ASHESI UNIVERSITY

DRIVER RISK CLASSIFICATION IN AUTO INSURANCE: USING DEEP NEURAL NETWORKS AND IN-VEHICLE CAN BUS DATA

APPLIED PROJECT

B.Sc. Management Information Systems

Winston Best-Ezeani

2020

DRIVER RISK CLASSIFICATION IN AUTO INSURANCE: USING DEEP NEURAL NETWORKS AND IN-VEHICLE CAN BUS DATA

APPLIED PROJECT

Applied Project submitted to the Department of Computer Science, Ashesi University College in partial fulfilment of the requirements for the award of Bachelor of Science degree in Management Information Systems

Winston Best-Ezeani

May 2020

DECLARATION

I hereby declare that this Applied Project is the result of my own original work and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature:

………………………………………………………………………………………………

Candidate's Name:

………………………………………………………………………………………………

Date:

………………………………………………………………………………………………

I hereby declare that preparation and presentation of this Applied Project were supervised in accordance with the guidelines on supervision of Applied Project laid down by Ashesi University College.

Supervisor's Signature:

………………………………………………………………………………………………

Supervisor's Name:

………………………………………………………………………………………………

Date:

………………………………………………………………………………………………

Acknowledgements

I would like to acknowledge the almighty God and my Family, without whom all of this would not have been possible. Also, many thanks to the numerous friends, staff, and faculty at Ashesi who listened to my ideas and helped flesh them out, including my supervisor Dr. Stephan Nwolley.

I would like to specifically recognize people whose inputs were instrumental. David Sasu for your special insight in Artificial Intelligence, which was consulted during various steps in this project, Lydia Adobea Dampare Addo for your support and providing a network of people in the Ghanaian Insurance space, Dr Esi Ansah for continuous advisory and creating many leads in this project.

Abstract

Among the numerous ways auto insurance companies can contribute to safety on the roads whilst creating more value for themselves and customers is by effectively analyzing individual driver risk level. Every driver's driving style is their fingerprint and when insurance companies can anticipate their style and the risk associated with it, they can create safer portfolios, reward good driving with lower premiums, customize customer offerings and penalize bad driving with higher premiums. A good driver risk assessment lies in identifying and analyzing behavioral patterns in driving. The challenge however with prevalent risk assessment methods in motor insurance are that they rely on non-precise data (age, occupation, address) and their assessment is merely descriptive, leaving little to no detail about the individual nature of the risk a driver might pose. In this project I design a web interface for users and use In-vehicle sensor data that identify driver behaviour patterns and a deep neural network with continuous learning capabilities to analyse and predict the driver's risk based on the data from their vehicle.

Table of Contents

Table of Figures

Chapter 1: Introduction

1.1 Background

Mobility and transportation have become essential economic inputs and are a key factor for economic growth [1]. Road transport constitutes a large proportion of current transport mediums. As a result, vehicle traffic accidents are a serious safety concern. According to a study in 2018 by the world health organisation (WHO), traffic accidents are a lead cause of death for people aged between 5 and 29 years and 1.35 million people die each year due to these accidents. Aside its fatality, problems of traffic congestion, productivity loss, medical, legal, emergency service, insurance and administration costs are immense, including adverse effects on fuel consumption and environmental conditions [2,3]. Government agencies have introduced comprehensive traffic laws, access to post-crash care and safer vehicle standards to improve traffic safety [2].

Auto Insurance and Driver Risk

Auto insurance companies (insurers) also do some heavy lifting by financially protecting policyholders against the costs incurred when these accidents occur. In return, the insurer gets paid a periodic premium for taking up the risk. Relatively, these measures do little to mitigate the risks of driver error – a top cause of traffic accidents [2]. In providing auto insurance, considerations include the type of vehicle, purpose of vehicle usage, vehicle features and category, value of vehicle, driver's age, occupation, demographic, residential area and the length of the insurance contract. After which, standardized models are used to calculate estimates of the premium to be paid [4]. These calculations ignore observable risks like road quality, time of day frequently driven and weather conditions. All these risk factors can be linked to the driver [2]. Hence, assessing the driver risk will significantly improve the precision of risk calculation and help more accurately insure policyholders against accidents.

Driver Behaviour

Driving risk describes the propensity of a driver to cause a crash [4]. This risk has been used in fields of vehicle manufacturing, transportation planning, telematics, insurance, policymaking etc. In usage-based insurance (UBI) and insurance pricing, orthodox calculation models are combined with driving behavior information to more precisely determine the price of auto insurance premiums for drivers with different driving styles [5,6]. In auto insurance, information like a driver's age, occupation and experience were key in analyzing driver risk. However, there is a question of how objective and relevant this driver information is, since adequate representation of driving features and the ability to differentiate between driving patterns is required to effectively assess risk. Good driver risk assessment lies in identifying behavioral patterns in driving [7,8,9]. Every driver has a unique fingerprint represented in this behavioral pattern. They help understand and recognize driver behaviour, identify aggressive driving styles and detect distracted driving [10,11]. Studies of driver behaviour patterns has seen application in numerous fields because they help to detect risky driving styles. In auto manufacturing, driving behaviour greatly influences the design of intelligent transportation systems (ITS) and helps build models for drivers to improve Advanced Driver Assistance Systems (ADASs) and improve vehicle safety.

1.2 Motivation

As use of advanced digital technologies increase, systems generate large amounts of user data. And companies are taking advantage of the high availability of data to improve and customize solutions for customers. This data generation is a motivation for this project. Since data that provides key insights of customer us is available, it would be prudent to analyze the data and use it to solve problems and offer better value. Another motivation is the presence of revolutionary technologies that are changing how we analyze data. One such technology is machine learning. Machine learning is a crucial part of Artificial intelligence and it describes the ability of a system to learn based on data it is trained on. This has birthed predictive and prescriptive technologies that help anticipate problems and profess solutions.

Therefore, the availability of data and machine learning are key motivations for this project and will feature in the proposed solution for the problems identified in the next section.

1.3 Problem Definition

Motor Insurance accounts for 50% of the entire non-life Ghanaian insurance industry and contributed in excess of $100m as at 2016 [12]. However, the industry has and continues to face numerous challenges. Increased road accidents and low penetration, sales of fake insurance policies and premium under-pricing have all undermined the industry's profitability [12,13]. These warranted the intervention of the National Insurance Commission (NIC) – the industry's licensed regulator. Since 2015, the NIC has made 3 major interventions that are changing the auto insurance landscape.

- In 2015, the NIC mandated that all drivers were required to take out third-party coverage [12].
- The NIC announced the Motor Insurance Database (MID) in January of 2020, that will link insurance companies directly with the Driver and Vehicle Licensing Authority –. This will make it possible to immediately look up the insurance status of a vehicle [14].
- In an attempt to introduce a more risk-based standards in Ghana, NIC has required all motor insurers to meet a minimum capital requirement of about $8.7m or a capital

level based on the level of risk on their books – whichever is the greater, by 2021[15].

Auto Insurance companies already assess vehicle risk, but with the introduction of the MID, mandatory coverage for all drivers and a near future risk-based insurance standard, there is a need to equally understand the driver's risk. Therefore, *insurers need a quick and effective way of currently evaluating a driver's risk level when they come to buy or renew their insurance coverage*. Now having the Motor Insurance Database, there is an opportunity to immediately analyse the driver risk and include it in the driver's profile for future references.

1.4 Proposed Solution and Benefits

With the understanding that volumes of driver data will be made available in centralized in the database, I propose a driver risk analysis system. *This system is a web platform that allows insurance companies quickly predict how risky a driver is. It analyses data from the driver's vehicle and classifies the driver on one of three levels – high risk, medium risk and low risk*. The company inputs driving data to the platform. This data is then fed into a deep learning algorithm, which extracts driver performance on all the features. The performance helps classify the driver in one of the risk levels. After which, the driver risk level is displayed. Benefits of the system include:

- Quick driver risk prediction, since the it includes a learning component.
- Optimizes insurance premiums
- Do better vehicle insurance risk assessment for drivers that ensure a safer portfolio for insurers over time.
- More personalized customer service, pricing and opportunity for usage-based insurance

- For auto insurance companies, this means probability of ensuring fewer accidents and claims, and, more importantly, more safety on the road.
- For policyholders, an opportunity to reduce cost of acquiring an insurance coverage.

1.5 Project Objectives

The focus of this contribution is to build an intelligent, analytical system that enables insurance companies identify different driver risk levels. Several important driving signals have been extracted from an automobile's controller area network (CAN Bus) data as driving indicators. Acceleration pedal pressure, boost pressure, brake pedal pressure, vehicle speed, steering wheel angle, and individual wheel speeds are the signals used in this study. The assumption is that classifying drivers into these risk categories will help insurance companies assess risk better and offer more optimized premiums and services to policyholders.

1.6 Overview of Chapters

The rest of this paper is structured as follows. The next chapter covers literature review and related work on modelling driver behaviour. Requirements specification for the proposed system is prepared in chapter 3. Chapter 4 describes the system architecture and data used for the study; testing and results are presented in chapter 5. Finally, chapter 6 presents conclusions and discusses future works.

Chapter 2: Literature Review and Related Work

2.1 Introduction

In this section, I will describe two different approaches researchers have taken in understanding driver behaviour and several applications of driver behaviour analysis in auto insurance.

2.2 Literature Review

This section explains what has been captured in literature. Driver behaviour and pattern study is not a new field of interest. Researchers usually take two broad approaches in understanding driver data – driver data sourcing and driver data modeling [16].

2.2.1 Driver Data Sourcing

Video-image data: this includes data obtained from observing visual parameters which may be images or audio. Modern Automotive technology uses this data for more efficient self-driving technologies. Using this data for learning is unfavored due to the poor amount of training data.

Simulation data: this involves data gotten from driving simulators. Researchers us them to observe driver behaviour under certain conditions. In these simulators, some conditions are kept constant so that researcher can isolate the condition under study. It has found major use in studying high-risk conditions that can result in fatalities or injuries. However, simulator studies are difficult to generalize, Also, as the same experiment is repeated with a participant in a simulator study, the effect of learning becomes a confounding factor [17].

Smartphone multi-sensor data: Smartphones come equipped with various sensors that are helpful in observing data about driver behavior. The sensors that are already exist in smartphones include motion sensor and position sensor types like accelerometer, gyroscope, linear accelerometer, GPS, manometer, and rotation. These sensor data have been crucial in mapping and routing applications and have also found use in applications for road mapping, road quality detection and telematics. It is therefore not far-fetched that it has been applied in studying driver behaviour, driving patterns. In [18], researchers considered the effect of gravity in modeling smartphone sensor data, which ensured good generalization ability for identifying driver behaviour.

In-vehicle multi-sensor data: Unlike smartphone sensor data that largely relies on device a driver carries; a rich source of information is processed by vehicle sensors and available for external services [19]. Due to its precision, availability and real-world application in the automotive industry, data from the in-vehicle sensors has been considered an efficient method for analyzing driver behaviour [7,19]. The data obtained from the CAN Bus of the vehicle. The data although having a high dimensionality can be filtered and diluted to extract the most meaningful focal points for driver behaviour recognition.

2.2.2 Driver Data Modelling

Machine learning models can help in quickly analyzing driver data. This is because they can easily recognize patterns and classify drivers based on these patterns. Examples of popular machine learning models that are used include: Convolutional Neural Networks (CNN), Multilayer perceptron (MLP), Decision trees (DT), Fuzzy-neural-network (FNN), K-nearest neighbor (KNN). Some statistical models like Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM) are also implemented in machine learning.

Researchers have using and mixed these different methods for modeling driving data to understand driver behaviour. In [20], *Nobuyuki Kuge, Tomohiro Yamamura and Osamu Shimoyama, Andrew Liu* showed that driver behavior modeling and recognition from lane changes is possible using hidden Markov models (HMMs). Also, in [18] *Jun Zhang,* *Zhongcheng Wu, Fang Li, Jianfei Luo, Tingting Ren, Song Hu, Wenjing Li, and Wei Li* fused deep Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) with an attention mechanism to automatically learn features of driving behaviors and model temporal features without professional knowledge in features modeling. Finally, *Pongtep Angkititrakul, Chiyomi Miyajima,and Kazuya Takeda* in [21] modeled driver behaviour in anticipating car-following by using a Gaussian mixture model (GMM) framework.

2.3 Related Works

This section cites examples of driver behaviour applications for auto insurance.

Cambridge Mobile Telematics (CMT) is a world-class company that leverages accurate data from IoT and mobile sensing technologies to identify phone distraction, classify drivers or passengers, recognize speeding and hard braking. Its mobile app, ***DriveWell*** is built to be user-friendly and engaging. They also found a creative way of getting driving data and helping drivers become better through gamification. With gamification, drivers get instant feedback about their driving and are exposed to gamification and incentives that change their behaviour [22].

Avis SafeDrive is an app by rental company, Avis Rent A Car that protects the driver by leverages functions like; ImpactAlert, panic button and weather alerts. They provide a Velcro strip smart sticker placed on the windscreen that interacts with the app. The sensors in the sticker collect driving data, which enables the company to understand more about the driver's style and behaviour during rental. Drivers who drive neatly, gain rewards through the app [23].

Other generic examples include automation and process improvement, improving underwriting, fraud detection and prevention, claims processing etc.

Chapter 3: Requirements Specification

3.1 Project scope and overview

This applied project is aimed at building a simple web platform, that outputs the risk level of drivers seeking an auto insurance coverage. The web application takes user inputs, analyzes and generates expected driver risk level. In order to efficiently design this system and address all the needs of its users, requirements must be elucidated. This chapter details the specific user and system requirements necessary for design and implementation of the web platform.

3.2 Requirements gathering

In gathering the requirements for this project, I relied on interviews with six industry experts and secondary research. Also, I consulted a publication on software requirements for machine learning applications published in ISO 26262 [24]. ISO 26262 is a risk-based safety standard for software applications. The requirements were answers to the following questions.

- How do auto insurance companies calculate driver risk level?
- How long does it take to do the driver risk assessment?
- Are the risk results reliable?

3.3 Requirements Analysis

3.3.1 Functional User Requirements

The user requirements describe expectation of the people who will use the proposed system. These requirements entail user operations and frontend interactions with the system. Some of these requirements were collected from the current process of assessing risk. The primary users of this web platform are auto insurance companies. There would be two types of user, the admin and a regular user. Admins are the people who work directly with the user data, whilst regular users any other staff in the insurance company that wants to use the data. The following are the user requirements:

- All users expect the system to be user friendly
- An admin user can login to the system
- An admin user can input driver information.
- An admin user upload driving data file.
- An admin user can generate and view a driver risk level
- An admin can export the driver risk level result
- A regular user can view drivers and their risk levels

Use Cases

The use case diagrams below show the visual interactions between the users and the system. Figure 2.1 outlines the activities of admins and Figure 1 outlines the list of activities the regular users engage in.

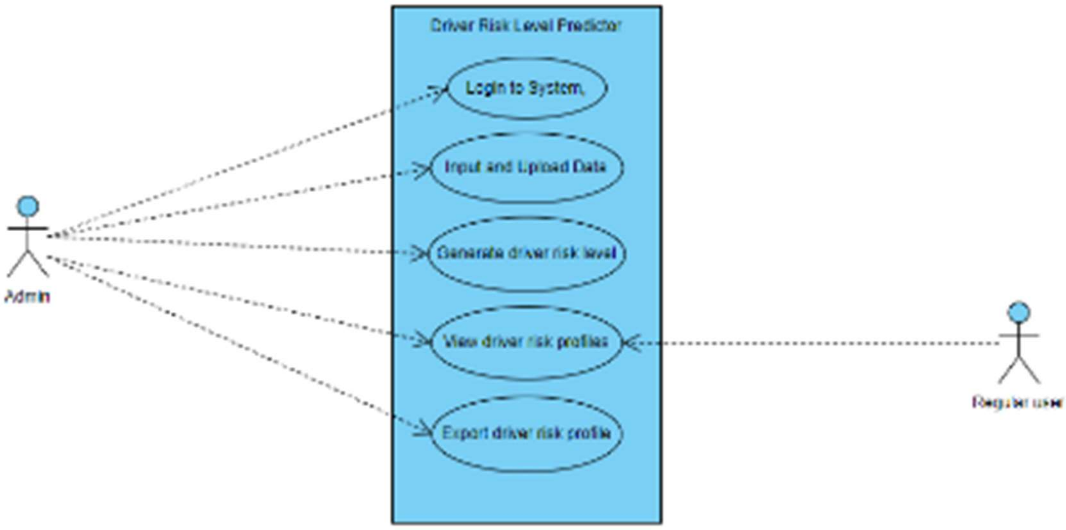

Figure 1: Use case diagram of Web Application

3.3.2 Functional System requirements

The system requirements represent technical features associated with the use of the system. The system defined here comprises of the web application and the deep learning model that performs the core function of predicting driver risk through analysis of driver input data and the applications of user requirements through the technologies. The system should:

- Support all kinds of browsers.
- Be able to store data persistently
- Allow input of driver information
- Control data access based on the user authorization
- Generate driver risk level
- Allow view of driver risk profiles

3.3.3 Non-functional requirements

Non-functional requirements are underlying attributes that are necessary for efficiency of the system. In [25], the non-functional requirements considered relevant for the insurance application and hence this project are accuracy, confidentiality, integrity, interoperability, security and usability. .

- Usability: The system should be implemented using a simple, user-friendly and responsive framework.
- Security: Authentication must be required to access certain system functions. Admin user passwords should be encrypted to ensure they are secure.

- Accuracy: The model used in the machine learning algorithm should ensures high levels of accuracy and allow for continuous learning and even further increase in accuracy of driver risk level predictions.
- Integrity: All driver numeric data must be accurate to a fixed decimal place for all entries.
- Interoperability: The system must be able to interact with any HTML (HyperText Markup Language) browser.
- Confidentiality: The system shall only allow assess to driver risk level data after the patient has bought an insurance coverage.

Chapter 4: System Architecture

This chapter explores the structure and architecture of the system. Details on the architecture model, system modules, database design and the user interactions are given.

4.1 System Architecture

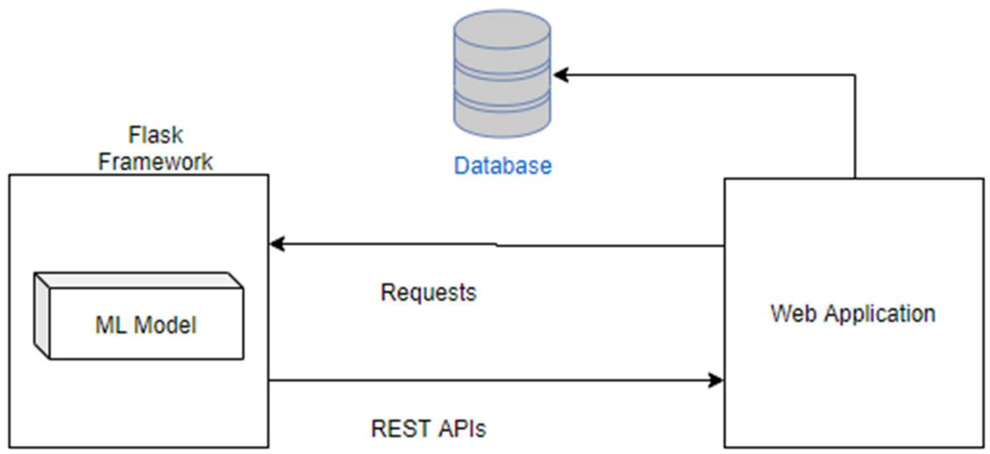

Figure 2: High Level System Architecture

This system architecture can be represented in an MVC architecture design and a data centered architecture design.

4.1.1 MVC Architecture

The MVC architecture has three logical components: The Model, View and the Controller. It is a standard architectural pattern that has been widely used in industry because it allows scalability. Also use of MVC saves time because developers can work on different modules at the same time, resources are used effectively, modification does not affect entire model, the platform is SEO friendly. The model interfaces with the database system and manages all the data stored in the system, the view displays and defines how the data retrieved from the database and presented to the users, the controller is the request handler that manages user interactions with the database operations. The use of MVC architecture helps with modularity during code build. The MVC architecture of this system is shown below.

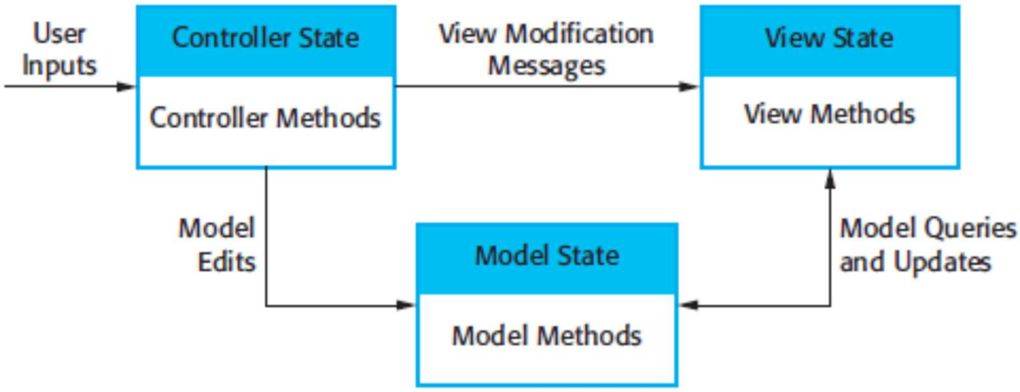

Figure 3: MVC Architecture Function

4.1.2 Data Centered Architecture

There are several modules in the driver risk prediction system. These modules are the data input module, the prediction module and the display module

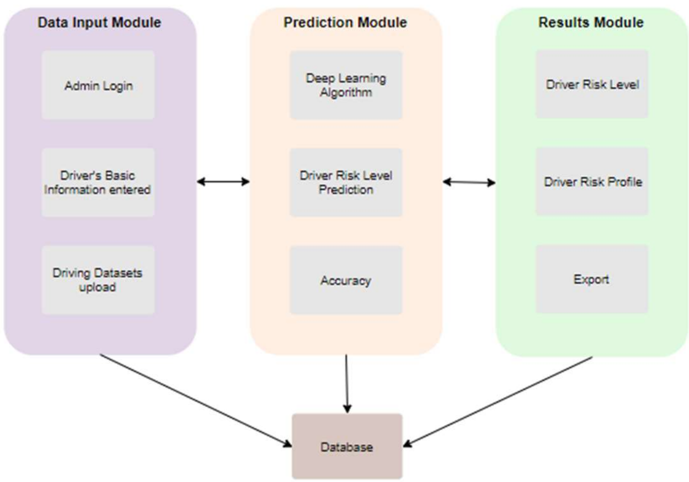

Figure 4: Data Centered Architecture

Data Input Module

Here the driving dataset is uploaded along with some driver basic information. The data input model sends the driver information to the database, where it is stored. And the driving dataset is passed to the prediction module.

Prediction Module

This module contains the trained deep neural network and learning algorithms. The driving dataset is passed into this module and the module outputs the driver risk level and accuracy. The prediction module also sends the risk level to the database, which creates a driver risk profile

Results Display Module

This module displays results from the prediction. From this module, the user can export and or view the result depending on their authorization level.

4.2 Database Design

Below is the entity relationship diagram of the database

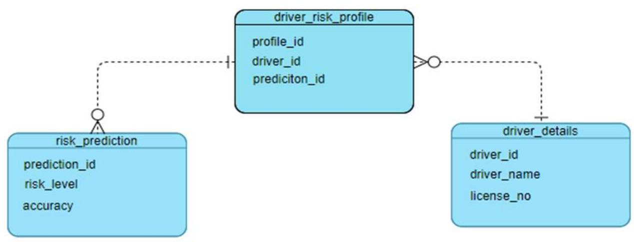

Figure 5: Driver Database Design

4.3 Activity Diagrams

The activity diagrams show the step by step flow of user interactions with the system. The driver risk prediction system has interactions with two users.

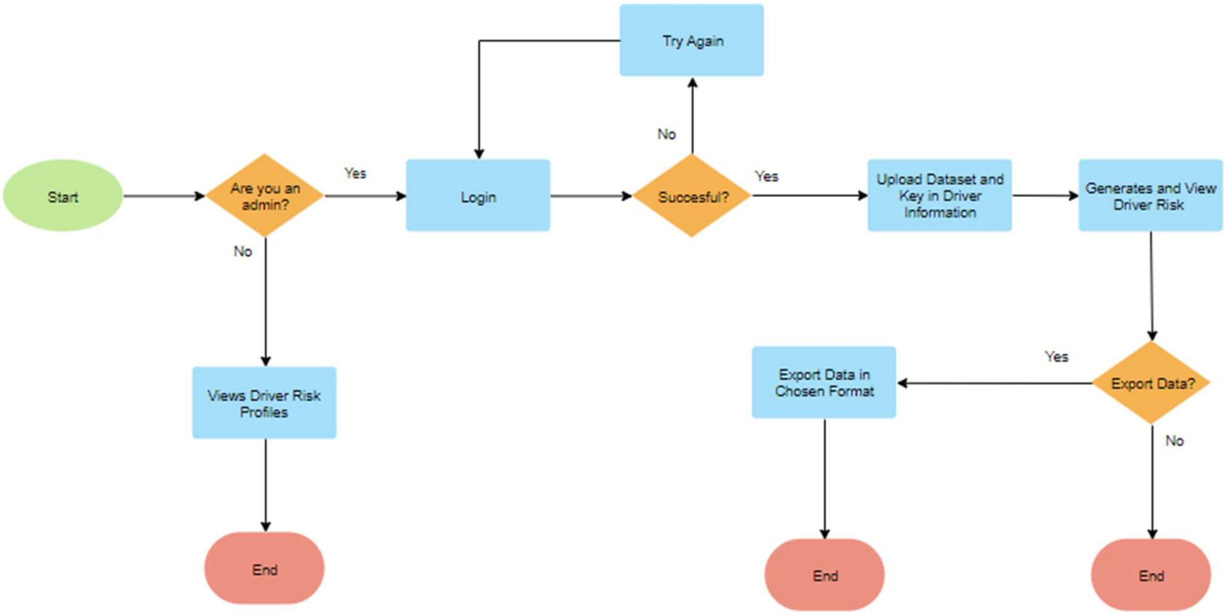

Figure 6: User Activity Flow Diagram

Admin activity

In the admin interface, the admin will be able to login and input some driver basic information and then upload their driving data. After data analysis, the driver risk is printed to the screen for the admin to view. The admin can choose whether to export the driver risk profile.

Regular user activity

Here, the user can only view the existing driver risk profile records from the database. When the user clicks view, all the records are printed to the screen.

Chapter 5: Implementation and Testing

5.1 Overview

This chapter describes in detail, the various implementation techniques and considerations that influenced the choice of technologies and tools used in the development of the system. It shows the interfaces and code snippets used in implementation. Also, the testing of the application is also detailed here. Below is a diagram of the technologies used for implementing the system.

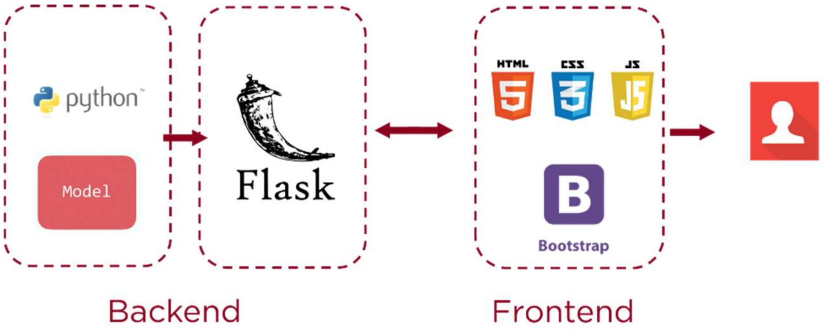

Figure 7: Implementing System Architecture

5.2 Implementation Tools and Technologies

This section contains details on all the software tools and technologies used to develop the prediction system. These include languages, frameworks, libraries and databases.

5.2.1 Languages

For this implementation, Python was a core technology because of its popularity in data analytic tools. Also, Python is lightweight and contains very useful libraries for machine learning applications including building deep neural networks.

- Python
- HTML
- CSS

- JavaScript

5.2.2 Frameworks

In this project, several key frameworks and libraries were used at all stages of creating the driver risk prediction web app. For the fronted, a simple Bootstrap front-end framework was employed. Bootstrap has proven to be efficient and resourceful in implementing aesthetic and responsive web apps.

- Bootstrap4
- Flask3
- REST API

5.2.3 Libraries

As stated earlier, Python was used extensively in this project because it contains various libraries for statistical analysis and machine learning applications. Other libraries used include

- JQuery and AJAX
- Pandas
- NumPy
- Matplotlib
- Scikit-learn

5.2.4 Database

- MySQL

5.3 Implementation

5.3.1 In-Vehicle CAN-BUS Sensor Data Collection

Using CAN bus data in driver behaviour analysis is not a new development. In [11], from driving behaviour, distraction detection from performing secondary tasks and driver identification from observed driving behaviour characteristics were observed by analyzing CAN Bus signals. Also, [10] used CAN Bus data to identify moderate and aggressive behaviors of drivers before a turn. The CAN Bus dataset used in this implementation is provided by the Virtual Vehicle Research Center, Graz, Austria. The data was sampled from a passenger vehicle in 20Hz. This data is obtained from the vehicle and contains several signals that are useful in interpreting a driver's risk.

5.3.2 Data Preprocessing

In this step, the CAN Bus data is prepared and cleaned by removing duplicates merging columns and encoding the required data types. Also, the relevant features (signals) are selected and processed data is exported in a csv file

Data Preparation

The CAN Bus data comes in a .hdf file format. This format stores data in a hierarchical structure and is used to store large amounts of data. The storage system is relative to a python dictionary whose values are also dictionaries. As it is, the data is highly dimensional.

```python

#The h5py library is the python library that is responsible for
#helping us to read the data in the .hdf file
import h5py
from itertools import zip_longest
import csv
import numpy as np

#The .hdf file that we are dealing with in this code has the filename 'Driver1.hdf'
filename = "Driver1.hdf"

#The following line of code is to read the data from the .hdf file.
with h5py.File(filename, "r") as f:
    a_group_key = list(f.keys())[1]

    print("step 1: file extraction complete")

```

Figure 8:Data preparation from the CAN Bus

Data Cleaning and Feature Selection

Once the data was obtained from the original data file, the needed signals were extracted, encrypted to the appropriate data type, combined in a matrix and the right features were selected.

```python

#1. This block prepares the chosen data signals into numpy arrays

    #PEDAL
    pedal_data = list(f[a_group_key]['AccPedal'])
    ped_array = np.array([])
    ped_array.astype('float32')
    for i in pedal_data:
        data = i[0]
        ped_array = np.append(ped_array,data)

```

```python
    #(Selected Features)
    # zip_longest() returns the longest and adds a fillvalue to empty slots
    combine_signal = zip_longest(ped_array, brake_array, vspeed_array, steer_array,
    wspeed_fl_array, wspeed_fr_array, wspeed_rl_array, wspeed_rr_array, fillvalue=0)

```

Figure 9: Extracted features

After the final data is obtained, it is exported into a csv file to be used in building the neural network for the machine learning model.

5.3.3 Building the Neural Network and Prediction Model

In the raw model build, relevant libraries are imported, the input data that was preprocessed in the previous section is split into training and testing sets, which measures the accuracy of the machine learning model. was created. The neural network is a deep neural network, which utilizes an MLPClassifier. The MLPClassifier implements a backpropagation algorithm that represents the learning ability of the neural network [26].

```
# Data loading
data = pd.read_csv('car2.csv')

df_x=data.iloc[:,:8]
df_y=data.iloc[:,-1]
#print(df_x)

# Splitting training and test set
x_train, x_test, y_train, y_test = train_test_split(df_x, df_y, test_size=0.2, random_state=4)

# Creating neural network model
risk_model=MLPClassifier(activation='logistic', solver='sgd',hidden_layer_sizes=(8,8,8),random_state=1, learning_rate='adaptive')

# Fitting model with training data
risk_model.fit(x_train, y_train)

# Predicting with testing data
predict_test = risk_model.predict(x_test)
```

Figure 10: Building the Neural Network basics

```
===========================================================================
The values were correctly predicted 0.8659397453674524% of the time
===========================================================================
```

Figure 11: Prediction Score

5.3.4 Integration with Web App

To integrate the model with a web app, all the logic for our model is stored in a python file and encapsulated into a python class. We then load the trained model into the class method. The pretrained model object is then serialized into a pickle object to convert it into a character stream so that we can later reconstruct the object in another python script. This object is what is used for prediction in the web application.

After this, two methods are created, predict_risk() and predict_risk_api(). These methods are the REST APIs that handle requests from the web application and are hosted on Flask framework. The return object from both REST APIs are json objects of the results from the machine learning model. These results are then sent back to the web application

5.3.5 Web App Pages

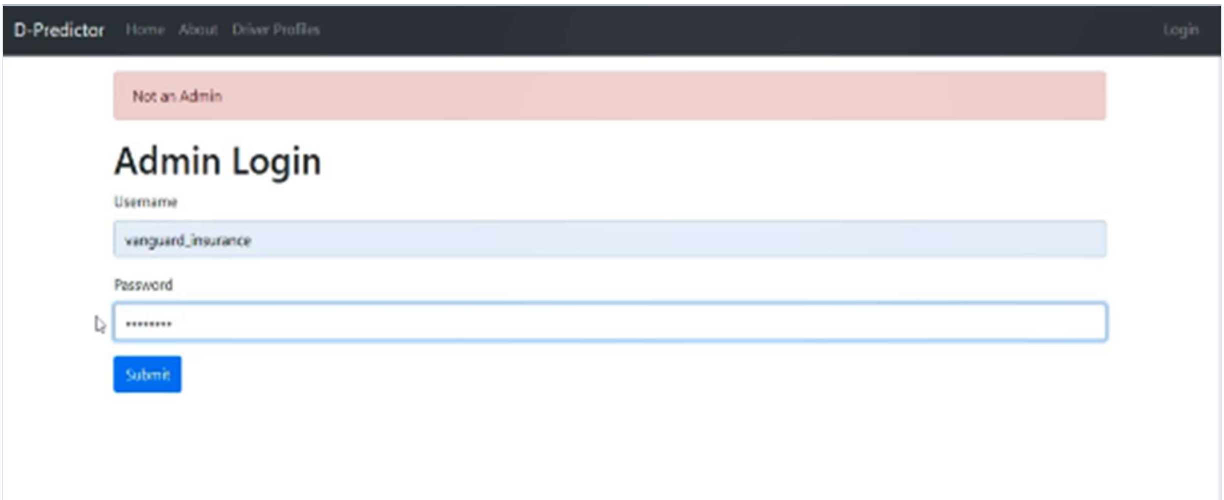

Figure 12: Admin login page

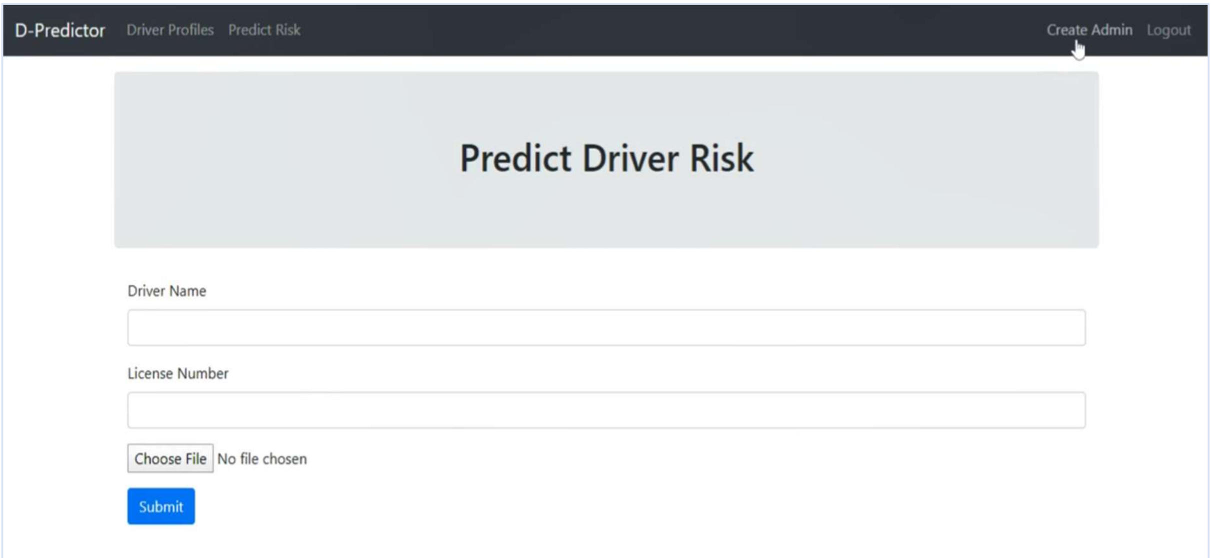

Figure 13: Driver data input page

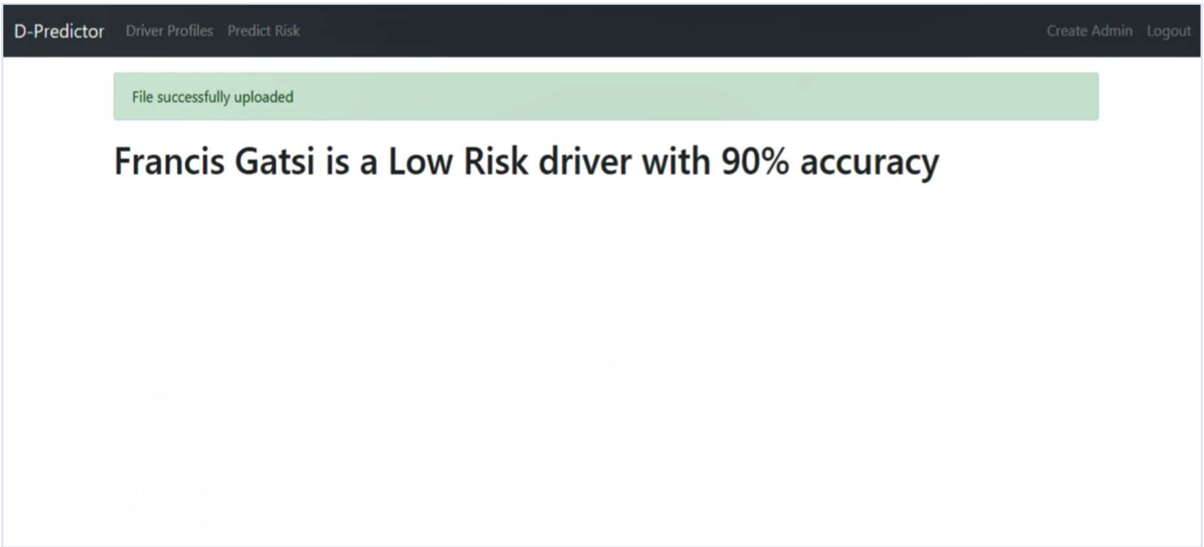

Figure 14: Risk prediction page

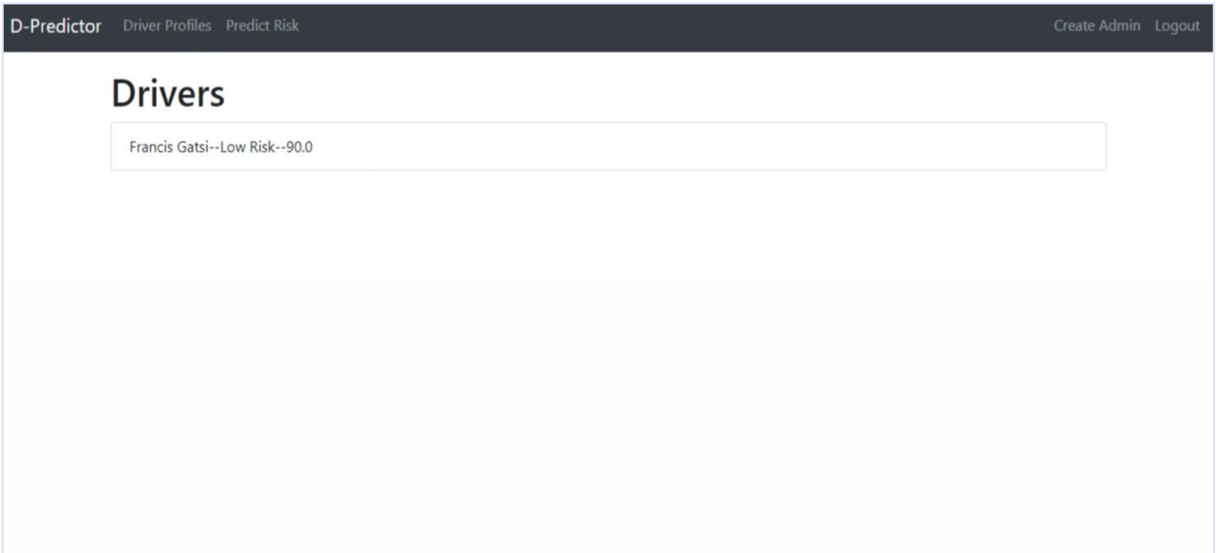

Figure 15: View driver risk profiles page

Chapter 6: User and Development Testing

6.1 Unit Testing

For unit testing, individual functions were tested with special case input parameters. For the web program, login functionality (email and password) was tested to make sure the required inputs were enforced. As the inputs were submitted, authentication was also tested to make sure that non-admin users could not bypass. Driver information input functionality was also tested. This was successful as well but will need some editing. This is because the license number field requires communications with the Ghana Driver and Vehicle Licensing Agency (DVLA). For the machine learning model, the trained model was tested with other driver data to ensure that the data was not overfitted. This was successful because prediction accuracies from the tested datasets recorded similar values with the original training data.

6.2 Component Testing

Component testing assessing the errors, bugs that surface as a result of integrating the various units of the application. Testing whether the model was predicting the right values, the REST API's were tested to see if the driver information was passed correctly passed. This test was successful. Component testing was also done to ensure that a driver cannot have two risk level predictions in the view page. The successful test shows that if a driver's risk level should be predicted and the driver exists in the database, the new prediction replaces the one that exists.

6.3 System Testing

Compatibility, interaction design and data flow were ensured during component testing. This was done by running the application on different browser, deploying the system on Heroku. Heroku is a cloud platform that enables developers operate and deploy apps. The Flask hosted web app successfully deployed to Heroku and the application ran well on 4 different web browsers.

6.4 Specific User Tests

During the user testing stage, the system was made available to an insurance firm. The company tested the system and commended the functionality. Security with login was tested and performed well. Prediction for sever

Chapter 7: Conclusion and Recommendations

The conclusions of this study are representative and useful for the Ghanaian motor insurance industry, but they do not present a generalized character, therefore they cannot be applied to all portfolio or insurance companies. On one side, this aspect is justified by the data used and the risk factors considered during the analysis process, meaning that every insurer can use different information on the insured to their benefit. On the other side, the used data is not obtained through a random selection related to the entire population of policyholders. The web platform is simple, and the neural network model effectively predicts a driver's risk level when secondary driver data is sourced. However, when primary car data is sourced, the risk level is not fully reflective of the driver's style and records accuracy levels of 60-70%. However, since every time a driver uploads driver data, the model is trained, the probability of this accuracy increasing as more datasets are uploaded is high.

Limitations and Constraints

There are several constraints with this system and some limitations of the system that can updated in future works.

1. The neural network model only accounts for driver risk when the vehicle is in motion. Any stationary activities like turning the steering wheel several times at a traffic light, revving the engine and excessive braking can skew the data and affect the credibility of the prediction.
2. This solution does not tackle the lack of education and misunderstanding of Auto Insurance Policies.
3. This solution does not offer a pricing formula. It just helps objectively analyses the risks to make pricing more precise and based on the driver.

Future Work

In future, this platform can use the neural network model on mass driver analytics since the Ghanaian government is working on a Motor Insurance Database to harmonize vehicle insurance data. Create a centralized database system that includes data from DVLA, the police and Insurance companies. Also, there are opportunities for future implementation of this system in fleet management. This project can be further enhanced in the future by:

- Including other sources of driver data; like smartphone sensor data.
- Using a LSTM machine learning model for more exploratory analytics of the CAN Bus data.
- Building a mobile app/USSD component of this prediction system.

References

[1] Filip, Nistor & Popa, Catalin. (2014). The Role of Transportation in Economic Development. "Mircea cel Batran" Naval Academy Scientific Bulletin. XVII. 10-12

[2] Global status report on road safety 2018 https://www.who.int/publications-detail/global-status-report-on-road-safety-2018

[3] The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised) https://trid.trb.org/view/1311862

[4] Mihaela David. 2015. Auto Insurance Premium Calculation Using Generalized Linear Models. Procedia Economics and Finance 20, (2015), 147–156. DOI:https://doi.org/10.1016/S2212-5671(15)00059-3

[5] Frederick F. Cripe and Stephen Fiete. 2011. Usage-based insurance cost determination system and method. Retrieved May 11, 2020 from https://patents.google.com/patent/US7937278B1/en

[6] Siniša Husnjak, Dragan Peraković, Ivan Forenbacher, and Marijan Mumdziev. 2015. Telematics System in Usage Based Motor Insurance. Procedia Engineering 100, (2015), 816–825. DOI:https://doi.org/10.1016/j.proeng.2015.01.436

[7] Saad Ezzini, Ismail Berrada, and Mounir Ghogho. 2018. Who is behind the wheel? Driver identification and fingerprinting. *J Big Data* 5, 1 (December 2018), 9. DOI:https://doi.org/10.1186/s40537-018-0118-7

[8] Jun Zhang, ZhongCheng Wu, Fang Li, Chengjun Xie, Tingting Ren, Jie Chen, and Liu Liu. 2019. A Deep Learning Framework for Driving Behavior Identification on In-Vehicle CAN-BUS Sensor Data. Sensors 19, 6 (March 2019), 1356. DOI:https://doi.org/10.3390/s19061356

[9] Pengfei Li, Jianjun Shi, and Xiaoming Liu. 2017. Driving Style Recognition Based on Driver Behavior Questionnaire. OJAppS 07, 04 (2017), 115–128. DOI:https://doi.org/10.4236/ojapps.2017.74010

[10] M. Zardosht, S. S. Beauchemin, and M. A. Bauer. 2018. Identifying Driver Behavior in Preturning Maneuvers Using In-Vehicle CANbus Signals. Journal of Advanced Transportation 2018, (November 2018), 1–10. DOI:https://doi.org/10.1155/2018/5020648

[11] SangJo Choi, JeongHee Kim, DongGu Kwak, Pongtep Angkititrakul, and John H L Hansen. Analysis and Classification of Driver Behavior using In-Vehicle CAN-Bus Information. 6.

[12] Oxford Business Group. 2019. Economic expansion supports growth in Ghanaian insurance market. *Oxford Business Group*. Retrieved May 8, 2020 from https://oxfordbusinessgroup.com/overview/premium-growth-sector-well-positioned-grow-back-continued-economic-expansion

[13] Middle East Insurance Review. 2020. Ghana: Digital motor database stems motor underpricing. *Middle East Insurance Review*. Retrieved May 8, 2020 from https://www.meinsurancereview.com/News/View-NewsLetter-Article?id=49961&Type=Africa

[14] National Insurance Commision Ghana. 2019. NIC to have Central Database for Authenticating Insurance Policies. Retrieved May 8, 2020 from https://nicgh.org/news/nic-to-have-central-database-for-authenticating-insurance-policies/

[15] Middle East Insurance Review. 2020. Ghana: Insurance regulator increases minimum capital requirements. *Middle East Insurance Review*. Retrieved May 8, 2020 from https://www.meinsurancereview.com/News/View-NewsLetter-Article?id=47208&Type=Africa

[17] Panel on Research Methodologies and Statistical Approaches to Understanding Driver Fatigue Factors in Motor Carrier Safety and Driver Health. 2016. Commercial Motor Vehicle Driver Fatigue, Long-Term Health, and Highway Safety: Research Needs. National Academies Press, Washington, D.C. DOI:https://doi.org/10.17226/21921

[18] Jun Zhang, Zhongcheng Wu, Fang Li, Jianfei Luo, Tingting Ren, Song Hu, Wenjing Li, and Wei Li. 2019. Attention-Based Convolutional and Recurrent Neural Networks for Driving Behavior Recognition Using Smartphone Sensor Data. *IEEE Access* 7, (2019), 148031–148046. DOI:https://doi.org/10.1109/ACCESS.2019.2932434

[19] Gorkem Kar, Shubham Jain, Marco Gruteser, Jinzhu Chen, Fan Bai, and Ramesh Govindan. 2017. PredriveID: pre-trip driver identification from in-vehicle data. In *Proceedings of the Second ACM/IEEE Symposium on Edge Computing - SEC '17*, ACM Press, San Jose, California, 1–12. DOI:https://doi.org/10.1145/3132211.3134462

[20] Nobuyuki Kuge, Tomohiro Yamamura, Osamu Shimoyama, and Andrew Liu. 2000. A Driver Behavior Recognition Method Based on a Driver Model Framework. 2000-01–0349. DOI:https://doi.org/10.4271/2000-01-0349

[21] Pongtep Angkititrakul, Terashima Ryuta, Toshihiro Wakita, Kazuya Takeda, Chiyomi Miyajima, and Tatsuya Suzuki. 2009. Evaluation of driver-behavior

models in real-world car-following task. In 2009 IEEE International Conference on Vehicular Electronics and Safety (ICVES), 113–118. DOI:https://doi.org/10.1109/ICVES.2009.5400201

[22] How It Works. Cambridge Mobile Telematics. Retrieved May 11, 2020 from https://www.cmtelematics.com/safe-driving-technology/how-it-works/

[23] NewsDesk. Avis partners with Discovery Insure to launch SafeDrive. Mynewsdesk. Retrieved May 11, 2020 from http://www.mynewsdesk.com/za/discovery-holdings-ltd/pressreleases/avis-partners-with-discovery-insure-to-launch-safedrive-2088272

[24] Rick Salay and Krzysztof Czarnecki. Using Machine Learning Safely in Automotive Software: An Assessment and Adaption of Software Process Requirements in ISO 26262. 56.

[25] Vikas Bajpai and Ravi Prakash Gorthi. 2012. On non-functional requirements: A survey. *2012 IEEE Students' Conference on Electrical, Electronics and Computer Science* (2012). DOI:https://doi.org/10.1109/SCEECS.2012.6184810