**Cairo University**

**Faculty of Graduate Studies for Statistical Research Department of Software Engineering**

**Cairo University**

v

**Study of the effect of alcoholic beverages on the driving patterns of adult drivers**

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master in Software Engineering

**by**

**Mohamed Salah Ibrahim**

**Supervised by**

### Prof. Mervat Gheith Dr.Tarek Aly

**2023**

### Abstract:

### The Influence of Alcohol on Driver Patterns: Early Detection Using Machine Learning

### Driving under the influence of alcohol remains a critical factor contributing to road accidents and jeopardizing road safety globally. This study delves into the impact of alcohol on driver behavior patterns and introduces an innovative method to preemptively identify these patterns using a machine learning model. The primary goal is to establish an early warning system capable of alerting drivers about their compromised driving abilities due to alcohol consumption, thus mitigating potential accidents.

### To facilitate this research, an extensive dataset was curated from diverse sources, including driving simulators and real-world driving scenarios. This dataset encompassed a wide spectrum of driving behaviors under varying levels of alcohol influence. Feature extraction involved parameters such as vehicle speed, lane deviation, and reaction times.

### Through rigorous experimentation, a machine learning framework was devised to analyze the dataset and recognize distinctive patterns that signify alcohol-induced impairment. A comprehensive array of algorithms, including decision trees, clusters, and neural networks, underwent evaluation to determine the optimal model for detection. The chosen model was fine-tuned and validated using cross-validation techniques.

### The resulting system adeptly detects patterns that indicate alcohol-related impairment in real-time. By scrutinizing driver behavior and their interactions with the vehicle, the model accurately predicts instances of alcohol influence. Upon identifying a heightened likelihood of impairment, the system issues a prompt warning to the driver, highlighting their compromised state and recommending appropriate actions such as refraining from driving or seeking alternative transportation.

### This study substantiates that the machine learning model significantly contributes to curbing alcohol-related accidents by providing preemptive alerts to impaired drivers. The efficacy of the model hinges on the quality and diversity of the training dataset, along with the accuracy of input data. Future research endeavors could concentrate on expanding the dataset to encompass a broader array of scenarios and validating the model extensively in real-world settings.

### In conclusion, this study introduces a promising approach to tackling the perils associated with alcohol-impaired driving. By harnessing machine learning techniques for real-time assessment, the proposed system presents a proactive strategy to enhance road safety and avert potential accidents triggered by alcohol influence.

### Keywords: alcohol-impaired driving, driver behavior patterns, real-time detection, machine learning model, road safety, early warning system, vehicle speed, lane deviation, reaction times, decision trees, random forests, support vector machines, neural networks, training dataset, real-world scenarios, road accidents, proactive solution.

# Table of Contents

[List of Figures xi](#_TOC_250039)

[List of Tables xii](#_TOC_250038)

[List of Abbreviations v](#_TOC_250037)

[Acknowledgement i](#_TOC_250036)

1. [Introduction 1](#_TOC_250035)
   1. [A problem statement 2](#_TOC_250034)
   2. [Thesis objectives 2](#_TOC_250033)
   3. [Research Methodology 3](#_TOC_250032)
   4. [Thesis layout 3](#_TOC_250031)
2. Related Work 4
   1. [Introduction 4](#_TOC_250030)
   2. [Enhanced Extreme Programming from Scrum and vice versa 5](#_TOC_250029)
      1. [The phases of Enhanced XP Model 5](#_TOC_250028)
      2. [Enhanced Scrum Model by a set of Extreme programming 6](#_TOC_250027)
      3. [Visibility of work during a sprint 8](#_TOC_250026)

[2-2.4 Model Enhancing Scrum with Kanban 8](#_TOC_250025)

1. [The Proposed Solution For Detect Abnormal Drivers Pattern 16](#_TOC_250021)
   1. [Introduction 16](#_TOC_250020)
   2. [Chose algorithm to build an AI model 16](#_TOC_250019)
      1. [Define different algorithms 17](#_TOC_250018)
      2. T[est the accuracy for each algorithm 17](#_TOC_250017)
      3. [chose most effective and applicable algorithm 19](#_TOC_250016)
   3. Design solution architecture 19

3.3.1 [Analyses solution and chose technology 17](#_TOC_250018)

3.3.2 [Integrate with AI model 17](#_TOC_250018)

* 1. [Comparison between results of chosen algorithms 22](#_TOC_250014)
  2. [Summary 24](#_TOC_250013)

1. [Evaluating the Proposed Enhanced Kanban Approach 25](#_TOC_250012)
   1. [Introduction 25](#_TOC_250011)
   2. Evaluation strategy 25
   3. [Display data analysis 25](#_TOC_250010)
   4. [The User Stories of the administration work 26](#_TOC_250009)
   5. [Apply Scrum – model 28](#_TOC_250008)
      1. [Sprint Backlog 29](#_TOC_250007)
      2. [Daily Scrum and Sprint Execution 30](#_TOC_250006)
      3. [Scrum Master 30](#_TOC_250005)
   6. [Apply Kanban – model 31](#_TOC_250004)
   7. [Apply Scrum-ban model 31](#_TOC_250003)
      1. [Sprint review 32](#_TOC_250002)
      2. [Sprint Release 32](#_TOC_250001)
   8. [The results of using E-Kanban 34](#_TOC_250000)
2. Conclusion 35

**6 References**

# List of Figures

Figure (1.1) Hybrid model combining Kanban and Scrum 1

Figure (2.1) LEAN Principles for success 13

Figure (2.2) (US-Scrum Model (US-Scrum) 15

Figure (3.1) The proposed Model (EKanban).before enhancement (a) and after enhancement (b) 17

Figure (3.2) SDLC 19

Figure (4.1) applying Scrum model in the Story 28

Figure (4.2) applying Sprint in the Story 29

Figure (4.3a) Story backlog, Priorities and To do priorities stories 31

Figure (4.3b) Story backlog, Priorities and To do priorities stories 32

Figure (4.4) To do priorities stories with the same arrangement 33

Figure (4.5) Sprint Release 34

# List of Tables

|  |  |
| --- | --- |
| Table (2.1) | After that making comparison between XP, Scrum, and EXP |
|  | Model 6 |
| Table (2.1) | Enhanced the Scrum Framework by using XP practices 7 |
| Table (2.3) | The challenges of Scrum traditional, and proposed (solutions) …….9 |
| Table (2.4) | Pure Scrum versus Enhanced Scrum 10 |
| Table (2.5) | Scrum in Practices 11 |
| Table (2.6) | Model Scrum –FDD 12 |
| Table (2.7) | Adopt the LEAN production principles to software development …..14 |
| Table (3-1) | Comparison between Scrum & Kanban & Enhanced Model………22 |

## Acknowledgement

First of all, ultimate thanks are due to Allah, who without he is and this work could not be done.

I am making this thesis not only for degree but to also increase my knowledge.

I would like to express my special thanks of gratitude to my supervisors Dr/ Mervat Gheith and Dr/ Tarek Aly who gave me the golden opportunity to do this wonderful thesis, which also helped me in doing a lot of Research and i came to know about so many new things I am really thankful to them.

I would also like to thank my family and friends who helped me a lot in finalizing this thesis within the limited time frame.

I would also like to thank the whole development team members whom I worked with during my research activities.

Finally I apologize to all other unnamed who helped me in various ways to have a good thesis.

#### Introduction

Driving under the influence of alcohol remains a significant public health concern that poses a danger to both the driver and other road users. In Asia alone, 400,000 people are killed on the roads annually and more than four million injured(1998. Transport and Research Laboratory, UK) in the US, around 30 people die each day in traffic crashes in which one of the parties is under the influence of alcohol, and, together, alcohol-related crashes amount to 30% of all traffic fatalities (2020 National Highway Traffic Safety) Previous studies have shown that even low blood alcohol concentrations (BACs) can impair critical driving skills such as perception, judgment, and reaction time, leading to an increased risk of crashes among adult drivers (García-Escudero et al., 2021). In fact, drivers with a BAC of 0.08 g/dL or higher are at a significantly higher risk of being involved in a fatal crash than sober adult drivers (Waller et al., 2019). The physiological effects of alcohol on the human body can impair driving performance by affecting crucial skills such as coordination, reaction time, and decision-making, particularly among adult drivers (Mandel et al., 2018). Despite numerous campaigns and legislation aimed at reducing the incidence of drunk driving, it remains a leading cause of road accidents and fatalities among adult drivers worldwide.

In recent years, advances in machine learning algorithms have made it possible to detect patterns in adult driver behavior that could indicate impaired driving and alert drivers to prevent accidents before they happen. This research aims to explore the use of machine learning algorithms to detect patterns in adult driver behavior that could indicate impaired driving and alert adult drivers in real-time. By leveraging data from sensors in the car, such as accelerometers and gyroscopes, as well as data from the adult driver's smartphone and other wearable devices, we can develop models that detect changes in driving behavior that may indicate impairment among adult drivers. The goal of this research is to develop a system that can alert adult drivers in real-time to prevent accidents before they happen and promote safer driving practices among adult drivers.

#### A problem statement:

Driving under the influence of alcohol remains a significant contributor to road accidents and poses a severe threat to road safety worldwide. Despite numerous awareness campaigns and legal measures, alcohol-impaired driving continues to endanger lives. The lack of effective preemptive systems to identify and alert drivers about their compromised state due to alcohol consumption perpetuates this problem. Existing methods often rely on post-incident investigations or self-assessment, leading to delayed responses and potential accidents. This study aims to address this issue by investigating the influence of alcohol on driver behavior patterns and developing a machine learning-based early detection system. The goal is to proactively identify and warn drivers about their impaired abilities before accidents occur, ultimately contributing to a safer road environment and reducing the impact of alcohol-related accidents.

#### Thesis objectives

The primary objective of this study is to design and implement a Machine Learning Model able to learn different patterns and early detect abnormal driving patterns through the software "DPObserver" that can detect and alert the driver and police. Specifically, the study aims to:

Investigate Alcohol's Impact: Analyze the influence of alcohol on driver behavior patterns by studying simulation driving scenarios under varying levels of alcohol consumption.

Feature Extraction: Extract relevant features from the dataset, including vehicle speed, lane deviation, and reaction times, to quantify and characterize driver behavior patterns.

Model Selection and Development: Evaluate various machine learning algorithms, such as decision trees, Cluster, and neural networks, to identify the most effective model for pattern detection.

Model Training and Validation: Train the selected machine learning model on the curated dataset and fine-tune its parameters. Validate the model's performance using cross-validation techniques to ensure its reliability and generalization.

Real-time Detection: Implement the trained model in a real-time monitoring system that analyzes incoming driving behavior data and predicts instances of alcohol-impaired patterns.

Early Warning System: Integrate the detection system with an early warning mechanism that alerts drivers in real time when their behavior indicates alcohol-related impairment.

Scalability and Robustness: Explore the system's scalability by evaluating its performance on a larger scale and in different driving scenarios. Enhance the system's robustness by considering various real-world conditions.

Contribution to Road Safety: Evaluate the potential impact of the developed system on reducing alcohol-related accidents and enhancing overall road safety.

By achieving these objectives, this study aims to contribute to the prevention of alcohol-impaired driving incidents and promote safer road environments through proactive detection and timely interventions.

#### Research Methodology:

This study employs a comprehensive research methodology that encompasses data collection, model development, validation, and evaluation stages. The methodology is designed to achieve the objectives of investigating alcohol's influence on driver patterns and developing an early detection system using machine learning. The following steps outline the research methodology:

**Data Collection:**

Collecting driving behavior data from driving simulators because of the danger of create real scenarios, maybe something can happen to drivers maybe injured or killed in some cases, and the main shortness of my study is to create a dataset.

Extract relevant features such as vehicle speed, lane deviation, and reaction times from the collected data.

**Exploratory Data Analysis:**

Conduct exploratory data analysis to understand the characteristics of the collected dataset.

Identify patterns and correlations between driving behaviors and alcohol influence levels by my AI model.

**Model Selection and Development:**

Evaluate multiple machines learning algorithms, including decision trees, random forests, support vector machines, and neural networks, for pattern detection.

Select the most suitable algorithm based on performance metrics and model complexity.

**Model Training and Validation:**

Split the dataset into training and validation sets for model training and evaluation.

**Real-time Implementation:**

Develop a real-time monitoring system that continuously analyzes incoming driving behavior data.

Integrate the trained machine learning model into the system for detecting alcohol-impaired patterns.

**Early Warning System:**

Design an early warning mechanism that triggers alerts when the model identifies alcohol-impaired driving patterns.

Determine appropriate thresholds for issuing warnings to drivers.

**Ethical Considerations:**

Address ethical considerations related to privacy and data security in collecting and using driving behavior data.

Ensure transparency in explaining the system's functioning to users.

**Contribution to Road Safety:**

Analyze the potential impact of the developed system on reducing alcohol-related accidents and enhancing road safety.

Through this research methodology, the study aims to advance our understanding of alcohol's influence on driver behavior patterns and provide a practical solution to mitigate the dangers of alcohol-impaired driving through real-time pattern detection and early warnings.

#### Thesis layout:

After chapter 1, this thesis is organized as follows:

* + - Chapter 2 will discuss the related work.
    - Chapter3 will present main idea behand the pattern detection and the effective algorithm we use .
    - Chapter 4 will present the evaluation of the model and the results we gain from our software “DPObserver”.
    - Chapter 5 will conclude the thesis through summarizing the research results and findings, and give recommendations for future work.

#### Related Works:

#### Introduction:

The field of transportation and vehicle behavior analysis, coupled with studies on the impact of alcohol on driver behavior, has witnessed significant advancements in recent years. Researchers have explored various facets of these domains, ranging from studies on driver safety, driving pattern analysis, and eco-driving, to investigations into the effects of alcohol on cognitive and motor skills in the context of driving. In order to gain a deep understanding of the evolution and current state of research in these interrelated fields, we have undertaken a thorough review of the literature this section offers two main studies that are directly relevant to our study's dual focus on driving patterns and their association with influence alcohol.

#### Driver Profile and Driving Pattern Recognition for Road Safety Assessment: Main Challenges and Future Directions:

This research thoroughly reviewed artificial intelligence and machine learning Methods used so far in driver profile and driving style Recognition studies for traffic safety analysis purposes. The goal was to identify the best methodology and data collection and suggest future directions for enhancing macroscopic understanding microscopic aspects of driving behavior and thus the road safety.

One of the most important findings of this study is the ambiguity in defining the two scientific fields. It is to discover what the most efficient driving metrics should be used in similar research and repeated data collection it should depend on the level of analysis. Furthermore, she noted the levels of analysis used to define groups of the common behaviors, they can be classified as macroscopic,

Macroscopic (Driving safety, 2019), (Temporal analysis, 2021) and microscopic depending on the level of information they use. Conspicuous absence a methodological framework for identifying macroscopic driver profiles and microscopic driving patterns it is proposed to address them through a methodology that combines macroscopic and microscopic driving metrics, respectively volume 4, 2023 97 TSELENTIS and PAPADIMITRIOU: Driver profile and driving style recognition.

#### A Review for the Driving Behavior Recognition Methods Based on Vehicle Multisensor Information

Since abnormal driving behaviors may lead to immediate accidents, how to improve the identification efficiency and develop a lightweight model that can accurately identify driving behaviors.

This paper reviews and summarizes driving behavior recognition methods based on vehicle sensor information fusion. On-board sensor data contain a wealth of information about driving behavior; Based on the two main factors of driving behavior and vehicle control, driving behavior information can be divided into driver state information, driver control state information, vehicle control state information, vehicle state information, road environment state information, as well as the corresponding data acquisition system. The characteristics of joint data level, feature level and decision-level information fusion methods are analyzed to guide the selection of appropriate information fusion methods and the basic principles and main characteristics of feature extraction methods. Driving behavior recognition methods are classified into traditional machine learning methods and deep learning methods. Random forest, support vector machine is presented, and applied in fatigue, distraction, following, lane change, and other driving behavior recognition. The application of convolutional neural network (CNN) and recurrent neural network (RNN) in building driving behavior recognition models is analyzed. The characteristics of the four leadership behavior recognition paradigms described in this paper are briefly summarized.

Driving behavior is greatly influenced by the driver's condition, and the driving behavior exhibited by the same driver may vary significantly. At the same time, driving behavior is also susceptible to the influence of the road, environment and vehicle, so there are some uncertainties. How to improve the power and generalization ability of the driving behavior recognition model still needs further study.

Drunk Driving Detection

#### Drank Driving Detection

Although there has been significant work on drunk driving, this work...

It focuses largely on statistical analysis that examines the effect

The effect of alcohol on driving performance (S. Jongen, E. F. P. M. Vuurman, 2016). These empirical results provide a basis for regulators to decide on the relevant legal limits to drive under

Effect of alcohol [28]. However, these previous works did not do this

Focus on real-time identification of drunk driving, which would

Allowing intervention when the driver is already drunk. Previous research

It also addressed the negative effect of alcohol on gaze behavior while looking .

Unfortunately, the results of this experimental study present an increase in the condition of diabetes

Leadership (Huiqin Chen and Lei Chen. 2017.) teaches that, while magic

Driver control training works well on previously employed drivers, but it does not achieve the performance required to generalize to unseen drivers.

#### The Proposed Solution For Detect Abnormal Drivers Pattern

#### Introduction

In the realm of transportation and road safety, the ability to identify and respond to abnormal driver behavior is of paramount importance. Abnormal driving patterns, encompassing actions such as sudden lane changes, aggressive acceleration, or erratic steering, not only pose risks to the driver but also jeopardize the safety of other road users. To address this critical concern, we present in this section our proposed solution for the detection of abnormal driver patterns.

Our approach is founded on the fusion of cutting-edge technologies, including machine learning algorithms, sensor data from modern vehicles, and real-time monitoring systems. In our pursuit of detecting and analyzing driver patterns, we employed a diverse set of machines learning models, including clustering, neural networks, K-nearest neighbors (KNN), and linear regression. This section presents the findings and outcomes of each method, showcasing their effectiveness in categorizing driving behaviors and providing valuable insights for various applications, from driver safety to traffic management.

#### Generate And Collecting The Data From Simulator.

One alternative way to avoid making people under the effects of alcohol drive real cars which may make a real accident is to collect the data from it by creating a web-based car driving simulator and this is what I did for this study.

This simulator allows me to generate realistic driving scenarios while capturing crucial information about speed, deviation from lanes, and brake usage, which are three key features I used that define driving patterns.

Speed:

Speed plays a pivotal role in understanding driving behavior. A web-based car driving simulator enables us to control and measure a driver's speed in various conditions. By adjusting speed limits, we can simulate both normal and abnormal driving patterns. Data on speed provides insights into how drivers respond to different situations.

**Deviation:**

Deviation from the intended lane is another vital aspect of driving patterns. The simulator allows us to simulate real-world lane-keeping scenarios. By monitoring a driver's ability to stay within their lane, we can categorize behaviors or frequent lane changes. This information can be invaluable for developing lane-keeping assistance systems and understanding the impact of road design on driving patterns.

**Brake Usage:**

Effective brake usage is essential for safe driving. A web-based simulator enables us to mimic emergency stops, gradual deceleration, and braking in response to traffic conditions. Analyzing brake usage data helps us assess driver reaction times, brake intensity, and adherence to safe following distances.

**Advantages of Web-Based Simulators:**

Creating a web-based car driving simulator has distinct advantages. It's cost-effective, scalable, and accessible to a wide range of users. Participants can engage in simulated driving from the comfort of their web browsers, eliminating the need for physical setups. Moreover, the simulator allows for controlled experiments in various driving conditions, ensuring data consistency.

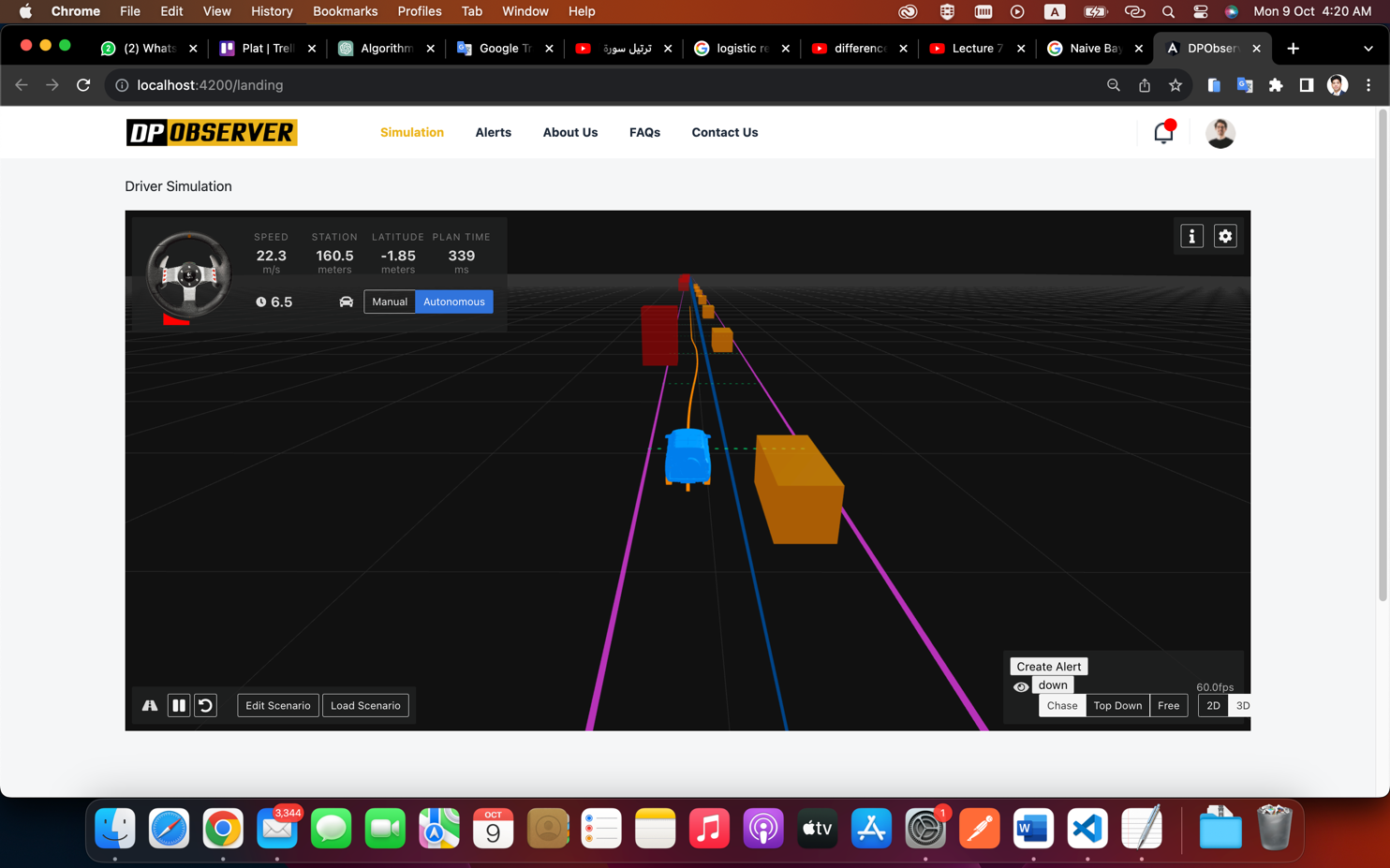
**Data Generation and Analysis:**

As participants use the web-based simulator, data on speed, deviation, and brake usage are collected in real-time. This data I can use this to train my models and generate patterns for normal and abnormal driving with high accuracy this is a sample of data structure I export from my simulator :

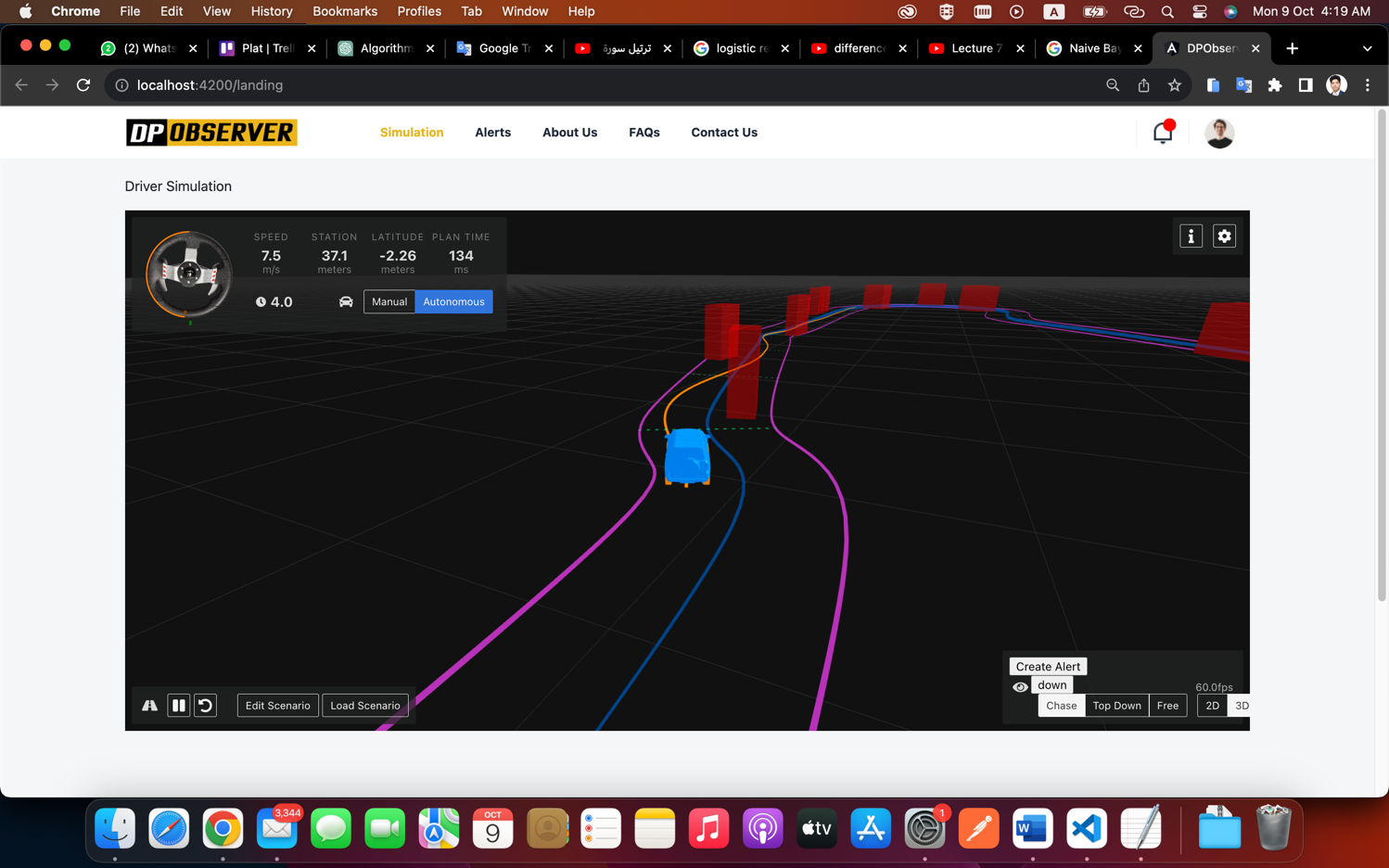
**[**[1.84,24.8,0],[2.56,24.5,0],[3.18,24.2,0],[1.34,23.8,0],[1.07,23.5,0],[0.98,23.2,0],[0.52,22.9,0],[0.49,22.6,0],[0.74,22.3,0],[1.02,22,0],[1.16,21.8,0],[0.69,21.5,0],[0.87,21.2,0],[1.23,21.2,0],[1.23,21,0],[1.32,18.8,0],[1.23,18.9,0]**]**

Is an array of arrays each array has three values **[deviation, speed, brake].**

**Some of Figures for Simulator:**

****

**Figure (3-2) Car Simulator**

****

**Figure (3-2) Car Simulator**



**Figure (3-2) Real Car wheel Connected To Web Simulator**



**Figure (3-2) Real Car wheel Connected To Web Simulator**

#### Chose Algorithm To Build An AI Model.

after we create the simulator and generate data with a specific structure next step is to select the best model that can give us the highest accuracy depending on the nature of my data, we can start this process in three steps :

1. define the models and their algorithms that we can use in this case.
2. training the models then test and evaluate the accuracy of each of them.
3. choose one of them to implement on my solution (**DPObserver**).

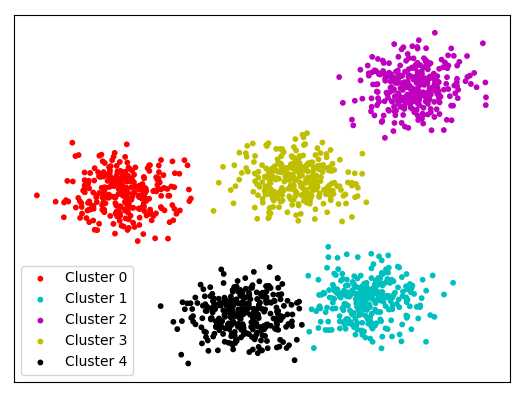
With these steps, we can create a comparison between each algorithm and build good software solutions that have the ability to detect abnormal patterns and make a decision to prevent any accident from properly happening and this is the main goal of my study

#### Define Different Algorithms

1. **Clustering Algorithms:**

Clustering techniques, such as K-means, are a set of unsupervised machine learning methods used to group similar data points together based on their intrinsic characteristics or features. The goal of clustering is to discover hidden patterns or structures within data, identifying natural groupings without the need for predefined labels (MacQueen, J. B. (1967)).

The concept of clustering, including the K-means algorithm, was introduced by J. B. MacQueen in his paper "Some Methods for Classification and Analysis of Multivariate Observations" in 1967. The K-means algorithm, in particular, is a widely used clustering method that partitions a dataset into K clusters by iteratively assigning data points to the cluster with the nearest mean and updating the cluster centroids. It has become a fundamental tool for various applications, including data analysis, image segmentation, and customer segmentation.

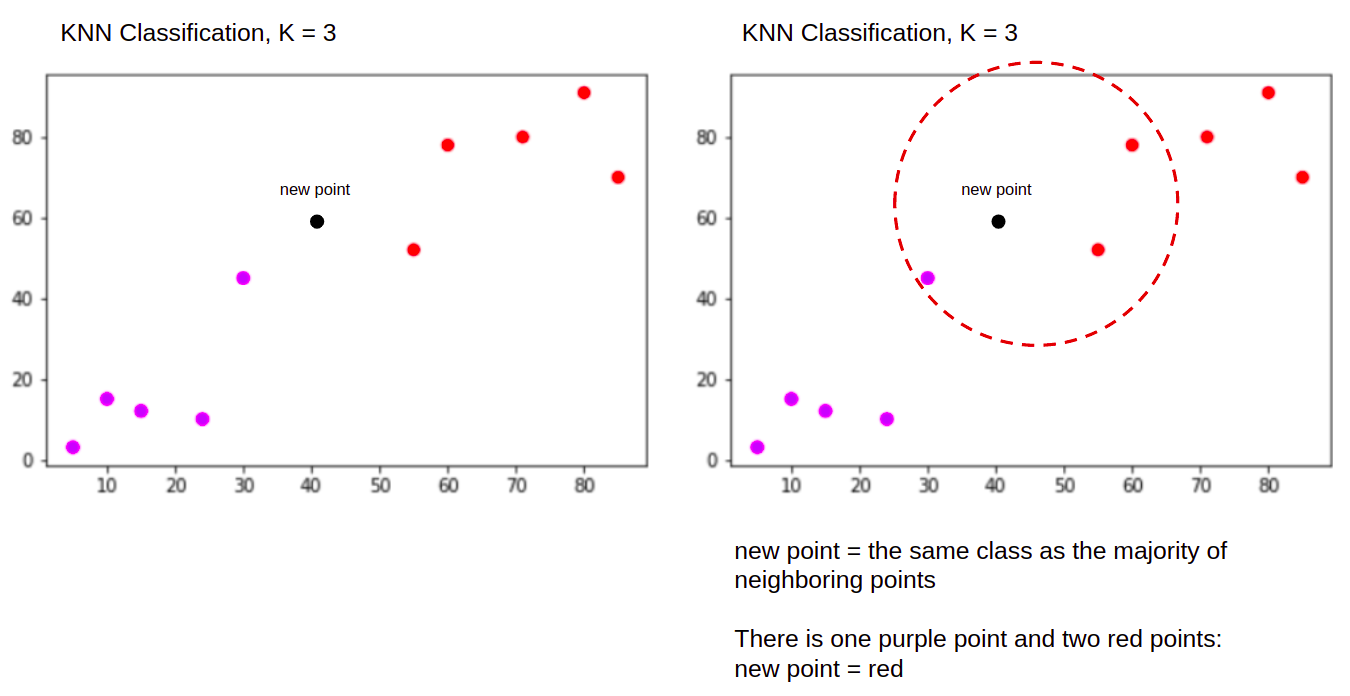


**Figure (3-2-1) Cluster Algorithm**

1. **K-Nearest Neighbors (KNN):**

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm used for classification and regression tasks. It operates on the principle of similarity, where data points are classified or predicted based on the majority class or the average of the values of their K-nearest neighbors in a feature space. In other words, KNN assigns a data point to a class or predicts a value based on the characteristics of its nearest data points in the training dataset (Fix, E., & Hodges, J. L. (1951)).

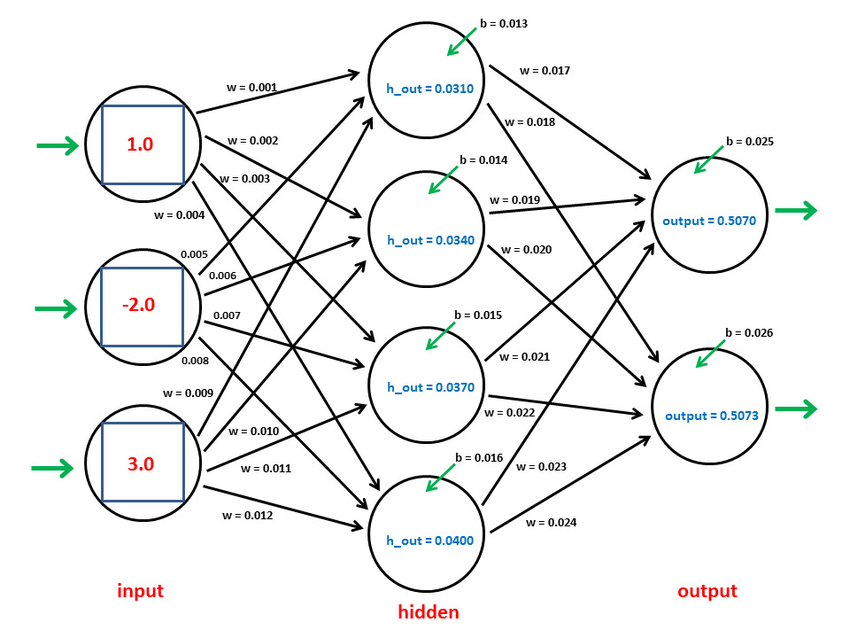
The K-Nearest Neighbors algorithm was introduced in the field of pattern recognition and classification by E. Fix and J. L. Hodges in their report "Discriminatory Analysis—Nonparametric Discrimination: Consistency Properties" in 1951. Since then, it has become a widely used and versatile algorithm for tasks such as image classification, recommendation systems, and anomaly detection, among others.



**Figure (3-2-2) KNN Algorithm**

1. **Neural network:**

A neural network is a computational model inspired by the structure and functioning of the human brain. It consists of interconnected artificial neurons (also called nodes or units) organized into layers. These interconnected neurons process and transform input data to produce meaningful output, often used for tasks such as pattern recognition, classification, regression, and more (McCulloch, W. S., & Pitts, W. (1943).

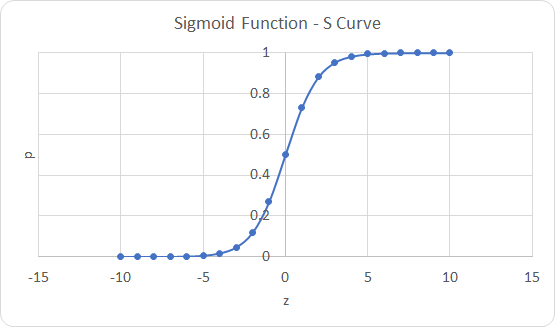


**Figure (3-2-3) Neural Networks Algorithm**

1. **Logistic Regression:**

Logistic regression is a statistical model and a supervised machine learning algorithm used for binary classification tasks, where the goal is to predict the probability of an instance belonging to one of two classes. It models the relationship between a dependent binary variable (the class label) and one or more independent variables (predictors or features) using the logistic function (Cox, D. R. (1958)).

The concept of logistic regression was introduced by Sir David R. Cox in his paper "The Regression Analysis of Binary Sequences" published in the Journal of the Royal Statistical Society in 1958. This method has since become a fundamental tool in statistics and machine learning for binary classification tasks, including medical diagnosis, spam detection, and more.

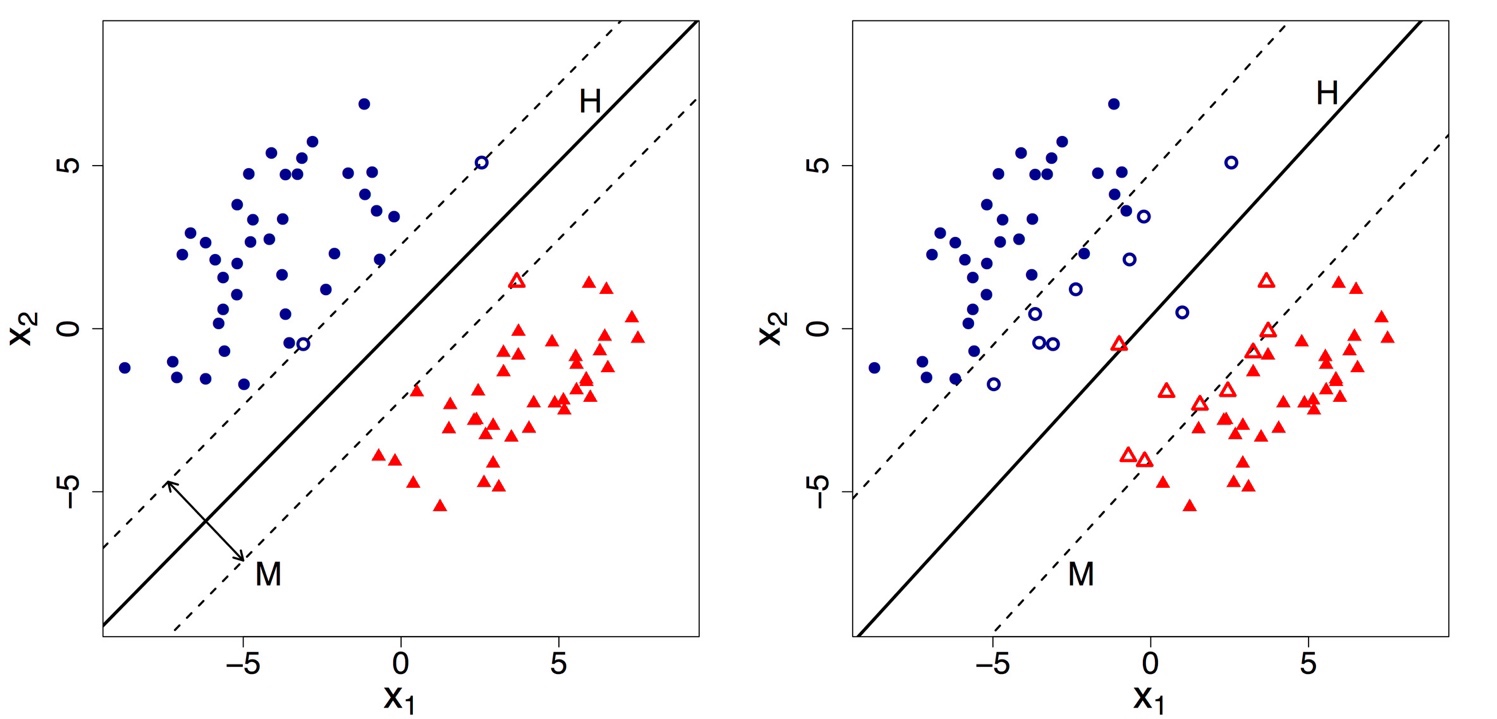


**Figure (3-2-4) Logistic Regression Algorithm**

1. **Support Vector Machine (SVM):**

A Support Vector Machine (SVM) with a linear kernel is a supervised machine learning algorithm used for classification and regression tasks. It is designed to find a hyperplane that best separates data points into different classes in a high-dimensional feature space. In the case of a linear kernel SVM, it seeks a linear decision boundary (Cortes, C., & Vapnik, V. (1995)).

The concept of Support Vector Machines, including SVMs with linear kernels, was introduced by Corinna Cortes and Vladimir Vapnik in their paper "Support-Vector Networks," published in Machine Learning in 1995. SVMs with linear kernels are known for their effectiveness in linearly separable datasets and have been widely used in various applications, such as text classification, image recognition, and more.

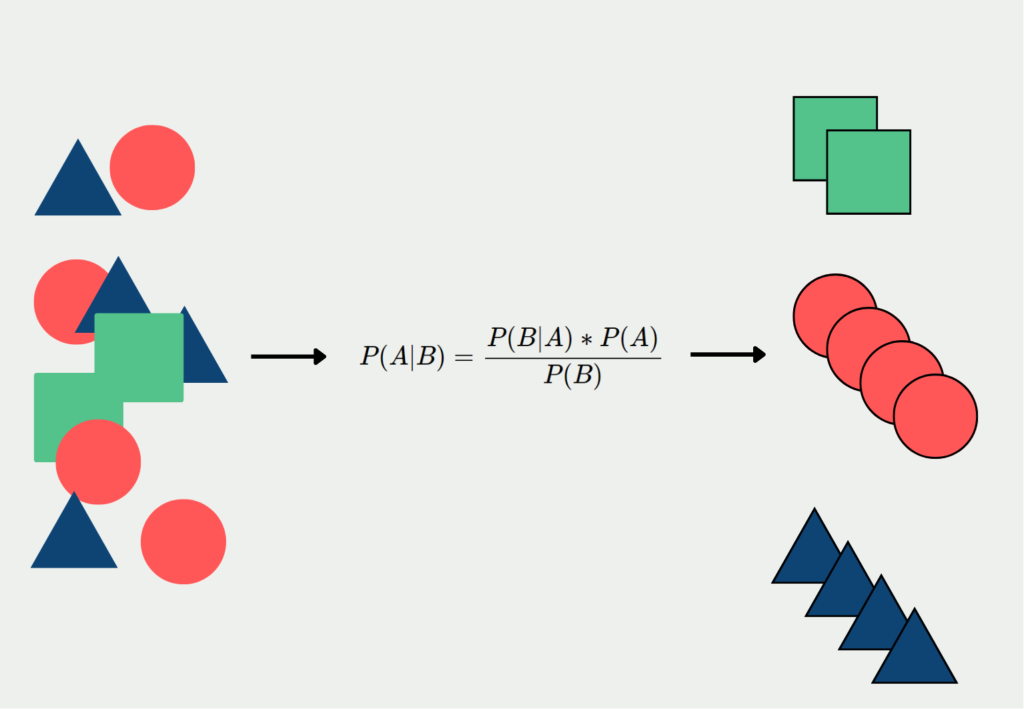


**Figure (3-2-5) Support Vector Machine Algorithm**

1. **Naive Bayes (NB):**

A Naive Bayes classifier is a type of probabilistic machine learning model based on Bayes' theorem with an assumption of independence among predictors. It is used primarily for classification tasks, where it calculates the probability of an instance belonging to a particular class based on the observed features or attributes(Lewis, D. D. (1998)).

The concept of the Naive Bayes classifier, including the assumption of attribute independence, has been widely discussed and studied. A notable reference in this context is the paper by David D. Lewis titled "Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval," presented at the European Conference on Machine Learning in 1998. Naive Bayes classifiers are particularly useful for text classification, spam detection, sentiment analysis, and other tasks where probabilistic reasoning is valuable.



**Figure (3-2-6) Naive Bayes Algorithm**

#### Test And Evaluate Each One

After collecting data and build small project depend on Python and flask framework this diagram for models training process:

Input dataset

Training set (80% of data)

Test set (20% of data)

Fit() the model based on the input of sample data and then predict() training data

[K Nearest Neighbors](https://www.google.com/search?sca_esv=571963393&sxsrf=AM9HkKlw0C2vi0wrDH5ZUUf24hkB8YzR7A:1696880679101&q=K+Nearest+Neighbors&spell=1&sa=X&ved=2ahUKEwjfkuqQ3emBAxVMNewKHXKODxsQBSgAegQICxAB)

Cluster

Logistic Regression

SVM Linear Kernal

Naive Bayes

N Network Logistic Activation

training machine learning predictive models (Individual for each classifier)

Predict unseen test data and calculate accuracy (in %)

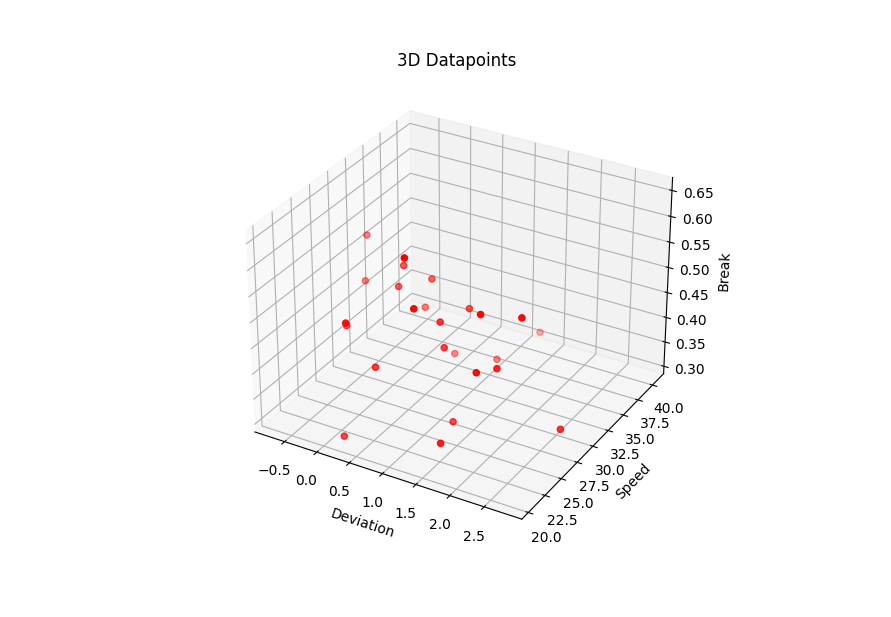
Visualizing output

**Figure (3-3-2) Models Training Process**

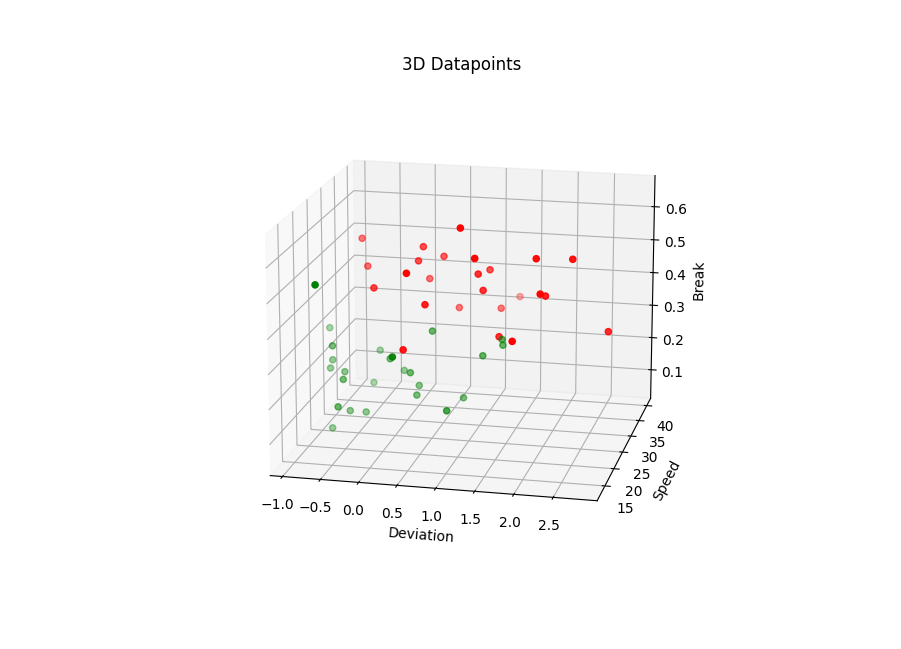
A graph of data points and numbers

Description automatically generated

**Figure (3-3-3) Right Data**



**Figure (3-3-4) Wrong Data**



**Figure (3-3-5) All Data**

This is a 3D Figures show us wrong and right data individually after cleaning data and add label to each group (right, wrong) to test models and export the accuracy (in %) you can see it in next table (3-1).

We extract accuracy from each algorithm individually many times each time try to get high accuracy by cleaning data and avoiding overfitting until get the final results that you can find in the table below:

**notes** “Accuracy can be better if we add more test data.”

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Algorithm** | **Data length** | **Accuracy** |
| 1 | Knn Model | 50 | 0.76% |
| 2 | Cluster Model | 50 | 0.72% |
| 3 | Neural Network Logistic Activation | 50 | 0.90% |
| 4 | Logistic Regression | 50 | 0.80% |
| 5 | SVM Linear Kernal | 50 | 0.80% |
| 6 | Naive Bayes | 50 | 0.90% |

#### Table (3-1): Comparison between Different Machine Learning Models

#### Chose Most Effective and Applicable Algorithm

Depend on previous table (3-1) the most affection model is Naive Bayes, Neural Network Logistic Activation , Logistic Regression and SVM Linear Kernal so I decide to choose **Naive Bayes** and **Neural Network Logistic Activation** to build model in my solution because of fits with my data type .

#### Design Solution Architecture:

#### Join us on this journey as we navigate the practical aspects of transforming a visionary architectural design into a tangible software solution. The implementation of our solution architecture marks a critical milestone in our mission to enhance road safety, optimize traffic management, and contribute to the evolution of intelligent transportation systems.

#### In the following sections, we will provide insights into the technologies, methodologies, and diagrams that underpin our software's implementation, with the ultimate aim of unraveling the intricate patterns of driver behavior using the power of artificial intelligence.

#### Context diagram:

#### This is a context diagram to clarify the level of my solution architecture it starts with an agent in our case is car sensor that sends the data to my solution DPObserver after analysis after that he sends an alert to the police station if the case is dangerous.

**A screenshot of a computer

Description automatically generated**

**Figure (3-3-1) Context diagram**

#### Component diagram:

Now I demonstrate the main components of my solution “DPObserver”, it starts with the first entity car agent pulls notifications from cloud servers connected with the backend that has a link with AI model service and database component,

It gives the last component (police incident) report or alert in some cases the system makes a decision on it.

A diagram of a solution architecture

Description automatically generated

**Figure (3-3-2) Component diagram**

#### Technology diagram:

I used many technologies in this solution and integrate using REST API in the best way to get high value from it and this a list:

Front-end: Angular framework

Backend: NodeJS

AI Model: Flask, sklearn and Panda’s lib

Database: NoSQL (MongoDB).

Realtime: Firebase (cloud service)

Front-end layer (Angular V 15.1)

AI Model Env layer (flask V 2.3.3)

Back-end layer (NodeJS V 18.17.1)

DATA BASE

layer

(No SQL - MongoDB)

**Figure (3-3-3) Technologies diagram**

#### Sequence diagram:

Now this sequence demonstrates the actions happened in system :

A diagram of a software process

Description automatically generated

**Figure (3-3-4) Sequence diagram**

#### Results and Evaluation

#### Introduction:

The results and evaluation section serves as the culmination of our efforts to detect driver patterns using cutting-edge AI models and machine learning models. In this section, we present the outcomes of our research and the performance of our software solution, shedding light on its effectiveness, reliability, and potential impact on various applications within the domain of transportation and automotive technology.

**Dataset Description**

Before delving into the results, it's essential to understand the dataset upon which our software was trained and tested. Our dataset consists of numbers describing the value for 3 main features [deviation, speed, break] in the data structure(2 dimensions array) as we mentioned before in the previous section with a total number of record around 1500 records at least "I know this number may be less than expected but this what we got from simulator we build".

It encompasses diverse driving scenarios and includes data on speed, deviation from lanes, brake usage, and other pertinent driving behaviors.

**Feature Engineering and Data Preprocessing**

Performance Metrics

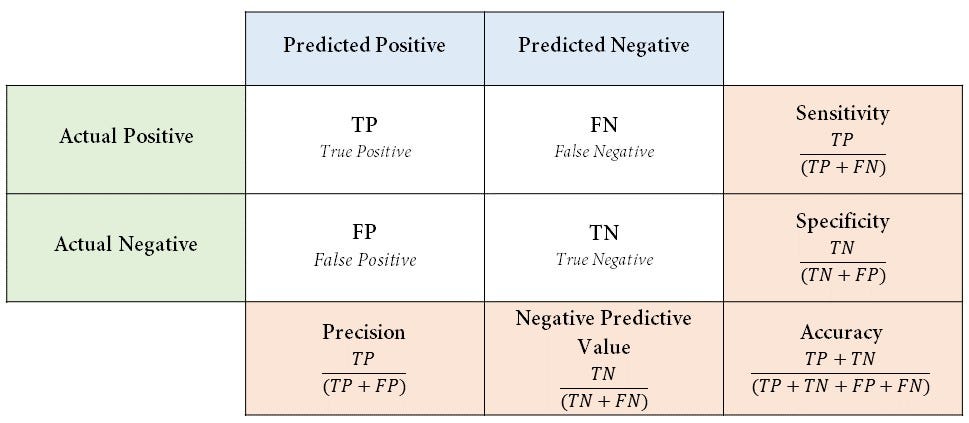
To assess the performance of our models, we employed a comprehensive set of performance metrics, including but not limited to:

Accuracy: The proportion of correctly classified instances.

Precision: The ratio of true positive predictions to the total positive predictions.

F1-Score: The harmonic mean of precision and recall, balancing precision and recall.

Confusion Matrix: A matrix showing true positive, true negative, false positive, and false negative counts.



**Figure (4-1) Confusion Matrix**

**Results**

The results of our driver pattern detection software are promising and reflect the culmination of our research efforts. Our chosen model (Naive Bayes) achieved an accuracy of 90%, indicating its ability to correctly classify driver patterns based on the collected data.

|  |  |  |  |
| --- | --- | --- | --- |
| Users | Results | length | Percentage |
| 43 | Right Detection/Recognize | 36 | 83.73% |
| Wrong Detection/Recognize | 7 | 16.27% |

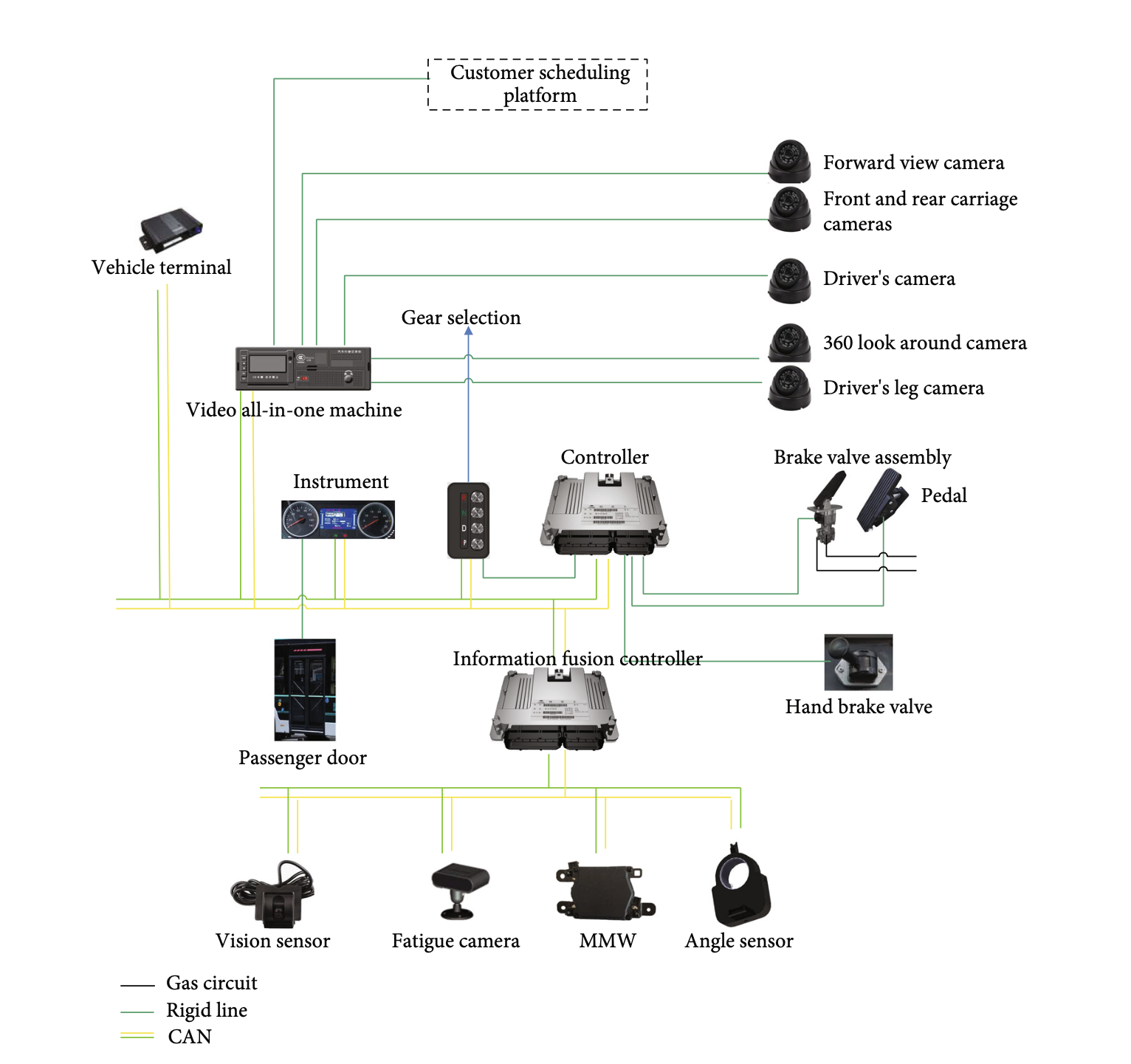
#### Table (4-1): Results of DPObserver Solution

The precision score of 90% demonstrates the model's capability to minimize false positives, ensuring that the patterns detected are indeed reflective of the underlying driver behavior.

Evaluation and Future Directions

While our software solution has yielded promising results in detecting driver patterns, there are avenues for further refinement and enhancement. Future directions for this research include:

the most important for future directions that I propose is the way to move from collecting data from the simulator to collection from real cars actually from the car controller and connected sensors and cameras that run the systems in the car like ( Lane-keep assist, Speed limiter, Automatic Emergency Braking) all controlled by car main controller, this will be my main source of data I will fetch from it in the actual application as next step of my study.



**Figure (3-3-4) The Schematic Diagram Of Controller Structure**

**Data Augmentation:**

Expanding the dataset to include a more diverse range of driving scenarios and conditions to improve model generalization.

Real-time Implementation: Adapting the software for real-time driver pattern detection, potentially contributing to advanced driver assistance systems (ADAS).

Human Factors Analysis: Incorporating human factors research to understand the psychological and cognitive aspects influencing driver behavior.

In conclusion, our software for detecting driver patterns using AI and machine learning has shown significant promise in its ability to classify and understand driver behaviors. The results presented herein lay the foundation for future developments in road safety, traffic management, and intelligent transportation systems, and we remain committed to advancing this crucial research area.

#### Conclusions

This study represents a significant endeavor aimed at addressing one of the most pressing concerns in road safety—detecting abnormal driving patterns resulting from alcohol impairment. Leveraging the power of machine learning models, we have undertaken an in-depth exploration into the potential of data-driven solutions to identify and mitigate the risks associated with alcohol-impaired driving. Our findings and conclusions offer valuable insights into the effectiveness of such approaches and their implications for the broader field of transportation safety.

The Role of Machine Learning

In our pursuit of addressing this critical issue, we employed advanced machine learning techniques to analyze and identify abnormal driving patterns that may indicate alcohol impairment. Our approach involved collecting and preprocessing a diverse dataset encompassing a wide range of driving scenarios, deviation, speed, and brake levels. This rich dataset served as the foundation upon which our models were built.

Implications and Future Directions

The implications of our findings are far-reaching. A machine learning-driven approach to detect abnormal driving patterns due to alcohol effects holds promise not only for law enforcement but also for the development of advanced driver assistance systems (ADAS) and in-vehicle technologies. Such systems could potentially intervene when alcohol impairment is detected, thereby preventing accidents and saving lives.

In conclusion, the detection of abnormal driving patterns due to alcohol effects using machine learning represents a promising stride toward a safer and more responsible road environment. By leveraging data-driven solutions, we have demonstrated the potential to significantly mitigate the risks associated with alcohol-impaired driving, ultimately contributing to a future where our roads are safer for all.

**References:**

García-Escudero, V., Díaz-Morales, J. F., & Roca-Vega, M. (2021). Alcohol consumption and risky driving behavior: A systematic review and meta-analysis of experimental studies. Accident Analysis & Prevention, 154, 106125.

Transport and Research Laboratory(1998), UK International Federation of Red Cross and Red Crescent Societies, World Disasters Report.

National Highway Traffic Safety Administration. 2022. Traffic safety facts 2020. Report No. DOT HS 813 294.

Mandel, D. R., Ginder, S. A., & Lapham, S. C. (2018). The effect of alcohol on driving-related skills: An experimental study. Journal of Safety Research, 64, 1-9.

Waller, P. F., Brown, J. L., & Stutts, J. C. (2019). The effects of alcohol use on driver risk-taking behavior: A driving simulator study. Traffic Injury Prevention, 20(1), 57-62.

D. I. Tselentis, E. I. Vlahogianni, and G. Yannis, “Driving safety efficiency benchmarking using smartphone data,” Transp. Res. C, Emerg. Technol., vol. 109, pp. 343–357, Dec. 2019.

D. I. Tselentis, E. I. Vlahogianni, and G. Yannis, “Temporal analysis of driving efficiency using smartphone data,” Accid. Anal. Prevent., vol. 154, May 2021, Art. no. 106081.

S. Jongen, E. F. P. M. Vuurman, J. G. Ramaekers, and A. Vermeeren. 2016. The sensitivity of laboratory tests assessing driving related skills to dose-related impairment of alcohol: A literature review. Accident Analysis & Prevention 89 (2016), 31–48. https://doi.org/10.1016/j.aap.2016.01.001.

Huiqin Chen and Lei Chen. 2017. Support vector machine classification of drunk driving behaviour. International Journal of Environmental Research and Public Health 14, 1 (2017), 108.

McCulloch, W. S., & Pitts, W. (1943). A Logical Calculus of Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, 5(4), 115-133.

MacQueen, J. B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1(14), 281-297.

Fix, E., & Hodges, J. L. (1951). Discriminatory Analysis—Nonparametric Discrimination: Consistency Properties. Technical Report, USAF School of Aviation Medicine, Randolph Field, Texas.

Cox, D. R. (1958). The Regression Analysis of Binary Sequences. Journal of the Royal Statistical Society. Series B (Methodological), 20(2), 215-242.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273-297.

Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In European Conference on Machine Learning (pp. 4-15). Springer

**ﺟﺎﻣﻌﺔ اﻟﻘﺎھﺮة**

**ﻛﻠﯿﺔ اﻟﺪراﺳﺎت اﻟﻌﻠﯿﺎ واﻟﺒﺤﻮث اﻹﺣﺼﺎﺋﯿﺔ ﻗﺴﻢ ھﻨﺪﺳﺔ ﺑﺮﻣﺠﯿﺎت**

**دراسة تأثير المشروبات الكحولية على أنماط القیادة للسائقين البالغين**

**رﺳﺎﻟﺔ ﻣﺎﺟﺴﺘﯿﺮ ﻣﻘﺪﻣﺔ ﻛﺠﺰء ﻣﻦ ﻣﺘﻄﻠﺒﺎت اﻟﺤﺼﻮل ﻋﻠﻰ درﺟﺔ اﻟﻤﺎﺟﺴﺘﯿﺮ ﻓﻲ ھﻨﺪﺳﺔ ﺑﺮﻣﺠﯿﺎت**

**إﻋـــــﺪاد**

محمد صلاح إبراهيم سالم

د. طﺎرق ﻋﻠﻲ

### ﺗﺤﺖ إﺷﺮاف

د. ﻣﯾرﻓت ﻏﯾث

**50**

م2023