



SYSTEM FOR PREDICTING TRAFFIC VIOLATIONS USING MACHINE LEARNING

¹Glen P Jose, ²Judin Joseph, ³Oliviya S Paimattam, ⁴Riya Mol Joseph, ⁵Naveena Tresa Joseph

¹²³⁴Student, ⁵Assistant Professor

¹Department of Information Technology

¹Amal Jyothi college of Engineering

¹Kottayam, India

Abstract: This system is designed to predict traffic violation of a driver on the basis of their behaviour. As road accidents shoot up day by day, there is an increased chance of neglecting enforced rules by drivers. This temptation in drivers can be due to physical, mental and emotional states. This is analysed by comparing different traits like demographic information, psychological problems, addiction, personality trait, aggression and stress. Prediction is made based on behavioural analysis of drivers. Traffic violation and human behaviour is highly related. Researches show that some victims of accidents are suffering from personality and behaviour defects. In such a scenario a prediction system using human behaviour in case of traffic violation is very relevant. In addition, different machine learning models are compared to identify the accuracy score and the best one is implemented in the system. Algorithms used for comparison are Logistic Regression, Random Forest and SVM.

Index Terms - Logistic Regression, Random Forest, Support Vector Machine, Machine Learning, Human Behaviour, Traffic Violation Prediction

1 INTRODUCTION

Traffic violation is major cause of road accident in India. Traffic violations or road accidents depend on many factors like culture, age, gender, physical condition of roads, education, marital status, personality, mental illness, addiction [2]. Human factors that have impact on traffic violations are fatigue, drowsiness, mental disorder, disability, driving style, lack of training for traffic safety and inappropriate method to obtain driving licence. This system for traffic violation prediction helps to identify which trait is common in people who violate traffic rules and in what group of people causes more traffic violations. This helps the officials to implement preventive measures and ensure traffic safety in our roads. Machine learning models such as Logistic Regression, Random Forest and SVM are used for predicting traffic violations using human behaviour. Different traits of humans are analysed by preparing questionnaire and circulating it among participants. The data collected are analysed and the accuracy of different models in prediction is identified. Also, this system predicts whether a participant violates traffic rules on the basis of response to the survey.

2 RESOURCE ALLOCATION

2.1 Need of the Study

The system is designed user-friendly and is accessible to wide range of people. It uses most modern technology with simple language. The system is built using python programming language. It is a prediction system thus it takes input values and interprets the results that whether a person violates or do not violate traffic rules.

2.2 Tools Used

The software will be built using python using different platform like Psytoolkit and Kaggle are used. Libraries such as NumPy and pandas are used in the implementation of the code.

Python: Python is a high – level programming language used for scientific computing using different libraries such as NumPy, SciPy and Matplotlib. Python libraries are used commonly in data science, machine learning and scientific computing. Python is considered as interpreted language because Python programs are executed by interpreter. Python is used in data science for data analytics, used to build websites, artificial intelligence, servers and back-end development.

PsyToolkit: PsyToolkit is a software package to program and run psychological surveys and experiments, especially those common in cognitive experimental psychology. It is a useful tool for students and researchers in the field of psychology who want to conduct online experiments and surveys. PsyToolkit provides a range of features, including the ability to create and customize experiments, collect data, and analyse results. It is a user-friendly platform that can be used by non-programmers and does not require any coding experience.

Kaggle: Kaggle provides an official API that can be accessed using a command line tool implemented in Python. Kaggle is a great platform for learning Python, as it provides competitions, kernels, and other opportunities for immersive learning. Kaggle's platform is designed for data science and machine learning competitions, and it is the world's largest community of data scientists

3 METHODOLOGY

3.1 Preparing Questionnaire

PsyToolkit is used to prepare the questionnaire. It is a platform offered by python for conducting psychological questionnaire. This survey collects data on human behaviour that is used to create a predictive system that can identify people who are more likely to violate traffic rules. Three different questionnaires—the BFI-S, the Driving Behaviour Survey, and the traffic questionnaire—were used in this study for the survey. Five personality traits, namely openness, extraversion, conscientiousness, neuroticism, and agreeableness, are measured using the BFI-S questionnaire. Driving behaviour survey is used to assess three constraints: anxiety-based performance deficits, exaggerated safety/caution behaviour and hostile/aggressive behaviours. Traffic questionnaire includes necessary demographic information such as name, age and gender.

3.2 Data Collection

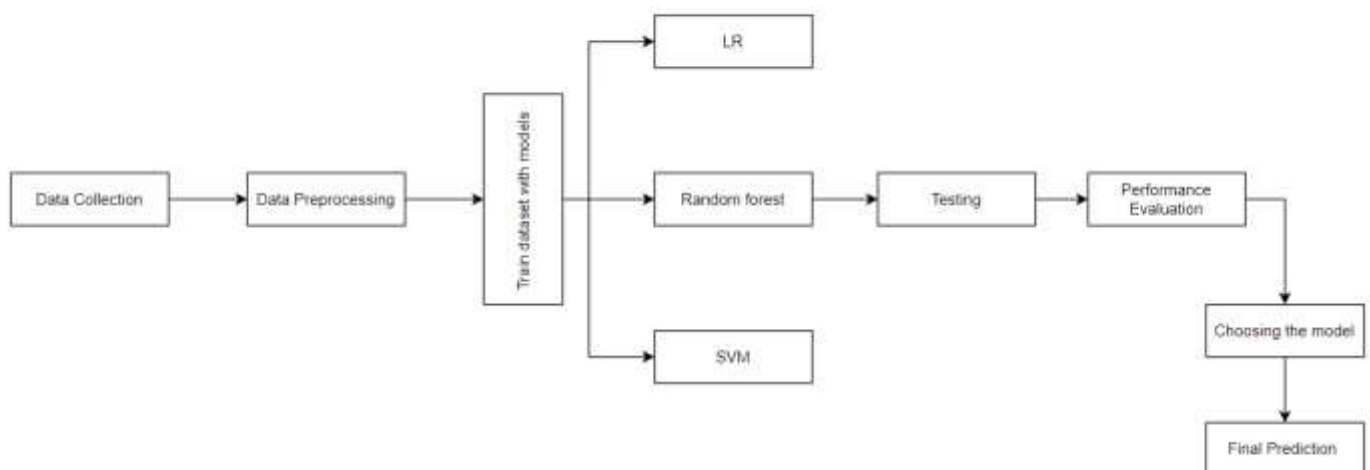
The survey is conducted using PsyToolkit and is circulated through WhatsApp. The questionnaire prepared in PsyToolKit is copied to a link and it is shared among random people. Responses made by participants are collected in excel sheets and further analysis can be done. PsyToolkit provides a generalized summary of the survey data. It provides information about each question and responses to it.

3.3 Dataset

Data is collected from random people by circulating the PsyToolkit link through WhatsApp. 226 participants have taken part in the survey and their responses are recorded in excel sheets. Among the 226 participants, 114 are men and 112 are women. Participants between the ages of 18 and 30 answered in greater numbers to the survey, that was performed among a random population. 126 participants are between the ages of 18 and 30; 57 are between the ages of 30 and 50; and 43 are between the ages of 50 and 70. Among 226 participants 115 does not violate traffic rules and 111 violates traffic rules.

3.4 System Architecture

Data is collected from participants by circulating a survey prepared using Psytoolkit. Collected data is preprocessed to find missing values and eliminate false data. The preprocessed dataset is split for testing and training. Three machine learning algorithms are used in this study such as Logistic Regression, random forest and SVM. The performance of each model is evaluated and the model with high performance is used to build the predictive system.



3.4.1 Data Preprocessing

Data preprocessing refers to the preparation and transformation of raw data into understandable form of data. It makes the data suitable for analysis and modelling. It includes data cleaning, data integration, data transformation and data reduction. These preprocessing steps are necessary for ensuring the quality, consistency and efficiency of the models.

3.4.2 Modelling

Modelling is the development and implementation of a predictive model. Algorithms used for modelling are Logistic Regression, Support Vector Machine and Random Forest. Among these three, the model with greater performance is chosen to build the predictive system. Training and testing of dataset is done for Logistic Regression, Support Vector Machine and Random Forest. The performance of trained model is evaluated using test data. Model is build using the three machine learning algorithms and the performance is compared and evaluated. Model with higher accuracy and performance is used to build the predictive system.

3.4.3 Model Evaluation

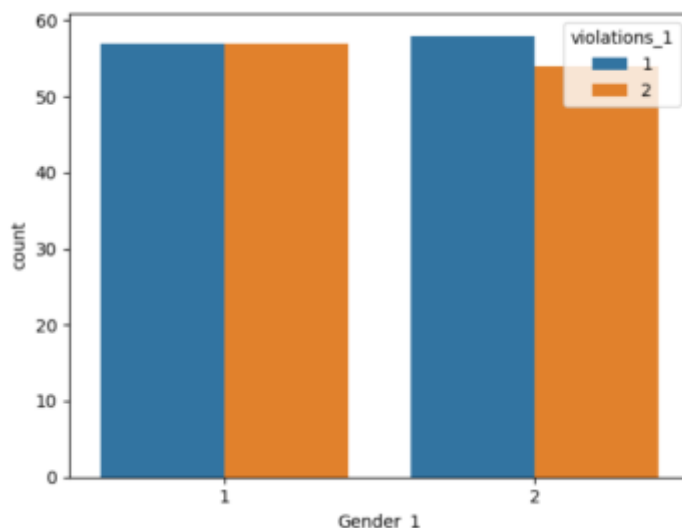
Model evaluation is the process of measuring the performance and evaluating the accuracy of the model. In this study, Logistic Regression, Support Vector Machine and Random Forest are used for model fitting. The dataset is split into two parts, 80 and 20 percentage. 80% of data is used as training model and the remaining 20% is used for testing.

4 Result

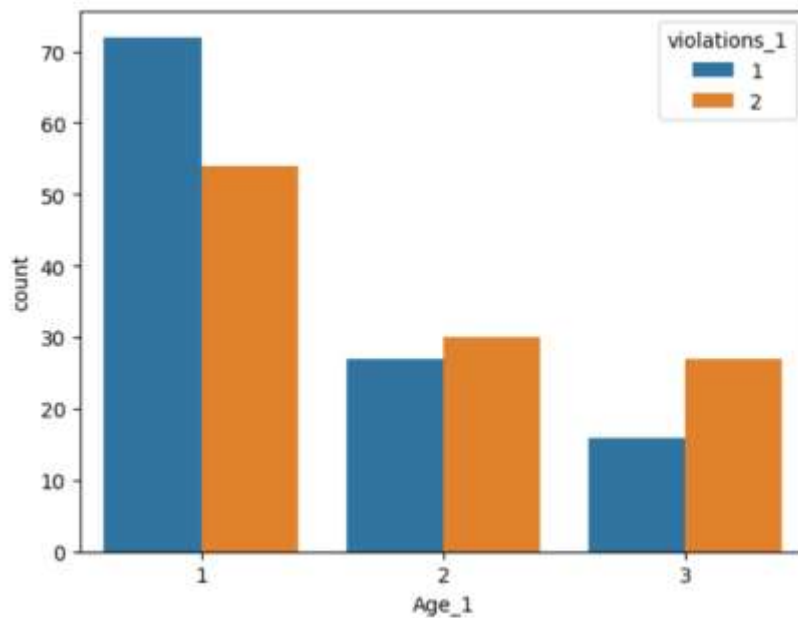
Models	Accuracy
Support Vector Machine	71%
Logistic Regression	69%
Random forest	76%

From the study, it is obtained that Random Forest is having high accuracy compared to other models. Support Vector Machine, Logistic Regression and Random Forest have accuracy 71%, 69% and 76% respectively. Therefore, Logistic Regression is used to build the predictive system. The prediction system accepts input values and predict the output that whether a person has chance to violate traffic rules.

Among 226 participants took part in the survey, 114 are male and 112 are female. Majority of both the sex possess more chance to violate traffic rules. Female participants possess very high chance in violation compared to male participants.



This study was conducted among random group of people between age limit 18-70. According to the responses to survey and data analysis it shows that participants of age group 18-30 has very high chance of violating traffic rules based on their behavioural traits like anxiety and aggressive behaviour. Participants of age group 30-50 and 50-70 possess low chance of violation due to behavioural deficits. Participants taken part in the survey are majority of age group 18-30.



5 Conclusion and Future Work

The aim of the study is to predict traffic violation using an accurate machine learning model. In this study, different machine learning models are used to evaluate the performance and build a predictive system using model with high performance. Linear Regression possess greater accuracy and predictive system is build using the same. Also, analysis of different traits on participants are also evaluated. This study shows that participants of age group 18-30 possess more aggressive behaviour while driving. Most of the participants are high or medium in the BFI-S traits. This means most of the participants are social and open minded. This system can contribute to raising awareness and encouraging responsible driving behaviour. By proactively identifying potential violators, law enforcement agencies can take preventive measures such as targeted awareness campaigns and educational initiatives. The traffic violation prediction system has the potential to enhance road safety, optimize law enforcement efforts, and promote a culture of responsible driving, contributing to a safer and more efficient transportation system for everyone.

Considering the future scope of this system, few initiatives can be included. More parameters can be added to the survey and a larger dataset can be collected. Also, modelling can be done using advanced machine learning techniques to improve the performance of the system. The real time implementation of the system can be achieved by switching from software to a mobile application.

REFERENCES

- [1] Modeling effects of driver safety attitudes on traffic violations in China using the theory of pLogistic Regressioned behavior, Chaopeng Tan, Yujia Shi, Lin Bai , Keshuang Tang , Kazufumi Suzuki, Hideki Nakamura, 19 February 2022, <https://doi.org/10.1016/j.ijatssr.2022.02.001>.
- [2] Traffic Violations Prediction System on the Basis of Human Behaviour, Deepti Goel, Rajesh Bhatia, and Kashish Bhatia, https://doi.org/10.1007/978-981-16-1295-4_3.
- [3] S-TVDS: Smart Traffic Violation Detection System for Indian Traffic Scenario , Aman Kumar, Shakti Kundu, Santosh Kumar, Umesh Kumar Tiwari, Jasmeet Kalra , March 2269, 10.35940/ijitee.D1002.0384S319.
- [4] An experimental study on errors regarding the driving behavior of young males caused by temporal urgency on open roads: A Bayesian estimation, Toshiaki Kimura , Yasuo Imai , Shingo Moriizumi , Asako Yumoto , Nozomi Taishi, Hiroshi Nakai , Kazumi Renge , 15 September 2021, <https://doi.org/10.1016/j.ijatssr.2021.09.001>.
- [5] Contribution of traffic behavior surveys for monitoring safety performance indicators in Germany: Case of mobile phone use while driving, Sophie Kröling , Tina Gehlert, Axel Wolferm Logistic Regression , online 1 November 2021, <https://doi.org/10.1016/j.ijatssr.2021.10.003>.
- [6] Alavi SS, Mohammadi MR, Souri H, Kalhori SM, JLogistic Regressionatifard F, Sepahbodi G (2267) Personality, driving behavior and mental disorders factors as predictors of road traffic accidents basedon logistic regression. Iran J Med Sci 42(1):24
- [7] Balasubramanian V, Sivasankaran SK (2269) Analysis of factors associated with exceedinglawful speed traffic violations in Indian metropolitan city. J Transp Saf Sec 1–17
- [8] Koehrsen W (2268) Modeling: teaching a machine learning algorithm to deliver business value. [Online]. Available <https://towardsdatascience.com/modeling-teaching-a-machinelearning-algorithm-to-deliver-business-value-ad0205ca4c86>. Accessed 22 June 2020.

[9] Yang J, Du F, Qu W, Gong Z, Sun X (2263) Effects of personality on risky driving Behavior and accident involvement for Chinese drivers. *Traffic Inj Prev* 14(6):565–571

[10] Sani SRH, Tabibi Z, Fadardi JS, Stavrinou D (2267) Aggression, emotional self-regulation, attentional bias, and cognitive inhibition predict risky driving behavior. *Accid Anal Prev* 109:78–88

