Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 7 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i07.47763

PAPER

Machine Learning Classification Algorithms for Traffic Stops—A Comparative Study

Mërgim H. Hoti(✉), Elvir Misini, Uran Lajçi, Lule Ahmedi

University of Prishtina, Prishtine, Republic of Kosovo

mergim.hoti@uni-pr.edu

ABSTRACT

The application of machine learning algorithms across various fields is gaining momentum, and the results increasingly emphasize the need for further testing and implementation. This is driven by the potential to streamline and expedite numerous processes. In this paper, we have employed five algorithms: KNN, Decision Tree, Random Forest, Logistic Regression, and Naive Bayes, and these algorithms have been tested in three large datasets. On average, their performance ranges from a minimum of 80% to a maximum of 90%. Data preprocessing has been completed, and concurrently, we have implemented the SMOTE algorithm to address the challenge of unbalanced data in this research. Simultaneously, the Naïve Bayes algorithm yields the most favorable results of Accuracy, Precision, Recall, and F1 Score, for the "is_arrested" class. Furthermore, to assess the performance of each algorithm, we employed metrics including Accuracy, Precision, Recall, and F1 Score. These metrics allowed us to decide which algorithm achieved the most effective classification.

KEYWORDS

supervised algorithms, classification algorithms, traffic stops, Accuracy, Precision, etc

1 INTRODUCTION

Every day, we see a growing trend of increasing traffic offenses and mistakes, despite advancements in vehicle self-control capabilities, ranging from simpler situations to interventions in more critical moments. Disregarding traffic signs directly contributes to problems that can result in the fatalities of traffic participants.

Hence, the utilization of machine learning algorithms in this context serves as an additional factor for enhancing performance in efforts to reduce instances of traffic accidents and violence. Furthermore, through the accurate identification of these factors, regulations and law enforcement can be tailored and refined to ensure the safety of every traffic participant.

In this paper, various cases are examined using classes from three datasets comprising 386,452 rows, sourced from [1]. This data has been gathered from the states

Hoti, M.H., Misini, E., Lajçi, U., Ahmedi, L. (2024). Machine Learning Classification Algorithms for Traffic Stops—A Comparative Study. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(7), pp. 18–29. https://doi.org/10.3991/ijoe.v20i07.47763

Article submitted 2024-01-04. Revision uploaded 2024-02-13. Final acceptance 2024-02-14.

of Stockton, Durham and Burlington, which encompass varying amounts of data with classes including age, gender, and their arrest status, among others. In this study, we address five distinct classes (age, race, sex, search conducted, outcome, and arrest) to assess their significance and evaluate how effectively the selected algorithms can predict them. At the same time, we have chosen five distinct algorithms: K-Nearest Neighbors (KNN), Decision Tree, Random Forest, Logistic Regression, and Naive Bayes.

The data are taken from the Stanford Open Policing Project [1], and these algorithms have been tailored according to the specific goal, which is the comparison of machine learning algorithms in traffic stops based on the classes we have selected. Simultaneously, this encompasses a series of data preparation processes, including handling missing values, normalization, and feature extraction, prior to the application of the chosen algorithms. As a result, we have applied cross-validation techniques to assess the behavior and performance of these algorithms. These techniques enable the partitioning of data into multiple subgroups, which are further divided into data for training models and evaluating their performance.

Overall, the aim of this paper is to utilize the selected algorithms to predict whether an individual has been arrested by the police, identify driver gender, and automatically classify the age group that they may be in. It's crucial to emphasize that the confidentiality and integrity of the detained individuals are maintained in accordance with the responsible unit's policies.

This research is structured in the following manner: Section 2 presents the state-of-the-art, Section 3 outlines the methodology, Section 4 presents and discusses the results, and finally, Section 5 offers conclusions regarding the obtained results.

2 RELATED WORK

The use of the classic machine learning algorithms is being directly and indirectly applied to the identification of traffic violations [2], [3], [4]. However, it is not the only approach used, as other techniques are also employed depending on the purpose, with some utilizing deep learning [5], [6], [7], [8], to ensure coherence and to test the performance of these algorithms. All of this aims to expedite the identification and automation of processes. To achieve this, various methods classify the data in a manner that enables the algorithms to be applied in a specialized manner, resulting in more accurate and efficient solutions.

Among related work to our contribution, next are listed works where authors use different datasets, i.e., different classes and predicting other relevant classes not covering our set of classes, although in the same domain of traffic violation.

In the paper [9], the authors address the risks posed by drivers through the evaluation of traffic violations using car language techniques. The data used in this study were extracted from the Ministry of Public Security's office, covering the years 2016 and 2017. The entry has 46 classes that encompass personal data, illegal actions, traffic violations, and traffic accidents throughout the year 2016. In the 2017 data subset, the classes are categorical and divided into two classes: positive or negative, based on driver involvement in traffic incidents. To assess the model's performance, the authors employed the Kolmogorov-Smirnov (KS) and ROC curve models, which are commonly used for visualizing binary classifiers. Additionally, a linear regression model was selected and trained to classify and confirm on a separate testing dataset.

The results indicate that by using the predictive model, it was possible to identify the top 3% of high-risk drivers effectively. Also, in paper [10] are used algorithms such as K Nearest Neighbors (KNN), Support Vector Machine (SVM), and CN2 Rule Inducer where the data were taken from Luzhou China.

The data used in this research was obtained from the traffic department in Luzhou, China, specifically from two highways during the year 2016. The selected algorithms for this study were K Nearest Neighbors (KNN), Support Vector Machine (SVM), and CN2 Rule Inducer. Upon analyzing the descriptive data, it was observed that driving in the wrong direction was the most prevalent type of violation identified by the machine learning algorithms. The results showed that, among the chosen algorithms and data, the KNN algorithm, with the number of clusters set to $k = 7$, achieved the highest performance with an accuracy of 99%. This outcome surpassed the accuracy of SVM and CN2 Rule Inducer significantly.

In their paper [11], the authors focus on detecting traffic violations through visual images and presenting them to the relevant institutions. The goal is to predict whether a driver is likely to violate traffic rules or not. To achieve this, the authors utilize Convolutional Neural Networks (CNN) to extract features from the captured photos. Subsequently, the SVM algorithm is employed to classify and categorize the observed actions. The data used for analysis is obtained in real-time from Closed-Circuit Television (CCTV) cameras. By evaluating the extracted data, evidence is generated, pinpointing the specific locations where traffic violations occur. This process allows for proactive detection and intervention in potential violations.

Furthermore, in another study by the authors [12], they propose a system named YOLOV3, which uses object detectors for detecting traffic violations. The classes used in this system include the region and location of vehicles within the evaluation frame and the identification of traffic signals. The proposed system achieves an impressive accuracy rate of up to 97.67% for vehicle calculation and detection, while the detection of speed-related violations reaches an accuracy of 89.24%.

The authors propose a deep-learning method for identifying devices that violate traffic rules [13]. They use a dataset collected from multiple cameras distributed across Persia and apply a multi-stage approach within the YOLO network. The study includes tests conducted on three different datasets: HGV (Heavy Goods Vehicles), License Plate, and Character. The proposed model consists of four stages, and each stage is individually evaluated for accuracy.

The results indicate that the accuracy for each stage is over 85%. Furthermore, when considering the overall efficiency of the proposed intelligent system across all stages combined, it achieves an accuracy of 70%. These findings show the effectiveness of the authors' proposed deep-learning approach in showing devices that violate traffic rules, particularly when considering the collective performance of all stages in the model.

Also, authors [1], in their research, give an important test case by delving into extensive empirical research on racial disparities in police stops, supplying a comprehensive overview of the patterns and implications of such stops across the United States. However, this study primarily focuses on racial disparities and does not engage in a comparative study of machine learning classification algorithms based on various features.

Also, in this way, research by [14] has discussed about the visualization techniques which could help our results to define what kind of visualization forms are suitable for.

Another significant contribution in this line of experimenting with a classic algorithm is [15], which introduces innovative methods for detecting discrimination through threshold tests. This paper emphasizes the development and application of threshold tests but does not explore the comparative efficiency and accuracy of different machine learning classification algorithms in the context of traffic stops.

Lastly, authors in [16] address the challenges posed by infra-marginality in outcome tests, offering insights into the complexities of interpreting discrimination tests. While this study sheds light on the intricacies of discrimination detection, it does not compare the performance of various machine learning algorithms using the data from the Stanford Policing Project. In contrast to the aforementioned studies, our research uniquely uses the data provided by the Stanford Policing Project to conduct a comparative analysis of machine learning classification algorithms based on specific features, filling a gap in the existing literature by not only applying machine learning to analyze traffic stops but also by comparing the effectiveness of different algorithms in this context.

3 METHODOLOGY AND DESIGN

The application and testing of the identified factors demand a contemporary approach to yield concrete results for optimal improvement. Consequently, data for application of the selected algorithms was initially sourced from [1].

About this we have increased some research questions which has helped us to define and orient what kind of direction this research will take.

The following research questions are:

1. How do different machine learning algorithms perform in predicting essential classes such as driver gender, arrest status, and age group in the context of analyzing traffic stops?
2. How effective is the SMOTE algorithm in addressing the challenge of imbalanced data when predicting the "is_arrested" class, and how does it impact the performance of different machine learning algorithms?
3. How do the performance metrics of machine learning algorithms vary across different datasets, such as Stockton, Durham, and Burlington, and what insights can be drawn from these variations?

Various cases are examined using classes from three datasets comprising 386,452 rows, sourced from [1]. This data has been gathered from the states of Stockton, Durham and Burlington, while used data are taken from the Stanford Open Policing Project [1]. Furthermore, these datasets are extensive, including over 200 million records, primarily collected from various locations across the United States. To identify the most favorable outcomes, we subjected this data to preprocessing to render it suitable for subsequent application. The preprocessing process involves several stages, starting with the choice of a specific class. In this case, we have chosen to work with classes: age, race, sex (driver_gender), search_conducted, outcome (stop_outcome), and arrest (is_arrested). This is because we will not work with all available data.

The classes that we predicted are driver_gender with only two values: Male and Female, is_arrested with two values: True and False, stop_outcome with three values: warning, citation, arrest; and the class age_group has three values: 10–30 years, 31–65 years and 66–100 years. It is noteworthy that the "stop_outcome" class, which essentially conveys whether an arrest took place or not, was not included in the feature selection process when predicting "is_arrested."

Subsequently, in the next phase, we performed data cleaning by eliminating all the lines that contained missing data, based on our filtering criteria. This action was necessary as the data that contained missing values did not provide sufficient information for analysis, and accuracy in this regard was compromised due to the absence of data.

Simultaneously, we performed age class discretization by dividing it into several segments to obtain more precise values and fines data. This allowed us to identify the optimal age range that receives the highest number of fines. Ultimately, the data was further refined by eliminating all duplicate rows that appeared more than once in the dataset.

Following the data preparation, we proceeded to evaluate five selected algorithms, which included K-Nearest Neighbors (KNN), Logistic Regression, Random Forest, Decision Tree, and Naive Bayes. The selection of these algorithms was based on their performance in classification tasks and their ability to effectively discern data patterns across various datasets.

This approach involved the application of two separate dataset portions, achieved by partitioning the primary dataset into a training and a testing dataset. We allocated 70% of the dataset for training and 30% for testing. Additionally, we explored other scenarios where the dataset was divided into 80% for training and 20% for testing. This was done to assess how the algorithms perform and the results they produce under varying dataset sizes. In the final phase, we conducted prediction analysis of the models and evaluated their performance using standard metrics, including accuracy, precision, recall, and F1 Score. Accuracy gauges overall correctness, serving as a fundamental benchmark.

Also, as we mention using standard metrics, in our research, each of them has its own contribution to proving generated results. Precision emphasizes accuracy in positive predictions, vital when false positives are consequential. Recall assesses the model's ability to capture all positive instances, crucial when missing positives has significant repercussions. F1 score balances precision and recall, ideal for reconciling the trade-off between false positives and false negatives. These metrics supply a comprehensive evaluation tailored to the task's nuances and implications.

4 RESULTS AND DISCUSSIONS

Based on the related work and the methodology defined during the process of generating our results, we now present the findings of our extensive analysis. Here are shown results of employing a suite of machine learning algorithms to predict essential classes like driver_gender, age_group, and is_arrested, based on a dataset comprising traffic stops. The following subsections elucidate our findings and their implications, considering evaluation metrics and visual aids.

The first aspect of our investigation pertains to the performance evaluation of the selected machine learning algorithms. For each class we predicted, we assessed the algorithms' predictive accuracy, precision, recall, and the generation of confusion matrices. These metrics supply valuable insights into the models' effectiveness and their ability to make accurate predictions. To ensure the reliability of our results, we also employed cross-validation techniques to guard against overfitting and optimize generalization.

Throughout this section, we delve into a detailed discussion of the results obtained, drawing connections between algorithm performance. Our analysis aims to shed light on the strengths and limitations of the employed machine learning algorithms in the context of predicting driver classes and traffic stop outcomes.

Table 1. Results of all selected algorithms and metrics performance

Algorithms	Metrics	Classes		
		Driver Gender	Is Arrested	Age Group
Logistic Regression	Accuracy	0.68	0.84	0.42
	Precision	0.60	0.98	0.54
	Recall	0.68	0.84	0.42
	F1 Score	0.57	0.90	0.46
K-Nearest Neighbors (k = 5)	Accuracy	0.62	0.80	0.51
	Precision	0.58	0.98	0.49
	Recall	0.62	0.80	0.51
	F1 Score	0.60	0.88	0.49
Decision Tree	Accuracy	0.69	0.85	0.41
	Precision	0.54	0.98	0.54
	Recall	0.69	0.85	0.41
	F1 Score	0.56	0.90	0.46
Random Forest	Accuracy	0.69	0.85	0.41
	Precision	0.53	0.98	0.54
	Recall	0.69	0.85	0.41
	F1 Score	0.56	0.90	0.56
Naive Bayes	Accuracy	0.47	0.99	0.55
	Precision	0.67	0.97	0.55
	Recall	0.47	0.99	0.55
	F1 Score	0.46	0.98	0.53

The "is_arrested" class in the dataset was unbalanced because it had 40,657 False values and 555 True values from the 41,202 total data points. To handle the imbalance, we applied the SMOTE algorithm to the training data, which is a perfect method of balancing data, especially in our case. The results shown in Table 1, for the "is_arrested" class after we applied SMOTE.

The best performance shown in this class is the results of execution of the Naive Bayes algorithm that had the highest results for accuracy, recall, and F1 score. All the other algorithms perform better than Naive Bayes in the Precision metric with a 98% score. Overall, all algorithms in all metrics performed very well, while the lowest score is 80% and the highest is 99%.

The "stop_outcome" class was not included in the prediction of "is_arrested" because of the high correlation between these two classes. Hence, we did not include them just to have clear results in this way.

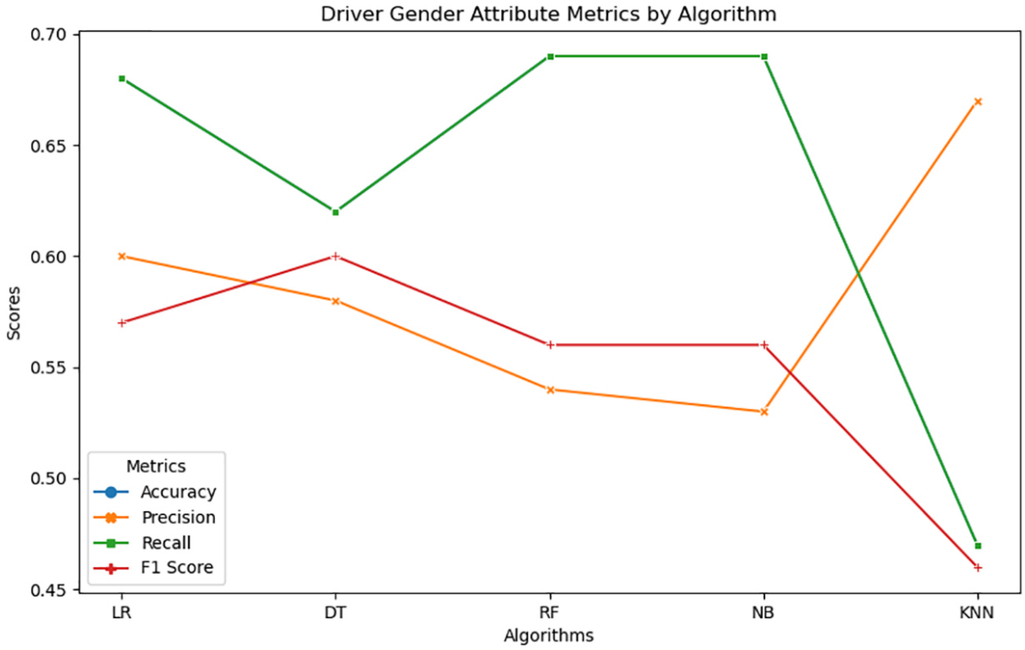

Fig. 1. Visualization of metrics scores on the algorithm for driver gender class

The analysis of the algorithms applied to the "Driver Gender" class yields some discerning insights. The accuracy and recall metrics shown in Figure 1 mirrored each other, with their lines overlapping, pointing to symmetry in the correct identification and retrieval of true cases for each class. Given that the driver's gender class is not unbalanced in the dataset, it was prudent to use the data in its existing state. Looking deeper into the performance of individual algorithms, we observe a nuanced depiction of their capabilities.

The Random Forest and Decision Tree classifiers stood out, commanding the highest scores in accuracy and recall, hovering around a commendable 69%. This performance slightly edges out Logistic Regression, which also proved a robust performance in these categories. Interestingly, Naive Bayes, despite its lower accuracy and recall scores, around 47%, managed to attain the highest precision at 67%. This shows the model's prowess in minimizing false positive rates, thus indicating a reliable performance in predicting true positive instances correctly.

The K-Nearest Neighbors (KNN) algorithm, however, presents itself as a balanced contender, highlighting a harmonized performance across all metrics. Notably, it bagged the highest F1 score, which is approximately 60%, indicating a favorable balance between precision and recall. This trait suggests KNN's potential to offer a reliable performance where minimizing both false positives and false negatives is a priority.

Furthermore, the F1 scores revealed a competitive edge for Logistic Regression and both the Decision Tree and Random Forest classifiers, all demonstrating a well-rounded performance with scores floating around the mid to high 50s. This paints a picture of a closely fought battle where the optimum choice of algorithm might boil down to specific nuances of a problem statement.

The "age_group" class was discretized into three distinct categories: "10–30 years", "31–65 years" and "66–100 years" This discretization facilitated the prediction of age-related classifications. As a preliminary step, we addressed the issue of class imbalance within the dataset by employing the Synthetic Minority Over-sampling Technique (SMOTE) algorithm. SMOTE allowed us to augment the representation of minority age groups, thereby mitigating the imbalance concern.

Among the models examined, Naive Bayes emerged as the most proficient predictor when considering its predictive performance in comparison to alternative algorithms. The tests are done in other datasets like Winston and Raleigh that are from the same source and show similar results, which tells that the models generalize well on new data.

Table 2. Results of all selected algorithms and metrics performance for the Durham dataset

Algorithms	Metrics	Classes		
		Driver Gender	Is Arrested	Age Group
Logistic Regression	Accuracy	0.64	0.94	0.32
	Precision	0.4	0.98	0.34
	Recall	0.64	0.94	0.32
	F1 Score	0.49	0.96	0.32
K-Nearest Neighbors (k = 5)	Accuracy	0.58	0.97	0.44
	Precision	0.55	0.96	0.38
	Recall	0.58	0.97	0.44
	F1 Score	0.56	0.97	0.41
Decision Tree	Accuracy	0.64	0.94	0.32
	Precision	0.4	0.98	0.45
	Recall	0.64	0.94	0.32
	F1 Score	0.49	0.96	0.34
Random Forest	Accuracy	0.64	0.94	0.32
	Precision	0.4	0.98	0.45
	Recall	0.64	0.94	0.32
	F1 Score	0.49	0.96	0.34
Naive Bayes	Accuracy	0.48	0.93	0.44
	Precision	0.65	0.97	0.46
	Recall	0.48	0.93	0.44
	F1 Score	0.44	0.94	0.37

To validate the suitability of our model, we have employed two more datasets to see how they perform with the same classes. In this scenario, the results from the Durham dataset shown in Table 2, underscore a remarkable level of generalization, especially in the classes "***Driver Gender***" and "***Is Arrested***." The performance metrics demonstrate strong consistency across multiple algorithms.

For example, the Logistic Regression, Decision Tree, and Random Forest algorithms all exhibited high accuracy and precision scores, particularly when predicting the "***Is Arrested***" class, with accuracy scores exceeding 90%. Furthermore, the F1 scores, which consider both precision and recall, are also remarkably high, suggesting a balanced and well-generalized model performance.

On the other hand, the data concerning the "***Age Group***" class shows a reasonable level of generalization, with an accuracy score of 44%.

Table 3. Results of all selected algorithms and metrics performance for the Burlington dataset

Algorithms	Metrics	Classes		
		Driver Gender	Is Arrested	Age Group
Logistic Regression	Accuracy	0.61	0.7	0.33
	Precision	0.37	0.98	0.45
	Recall	0.61	0.7	0.33
	F1 Score	0.46	0.81	0.33
K-Nearest Neighbors (k = 5)	Accuracy	0.57	0.99	0.5
	Precision	0.54	0.98	0.43
	Recall	0.57	0.99	0.5
	F1 Score	0.54	0.98	0.38
Decision Tree	Accuracy	0.61	0.7	0.33
	Precision	0.55	0.98	0.46
	Recall	0.61	0.7	0.33
	F1 Score	0.47	0.81	0.33
Random Forest	Accuracy	0.61	0.7	0.33
	Precision	0.56	0.98	0.46
	Recall	0.61	0.7	0.33
	F1 Score	0.47	0.81	0.33
Naive Bayes	Accuracy	0.46	0.97	0.5
	Precision	0.61	0.98	0.41
	Recall	0.46	0.97	0.5
	F1 Score	0.39	0.97	0.38

In the case of the Burlington dataset results shown in Table 3, it indicates a commendable level of generalization across different algorithms. The "***Is Arrested***" class stands out with consistently high accuracy and precision scores above 70% across multiple algorithms.

This clearly indicates that the algorithms can generalize well to new, unseen data, maintaining a high level of reliability and predictive power. Moreover, the K-Nearest Neighbors algorithm shows a notable improvement in the "***Age Group***" class with balanced performance, as indicated by an accuracy score of 50% and a recall score of 50%.

Furthermore, the consistently high precision scores observed across the board in predicting the "***Is Arrested***" class are particularly encouraging, highlighting the algorithms' capability to generalize effectively with a minimal rate of false positives. The Naive Bayes algorithm, for example, consistently maintains a precision score above 97% in predicting the "***Is Arrested***" class across all three selected datasets.

The variations in these metrics across different algorithms may be due to differences in how each algorithm handles the data and the nature of the classification task. For example, logistic regression tends to provide balanced results, while Decision Tree and Random Forest might be more prone to overfitting. The choice of algorithm has been made based on the specific requirements and trade-offs between precision and recall for each prediction task.

5 CONCLUSIONS

Nowadays, the application of machine learning algorithms holds a significant role, especially in situations involving the sensitive identification and classification of points or issues. In our case, the selected algorithms demonstrate remarkable potential for capturing the humor evident in traffic stops and everyday cases, irrespective of the growing digitalization of our surroundings.

Through this work, the results are presented, and we compare the performance of a total of (5) different algorithms in classifying cases when we encounter various classes. The selected algorithms all exhibit impressive performance, with the lowest value at 80% and the highest at 90%, according to the metrics we employed. The metrics we applied to ascertain the quality of the algorithms' outcomes include Accuracy, Precision, Recall, and F1 Score.

All these metrics have been applied to all algorithms, and their performance varies based on the classes that have been selected and the relationships they share with each other, ranging from the lowest to the highest. This variation is because some classes possess imbalanced data, and for this, we applied the SMOTE algorithm, which yielded highly favorable results in our context. As evident from the results presented in the table and visually, when we tested Naïve Bayes using Laplace correction, it produced significantly superior predictive outcomes compared to all other algorithms (but this has an impact in cases where the SMOTE algorithm is not used). This is primarily because Laplace correction has a highly beneficial impact on enhancing stability and managing discrete data, particularly when dealing with classes representing age groups.

Finally, the accuracy and recall results are 99%, with an F1 Score of 98% and Precision of 97%. This indicates that when working with this dataset, Naïve Bayes is well-suited for classification and analysis. Furthermore, in both datasets (Winston and Raleigh datasets) employed for model training and testing using the same algorithms as in the initial case, the performance is promising. Specifically, for the Logistic Regression, Decision Tree, and Random Forest algorithms, accuracy and precision, particularly for the "is arrested" class, surpass the 90% mark.

Additionally, the F1 Score, which supplies a comprehensive assessment of precision and recall, attains a promising performance as well. Meanwhile, the Naïve Bayes algorithm consistently achieves a precision rate exceeding 97% when predicting the "Is Arrested" class across all datasets used. The challenges and issues identified during this research primarily pertain to the handling of unbalanced and missing data.

6 REFERENCES

[1] E. Pierson *et al.*, "A large-scale analysis of racial disparities in police stops across the United States," *Nat Hum Behav,* vol. 4, no. 7, pp. 736–745, 2020. https://doi.org/10.1038/s41562-020-0858-1

[2] M. Nama, A. Nath, N. Bechra, J. Bhatia, S. Tanwar, M. Chaturvedi, and B. Sadoun, "Machine learning-based traffic scheduling techniques for intelligent transportation system: Opportunities and challenges," *Int J Commun Syst,* vol. 34, no. 9, pp. 1–25, 2021. https://doi.org/10.1002/dac.4814

[3] S. A. Mostafa, A. Mustapha, A. A. Ramli, M. F. M. D. Fudzee, D. Lim, and S. A Kashinath, "A machine learning-based data fusion model for online traffic violations analysis," in *International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing,* 2022, pp. 847–857. https://doi.org/10.1007/978-981-16-3071-2_67

[4] A. C. Uy, R. A. Bedruz, A. R. F. Quiros, A. A. Bandala, and E. P. Dadios, "Machine vision for traffic violation detection system through genetic algorithm," in *2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM),* Cebu, Philippines, 2015. https://doi.org/10.1109/HNICEM.2015.7393241

[5] Y.-W. Wu and T.-P. Hsu, "Mid-term prediction of at-fault crash driver frequency using fusion deep learning with city-level traffic violation data," *Accident Analysis and Prevention,* vol. 150, p. 105910, 2021. https://doi.org/10.1016/j.aap.2020.105910

[6] S. Asadianfam, M. Shamsi, and A. R. Kenari, "VD-MRDL: Traffic violation detection system using MapReduce-based deep learning for large-scale data," *Multimed Tools Appl,* vol. 80, no. 2, pp. 2489–2516, 2021. https://doi.org/10.1007/s11042-020-09714-8

[7] S. Usmankhujaev, S. Baydadaev, and K. J. Woo, "Real-time, deep learning based wrong direction detection," *Applied Sciences (Basel),* vol. 10, no. 7, p. 2453, 2020. https://doi.org/10.3390/app10072453

[8] S. Raj Anand, Naveen Kilari, and D. Udaya Suriya Raj Kumar, "Traffic signal violation detection using artificial intelligence and deep learning," *International Journal Advanced Research Engineering and Technology (IJARET),* vol. 12, no. 2, pp. 207–217, 2021. https://doi.org/10.34218/IJARET.12.2.2021.019

[9] A. Fang, C. Qiu, L. Zhao, and Y. Jin, "Driver risk assessment using traffic violation and accident data by machine learning approaches," in *3rd International Conference on Intelligent Transportation Engineering,* Singapore, 2018. https://doi.org/10.1109/ICITE.2018.8492665

[10] M. Zahid, Y. Chen, A. Jamal, K. Al-Ofi, and H. M. Al-Ahmadi, "Adopting machine learning and spatial analysis techniques for driver risk assessment: Insights from a case study," *International Journal of Environmental Research and Public Health,* vol. 17, no. 14, p. 5193, 2020. https://doi.org/10.3390/ijerph17145193

[11] P. Srinivas Reddy, T. Nishwa, R. Shiva Kiran Reddy, Ch. Sadviq, and K. Rithvik, "Traffic rules violation detection using machine learning techniques," in *Proceedings of the 6th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatre, India, 2021. https://doi.org/10.1109/ICCES51350.2021.9488998

[12] M. Ruben and J. Franklin, "Traffic signal violation detection using artificial intelligence and deep learning," in *5th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2020. https://doi.org/10.1109/ICCES48766.2020.9137873

[13] A. Omidi, A. Heydarian, A. Mohammadshahi, B. A. Beirami, and F. Haddadi, "An embedded deep learning-based package for traffic law enforcement package for traffic law enforcement," in *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada, 2021, pp. 262–271. https://doi.org/10.1109/ICCVW54120.2021.00034

[14] J. Vasilev, P. Petrov, and J. Jordanov, "A practical approach of data visualization from geographic information systems by using mobile technologies," *International Journal of Interactive Mobile Technologies (iJIM),* vol. 18, no. 3, pp. 4–15, 2024. https://doi.org/10.3991/ijim.v18i03.46655

[15] E. Pierson, S. Corbett-Davies, and S. Goel, "Fast threshold tests for detecting discrimination," in *International Conference on Artificial Intelligence and Statistics,* PMLR, 2018, pp. 96–105. https://doi.org/10.48550/arXiv.1702.08536

[16] C. Simoiu, S. Corbett-Davies, and S. Goel, "The problem of infra-marginality in outcome tests for discrimination," *The Annals of Applied Statistics,* vol. 11, no. 3, pp. 1193–1216, 2017. https://doi.org/10.1214/17-AOAS1058

7 AUTHORS

Mërgim H. Hoti is a PhD candidate at the South East European University in Computer Science. Also, currently, he is Teaching Assistant at the University of Prishtina "Hasan Prishtina", Faculty of Electrical and Computer Engineering and his fields of interest are Data Science, AI, Machine Learning and Cyber Security. He has several publications in these fields. This author can be contacted at email: mergim.hoti@uni-pr.edu.

Elvir Misini is an MSc candidate in Computer and Software Engineering at the University of Prishtina. Currently, Elvir works as a Backend Node.js developer, specializing in various databases such as MySQL, PostgreSQL and MongoDB. You can reach out to Elvir at email: elvir.misini@student.uni-pr.edu.

Uran Lajçi is an MSc candidate at the University of Prishtina in Computer and Software Engineering. Also, currently, he is working as a Data Scientist at Solaborate, and his fields of interest are Data Science, Machine Learning, Artificial Intelligence, Algorithms and building Software. This author can be contacted at email: uran.lajci@student.uni-pr.edu.

Lule Ahmedi is Prof. Dr. at the University of Prishtina "Hasan Prishtina", Faculty of Electrical and Computer Engineering and his fields of interests are Web Semantic, Recommender Systems, Social Network Analysis. She has several papers published in credible databases. This author can be contacted at email: lule.ahmedi@uni-pr.edu.