A Quantitative Approach to Road Safety in Morocco: Reducing Accidents through Predictive Modeling

Marwane Bel-lahcen[1], *El Arbi* Abdellaoui Alaoui[2,⋆], *Stéphane Cédric* Tékouabou Koumétio[3,4], and *Othmane NAGGAR* Naggar[5]

[1]*Laboratory ADMIR, Higher National School of Computer Science and System Analysis (ENSIAS), University of Mohamed V, Rabat, Morocco*
[2]*IEVIA team, IMAGE laboratory, Department of Sciences, ENS, Moulay Ismail University of Meknes, Morocco*
[3]*Center of Urban Systems (CUS), Mohamed VI Polytechnic University (UM6P), Lot 660, Hay Moulay Rachid, 43150, Ben Guerir, Morocco.*
[4]*LITE, Department of Computer and Educational Technology, Higher Teacher Training College, University of Yaoundé 1, Yaoundé, Cameroon.*
[5]*Moroccan Foundation for Advanced Science Innovation and Research (MAScIR), Rabat Design Center, Rue Mohamed Al Jazouli, Madinat Al Irfane, 10100 Rabat, Morocco.*

Abstract. This paper uses machine learning to predict road accidents in Morocco, a country marked by high annual accident rates. Our model employs data such as weather, time of day, and road conditions, derived from historical accidents and environmental records. Findings suggest that such predictive modeling can enable traffic authorities to anticipate high-risk situations and enact pre-emptive safety measures, contributing to significant reductions in road accidents. This study provides a data-driven approach towards policy implementation for road safety, with insights applicable to global road safety initiatives.

1 Introduction

Road safety is a critical issue that affects societies worldwide, not just in terms of human lives lost, but also through substantial economic consequences [1]. With the increasing rate of accidents, it has become paramount to explore effective strategies to manage and reduce their occurrence. Understanding the dynamics and factors contributing to road accidents could significantly help in mitigating these incidents, making our roads safer [2–4].

In this era of data-driven decisions, machine learning algorithms offer an innovative approach to solve complex problems. In the context of road safety, these algorithms provide a tool for modelling and predicting accident likelihood based on numerous factors. Through pattern recognition and predictive modelling, machine learning could be instrumental in identifying high-risk scenarios and thus enabling proactive safety measures [5–7]. This study explores this potential, demonstrating how technology could be leveraged in managing the pressing issue of road accidents.

In this paper, we provide a comprehensive quantitative analysis of road safety in Morocco, aiming to derive strategic insights to mitigate traffic accidents and enhance road safety. Morocco, with its significant yearly road accidents, represents an ideal case study for examining the various factors contributing to road mishaps and exploring potential solutions through predictive modeling.

The study leverages machine learning algorithms to predict the likelihood of accidents, based on various factors such as weather conditions, time of day, type of vehicle, and road conditions, among others. This model uses a robust dataset, including historical accident data and corresponding environmental conditions, gathered from multiple sources across the country.

The model's predictive capacity allows for pre-emptive measures, with potential application by traffic management authorities to foresee high-risk scenarios and implement necessary safety precautions. The findings indicate that predictive modeling can contribute substantially to reducing road accidents in Morocco by identifying patterns and risk factors that might otherwise remain unnoticed.

Through our analysis, this paper serves as a groundbreaking step towards data-driven policy implementation for road safety in Morocco and, by extension, offers critical insights that could be beneficial to road safety programs worldwide.

2 Methodology

In this section, we are going to write the different mechanisms, which we will use to reach the objectives of this study. The methodology of our study to build a machine

⋆e-mail: e.abdellaouialaoui@umi.ac.ma

learning model for predicting road accidents in Morocco follows a structured workflow divided into six main stages:

1. Data Collection: The first step is the accumulation of historical data related to road accidents in Morocco. This process involves gathering relevant information like weather conditions, time of day, type of vehicles, road conditions, and other significant factors from various data sources.

2. Preprocessing: Once the data is collected, it undergoes preprocessing, which includes cleaning, normalization, and feature selection. This step aims to prepare the data for machine learning algorithms, eliminating noise and irrelevant information, ensuring that the data fed into the model is of the highest quality.

3. Building the Model: Based on the processed data, a suitable machine learning model is developed. The model selection is guided by the nature of our problem and the type of data we have.

4. Training, Cross-Validation, and Testing: After model development, the data is split into training and testing sets. The model is trained on the training set and cross-validated to fine-tune its parameters. Subsequently, the model is tested to evaluate its predictive performance.

5. Evaluation: The model's performance is assessed through relevant evaluation metrics, such as accuracy, precision, recall, and F1 score. These metrics provide insights into how well the model can predict road accidents in Morocco.

6. Prediction: Finally, once the model is deemed satisfactory based on evaluation metrics, it is used to predict the likelihood of road accidents under varying conditions. These predictions enable traffic authorities to anticipate high-risk scenarios and enforce appropriate safety measures.

2.1 Models

The machine learning algorithms used in this study include K-Nearest Neighbors (KNN), Polynomial Regression (PR), Support Vector Regression (SVR), Decision Tree (DT), Random Forest (RF), and XGBoost.

1. K-Nearest Neighbors (KNN): KNN is a simple, instance-based learning algorithm that classifies instances based on their similarity to instances in the training dataset. In the context of this study, it would predict the number of road accidents by considering the 'k' most similar instances from the training data and taking the average of their outcomes. The 'k' parameter needs to be carefully selected to ensure a good balance between underfitting and overfitting.

2. Polynomial Regression (PR): Polynomial Regression is a type of regression analysis that models the relationship between the independent variable(s) and the dependent variable as an nth degree polynomial. In this study, it's used to identify potential non-linear relationships between our predictors and the number of road accidents.

3. Support Vector Regression (SVR): SVR is a variant of Support Vector Machines (SVM), a powerful machine learning model used for classification and regression tasks. In SVR, we try to fit the best line within a predefined or threshold error value. The objective of SVR is to minimize the coefficients — not to eliminate the error. SVR delivers a unique prediction that ignores the noise.

4. Decision Tree (DT): Decision Trees are simple yet effective machine learning models that predict the outcome based on decision rules inferred from the data features. In this context, it would model the decisions leading to a specific number of road accidents based on the various conditions (e.g., weather, time of day, road conditions, etc.).

5. Random Forest (RF): Random Forests are an ensemble learning method, which operates by constructing multiple decision trees during the training phase. The decision of the majority of the trees is chosen as the final decision. It is one of the most effective machine learning models for prediction, as it mitigates the overfitting problem common in single decision trees.

6. XGBoost: XGBoost stands for eXtreme Gradient Boosting, a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. It's an efficient and powerful model, known for its speed and performance. In the context of our study, it serves to sequentially correct the mistakes of weak classifiers and improve accuracy with each step.

Each of these algorithms was implemented and evaluated in this study, and their performance metrics were compared to identify the most effective one for predicting road accidents in Morocco.

3 Results analysis

In this study, we use six models, KNN, polynomial regression, SVR, decision tree, Random Forest and XGBoost for prediction the number of road accident in Morocco using a 34789 for the train and 8585 for test. Then, we used the shape value to explain the prediction value. The results of performance for the train and test of each previous model are presented in the table 1.

The table 1 provides an overview of the predictive performance of six different machine learning algorithms used to forecast road accidents in Morocco. These include K-nearest neighbors (KNN), Polynomial Regression (PR),

Figure 1. Flowchart for this study

Table 1. Evaluation of the performance prediction of our system with ensemble methods in terms of MAE, MSE and R^2.

Algorithm	R^2		MAE		RMSE	
	Train score	Test score	Train score	Test score	Train score	Test score
KNN	0.71	0.55	5.02	3.72	7.94	10.38
Polynomial Regression (PR)	0.24	0.21	8.59	8.05	12.93	13.85
SVR	0.06	0.07	5.9	5.75	15.30	16.17
Decision tree(DT)	0.89	0.89	2.00	1.91	5.06	5.02
Random Forest(RF)	0.98	0.91	1.73	0.58	1.53	4.46
XGBoost	0.93	0.92	1.40	1.57	3.32	3.45

Support Vector Regression (SVR), Decision Tree (DT), Random Forest (RF), and XGBoost.

The performance metrics include R-Squared (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) for both the training and testing datasets. These metrics allow us to assess the accuracy and robustness of each model.

1. K-Nearest Neighbors (KNN): With an R^2 score of 0.71 on training data and 0.55 on testing data, the KNN model seems to have moderate predictive performance. The relatively higher errors (MAE and RMSE) indicate room for improvement in its ability to predict road accidents accurately.

2. Polynomial Regression (PR): PR has a considerably low R^2 score (0.24 for training and 0.21 for testing), suggesting a weak fit to the data. This is confirmed by the highest MAE and RMSE values amongst all models, implying PR may not be the best choice for this problem.

3. Support Vector Regression (SVR): Similar to PR, SVR also shows a low R^2 score (0.06 for training and 0.07 for testing), indicating a poor fit. The error metrics are slightly better than PR but still high compared to other models.

4. Decision Tree (DT): The DT model exhibits good performance with high R^2 scores (0.89 for both training and testing), implying a good fit to the data. It also has relatively low MAE and RMSE, indicating a high predictive accuracy.

5. Random Forest (RF): RF outperforms all other models, boasting the highest R^2 score (0.98 for training and 0.91 for testing) and the lowest error metrics. This suggests that RF provides a highly accurate prediction with a robust fit to the data.

6. XGBoost: XGBoost also shows promising results with a high R^2 score (0.93 for training and 0.92 for testing) and lower error metrics, marking it as an efficient predictive model, though slightly less effective than RF.

In summary, both Random Forest and XGBoost models offer the best performance in predicting road accidents in Morocco, with RF showing a slight edge. Meanwhile, KNN provides moderate results, and the Polynomial Regression and SVR models may need to be reconsidered due to their lower performance.

4 Conclusion

In this study, we have employed six distinct machine learning algorithms to predict the occurrence of road accidents in Morocco based on historical data. The application of these models offers a data-driven approach to address the pressing issue of road safety, which is a significant concern worldwide. Our rigorous methodology ensured a robust and repeatable research process, enabling us to identify the most effective predictive models accurately.

The evaluation of the models revealed that the Random Forest and XGBoost algorithms outperformed others, offering the highest R^2 scores and the lowest error metrics on both the training and testing datasets. These models proved to be highly efficient in predicting road accidents in Morocco, which can be a powerful tool for the authorities in implementing preventive measures and improving road safety.

Meanwhile, algorithms like Polynomial Regression and Support Vector Regression did not perform as well, indicating that these may not be the optimal choices for this particular task.

In summary, our research underscores the potential of machine learning in addressing road safety issues. It provides valuable insights that can help in the development of efficient strategies for accident prevention and contributes to the larger goal of saving human lives. Future research could focus on optimizing the existing models further or explore other promising machine learning algorithms.

References

[1] S.A. Raza, A.W. Siddiqui, F.M. Butt, M.A. Elahi, K.S. Minhas, Data in Brief **44**, 108502 (2022)

[2] S. Colagrande, Transportation Research Procedia **60**, 504 (2022), new scenarios for safe mobility in urban areasProceedings of the XXV International Conference Living and Walking in Cities (LWC 2021), September 9-10, 2021, Brescia, Italy

[3] S. Ahmed, M.A. Hossain, S.K. Ray, M.M.I. Bhuiyan, S.R. Sabuj, Transportation Research Interdisciplinary Perspectives **19**, 100814 (2023)

[4] A.W. Siddiqui, S. Arshad Raza, M. Ather Elahi, K. Shahid Minhas, F. Muhammad Butt, Accident Analysis Prevention **174**, 106767 (2022)

[5] P. Gorzelanczyk, Results in Engineering **17**, 100981 (2023)

[6] P. Gorzelanczyk, Heliyon **9**, e12767 (2023)

[7] A. Gatera, M. Kuradusenge, G. Bajpai, C. Mikeka, S. Shrivastava, Scientific African **21**, e01739 (2023)