| Correctly Classified Instances   | 21       | 95.4545 % |
| Incorrectly Classified Instances | 1        | 4.5455 %  |
| Kappa statistic                  | 0.9474   |           |
| Mean absolute error              | 0.0101   |           |
| Root mean squared error          | 0.1005   |           |
| Relative absolute error          | 5.1214 % |           |
| Root relative squared error      | 31.813 % |           |
| Total Number of Instances        | 22       |           |


```

Figure 41: Accuracy of C4.5 decision tree algorithm for 80-20 dataset

```
==== Confusion Matrix ====  
  
a b c d e f g h i    <-- classified as  
4 0 0 0 0 0 0 0 0 | a = child  
0 3 0 0 0 0 0 0 0 | b = teen  
0 0 3 0 0 0 0 0 0 | c = adult  
0 0 0 1 0 0 0 0 0 | d = senior  
0 0 0 0 2 0 0 0 0 | e = sedan  
0 0 0 0 0 3 0 0 0 | f = Heavy vehicle  
0 0 0 0 0 0 0 0 0 | g = small  
0 0 0 0 0 0 0 2 0 | h = medium  
0 0 0 0 0 0 0 1 3 | i = large
```

Figure 42: Confusion matrix for C4.5 decision tree algorithm on 80-20 dataset

Again, as is apparent from the confusion matrix above, there is only one misclassification wherein a large sized animal is classified as a medium sized animal.

#### *Discussion on the results of C4.5 Decision Tree*

In this section, we provide a graphical representation of the accuracy of C4.5 decision tree algorithm on different dataset ratios.

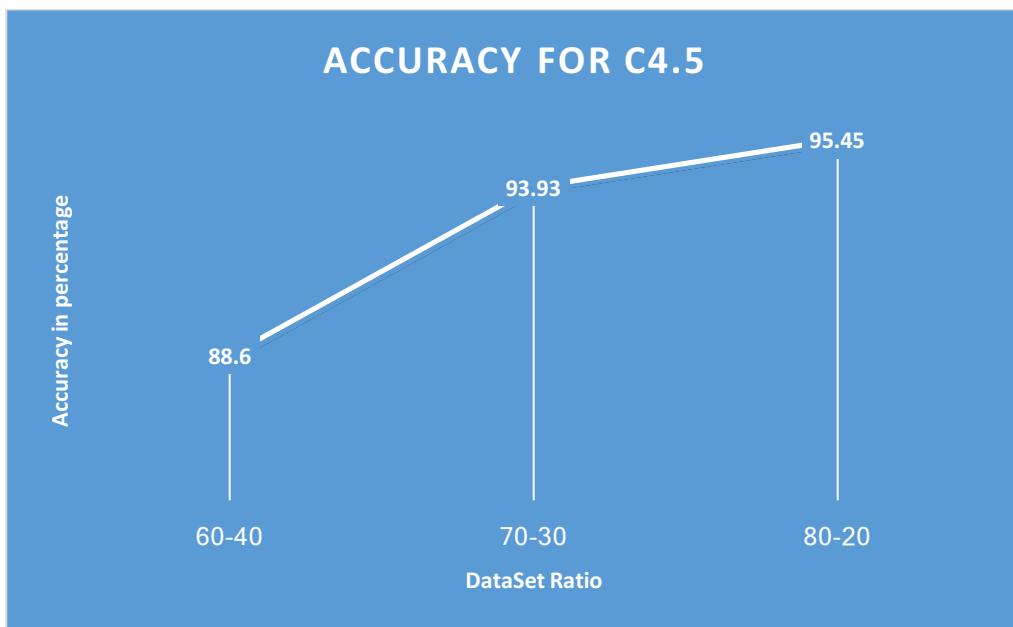


Figure 43: Graph showing the accuracy for different dataset ratio for C4.5

Here, the graph shows the accuracy as a function of the training size. The x-axis represents the Dataset Ratios of 60-40, 70-30 and 80-20 and the y- axis represents the accuracy in percentages and as is apparent from the graph, the accuracy of C4.5 algorithm increases with an increase in the training data. Also, it is just intuitive that when the training data increases from 60 to 80 percent of the total data points then the size of the tree is also growing in size. Hence, with increase in training data the accuracy with which the algorithm classifies the test data increases. A point here to note is that a large size of tree is not always the most accurate, but a small pruned tree can also give the same results. However, weka generates the pruned tree instead of a large tree so the accuracy is maximum at 80-20. Also, we can see that from confusion matrix that with increase in

training data the classifier's confusion between heavy vehicles and sedan was completely removed and that the confusion between medium and large sized animals was reduced to only one instance getting misclassified. So, it is not wrong to say that increasing training data improves the accuracy and reduces the confusion.

### 6.2.3 C5.0 algorithm application: Results and Discussion

In this section, we will provide results in terms of accuracy and confusion matrix obtained by applying the C5.0 algorithm using Rstudio. We have tested the results on different dataset ratios ranging from 60-40 to 80-20 which means we first have divided the data into 60 percent being treated as training data and the rest 40 percent as testing data and the same convention is followed for 70-30 and 80-20 ratios.

#### *60-40 ratio*

As stated above, 60 percent of the entire dataset is used for training and the rest is used for testing. The results are attached below:

Overall Statistics	
Accuracy : 1	

Figure 44: Accuracy of C5.0 decision tree algorithm for 60-40 dataset

pred		true								
		adult	child	Heavy vehicle	large	medium	sedan	senior	small	
adult		4	0		0	0	0	0	0	0
child		0	2		0	0	0	0	0	0
Heavy vehicle		0	0		5	0	0	0	0	0
large		0	0		0	0	0	0	0	0
medium		0	0		0	0	1	0	0	0
sedan		0	0		0	0	0	2	0	0
senior		0	0		0	0	0	0	2	0
small		0	0		0	0	0	0	0	4
teen		0	0		0	0	0	0	0	0

Figure 45: Confusion matrix for C5.0 decision tree algorithm on 60-40 dataset

As can be seen from the accuracy and confusion matrix, the results of C5.0 on 60-40 ratio of dataset was 100 percent with no confusion at all but this might hint towards overfitting.

### *70-30 ratio*

The same convention is followed where we divide the 70 percent of data for training the classifier and the rest for testing it. The results in terms of accuracy and confusion matrix is attached below.

Overall Statistics									
Accuracy : 1									

Figure 46: Accuracy of C5.0 decision tree algorithm for 70-30 dataset

pred	true									
	adult	child	Heavy vehicle	large	medium	sedan	senior	small	teen	
adult	5	0	0	0	0	0	0	0	0	
child	0	5	0	0	0	0	0	0	0	
Heavy vehicle	0	0	5	0	0	0	0	0	0	
large	0	0	0	2	0	0	0	0	0	
medium	0	0	0	0	2	0	0	0	0	
sedan	0	0	0	0	0	3	0	0	0	
senior	0	0	0	0	0	0	4	0	0	
small	0	0	0	0	0	0	0	0	4	
teen	0	0	0	0	0	0	0	0	0	

Figure 47: Confusion matrix for C5.0 decision tree algorithm on 70-30 dataset

As is apparent from the images above, on changing the dataset ratio for training and testing, there is absolutely no change in the accuracy. And a 100% accuracy is a suspicious accuracy hinting towards overfitting or errors in the data. It may also mean that there might be 1:1 correlation between one or more features and the target variable.

### *80-20 ratio*

Following the same convention as stated above we divide the dataset in 80-20 ratio for training and testing respectively. The results are illustrated below.

Overall Statistics									
Accuracy : 1									

Figure 48: Accuracy of C5.0 decision tree algorithm for 80-20 dataset

		true								
		adult	child	Heavy vehicle	large	medium	sedan	senior	small	
pred		adult	4	0	0	0	0	0	0	0
	adult	0	2	0	0	0	0	0	0	0
	child	0	0	5	0	0	0	0	0	0
	Heavy vehicle	0	0	0	0	0	0	0	0	0
	large	0	0	0	0	0	0	0	0	0
	medium	0	0	0	0	1	0	0	0	0
	sedan	0	0	0	0	0	2	0	0	0
	senior	0	0	0	0	0	0	2	0	0
	small	0	0	0	0	0	0	0	0	4
	teen	0	0	0	0	0	0	0	0	0

Figure 49: Confusion matrix for C5.0 decision tree algorithm on 80-20 dataset

#### *Discussion on the results of C5.0 Decision Tree algorithm*

In this section, we provide a graphical representation of the accuracy of C5.0 decision tree algorithm on different dataset ratios.

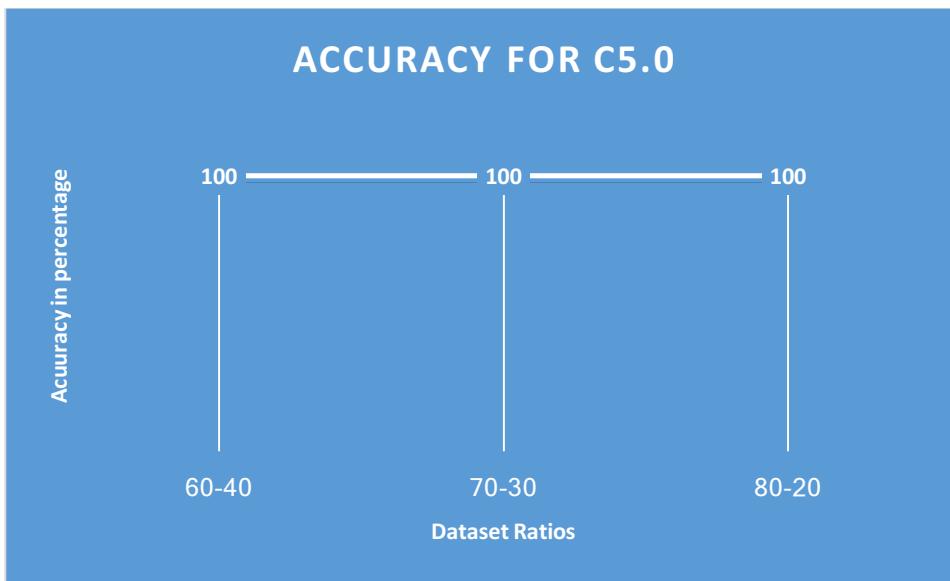


Figure 50: Accuracy of C5.0 Decision tree algorithm on different dataset ratios

The graph is representative of the accuracy of the C5.0 on different dataset ratios of 60-40, 70-30 and 80-10. The x-axis represents the dataset ratios and the y-axis is representative of the accuracy in percentages. And it can be seen from the graph that on increasing the training data, there is absolutely no change in the accuracy achieved by the classifier and the accuracy remains at a constant value of 100 percent, which is giving rise to suspicion as it might hint towards a chance of overfitting or an error. It could also hint towards a 1:1 correlation between one or more features and the target value. The 100 percent accuracy could also mean that there might be some errors in the dataset. So, overall the best accuracy

doesn't mean the best results at all times because in real-life scenario even the smallest bit of error could lead to a life and death scenario when relating to the crash scenario of the autonomous vehicle.

#### 6.2.4 Random Forest algorithm application: Results and Discussion

In this section, we will provide results in terms of accuracy and performance of the Random Forest algorithm with reference to different dataset ratios raging from 60-40 to 80-20.

##### *60-40 ratio*

Here, we divide our data set into training and testing data. We will treat 60 % of our dataset as training data and the rest will be used for the testing purposes. We will attach the results in terms of accuracy and confusion matrix below.

```
RandomForest
Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.15 seconds

== Evaluation on test split ==

Time taken to test model on test split: 0.01 seconds

== Summary ==

Correctly Classified Instances      43          97.7273 %
Incorrectly Classified Instances   1           2.2727 %
Kappa statistic                   0.9741
Mean absolute error               0.0352
Root mean squared error          0.0974
Relative absolute error          17.8745 %
Root relative squared error     30.8099 %
Total Number of Instances        44
```

Figure 51: Accuracy of Random Forest Algorithm on 60-40 dataset

```

== Confusion Matrix ==

a b c d e f g h i    <-- classified as
7 0 0 0 0 0 0 0 0 | a = child
0 5 0 0 0 0 0 0 0 | b = teen
0 0 6 0 0 0 0 0 0 | c = adult
0 0 0 4 0 0 0 0 0 | d = senior
0 0 0 0 5 0 0 0 0 | e = sedan
0 0 0 0 0 5 0 0 0 | f = Heavy vehicle
0 0 0 0 0 0 1 1 0 | g = small
0 0 0 0 0 0 0 4 0 | h = medium
0 0 0 0 0 0 0 0 6 | i = large

```

Figure 52: Confusion matrix for random forest classification on 60-40 dataset

As we can see from the confusion matrix, there is only one incorrectly classified instance and that the algorithm confused between the small sized animal and medium sized animal.

#### *70-30 ratio*

The same convention is followed as mentioned above except that we divide the dataset here into 70-30 ratio which is 70% for training and the rest for testing.

```

RandomForest

Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.17 seconds

== Evaluation on test split ==

Time taken to test model on test split: 0.01 seconds

== Summary ==

Correctly Classified Instances      33          100      %
Incorrectly Classified Instances   0           0        %
Kappa statistic                   1
Mean absolute error               0.0329
Root mean squared error          0.091
Relative absolute error          16.6444 %
Root relative squared error     28.6698 %
Total Number of Instances        33

```

Figure 53: Accuracy of Random Forest Algorithm on 70-30 dataset ratio

```

== Confusion Matrix ==

a b c d e f g h i    <-- classified as
6 0 0 0 0 0 0 0 0 | a = child
0 5 0 0 0 0 0 0 0 | b = teen
0 0 5 0 0 0 0 0 0 | c = adult
0 0 0 1 0 0 0 0 0 | d = senior
0 0 0 0 3 0 0 0 0 | e = sedan
0 0 0 0 0 4 0 0 0 | f = Heavy vehicle
0 0 0 0 0 0 0 0 0 | g = small
0 0 0 0 0 0 0 3 0 | h = medium
0 0 0 0 0 0 0 0 6 | i = large

```

Figure 54: Confusion matrix for Random Forest algorithm on 70-30 dataset ratio

It is apparent from the accuracy and confusion matrix that there is no confusion and all the testing instances are classified correctly.

#### *80-20 ratio*

Here, everything remains the same only the dataset ratio changes to 80 and 20 percent for training and testing respectively.

```

RandomForest

Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.12 seconds

== Evaluation on test split ==

Time taken to test model on test split: 0 seconds

== Summary ==

Correctly Classified Instances      22          100      %
Incorrectly Classified Instances   0           0         %
Kappa statistic                   1
Mean absolute error               0.0302
Root mean squared error          0.0893
Relative absolute error          15.2995 %
Root relative squared error     28.2536 %
Total Number of Instances        22

```

Figure 55: Accuracy of Random Forest algorithm on 80-20 dataset

```
== Confusion Matrix ==
```

a	b	c	d	e	f	g	h	i	<-- classified as
4	0	0	0	0	0	0	0	0	a = child
0	3	0	0	0	0	0	0	0	b = teen
0	0	3	0	0	0	0	0	0	c = adult
0	0	0	1	0	0	0	0	0	d = senior
0	0	0	0	2	0	0	0	0	e = sedan
0	0	0	0	0	3	0	0	0	f = Heavy vehicle
0	0	0	0	0	0	0	0	0	g = small
0	0	0	0	0	0	0	2	0	h = medium
0	0	0	0	0	0	0	0	4	i = large

Figure 56: Confusion matrix for Random Forest algorithm on 80-20 dataset

#### *Discussion on the results of Random Forest algorithm*

In this section, we have presented the graphical representation of the random forest algorithm for classification.

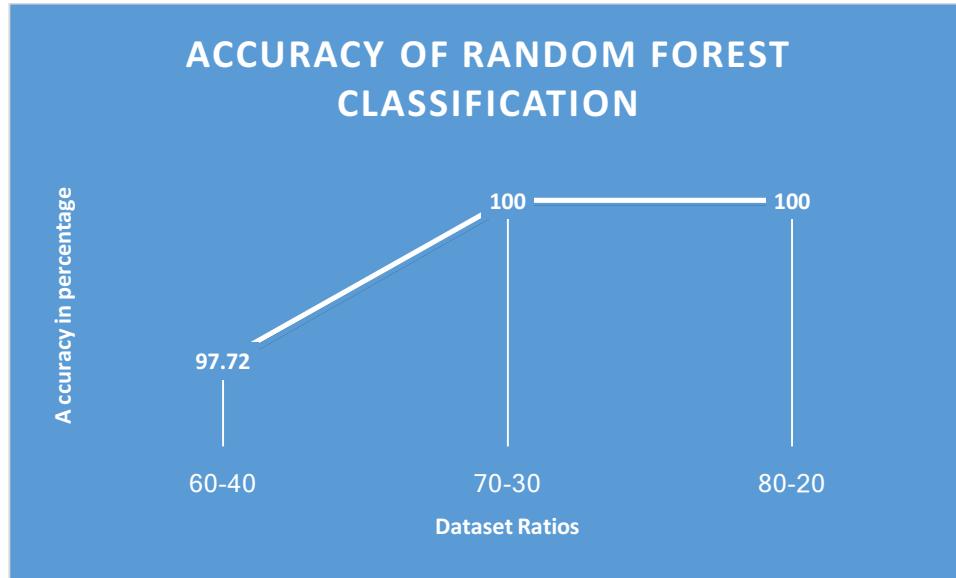


Figure 57: Accuracy of Random Forest algorithm on different dataset ratios.

As it can be seen from the graph, the x and y axis are representative of dataset ratios and the corresponding accuracies respectively. Also, we see that when the dataset was split in 60-40 ratio the accuracy was 97.72 percent with only one instance being incorrectly

classified and that was a small animal classified into the medium sized animal. However, with increase in the training data, the accuracy improved and became a perfect one of 100 percent. However, this might be suggestive of overfitting since dataset is too closely fit and has a limited set of data points even though the chances of overfitting are lesser in random forest but we have to keep in mind that the number of features too is less in our dataset. And random forest is generally applied to very large datasets with thousands of features to yield best results. So, overfitting should be avoided at all costs. Also, as can be seen from above attached Figures, the time required to train and build the model in Naïve Bayes' and C4.5 and C5.0 algorithm was always around 0.01 however in case of random forest the same is 0.12seconds.

### **6.3 Summary of the experimental results**

As is apparent from the above results, Naïve Bayes' works best with small datasets and that its accuracy diminishes with increase in the training data as it is not able to build a model that is representative of the actual dataset. And we saw that on increasing the training data from 60 percent of total data to 80 percent, the confusion matrix showed an increase in the confusion between child and teen, and different sized animals. It showed a maximum accuracy of 84.09% which is not bad but considering the fact that we have considered a very small world and just have built a framework that could be used in future for a very large dataset and that with a large number of features and different categories of objects, using Naïve Bayes for classification might not be best approach. Also, when it comes to life and death scenario as might be the case in future with the introduction of autonomous vehicles, accuracy is of prime importance because that would affect the course of action taken by the AV.

Next in C4.5 algorithm, we were able to visualize the pruned tree and the accuracy seemed to improve with more training data. We saw an increase in accuracy from 88.6% to 95.45% when increasing the training data from 60 percent to 80 percent. And the results of confusion matrix showed a reduction in confusion between heavy vehicles and sedan to zero and the confusion between medium and large sized animals being reduced to only one misclassified instance. Thus, we can say that an increase in training data helps improves

the accuracy and also reduce the confusion. Also, when using C4.5 or for that matter any decision tree algorithm we will be able to visualize the tree and see the classification rules of the model which would not help programmers see the efficiency of the approach but also would be understandable by the user if he is allowed to see the criteria for object classification and this could help him decide in a better way how his ethics be replicated. Like the definition of a heavy vehicle might/might not be clear to a person but when he sees the criteria of putting a particular vehicle into the heavy vehicle category then it might help him take a much informed decision. Thus, using C4.5 or C5.0 in the classification algorithm of the AV to help identify and classify a particular detected object into the right category would be a beneficial decision. However, there exists certain other machine learning algorithms which could yield better results and could also avoid the drawback of overfitting as decision trees are more prone to overfitting than the random forest counterparts.

In C5.0 algorithm, we saw that the accuracy remained constant with changing ratios of training and testing data and there was no confusion as can be seen from the confusion matrix. However, a result of 100 percent accuracy doesn't always mean that it is the best algorithm because the hundred percent accuracy could be a result of overfitting and some errors in the data set. It may also mean that there could be a one to one correlation between one or more features of the dataset and the target variable or class. And this could lead to unwanted results. When it comes to real life scenarios of life and death in case of crash scenarios of the autonomous vehicles, it is better not to rely on algorithms that might give perfect accuracy owing to errors in the dataset. Although, C5.0 is said to be more accurate and in some terms better than the C4.5 however, we do not want to take a chance with our customized ethics because one has to be legally and morally responsible for the life of humans or animals and any error could actually change the entire scenario and results. Therefore, we did not use C5.0 in the implementation of our proposed work.

And finally, we tested our dataset using random forest algorithms and it also showed an increase in accuracy when the training data was increased from 60 percent to 80 percent. The accuracy reached a perfect value of 100 percent which is a great score and also could

be deemed the best. However, it could be suggestive of overfitting in case of our dataset. But, when considering the real world scenario with thousands of objects that could be encountered in a road environment together with their hundreds of features, using random forest would be the safest bet. For our dataset, we will use C4.5 as our classification algorithm as it yielded us the result of 95.45% accuracy and the time to build the model in each case was also around 0.01 seconds which was lesser than the time required to build the model in case of random forest (0.17 seconds). Hence, we will use C4.5 in our work. However, ours is the framework for the customized ethics AVs and this should be kept in mind and other machine learning algorithms for classification can be tested in future and one that yields the best results should be used.

## **CHAPTER 7      CONCLUSION**

The introduction of autonomous cars will open a plethora of opportunities and advantages for the common man as well as for the society as whole. It will altogether bring down the accident fatalities and injuries by about 90%. [7,9,19,21] Also , it will contribute towards the development of a cleaner environment by reducing pollution [9] and reducing the fuel consumption [7]. However, they will also bring about various ethical and social dilemmas as mentioned in the works of Rehwan et al., Lin and Goodall and many other authors that it will lead to social and ethical dilemma when an autonomous vehicle would be in a crash scenario, then who gets to decide whether to kill or save someone, or to kill or be killed. This raised the question of ethics and many researchers derived an analogy between the famous thought-experiment of philosophy called as ‘the trolley problem’ and the accident algorithms of the autonomous vehicles. [13,17,18,31,35].

Also, this led to the discussion about the different kind of ethics such as those of deontological and utilitarian ethics as to which one should the car follow in an accident scenario and studies were conducted by Rehwan et al. [9] which led to the results that people want that the autonomous vehicles follow the utilitarian ethics of reducing the harm and contributing towards greater good but won’t buy the car themselves and would want their car to save them at all cost which led us to the idea of customizing the ethics of the AV according the consumer who buys the car. Also, the studies revealed that people would not want to buy vehicle whose ethics have been regulated by the government. Hence, we came up with the idea of having adaptable ethics. Our work will try to capture the ethics of the person by asking him questions however, since there could be millions of scenarios in real life so only a small set of questions would not be able to capture the ethics. So, we created a small world with only a few different types of objects that could be encountered in a road environment and then would ask questions from users. However, at any point if the user is not able to answer all the questions at one go, and would still want to ride in his vehicle, therefore we gave the option of default ethics.

Default or pre-defined ethics would be a set of ethics that would be followed only when the objects detected in an accident scenario do not appear in the priority list (ethics) of the person. This further would solve the daunting question of the moral and legal responsibility of the crash. As discussed by Hevelke et al. [19] that who would bear the responsibility of such crashes, the default and user defined ethics can solve the issue in the following way. In cases, where the user ethics is followed, the user will be responsible and in scenarios where the default ethics is followed it would be the collective responsibility of the OEM and the government and other stakeholders.

And, our proposed work also solves the problem of having to design universal ethics as that might not be acceptable to all the people.

Nyholm et al. [21] in their work identified the basic differences between the trolley problem and the accident algorithms of the autonomous vehicles. Also, Goodall has very well provided response to some criticism relating to the driverless car like those of “Driverless cars will never crash” etc.

So, we proposed a framework for the formalization of ethics in autonomous vehicles which has answered some of the unanswered questions relating to ethics of the autonomous technology and also has provided food for thought for future researches in the domain. The proposed approach captures the ethics of the user and reflects it as the ethics of the autonomous vehicle that the user owns. Thus, solving the issue of having to design universal ethics or forcing users to buy autonomous vehicle with utilitarian ethics. However, the ethics of the user would be regulated by the government to some extent. We created a dataset with some objects and found the classification algorithm that best catered to our needs in this proposed work. However, other algorithms can be tested in future and one with highest accuracy and better results can be selected. After selecting the algorithm, we then generated questions for the user to answer and based on that, generated a priority list (ethics) of the person. Now, we used our classification algorithm to help classify objects into the respective category and in an accident scenario, the objects detected would be classified and their priorities would be checked and based on the position of the detected

objects in the priority list, the required action will be taken. We also provided the default list since it might not be possible for the user to answer all the questions at one go. This shall answer the questions relating to responsibility of crashes. We have tried to provide solutions for some of the questions relating to the autonomous technology however, there are a few limitations to our work which we would elaborate in the next section. Other applications of the proposed ethical framework would be in the field of robotics especially in case of home robots and in battle grounds with autonomous drones, autonomous aircrafts and unmanned autonomous systems which are not yet common in military operations. A perfectly ethical autonomous robot/drone or robot can perform better than the human beings in battlefield and could take the right decision at the right time. Also, an ethical home robot could prove to be of great help in taking care of the elderly and providing company to them.

## **7.1 Limitations**

The proposed work has only tested four classification algorithm and selected the one that catered best to our needs however, there exists various other algorithms that could yield better results than our results. We have used machine learning as a pilot study and an in depth analysis is not done so this could be done in future so as to improve the performance and choose a better and faster machine learning algorithm. Also, our work just provides a framework for the adaptable feedback mechanism of customized ethics and assumes a very small world scenario comprising of only a few objects. However, in a real-life scenario there could be a large number of objects with so many different features so this is yet another limitation of our work. Also, we have not considered the scenario of prioritizing self or self and passengers. And this could be the future work of our proposed work. Also, we have used the concept of selection sorting to generate the priority list/ ethics of a person however, a better and faster concept could be used in future.

## **7.2 Discussion and Future work**

Our proposed work has provided a food for thought for future research and better algorithms can be used in future. Also, we have carried out our experimentation and implementation using our laptop but we do not yet know how the proposed work would perform in real life scenario in an AV. Another future work would be to incorporate the above mentioned limitations to our work in the proposed work. Also, when one buys the car, then it not always that person who uses the car; members of his family too could use the car so another future work would be to have an identification and authentication algorithm in the AV which could change the ethics settings based on the person riding the car. A better approach to capturing the ethics of the user can also be defined in future which might be taking less time and could be based on faster algorithms. More accurate and faster algorithms could be used for classification and generation of the priority list and the action taken. Thus, there is still a lot of work remaining in the ethics domain before the AVs with customized ethics can hit the road.

## REFERENCES

- [1] I. Bae and S. Olariu, "A Tolerant Context-Aware Driver Assistance System for VANETs- Based Smart Cars", in Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE, Miami, FL, USA, 2017.
- [2] S. Thurston, "Bottom line with driverless cars: Will people buy them?", Tampa Bay Times, 2017. [Online]. Available: <http://www.tampabay.com/news/business/bottom-line-with-driverless-cars-will-people-buy-them/2166010>. [Accessed: 03- Jul- 2017].
- [3] L. Fancher, "Hard Drive: Self-Driving Cars Are Closer Than They Appear - By - February 19, 2014 - SF Weekly", SF Weekly, 2014. [Online]. Available: <http://www.sfweekly.com/news/hard-drive-self-driving-cars-are-closer-than-they-appear/>. [Accessed: 03- Jul- 2017].
- [4] D. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations", Transportation Research Part A: Policy and Practice, vol. 77, pp. 167-181, 2015.
- [5] "About DARPA", Darpa.mil, 2017. [Online]. Available: <https://www.darpa.mil/about-us/about-darpa>. [Accessed: 03- Jul- 2017].
- [6] "The DARPA Grand Challenge: Ten Years Later", Darpa.mil, 2017. [Online]. Available: <http://www.darpa.mil/news-events/2014-03-13>. [Accessed: 03- Jul- 2017].
- [7] A. Thierer and R. Hagemann, "Removing Roadblocks to Intelligent Vehicles and Driverless Cars", SSRN Electronic Journal, 2014.
- [8] A. Hars, "Autonomous cars: The next revolution looms Alexander Hars", Inventivio Innovation Briefs, vol. 04, 2010.
- [9] J. Bonnefon, A. Shariff and I. Rahwan, "The social dilemma of autonomous vehicles", Science, vol. 352, no. 6293, pp. 1573-1576, 2016.
- [10] "Technology –Waymo", Waymo, 2017. [Online]. Available:<https://waymo.com/tech/>. [Accessed: 03- Jul- 2017].
- [11] "Journey –Waymo", Waymo, 2017. [Online]. Available: <https://waymo.com/journey/>. [Accessed: 03- Jul- 2017].
- [12] X. Mosquet, T. Dauner, N. Lang, M. Rüßmann, A. Mei-Pochtler, R. Agrawal and F. Schmieg, "Revolution in the Driver's Seat: The Road to Autonomous Vehicles", [www.bcgperspectives.com](http://www.bcgperspectives.com), 2015. [Online]. Available: <https://www.bcgperspectives.com/content/articles/automotive-consumer-insight-revolution-drivers-seat-road-autonomous-vehicles/>. [Accessed: 03- Jul- 2017].

- [13] M. Maurer, J. Gerdes, B. Lenz and H. Winner, Autonomous driving. Berlin: Springer Open, 2016, pp. 69-85.
- [14] "Deontological Ethics", www3.sympatico.ca, 1967. [Online]. Available: <http://www3.sympatico.ca/saburns/pg0405a00.htm>. [Accessed: 03- Jul- 2017].
- [15] "Online Guide to Ethics and Moral Philosophy", Caae.phil.cmu.edu. [Online]. Available: <http://caae.phil.cmu.edu/Cavalier/80130/part2/sect9.html>. [Accessed: 03- Jul- 2017].
- [16] "Thought experiment", En.wikipedia.org. [Online]. Available: [https://en.wikipedia.org/wiki/Thought\\_experiment](https://en.wikipedia.org/wiki/Thought_experiment). [Accessed: 03- Jul- 2017].
- [17] T. Cathcart, The trolley problem, or, would you throw the fat guy off the bridge?. New York: Workman Pub. Co., Inc., 2013.
- [18] D. Edmonds, Would you kill the fat man?. Princeton: Princeton University Press, 2015.
- [19] A. Hevelke and J. Nida-Rümelin, "Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis", Science and Engineering Ethics, vol. 21, no. 3, pp. 619-630, 2014.
- [20] Marchant, G. E., & Lindor, R. A., "The coming collision between autonomous vehicles and the liability system", Santa Clara L. Rev., pp. 1321-1340, 2012
- [21] S. Nyholm and J. Smids, "The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem?", Ethical Theory and Moral Practice, vol. 19, no. 5, pp. 1275-1289, 2016.
- [22] A. Humayed and B. Luo, "Cyber-physical security for smart cars: taxonomy of vulnerabilities, threats, and attacks", in ICCPS '15 Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems, Seattle, Washington, 2017, pp. 252-253.
- [23] M. Chaudhry, C. Seth and A. Sharma, "Feasibility Analysis of Driverless Car Using VANETs", Discovery, vol. 15, no. 42, 2014.
- [24] J. Sun, Z-H. Wu and G. Pan, "Context-aware smart car: from model to prototype," Journal of Zhejiang University SCIENCE A, Vol. 10, No. 7, pp. 1049-1059, 2009.
- [25] F. Kröger, "Automated Driving in its Social, Historical and Cultural Contexts"Autonomous driving. Berlin: Springer Open, 2016, pp. 52-67.

- [26] M. Weber, "Where to? A History of Autonomous Vehicles | Computer History Museum", Computerhistory.org, 2014. [Online]. Available: <http://www.computerhistory.org/atchm/where-to-a-history-of-autonomous-vehicles/>. [Accessed: 03- Jul- 2017].
- [27] "Autonomous car", En.wikipedia.org, 2017. [Online]. Available: [https://en.wikipedia.org/wiki/Autonomous\\_car](https://en.wikipedia.org/wiki/Autonomous_car). [Accessed: 03- Jul- 2017].
- [28] "J3016A: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International", Standards.sae.org, 2016. [Online]. Available: [http://standards.sae.org/j3016\\_201609/](http://standards.sae.org/j3016_201609/). [Accessed: 03- Jul- 2017].
- [29] G. Yeomans, Autonomous vehicles Handing over control: Opportunities and Risks for Insurance. LLoyd's, 2014, p. 9.
- [30] NHTSA, "Preliminary Statement of Policy", pp. 7-8
- [31] P. Lin, "The Ethics of Autonomous Cars", The Atlantic, 2013.
- [32] E. Volokh, "Ford "Know[s] Everyone Who Breaks the Law" Using Cars They Made - Why Aren't They Doing Something About It? - The Volokh Conspiracy", The Volokh Conspiracy, 2014. [Online]. Available: <http://volokh.com/2014/01/10/ford-knows-everyone-breaks-law-using-cars-made-arent-something/>. [Accessed: 03- Jul- 2017].
- [33] E. Volokh, "Tort Law vs. Privacy - The Volokh Conspiracy", The Volokh Conspiracy, 2013. [Online]. Available: <http://volokh.com/2013/11/25/tort-law-vs-privacy/>. [Accessed: 03- Jul- 2017].
- [34] N. Goodall, "Machine ethics and automated vehicles", in Road vehicle automation, Springer, 2014, pp. 93-102.
- [35] Lin, Patrick, "Why ethics matters for autonomous cars.", in Autonomous Driving. Springer,Berlin Heidelberg, 2016, pp. 69-85.
- [36] M. Fox, "Driverless cars may not be ready for prime driving time, expert says", CNBC, 2016. [Online]. Available: <http://www.cnbc.com/2016/07/10/driverless-cars-arent-ready-for-the-road-expert-warns.html>. [Accessed: 03- Jul- 2017].
- [37] J. Gerdes and S. Thornton, "Implementable Ethics for Autonomous Vehicles", in Autonomous Driving, Berlin: Springer, 2016, pp. 97-105.
- [38] W. Wallach and C. Allen, Moral machines. New York: Oxford University Press, 2010.
- [39] B. van Arem, C. van Driel and R. Visser, "The Impact of Cooperative Adaptive Cruise Control on Traffic-Flow Characteristics", IEEE Transactions on Intelligent Transportation Systems, vol. 7, no. 4, pp. 429-436, 2006.

- [40] P. Gao, H. Russel and Z. Andreas, "A roadmap to the future for the auto industry", McKinsey Quarterly, Oct, 2014.
- [41] K. Spieser, K. Treleaven, R. Zhang, E. Fazzoli, D. Morton and M. Pavone, "Toward a Systematic Approach to the Design and Evaluation of Automated Mobility-on-Demand Systems: A Case Study in Singapore", in Road Vehicle Automation, Springer, 2014, pp. 229-245.
- [42] "Average Height to Weight Chart - Babies to Teenagers", Disabled World, 2017. [Online]. Available: <https://www.disabled-world.com/artman/publish/height-weight-teens.shtml>. [Accessed: 03- Jul- 2017].
- [43] Fryar CD, Gu Q, Ogden CL, Flegal KM, "Anthropometric reference data for children and adults: United States, 2011–2014", National Center for Health Statistics, Vital Health Stat 3(39), 2016.
- [44] Heavy truck weight and dimension limits for interprovincial operations in Canada resulting from the Federal-Provincial-Territorial Memorandum of Understanding on Interprovincial Weights and Dimensions. Ottawa: The Task Force, 1999, pp. 71-81.
- [45] M. Bay, "A careful look at different sedan dimensions", Carfinderservice.com, 2017. [Online]. Available: <https://www.carfinderservice.com/car-advice/a-careful-look-at-different-sedan-dimensions>. [Accessed: 03- Jul- 2017].
- [46] H. Zhang, "Exploring conditions for the optimality of Naïve Bayes", International Journal of Pattern Recognition and Artificial Intelligence, vol. 19, no. 02, pp. 183-198, 2005.
- [47] "Choosing a Machine Learning Classifier", Blog.echen.me, 2016. [Online]. Available: <http://blog.echen.me/2011/04/27/choosing-a-machine-learning-classifier/>. [Accessed: 03- Jul- 2017].
- [48] P. Tan, M. Steinbach and V. Kumar, Introduction to data mining. Dorling Kindersley: Pearson, 2015, pp. 145-195.
- [49] "Is C5.0 Better Than C4.5?", Rulequest.com, 2017. [Online]. Available: <http://www.rulequest.com/see5-comparison.html>. [Accessed: 03- Jul- 2017].
- [50] "Random forests - classification description", Stat.berkeley.edu, 2017. [Online]. Available: [https://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_home.htm](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm). [Accessed: 03- Jul- 2017].
- [51] "Weka (machine learning)", En.wikipedia.org. [Online]. Available: [https://en.wikipedia.org/wiki/Weka\\_\(machine\\_learning\)](https://en.wikipedia.org/wiki/Weka_(machine_learning)). [Accessed: 03- Jul- 2017].

- [52] "Weka 3 - Data Mining with Open Source Machine Learning Software in Java", Cs. waikato.ac.nz. [Online]. Available: <http://www.cs.waikato.ac.nz/ml/weka/>. [Accessed: 03-Jul- 2017].
- [53] Eibe Frank, Mark A. Hall, and Ian H. Witten, The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.
- [54] "About", RStudio, [Online]. Available: <https://www.rstudio.com/about/>. [Accessed: 03- Jul- 2017].
- [55] "Data Structures and Algorithms Selection Sort", [www.tutorialspoint.com](http://www.tutorialspoint.com). [Online]. Available: [https://www.tutorialspoint.com/data\\_structures\\_algorithms/selection\\_sort\\_algorithm.htm](https://www.tutorialspoint.com/data_structures_algorithms/selection_sort_algorithm.htm). [Accessed: 04- Jul- 2017].
- [56] G. Veruggio and F. Operto, "Roboethics: a bottom-up interdisciplinary discourse in the field of applied ethics in robotics", *International review of information ethics*, vol. 6, no. 12, pp. 2-8, 2006.
- [57] N. Sharkey, "COMPUTER SCIENCE: The Ethical Frontiers of Robotics", *Science*, vol. 322, no. 5909, pp. 1800-1801, 2008.
- [58] B. Stahl and M. Coeckelbergh, "Ethics of healthcare robotics: Towards responsible research and innovation", *Robotics and Autonomous Systems*, vol. 86, pp. 152-161, 2016.
- [59] C. Allen, W. Wallach and I. Smit, "Why Machine Ethics?", *IEEE Intelligent Systems*, vol. 21, no. 4, pp. 12-17, 2006.