CLASSIFICATION OF DRIVER BEHAVIOURS USING MACHINE LEARNING

RUQAYAH ALAA ZAIDAN

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

UNIVERSITI PENDIDIKAN SULTAN IDRIS

2021

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

CLASSIFICATION OF DRIVER BEHAVIOURS USING MACHINE LEARNING

RUQAYAH ALAA ZAIDAN

DISSERTATION SUBMITTED IN FULFILLMENT OF THE REQUIREMENT
FOR THE DEGREE OF MASTER OF SCIENCE
(MASTER OF RESEARCH)

FACULTY OF ART, COMPUTING AND CREATIVE INDUSTRIES
SULTAN IDRIS EDUCATION UNIVERSITY

2021

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

UPSI/IPS-3/BO 32
Pind: 00 m/s: 1/1

Please tick (✓)	
Project Paper	
Masters by Research	✓
Masters by Mix Mode	
Ph.D.	

INSTITUTE OF GRADUATE STUDIES
DECLARATION OF ORIGINAL WORK

This declaration is made on the 21/01/2020

i- Student's Declaration:

I Ruqayah Alaa Zaidan- M20161000970-Faculty of Art, Computing, and Creative Industry hereby declares that the dissertation/thesis for Master of Philosophy titled "Classification Of Driver Behaviours Using Machine Learning" is my original work. I have not plagiarized from any other scholar's work and any sources that contain copyright had been cited properly for the permitted meanings. Any quotations, excerpt, reference or re-publication from or any works that have copyright had been clearly and well cited.

_____

Signature of the student

ii- Supervisor's Declaration:

I Dr. Suliana Sulaiman hereby certify that the work entitled, "Classification of Driver Behaviour Using One R" was prepared by the above-named student, and was submitted to the Institute of Graduate Studies as a partial / full fulfillment for the conferment of the requirements for Master of Philosophy (By Research), and the aforementioned work, to the best of my knowledge, is the said student's work.

_____ _____

Date Signature of the Supervisor

UPSI/IPS-3/BO 31
Pind.: 01 m/s:1/1

INSTITUT PENGAJIAN SISWAZAH /
INSTITUTE OF GRADUATE STUDIES

BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK
DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM

Tajuk / *Title*: **Classification Of Driver Behaviours Using Machine Learning**

No. Matrik /*Matric No.*: **M20161000970**

Saya / *I* : **Ruqayah Alaa Zaidan**

mengaku membenarkan Tesis/Disertasi/Laporan Kertas Projek (Kedoktoran/Sarjana)* ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut:-
acknowledged that Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) reserves the right as follows:-

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik UPSI.
   The thesis is the property of Universiti Pendidikan Sultan Idris
2. Perpustakaan Tuanku Bainun dibenarkan membuat salinan untuk tujuan rujukan dan penyelidikan.
   Tuanku Bainun Library has the right to make copies for the purpose of reference and research.
3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.
   The Library has the right to make copies of the thesis for academic exchange.
4. Sila tandakan (√) bagi pilihan kategori di bawah / Please tick (√) from the categories below:-

☐	**SULIT/*CONFIDENTIAL***	Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub dalam Akta Rahsia Rasmi 1972. / *Contains confidential information under the Official Secret Act 1972*
☐	**TERHAD/*RESTRICTED***	Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan ini dijalankan. / *Contains restricted information as specified by the organization where research was done.*
☐	**TIDAK TERHAD / *OPEN ACCESS***	

_____

(Tandatangan Pelajar/ Signature)

Tarikh: _____

(Tandatangan Penyelia / *Signature of Supervisor*)

& (Nama & Cop Rasmi / *Name & Official Stamp*)

Catatan: Jika Tesis/Disertasi ini **SULIT @ TERHAD**, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan ini perlu dikelaskan sebagai **SULIT** dan **TERHAD**.

Notes: If the thesis is CONFIDENTAL or RESTRICTED, please attach with the letter from the related authority/organization mentioning the period of confidentiality and reasons for the said confidentiality or restriction.

ACKNOWLEDGEMENT

"بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ"

"*In the name of Allah the most gracious the most merciful*"

Alhamdulillah, first and foremost, praise be Allah, the Cherisher and Sustained of the World and to the Prophet Muhammad (Peace and Blessings of Allah Be Upon Him) who was sent by Allah to be a great teacher to the mankind.

I would like to express my special thanks to my husband, Dr. Bilal Bahaa Zaidan, who has been a constant source of support and encouragement during the challenges of graduate school and life. You are the one who let me finish my degree. I will keep on trusting you for my future am truly thankful for having you in my life.

Special thanks to my supervisor, Dr. Suliana Sulaiman, for his guidance and advice throughout the research, his patience, kindness, interjecting a healthy dose of common sense when needed.

My warmest appreciation to my beloved parents who support me with their love, do 'a support for both moral and financial for my study. Many thanks go to my father Alaa Zaidan Al-joubori for always s being there and never give up in supporting me.

I would like to express my very great appreciation to my brother, Dr. Abdulla Alamoodi who always encourage and support me, this work would not have been possible without his guidance and involvement, his support and encouragement on daily basis from the start of the project till date. Under his guidance I successfully overcame many difficulties and learnt a lot. . For all these, I sincerely thank his from bottom of my heart and will be truly indebted to his throughout my life time.

My heart overflows with gratitude for all my friends for being supportive and understanding. I would like to extend my appreciation to those who involved and give a helpful hand in ensuring the success of this research.

This research would not have come to fruition without all your help and supports.

Thank you. Allah blesses you

ABSTRACT

According to the Malaysian Institute of Road Safety Research (MIROS), over 500,000 car accidents occurred in 2016, making cars an unsafe means of transportation. This research aimed to collect driver behaviour-related data for Malaysian drivers to provide useful insights for Malaysian driving profile and to modulate machine learning for classification tasks. Twenty-one drivers (11 male and 10 female) were studied and compared for their driving style in Lebuhraya Behrang Stesen-tg malim (11 km per driver). Drivers were asked to drive naturally while considering their safety. Two analysis techniques were utilized (i.e. Statistical and Machine Learning-Based). Different conclusions were drawn from each analysis. The number of driving events for each driver was calculated (i.e. aggressive, normal and safe) and statistical tests (i.e. Mean, Standard Deviation, Correlation analysis, One-way ANOVA and T-test) presented significant differences between each driver from the same gender versus their peers from the opposite gender. The statistics were presented per driver, his/her group and a comparison with their peers. For a driver to be considered as aggressive or normal, a challenge was presented because no identification measure existed (i.e. threshold for driving event number to be considered aggressive or normal). However, each driving event was identified based on literature. Finally, it was determined that classifying drivers was possible through their gender but not based on their aggressiveness level. One R Machine learning classifier presented good accuracy at 95.24 % in comparison with j48DecisionTree, Naive Bayes, One R, and SMO-SVM. The implications of the findings of this study suggest male and female drivers tend to drive aggressively. A reason for such mortality can be because of the cadence of front-end car accidents, which is a clear outcome of aggressive driving behaviour (i.e. speeding, braking, etc.). Identifying such behaviour using ML will save lives domestically and internationally

KLASIFIKASI PENGENDALIAN PENGHARGAAN MENGGUNAKAN SATU R

ABSTRAK

Berdasarkan Institut Penyelidikan Keselamatan Jalan Raya Malaysia (MIROS), terdapat lebih daripada 500,000 kemalangan kereta setiap tahun yang menjadikan kereta sebagai pengangkutan yang tidak selamat. Di dalam skop kajian literatur, tidak ada artikel yang menjelaskan sebab di sebalik jumlah kemalangan ini. Kajian ini bertujuan untuk mengumpul data berkaitan tingkah laku pemandu di Malaysia ke arah memberikan pandangan berguna untuk profil memandu di Malaysia. Dengan tambahan untuk memodulasi pembelajaran mesin untuk tugas klasifikasi. Dua puluh satu pemandu (11 lelaki dan 10 wanita) telah dieksperimen untuk gaya memandu mereka di Lebuhraya Behrang Stesen-tg malim (11 KM setiap pemandu). Pemandu diminta memandu secara biasa dengan mempertimbangkan keselamatan mereka. Dua teknik analisis telah digunakan (iaitu Berdasarkan Pembelajaran Statistik dan Mesin). Kesimpulan yang berbeza diambil dari setiap analisis. Bilangan kejadian memandu telah dikira untuk setiap pemandu (contoh: berunsur agresif, normal dan selamat) dan ujian statistik membentangkan kepentingan perbezaan antara setiap pemandu dari kumpulan jantina yang sama berbanding rakan mereka dari jantina yang berbeza. Statistik ini dibentangkan mengikut setiap pemandu, setiap kumpulan dan dengan perbandingan dengan rakan mereka. Bagi pemandu yang dianggap sebagai agresif atau biasa, cabaran dibentangkan kerana tiada langkah pengenalan wujud (iaitu ambang untuk bilangan kejadian memandu yang dipertimbangkan sebagai agresif atau normal). Walau bagaimanapun, setiap kejadian memandu telah dikenal pasti berdasarkan literatur. Pada akhirnya, mengelaskan pemandu berdasarkan tahap agresif tidak dapat dilaksanakan tetapi mengikut jantina mereka. Satu pengelas pembelajaran (ML) Mesin R membentangkan ketepatan sebanyak 95.2381 % yang baik berbanding dengan j48DecisionTree, NaiveBays,OneR,danSMO-SVM.Implikasi kajian ini mencadangkan pemandu lelaki dan wanita yang cenderung memandu agresif. Alasan kematian sedemikian boleh disebabkan oleh turutan kemalangan di bahagian hadapan kereta, yang merupakan hasil yang jelas daripada tingkah laku memandu secara agresif. (contoh: memandu laju, menekan brek, dll.). Mengenal pasti tingkah laku seperti menggunakan ML akan menyelamatkan banyak nyawa di dalam negeri, dan mungkin berjuta-juta nyawa di peringkat antarabangsa.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

RESEARCH BACKGROUND

1.1 Introduction

This chapter explains the research background of this thesis and highlights the different areas and points that contribute to the understanding of the thesis topic. Among the points covered in this chapter are the research background, which discusses the origin of the topic, followed by the statement of the problem which discusses how the problem in this dissertation emerged. Other important highlights are addressed including research objectives, research questions and research scope.

1.2 Research Problem

Intelligent transportation system is considered one of the most important current topics due to several reasons, including improving road safety and reducing traffic congestion among others. This thesis focuses on driver behaviour, which is considered as an important factor in the intelligent transportation system.

Driver behaviour analysis is the science of understanding the behaviour of drivers while driving. This analysis can be conducted by employing statistical techniques to compare or identify the patterns (e.g. acceleration, brake pedal, deceleration, speed, overtake, manoeuvre etc.). Artificial intelligence (AI) is used to classify different behaviour patterns, such as normal and aggressive driving.

Driver behaviour has received considerable attention in several Western countries because understanding the behaviour can reduce road accidents, which is one of the main objectives of the intelligent transportation system (Amsalu & Homaifar, 2016; J. Carmona, F. García, D. Martín, A. d. l. Escalera, & J. M. Armingol, 2015b; Jiménez et al., 2016; Karimi, Zimmerman, Nawn, & Sutovsky, 2010; Kim et al., 2016a; Sato & Akamatsu, 2012; Tada et al., 2014). Understanding the behaviour of drivers can minimize driver distraction (A et al., 2010), make cars as a safe and human friendly product (Craye, Rashwan, Kamel, & Karray, 2016; R. Terada, H. Okuda, T. Suzuki, K. Isaji, & N. Tsuru, 2010), reduce car crashes/accidents (B. F. Wu, Chen, & Yeh, 2014) (Satzoda & Trivedi, 2015) (Shi, Wei, & Shi, 2012) (M. P. Philipsen et al., 2015), reduce the number of deaths and injuries (J. Carmona, Miguel, Martin, Garcia,

& Escalera, 2016) (K. Takeda et al., 2011), improve road safety and transportation efficiency (J. G. P. Rodrigues, F. Vieira, T. T. V. Vinhoza, J. Barros, & J. P. S. Cunha, 2010) (Das, Zhou, & Lee, 2012) (Ohn-Bar, Tawari, Martin, & Trivedi, 2014) (Albert, Musicant, Oppenheim, & Lotan, 2016), mitigate the safety of the cars following (J. Wang, Zhang, Zhang, & Li, 2013) , maintain safe distance headways (Kondyli, Sisiopiku, Zhao, & Barmpoutis, 2015), improve driving performance (Ramyar et al., 2015), analyse drivers' general characteristics in lane-keeping process (Jieyun, Jianqiang, Changchun, Meng, & Keqiang, 2014), reduce energy consumption and mitigate emissions (J. Wang, Xiong, Lu, & Li, 2015), educate drivers on the importance of the mentioned factors (Andria, Attivissimo, Di Nisio, Lanzolla, & Pellegrino, 2016) (J. Engelbrecht, M. J. Booysen, G. J. van Rooyen, & F. J. Bruwer, 2015b) (Castignani, Derrmann, Frank, & Engel, 2015b) and finally avoid uncontrollable driving behaviours (Rakotonirainy, Schroeter, & Soro, 2014).

An analysis of previous literature indicates that studies in this area can be categorised into driver behaviour analysis, data acquisition, review articles and frameworks. The main focus of this research is to analyse drivers' behaviours in Malaysia. Academic literature published in the past ten years shows a lack of studies in Malaysia in this area, but statistics suggest serious concerns need to be addressed for drivers in Malaysia.

1.3 Problem Statement

With the scope of driver behaviour, researchers face several challenges in terms of understanding its patterns. Most of these challenges involve data. The lack of or incompleteness of data is one of the burdens of researchers aiming to develop a universal driver behaviour analysis model (Piotr, Turek, Byrski, & Cetnarowicz, 2015). Meanwhile, the accuracy of the gathered information plays an important role in data analysis to avoid the issue of the lack of trust from the users' side (Bruwer & Booysen, 2015). Other problems include the use of smartphones, which results in low accuracy as compared to approaches that employ the OBD (AbuAli, 2015). Other researchers have suggested the use of smartphones due to its low complexity and cheaper approach when it comes to data collection. The availability of sophisticated products to measure driver behaviour analysis is limited in commercial vehicles. Therefore, driver monitoring and driver behaviour analysis can be quite challenging (Das et al., 2012). The task of analysing the patterns of driver behaviour require huge amounts of data and different analyses among different age groups, gender etc. (Piotr et al., 2015). The stated challenges discussed pertain to the analysis of driver behaviour in general and do not include the context of longitudinal studies wherein driver characterisation is also a challenge when the vehicle shared between different drivers (Wallace et al., 2016). In summary, the cost of data collection, availability of data, the accuracy of data, amount of data and complexity of data collection are the main challenges that researchers face when they aim to identify the patterns of drivers in a particular community.

Previous research articles have suggested using data acquisition method with reasonable cost, accuracy and complexity. However, the higher the value of the data, the better the analysis (J. Carmona et al., 2016). Articles have also suggested a larger number of drivers with different tasks, traffic conditions and roads (Jiménez et al., 2016) (J. Wang et al., 2013). These studies can reduce the deaths and injuries associated with driving and increase road safety. It can also help insurance companies assign different costs based on the drivers' score (Castignani, Frank, & Engel, 2013a).

In Malaysia, car accidents represent serious concerns. Based on MIROS (MIROS, 2017) studies, the average number of car accidents has reached more than 500,000 per year, which makes cars an unsafe transportation method. However, a review of the academic literature with the scope of our survey show that no article has analysed the patterns of Malaysian driver behaviours. Hence, this research is an attempt to develop a framework for driver behaviours in Malaysia. This framework consists of different modules, including data collection, data cleaning, feature extraction, driver profiling, statistical and descriptive analyses and pattern classification. This framework can be used in future research to analyse and understand related topics to Malaysian driver behaviours. To evaluate this framework, a case study on Malaysian drivers that involved longitudinal driving while the system collected their data for use in the analysis and feature extraction was conducted. The results of this study can be used as a guideline for identifying the following behaviours:

- To understand the safety issues, driver profiling (identify the speed, acceleration and deceleration normal and aggressive events) should be a required submodule.

- To understand the similarities and differences of drivers, statistical analysis is the required submodule.
- To understand the relationship between driver behaviours, correlation analysis is the required submodule.
- To modulate machine learning towards recognizing driving patterns, feature extraction and classification submodules are required.

1.4 Research Question

This research conducted to answer the below questions:

1. What is average ratio of aggressive event (e.g. over speed) per kilometre and/or per hour for Malaysian drivers?
2. How is the driving pattern among Malaysian driver behaviour dataset?
3. How to classify Malaysian driver behaviour dataset?
4. What is the accuracy parameter (Recall, Precision and Specificity) of the classifier on the task of Malaysian driver behaviour classification?

1.5 Research Objective

This research attempt to analyses Malaysian driver behaviour with different configuration. Towards this end, the below objectives are to be achieved during this research.

1. To analyses attribute for aggressive and normal events for Malaysian driver behaviour.
2. To analyses driving pattern among Malaysian drivers.

3. To classify male and female drivers based on driving pattern for Malaysian driver.
4. To evaluate the classification based on precision, recall and accuracy.

1.6 Scope of Study

- This research aimed to study the driver behaviour on the longitudinal setting other driving style such as literal driving is not scope for this research
- The selected age group in the study between (Young: 18-35 years and Young adults: 36-55 years).
- The rationale behind the selection of proton highway in Tanjug Malim due two reason one is its short distance were experiment can be made on highway setting, second due to the safety were no much car are in the particular road.
- For the collection of data, the OBD connected to the car and communicated to the mobile to send the data of the speed while the acceleration and deceleration are extracted from the speed features.
- All drivers were subjected to the same car, same trip, same destination, same environment and same variables.

1.7 Operational Definition

Some words and definitions might not be totally clear to some readers, and a clarification for such elements is good to allow the reader to grasp what this words or phrase is intended for. Therefore, this section aims to display and clarify terms and definitions used in this research, all of them are presented in Table 1.1.

Table 1.1

Operational Definitions

List of term and abbreviations	Definition
Intelligent Transportation System	An intelligent transportation system is an advanced application which aims to provide innovative services relating to different modes of transport and traffic management and enable users to be better informed and make safer, more coordinated, and 'smarter' use of transport networks
Driver Behaviour Analysis	It refers to the analysis for the behaviour of drivers by using the data that is collected from the connected vehicles
Universal Driver Behaviour Analysis Model	An analysis Model which can be applied to all car types across all countries
OBD	(On-Board Diagnostics) is a vehicle's self-diagnostic and reporting tool used for data collection and and diagnosis of car. OBD systems also enables access to the status of the various vehicle sub-systems
(PRISMA) Statement	PRISMA is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses.
Search Query	A search query or search term is the actual word or string of words that a search engine user types into the search box
Inclusion Criteria	Characteristics That the Downloaded Articles Must Have If They Are to Be Included in This Thesis
Exclusion Criteria	Characteristics That Disqualify Downloaded Articles from Inclusion in The Thesis
Taxonomy	The Process of Naming and Classifying Articles into Groups Within A Larger Mapping, According to Their Similarities and Differences
Feature Processing	Process of Extracting Features Used in This Thesis and Make It Ready for The Analysis
Driver Profiling	Understanding Drivers Characteristics Based on Their Analysed Data
Data Pre-processing	Steps Included Towards Making Data Ready for The Analysis
Windowing	Process of Average Extraction for Data in Different Level
Aggressive Events	Number of Times Driver Drives Recklessly
Safe Events	Number of Times Driver Drives Safely
TPR	Total Physical Response
Descriptive statistics	are used to describe the basic features of the data in a study.
Statistical analysis	is the science of collecting data and uncovering patterns and trends?
Correlation analysis	is a statistical method used to evaluate the strength of relationship between two quantitative variables
T. test	is one type of inferential statistics. It is used to determine whether there is a significant difference between the means of two groups
ANOVA	is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts

1.8 Limitations of the Study

Despite the high value of this research, it has certain limitations, which are summarized in the following:

- Financial restraints that hinder the ability to gather data on a larger scale
- The search only targeted specific behaviours related to acceleration and deceleration.

1.9 Importance of Research

This research can be considered as the first brick towards advancing transportation in Malaysia, not only from the drivers' perspective, also from the government, particularly in departments involved in designing roads, setting policies and employing different forms of interventions. Others can also benefit from this work, such as those in insurance companies. Academic researchers can also benefit by understanding different cultural contexts concerning drivers, draw casual conclusions, and conduct more studies to make transportation better, safer and more convenie¬t accident-free.

1.10 Summary

This chapter discussed a brief introduction of driver Behavior and intelligent transportation system to show the area in such domain of research. After that, some major components were discussed including the problem statement which addresses the issues and gaps this research is trying to deal with, followed by research objectives and questions which introduces major let points on the things this research will achieve and how so. Research significance was also discussed to show importance of such topic and how it can impact our lives. In addition, the scope of this research was also discussed to show the settings applied in this thesis.