RISK Analysis

(Driver Risk Assessment and Monitoring)

Poojitha Amin, Sneha Vadakkemadathil, Shivam Gupta, Rovin Singh Patwal
Software Engineering,
San Jose State University
San Jose, US
Poojitha.amin@sjsu.edu, Sneha.Vadakkemadathil@sjsu.edu, shivamkumar.gupta@sjsu.edu, rovinsingh.patwal@sjsu.edu

Advisor: Prof. Rakesh Ranjan

Abstract— The total number of automobile accidents and fatalities has been on the increase in most States. If a person has had many accidents in the past, car insurance companies and transformation network companies (like Uber, Lyft etc.,) logically assume that the person is likely to have more accidents in the future. The project aims to look at certain other data points along with the driving history and identify the risky and cautious drivers. It also facilitates monitoring the driving behavior and provide key signs of unsafe and aggressive driving practices.

I. INTRODUCTION

Nowadays, by the virtue of machine learning, we have computers with the ability to learn without explicitly being programmed. It is the best way to make progress towards human-level AI. Machine learning is very pervasive today and is applied in a variety of computing tasks, mainly focused on finding patterns and prediction making, with the use of computers. One such application is predicting the risk factor involved in the driving behavior of a person. This can contribute to the evaluation of insurance rates as well as understand the risk involved in employing a driver at transportation network companies.

The driving record is a major factor in determining one's insurance premium. But there are other factors as well, which should be considered for understanding the probability of the accidents and claims. They are age, gender, driving experience, marriage status, annual mile driven, vehicle type and credit score. For example, an 18-year-old boy with a sports car is more likely to have an accident than a 45-year-old man with a regular sedan. As per the recent survey and statistics, eight of ten younger drivers report speeding at least monthly on each road type. Six in ten drivers age 65 or older report speeding on all road types. Males are 50 percent more likely than females to drive over the posted speed limit. The higher the probability of an accident or traffic violation, the higher is the probability of a claim. With the help of this project, we will take-into-account all these factors and estimate the risk, broadly dividing into three categories – Low-Risk, Average-Risk and High-Risk. However, it is also known that every driver is unique and sometimes, prediction using historical data and by means of common physical laws, may not precise and accurate. For such cases, the project introduces the second feature, which would make observations while the driver is engaged in the actual driving task, in a real vehicle, operating on a real road. This is the way forward and is the valid method to conclude on the driving pattern. The project simulates this feature using the sample data.

II. TECHNOLOGIES

A. Machine Learning

For machine learning, the open source Python library scikit-learn is used. The algorithm used for classification is Linear discriminant analysis.

B. Back-end Server

For back-end we are using Flask-Python web framework to redirect the various website. Apache2 web server is used to host the code.

C. Client-Side

Bootstrap, Java script, JQuery, Flask, Angular JS, are used for implementing the client side interface.

D. Cloud platform

Amazon Web Service for cost effectiveness, scalability and high performance. We are using RDS MySQL database for connectivity over cloud.

E. IoT Platform

IBM Watson IoT platform, to consume and communicate data along with analysis in real time.

F. Wiring programming tool

IBM Node-RED is used for wiring together the components of the application. The browser based editor simplifies the work.

G. Tool for behaviour analysis

IBM Watson IoT driver behavior, to collect driver data from a simulated journey. Also provides intelligent analytics for large volumes of automotive data.

H. Service API's

IBM Watson IoT Context Mapping, to avail additional automotive solutions using the context mapping REST API.

I. Database

MySql is used for its reliability, security and high performance.

III. RISK ANALYSIS IMPLEMENTATION

The front-end application built is used for taking the user inputs on the specifications of the driver. The information is saved in the database and further pushed for prediction of the risk factor. The machine learning algorithm is applied and the final risk score predicted is returned to the client. For bulk uploads, the file is read, risk is evaluated for each entry and the bulk output is shared with the user via email.

A. Figures and Tables

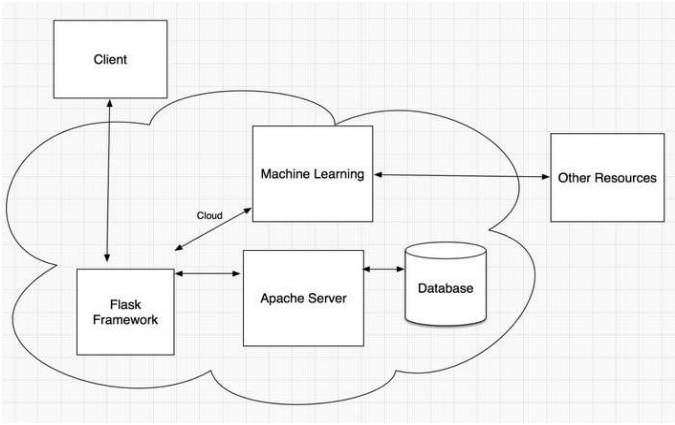
Table 1 shows the distribution of cases and controls amongst age and sex.

Table 1.
Participants in the Auckland Car Crash Injury Study, by age group and sex, Auckland, New Zealand, 1998–1999

Age group (years)	Cases (*n* = 571)				Controls (*n* = 588)			
	Females		Males		Females		Males	
	No.	%	No.	%	No.	%	No.	%
<25	53	26.8	142	38.1	37	16.4	54	14.9
25–34	42	21.2	91	24.4	50	22.1	75	20.7
35–44	36	18.2	49	13.1	61	27.0	93	25.7
45–54	17	8.6	44	11.8	36	15.9	71	19.6
55–64	21	10.6	18	4.8	28	12.4	52	14.4
65+	29	14.6	29	7.8	14	6.2	17	4.7
Total	198	34.7	373	65.3	226	38.4	362	61.6

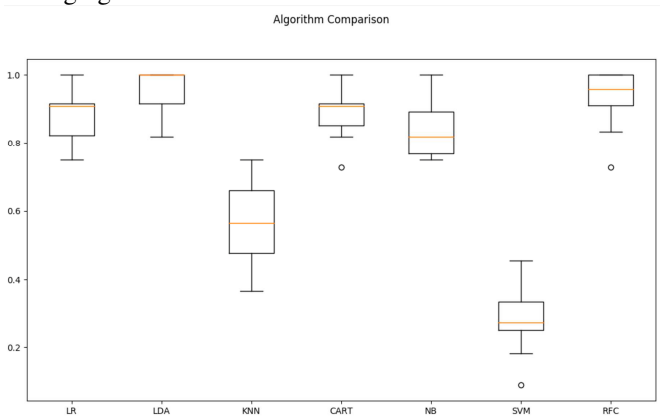
B. Architecture

As per our Architecture, the user must pass basic Driver's information to our application using the UI. Then our application is going to refer different databases to fetch details like DMV point, Vehicle Type, Credit Score, and other. Using this model we will apply Machine learning to predict a score for the user.



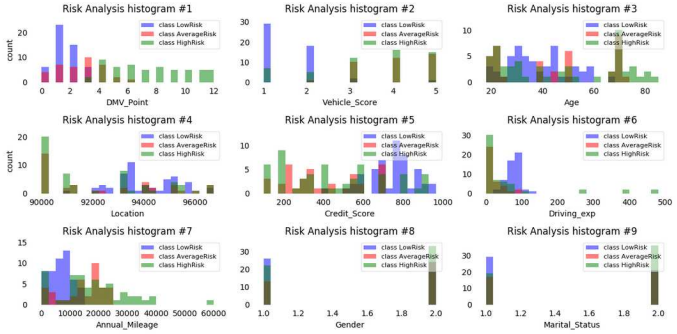
C. Machine Learning

Linear Discriminant Analysis is used for predicting the risk score of the driver. The statistical properties of the data are estimated and plugged into the LDA equation to make predictions. The model uses the Bayes Theorem to estimate the probability of each class and the likelihood of the data belonging to that class.



Above graph shows the camparison between different Machine Learning Algorithms prediction.

The below figure demonstrates the risk analysis histogram for the various attributes from the data set, contributing to the calculation of risk.

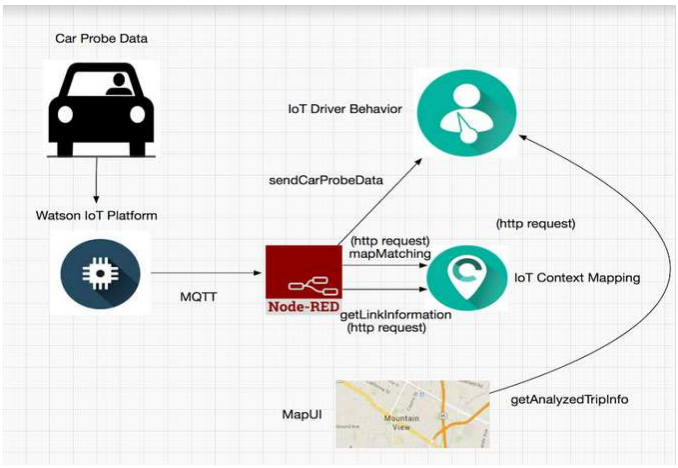


IV. DRIVER BEHAVIOR MONITOR IMPLEMENTATION

The setup gives an insight into the behavior of the driver through a simulated scenario. It mainly gives information about the following:

- Speeding
- Sharp turns
- Harsh Acceleration
- Harsh Braking
- Frequent Acceleration
- Frequent Braking

A. Architecture



1. The car is simulated by a Java client in this project.

2. After receiving the car probe data, IBM context mapping API is used to map matched data using mapMatching API and get road type data using the getLinkInformation API.

3. The car probe data is then sent for analysis using Driver Behavior's sendCarProbe API.

4. Job request is sent using the sendJobRequest API of Driver Behavior.

5. getAnalysedTrip SummaryList API is used for getting the analysed trip summary.

6. getAnalysedTripInfo API of Driver Behavior is used for getting the trip information by passing the trip UUID as input.

7. The map UI from the client is displayed in the end indicating the vehicle trace and driving behavior segments.

8. Placing the mouse over the segments shows the description of the behavior.

V. ADDITIONAL FEATURES

For additional features, we have added SSL certificate in our AWS instance to make it more secure, we have bulk upload option in our application so that user can just upload the data to our system, and we can calculate the score and let him check back. We are flask to make our application stronger against Cross Site Scripting.

VI. CHALLENGES

The most challenging part for this project is to get the data from third party systems. We must rely on the servers residing out of our system to get the other information. One of the other challenges which we see moving forward in this project are Marketing and Investment in the project, as Our Machine Learning Model will be as good as Data, and this amount of data does not come for free.

VII. GITHUB LINK

https://github.com/SJSU272LabS17/Project-Team-19

ACKNOWLEDGEMENT

The authors would like to thank Prof. Rakesh Ranjan for his ideas and guidance.

REFERENCES

[1] http://www.sciencedirect.com/science/article/pii/S0001457502001069

[2] http://ranjanr.blogspot.com/

[3] http://www.sciencedirect.com/science/article/pii/S0001457599000093

[4] https://catalog.data.gov/dataset/motor-vehicle-crashes-vehicle-information-beginning-2009

[5] https://catalog.data.gov/dataset/traffic-tickets-issued-beginning-2008

[6] https://catalog.data.gov/dataset/demographic-statistics-by-zip-code-acfc9

[7] https://catalog.data.gov/dataset/traffic-violations-56ddaR. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.