Experimental Investigation of Driver Behavior and an Approach for Implementation of Usage Based Insurance in India

Pawan Wawage[1], Yogesh Deshpande[2]

Vishwakarma University, Pune, India.

Abstract: *Usage Based Insurance has a large scope in India, but it is essentially non-existent elsewhere. As a next step, we have teamed up to help design a system that will assist us in overcoming this issue. The goal is to create a machine learning model that will allow us to generate driver/customer safety scores depending on where and how the vehicle has been operated. We have accomplished this by analyzing data obtained from sensors on our Smartphone. This data and categorization system can then be used to solve real-world issues, such as Usage Based Insurance (UBI). In many different nations, Usage Based insurance has been demonstrated to be effective and appealing to both insurers and insured. However, in India, this method is rarely used and almost non-existent, but it is on the horizon. As a result, our proposed works usage in India has a lot of promise.*

Keywords: **Driver Behavior; Machine Learning; Pay-As-You-Drive; Usage Based Insurance.**

1. Introduction

Having the appropriate kind of knowledge, but more importantly, data, is critical in today's world. It's not about how much information you can receive, but about what kind of information you can get and why you want to have it. That is why data is so crucial. Because the world is constructed on data, it makes sense to learn how to use information to improve the world in a variety of ways. As a result, the goal of our work is to collect data [1]. about driving a vehicle and create a categorization system that will provide us with information about the driver's behavior and profile while driving. This data and categorization system can then be used to solve real-world issues, such as Usage Based Insurance (UBI).

Usage-based insurance, also known as the Pay-As-You Drive (PAYD) model, is effective for giving cheaper insurance to consumers who have strong safety scores rather than relying solely on other statistical characteristics such as age and gender (age, gender, marital status etc.) [3]. This system will assist us in identifying driver behavior based on characteristics such as aggressive braking, sudden turns, speeding, and cornering, which will aid us in preventing vehicle accidents caused by road rage. The fundamental goal of UBI is to provide a more practical and equitable method for determining risk, with customers who have demonstrated aggressive driving behavior being required to pay a higher premium [7][8].

Usage-based insurance, sometimes referred as pay-as-you-drive insurance, is a type of auto insurance that, depending on the specific insurer's programme, can track how far, where, and how a vehicle is driven [7]. We plan to accomplish this by analyzing data obtained from sensors on our Smartphone [5]. The number of cell phones in use around the world continues to rise year after year, resulting in new signal processing applications.

The Smartphone is a ubiquitous gadget that comes pre-loaded with a variety of environmental sensing capabilities. Modern smart phone’s also have adequate processing power to clean data and run classification models [10]. In many different nations, usage-based insurance has been demonstrated to be effective and appealing to both insurers and insured. However, in India, this method is rarely used and almost non-existent, but it is on the horizon. As a result, our works use in India has a lot of promise.

The risk characteristics and variables being analyzed in the case of UBI are placed on top of normal rating methodologies including age, gender, marital status, credit and driving histories, and insurance ratings, all of which have been demonstrated to be predictive over time. UBI expands the scope of relevant data, allowing risk to be assessed based on real-world driving behavior, inclinations toward safe or risky driving behaviors, trip parameters, vehicle condition, location of operation, weather conditions, and more [1][16]. When combined with standard rating criteria, UBI demonstrates a more accurate evaluation of risk, allowing insurers to give the best drivers the biggest discounts while yet maintaining a healthy profit margin.

Planning, labour, technology and equipment, logistics, and support all contribute to the efficient deployment of UBI and analytical tools to record and analyze driving data. However, the financial benefits of usage-based insurance often justify these costs:

- Recruiting low-risk drivers
- Increasing customer retention
- Lowering the cost of claims
- Expanding the number of potential touch-points each year
- Providing insurance plans with individualized, revenue-generating value-added services in order to better satisfy consumer interests.

By adding low-risk drivers to their customer base, using driving data to improve pricing, improving the customer perception of the company as more technologically advanced, and strengthening long-term relationships through closer communication, telematics can help insurance carriers improve their market position [12]. When developing a telematics programme, specific strategies such as meticulous planning and testing, the use of value-added services, the manner in which customer interactions are handled, the pricing model, and effective use of data analytics within the programme can increase the chances of success. Overall, it's clear that the vehicle insurance industry is rapidly changing. Early adopters of usage-based insurance programmes will have a distinct advantage over those that wait and risk losing valuable clients to more imaginative competition.

Aggressive driving behavior and road rage incidents such as speeding in heavy traffic, lane cutting, and quick braking of drivers during short-term and long-term driving should be addressed, according to the survey's findings [11]. Fewer road accidents and insurance claims would come from efforts to recognize the risk posed by aggressive driving by taking into account a driver's behavioral and emotional components. The outcomes of this study will help motor insurance firms measure driving risk more accurately and propose a solution for computing tailored rates based on driving behavior, with a focus on risk prevention [4].

We came across a number of research papers on driver behavior and usage-based insurance. According to the results of the survey, the random forests algorithm [12][13] was deemed to be the best of the algorithms shown in the Table 1 below:

Table 1. Machine Learning Models and their accuracy.

Model name	**Accuracy**	**Purpose**
Support Vector Machine [6].	< 70%	Tested on all types of driving events
Bayesian Network [6].	< 70%	Tested on all types of driving events
Random Forests [6].	Worst – 88%, Best – 97%	Consistently good results among all types of events; Classify 3 distinct states: normal, drowsy, aggressive & Best and most consistent model;
Multi-Layer Perception [3].	94% - 99% for two cases	Performs exceptionally well for aggressive lane change events & not consistent results among all events
Simple RNN [3].	Max 70%	Tested on all types of events with accelerometer data
Long Short-Term Memory (LSTM) [6].	Max 95%	Tested on all types of events with accelerometer data; Good model but only works efficiently with 10 neurons
Gated Recurrent Unit (GRU) [9].	Max 95%	Tested on all types of events with accelerometer data
Naïve Bayes Classifier [9].	Max 90%	For Aggressive driving styles
Gaussian Mixture Model [15].	< 80%	Good result on training data but very limited in real world
Neural Networks [6].	Out-performs all	Very high computational power and load

2. Methodology

Because none of the other systems had cloud connectivity, our goal was to interface our system with any of the cloud platforms [12]. After evaluating numerous platforms, we discovered that the Google Cloud Platform was the most efficient with our system. Google App Engine was the most cost-effective way to install the machine learning model on the cloud platform. Visual data analysis was also employed to observe data trends in existing vehicle data sets. On the supplied raw-datasets, we applied Time-Series graphs and attempted to make sense of the resulting graphs. We used data from the smartphone's accelerometer and gyroscope sensors to compare the two.
This work seeks to create a system that includes multiple functionalities such as Smartphone sensors (accelerometer, gyroscope, and magnetometer) for data collection, driver behavior profiling (safe or aggressive), calculating safety score and more. For that excursion, assigning a safety score to the driver, using the above data for Usage Based Insurance premium on a monthly or quarterly basis, we used a Smartphone platform for data collection and cleaning. We deployed the model on an Android Smartphone and a cloud platform in order to handle large amounts of data, depending on the algorithm load.

2.1. Data Collection Environment

The purpose was to record smart phone sensor data as the driver performed specific driving actions in a realistic traffic environment [6]. Once the model driver (persona) has been determined, the next stage is to develop the experiment and the experimental conditions. The trial lasted for days, with the driver making two trips each day. Each one-way travel was about 4-5 kilometres long and took about 5-10 minutes on average. In this experiment, we used an Android application to record data from smart phone sensors (accelerometer, gyroscope, etc.). The smart phone was horizontally mounted with phone holder on the car's dashboard while the driver was executing driving activities. Table 2 shows the experimental setup in full.

Table 2. Experimental setup

Number of drivers	2
Phone model	Samsung Galaxy M20 and ONE Plus 8
Phone fitting	Dashboard phone holder/stand
Phone position	Horizontal
Video recording	Phone rear camera
Vehicle type	LMV
Vehicle model	Wagon-R, LXI
Road type	City
Traffic condition	Mild
Smartphone Sensors	Accelerometer and Gyroscope
Data recorded	Longitude, Latitude, Speed, Distance, Time, Heading.

2.2. Random Forests Algorithm

Machine learning algorithms have traditionally been employed in projects and datasets to forecast a goal value, attribute, or discover patterns. Data is utilized in these types of applications to assist machines in learning new patterns so that they can make accurate predictions based on newly added input data at different time intervals. Some of the most popular machine learning methods are Linear regression, Decision trees, Support Vector Machines (SVMs), Naive Bayes, Neural Networks, and Ensemble approaches [6]. The current study employs supervised machine learning to enable a training method using the labelled training dataset provided. The Random forests approach is evaluated in terms of the accuracy and loss of the supplied model.

Random forest is a supervised machine learning approach for solving classification and regression issues. It uses the majority vote for classification and the average for regression to generate decision trees from various samples. The Random Forest algorithm's ability to handle data sets with both continuous and categorical variables, as in regression and classification, is one of its most important features. It outperforms the competition when it comes to classification problems.

A decision tree is a type of tree that can be found in a forest. Random forests build decision trees out of randomly chosen data samples, get predictions from each tree, then vote on the best option. With the model, Scikit-learn includes an extra variable that displays the relative importance or contribution of each attribute to the prediction. During the training phase, it calculates the relevance score of each feature automatically. Steps involved in Random Forests algorithm:

Step 1: From a data set of k records, n random records are chosen at random in Random Forest.

Step 2: Individual decision trees are created for each sample.

Step 3: At the end of each decision tree, there is a result.
Step 4: The final result for classification and regression is based on Majority Voting or Averaging, as appropriate.

2.3. Proposed Architecture

The architecture shown in the Figure 1 is followed in our system. As shown in Figure 2, the function is sent into the system via the mobile app. The data collection will be handled by the mobile app. In the mobile app, the safety score and driver behavior will be presented. The safety score is calculated on a cloud server using APIs, and the entire model is delivered using the flask application.

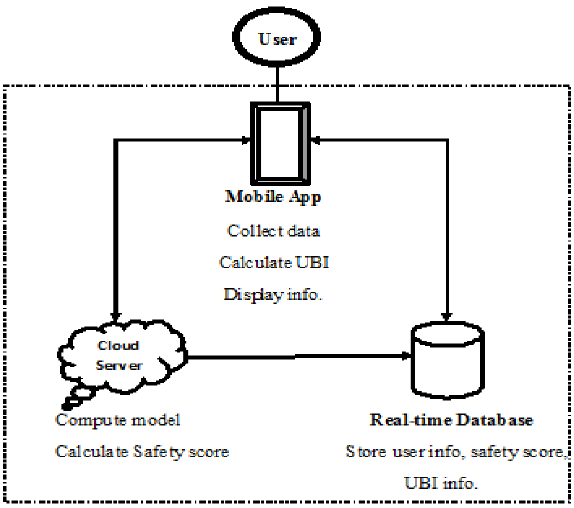

Figure 1. System architecture

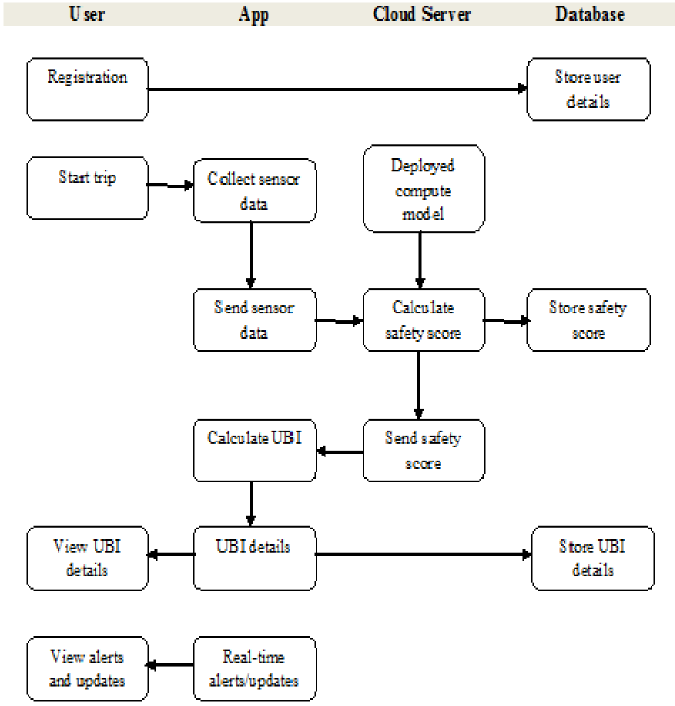

Figure 2. System flow

User: In order to use the system's services, users must first register in the app. Users have their own unique credentials to login to further access facilities once they have registered. The user will have access to UBI data as well as all alerts and updates.
App: The sensor data will be collected by the app, which will then be sent to the cloud.
Cloud server: The safety score will be calculated and sent to the app using the cloud server's machine learning model.
Database: The database will be in charge of storing the user's real-time data as well as other UBI-related information.

2.4. Proposed UBI Strategy

The proposed usage-based insurance system determines the amount of insurance premium paid by the insured. The safety score of the drivers will be generated based on the smartphone sensor data collected throughout all of the excursions. The steps to determine the safety score are listed below, and Table 3 displays the real logic:

- Every instance starts with a 10 out of 10 safety score.
- We get a list of 0's (Normal) and 1's (Exceptional) for every 5 min interval of data (Aggressive).
- We calculate the number of 1's in the list as a percentage.
- We deduct a percentage of 1's from the safety score based on the percentage of 1's.

Table 3. Safety score logic

Aggressive instances (% of 1's)	Deduction from 10
> 20%	-1
> 25%	-2
> 30%	-3
> 35%	-4
> 40%	-5
> 50%	-6

Data is collected in 5-minute intervals from the start to the completion of the trip. During the first interval, the score was calculated using the prior scoring procedure. The score for the next interval is then calculated, followed by the average (for example, if the second period's score is 7, the overall score becomes 6+7/2 = 6.5). The average of the previous overall score and the current interval score is then calculated. (For example, if the third period's score is 8, the overall score is 6.5+8/2 = 7.25). The same method can be used to calculate the score for each trip. This number can be averaged over all of the trips made in a week or month to get the average safety score for that week or month. We can average the weekly or monthly scores to determine the overall score, depending on whether the insurance payment is paid monthly or annually. The proposed approach in Table 4 can be used to provide a discount in the premium for the consumer based on these scores.

Table 4. Safety score and discount logic

Safety Score	Discount in premium
6	5%
7	10%
8	12.5%
9	15%
10	20%

3. Results and discussion

The method was evaluated using the following metrics: (i) accuracy; (ii) F1-score; (iii) precision; (iv) recall; and (v) support. The loss metric, which applies to both the training and testing sets, is determined as the total of the predicted value's distances from the real values. The number of correct predictions divided by the total number of predictions produced is the accuracy metric. The accuracy metric was applied to both the training and testing sets. The precision and recall numbers, as well as the loss and accuracy data, are used to calculate the F1-score. Precision and recall measurements represent the rate of true positives in anticipated positives and actual positives, respectively. As a result, the harmonic mean of precision is the F1-score and is indicated in Table 5 and Table 6 respectively.

Table 5. Result metrics of the tested algorithm on the dataset

	PRECISION	RECALL	F1-SCORE	SUPPORT
-1	1.00	1.00	1.00	5
0	0.80	0.86	0.83	1826
1	0.85	0.79	0.82	1808
ACCURACY	-	-	0.83	3639
MACRO AVG.	0.88	0.88	0.88	3639
WEIGHTED AVG.	0.83	0.83	0.83	3639

Table 6. Result metrics of the tested algorithm on the dataset

	PRECISION	RECALL	F1-SCORE	SUPPORT
-1	1.00	1.00	1.00	4
0	0.90	0.90	0.90	1692
1	0.91	0.91	0.91	1945
ACCURACY			0.91	3641
MACRO AVG.	0.94	0.94	0.94	3641
WEIGHTED AVG.	0.91	0.91	0.91	3639

3. Conclusion

We've successfully implemented a classification model with high accuracy, efficiency, and scalability. We were able to create an end-to-end system that adhered to new technologies and trends by deploying on a cloud server and connecting it with an Android app. To improve the accuracy and efficiency of the classification model, Deep Learning or ensemble methods might be used. Different types of training data, such as data vectorization and increased hyper parameter tweaking, can be used to further investigate the model. New features, such as sleepiness monitoring and real-time notifications, can be added to the app.

REFERENCES

[1] *Abdelrahman, A., Hassanein, H. S., & Abu-Ali, N. (2019, May). A cloud-based environment-aware driver profiling framework using ensemble supervised learning. IEEE International Conference on Communications (ICC) (pp. 1-6).*

[2] *Ameksa, M., Mousannif, H., Al Moatassime, H., Elassad, Z. Elamrani Abou (2020). Toward Flexible Data Collection of Driving Behavior. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, vol. 4, pp. 33-43.*

[3] *Bian Yiyang, Yang Chen, Zhao J., Leon Liang (2018). Good Drivers Pay Less: A study of Usage-based Vehicle Insurance Models. Transportation Research Part A: Policy and Practice, pp. 107.*

[4] *Eboli, L., Mazzulla, G., & Pungillo, G. (2017). How to define the accident risk level of car drivers by combining objective and subjective measures of driving style. Transportation research part F: traffic psychology and behaviour, 49, 29-38.*

[5] *Eleni Mantouka, Emmanouil Barmpounakis, Eleni Vlahogianni, John Golias (2021). Smartphone Sensing for Understanding Driving Behavior: Current Practice and Challenges. International Journal of Transportation Science and Technology, pp. 266-282.*

[6] *Ferreira, J., Carvalho, E., Ferreira, B. V., de Souza, C., Suhara, Y., Pentland, A., & Pessin, G. (2017). Driver behavior profiling: An investigation with different smartphone sensors and machine learning. PLoS one, 12(4), e0174959.*

[7] *Isaac Skog, Peter H'andel, Martin Ohlsson, and Jens Ohlsson (2013). Challenges In Smartphone-Driven Usage Based Insurance. IEEE Global Conference on Signal and Information Processing, pp. 1135, doi: 10.1109/GlobalSIP.2013.6737098.*

[8] *Isaac Skog, Peter H'andel, Martin Ohlsson, and Jens Ohlsson (2013). Smartphone-Based Measurement Systems for Road Vehicle Traffic Monitoring and Usage-Based Insurance. IEEE Systems Journal, vol. 8, no. 4, pp. 1238-1248, doi: 10.1109/JSYST.2013.2292721.*

[9] *Kanarachos Stratis, Christopoulos Stavros, Richard G., Chroneos Alexander (2018). Smartphones as an Integrated Platform for Monitoring Driver Behaviour: The role of Sensor Fusion and Connectivity. Transportation Research Part C: Emerging Technologies, pp. 95.*

[10] *Lattanzi, Emanuele & Castellucci, Giacomo & Freschi, Valerio (2020). Improving Machine Learning Identification of Unsafe Driver Behavior by means of Sensor Fusion. Applied Sciences, vol. 18.*

[11] *Liu Tianshi, Yang Guang, Shi Dong (2020). Construction of Driving Behavior Scoring Model Based on OBD Terminal Data Analysis. 5th International Conference on Information Science, Computer Technology and Transportation, ISCTT, pp. 24-27.*

[12] *Ping Peng, Qin Wenhu, Xu Yang, Miyajima Chiyomi, Takeda Kazuya (2019). Impact of Driver Behavior on Fuel Consumption: Classification, Evaluation and Prediction using Machine Learning. IEEE Access, vol. 7, pp. 78515-78532.*

[13] *Sarah Kadhim Alluhaibi, Munaf S. Najim Al-Din, Aiman Moyaid (2018). Driver Behavior Detection Techniques: A survey. Int. J. Appl. Eng. Res 13.11, pp. 8856-8861.*

[14] *Subramanian Arumugam and R. Bhargavi (2019). A Survey on Driving Behavior Analysis in Usage Based Insurance using Big Data. J Big Data, vol. 6, pp. 86.*

[15] *Vasconcelos, I., Vasconcelos, R. O., Olivieri, B., Roriz, M., Endler, M., & Junior, M. C. (2017). Smartphone-based outlier detection: a complex event processing approach for driving behavior detection. Journal of Internet Services and Applications, 8(1), 1-30.*

[16] *Warren Josh, Lipkowitz Jeff, Sokolov Vadi (2019). Clusters of Driving Behavior from Observational Smartphone Data. IEEE Intelligent Transportation Systems Magazine, vol. 3, pp. 171-180.*