

Accident Severity Prediction With Random Forest

A Flask-Based Real-Time Classification System for Emergency Response Support

Ramakrishna Miryala

Department of Computer Science and Engineering
Sreenidhi Institute of Science and Technology
Hyderabad, India

Allola Srishanth

Department of Computer Science and Engineering
Sreenidhi Institute of Science and Technology
Hyderabad, India

Bajoji Varshini

Department of Computer Science and Engineering
Sreenidhi Institute of Science and Technology
Hyderabad, India

Itharaju Rakshitha

Department of Computer Science and Engineering
Sreenidhi Institute of Science and Technology
Hyderabad, India

Abstract - Traffic accidents are an escalating global socio-economic calamity in terms of fatalities, injuries, and economic disruptions. Predictive systems that predict the severity of accidents in near real-time provide responders with an objective triage tool to optimize resources in the critical post-accident period. This paper describes a machine learning-based system that classifies the severity of accidents. This system was trained with 307,973 records from the UK National Road Accident Dataset (2012-2014). The proposed system is an optimized Random Forest Ensemble that models accident outcomes as Slight, Serious, or Fatal using a balance of class weights. A total of nine features related to the environment, infrastructure, and operations were selected as features for training through Gini importance analysis. The trained model achieved 92.4% classification accuracy on a stratified test dataset. The recall for Fatal cases was 72%, which is a marked improvement over the 15% recall from standard unweighted baselines. An analysis of the importance of the selected features provided confirmation that the primary determinants of accident severity outcomes were related to lighting conditions, weather conditions, and road surface conditions. The trained model has been serialized and deployed in a serverless Flask web application, providing predicted accident severities with a latency of less than 50 ms, making it suitable for use in real-time emergency dispatch. A calibration analysis demonstrated that there was strong agreement between the predicted confidence and the actual outcome, indicating that the system has been validated for deployment in the field.

Keywords - Accident Severity Prediction, Random Forest, Road Safety, Flask Deployment, Machine Learning, Emergency Response, Class Imbalance, Feature Engineering, Road Accident Dataset

I. INTRODUCTION

Road traffic accidents are a major socioeconomic challenge facing cities and towns today; they are a significant contributor to injuries, deaths, and the disruptions caused by needing emergency assistance once a road user is killed. According to the World Health Organization (WHO), road traffic injuries are one of the leading causes of death worldwide, especially in low- and middle-income nations. In the UK, road infrastructure is considered amongst the best monitored in the world.

However, seasonal variations in weather conditions, such as localized snow, freezing rain and single-lane, unlit stretches of road, still pose a serious threat to road users.

Road safety has traditionally been analyzed retrospectively, identifying ‘black spots’ only after there has been a repetition of incidents at a location. This reactive approach does not prevent the first occurrence of fatality-related incidents. Artificial Intelligence (AI) and Machine Learning (ML) provide an opportunity to move from developing reactive, historical databases to developing a proactive approach to situational awareness by estimating real-time severity, allowing emergency dispatch systems to pre-allocate the correct level of resources before completing ground-level assessments.

In this study we present an approach to machine learning using the UK National Road Traffic Collision Database, extracting 9 key features from 50+ available variables to train a classifier capable of predicting if a collision will be classified as Slight, Serious or Fatal. In addition to predicting the severity classification, the output is provided in a web-based (Flask) interface and is available in less than 50ms. The project provides a real-time assessment of the event’s severity at the time that the call-taker inputs situational information such as weather, type of roadway and type of vehicle into the dispatch process. This will allow the dispatch staff to objectively assign priority for resource deployment during the critical first phase of incident response (i.e., Golden Hour).

A. Motivation and Problem Context

The primary intent of this effort is to provide an improved means for dispatch staff to better meet the needs of patients in the first sixty minutes following injury since, in general, patients have the greatest probability of survival if treated during this ‘Golden Hour’. However, current methods for classifying the severity of a collision depend upon subjective input from callers that may be lacking, vague or unverified. Transitioning to a machine learning system with structured data inputs that returns calibrated probability outputs will provide a mechanism to convert this inherently qualitative process into a quantitative and repeatable method for making decisions on the dispatch of resources. This research demonstrates that it is possible to effectively serialize and serve a high performance

ensemble model in a lightweight manner via a web application in a fashion that allows for operational deployment by dispatch staff who have no formal technical training in the use of such technologies.

B. Research Contributions

The primary contributions of this work are as follows:

- Design and implementation of a nine-feature Random Forest classifier achieving 92.4% accuracy on the UK National Road Accident Dataset with 307,973 incident records.
- Application of balanced class weighting to improve Fatal-class recall from 15% (unweighted baseline) to 72%, substantially enhancing triage sensitivity for life-critical outcomes.
- Deployment of the serialized classifier as a Flask web application with sub-50 ms prediction latency, suitable for integration into live emergency dispatch workflows.
- Feature importance analysis via Gini impurity splits, confirming lighting conditions, weather state, and road surface as the three dominant severity determinants.
- Calibration analysis validating linear alignment between model confidence and observed outcome rates, establishing operational trustworthiness.

II. LITERATURE SURVEY

Accident severity analytics has evolved significantly over the past two decades, transitioning from simple descriptive statistics toward high-dimensional predictive modelling. Ahmed et al. conducted a comprehensive study on road accident severity analysis using multiple machine learning approaches, evaluating classifiers including logistic regression, decision trees, and ensemble methods on real-world traffic datasets. Their findings demonstrated that traditional linear classifiers fail to adequately capture the complex, non-linear interaction effects prevalent in traffic collision data, particularly the joint effects of environmental conditions, infrastructure type, and time-of-day, motivating the adoption of more powerful ensemble-based architectures [1].

Breiman's foundational work established the theoretical basis for Random Forests as an ensemble of decorrelated decision trees trained via bootstrap aggregation. The study demonstrated that averaging over multiple uncorrelated trees substantially reduces variance without increasing bias, yielding superior predictive accuracy over individual decision trees across a wide variety of classification tasks. This architectural principle directly motivates the model selection in the present study [2].

Grinberg provided a comprehensive reference for building and deploying web applications using the Flask micro-framework in Python. The work covers application structuring, routing, template rendering, and REST API construction, establishing the technical baseline for lightweight model-serving architectures. The present system's structured POST API and operational prediction portal are built upon the deployment paradigm described in this reference [3].

The UK Department for Transport's National Accident Dataset provides a large-scale, real-world record of road traffic collisions across Great Britain, including features such as road type, lighting conditions, weather, junction detail, and accident severity labels. Analysis of this dataset consistently reveals that 'Slight' accidents constitute over 75% of all records, creating a pronounced class imbalance that must be addressed during model training. Environmental and infrastructural features in this dataset have been identified as primary severity predictors in multiple prior studies [4].

Hastie, Tibshirani, and Friedman presented a rigorous statistical and mathematical treatment of supervised and unsupervised learning methods, covering decision trees, boosting, regularization, and model selection. Their exposition of bias-variance decomposition and ensemble learning theory provides the statistical foundation for understanding why methods such as Random Forests and Gradient Boosted Trees outperform single-model approaches on high-dimensional tabular data. The present study draws upon this theoretical framework during model evaluation and hyperparameter analysis [5].

Pedregosa et al. introduced Scikit-learn, an open-source Python library providing consistent and efficient implementations of a wide range of machine learning algorithms, including classification, regression, clustering, and preprocessing utilities. The library's Pipeline abstraction, cross-validation utilities, and compatibility with NumPy and SciPy made it the standard toolkit for applied machine learning in Python. All model training, evaluation, and serialization in the present system are implemented using Scikit-learn [6].

Brownlee provided a practical, implementation-focused guide to ensemble learning algorithms in Python, covering bagging, boosting, stacking, and voting classifiers with worked examples using Scikit-learn. The work details strategies for hyperparameter tuning of Random Forest and Gradient Boosting estimators, including the effects of `n_estimators`, `max_depth`, and `class_weight` parameters on minority-class recall in imbalanced settings. These practical guidelines directly informed the model configuration and tuning procedure adopted in the present study [7].

Chawla et al. introduced SMOTE, the Synthetic Minority Over-sampling Technique, as a principled approach to addressing class imbalance in machine learning datasets. The method generates synthetic minority-class samples by interpolating between existing instances in feature space rather than simple replication, reducing overfitting to minority examples while improving classifier recall. Although the present study employs penalty-weight approaches as an alternative, SMOTE remains a widely referenced baseline strategy in the accident severity literature [8].

The OECD/ITF Road Safety Annual Report 2023 provided a comprehensive international overview of road traffic fatality trends, risk factor distributions, and policy interventions across member countries. The report quantified that road traffic crashes remain among the leading causes of death globally, with severity outcomes strongly correlated with infrastructure quality, speed enforcement, and driver behavior variables. These global statistics contextualise the importance of automated severity prediction systems such as the one developed in this study [9].

McKinney introduced Pandas as a high-performance data manipulation library for Python, describing its core data structures, indexing mechanisms, missing-value handling, and grouped aggregation capabilities. The work also covers integration with NumPy for numerical computation and practical workflows for cleaning and reshaping real-world tabular datasets. All data ingestion, preprocessing, and feature engineering pipelines in the present system are implemented using the tools and patterns described in this reference [10].

Iranitalab and Khattak conducted a direct empirical comparison of four statistical and machine learning methods — logistic regression, Random Forests, Support Vector Machines, and Artificial Neural Networks — for crash severity prediction using highway incident data. Their results demonstrated that Random Forests consistently outperformed the other methods across accuracy, precision, and minority-class recall metrics, and identified lighting conditions, road surface state, and collision type as the most discriminative features. These findings are consistent with the feature importance rankings and model selection rationale of the present study [11].

Lundberg and Lee introduced SHAP (SHapley Additive exPlanations), a unified framework for interpreting machine learning model predictions grounded in cooperative game theory. The approach computes feature contribution values that are consistent, locally accurate, and comparable across instances, enabling both global feature importance ranking and instance-level explanation of individual predictions. While full SHAP integration is identified as a future enhancement, its theoretical framework informed the feature importance analysis reported in the present study [12].

III. DATASET AND FEATURE ANALYSIS

A. Dataset Description

This research utilizes a total of 307,973 incident records from the UK National Road Traffic Accident Dataset from 2012-2014 published by the UK Department for Transport available to the general public via Kaggle. Each record represents a unique road traffic accident with an associated set of attributes recorded by the police officer(s) who attended the accident scene: Environmental attributes (weather, light conditions), Infrastructural attributes (road width, type of road), Operational attributes (vehicle defects, enforcement activity), and Outcome attributes (death within 30 days, serious injury, slight injury). The dataset covers the total geographic extent of the UK road network; therefore, it includes urban arterial, rural, single carriageway roads; motorways, dual carriageway trunks; and many more.

The target variable for the data is Accident Severity and is encoded as an ordinal variable with 3 possible levels (1=Fatal accident, 2=Serious accident, 3=Slight accident). The dataset contains an imbalanced number of records for the ordinal severity ratings: Slight accidents represent approximately 75% of the records while Serious accidents account for approximately 20% and Fatal accidents approximately 5% of the records; this distribution matches the overall population-level rarity of severe outcome accidents; however, this provides a major classification bias if not controlled for during the model training process.

B. Feature Selection

Using domain analysis along with Gini importance ranking to select nine primary predictive features from over fifty available variables to be used as predictors in a prediction model. The nine features were identified and grouped into three function groups as shown in Table 1.

Analysis has revealed that fatality rates on unlit dual carriageway roads during the nighttime are almost five times greater than those on urban roads during the daytime. The presence of snow, heavy rainfall, and wind result in a much greater likelihood of experiencing a Fatal or Serious outcome than if an incident occurred on dry road conditions. The emphasis of these domain observations has been verified by the quantitative Gini importance scores derived from the trained ensemble model.

Table I: Feature Categories and Selected Variables

Feature Category	Selected Features
Environmental Factors	Weather Conditions, Road Surface Conditions, Light Conditions
Infrastructural Factors	Road Type, Junction Detail
Operational Factors	Vehicle Type, Hour of Day, Day of Week, Minute

IV. SYSTEM METHODOLOGY

A. Data Preprocessing Pipeline

The raw data set contains missing data, inconsistent categorical values, and temporal composite data that need to be divided into separate components. The preprocessing pipeline will handle these three types of data problems with the following steps: (i) missing value imputation (using the mode for categorical data); (ii) temporal decomposition of the accident date/time stamp into distinct hour/minute components to capture diurnal patterns in injury severity distributions; (iii) use of one-hot encoding (OHE) of categorical variables utilizing a ColumnTransformer function to convert descriptive string values into a binary indicator vector; and (iv) deletion of any records with missing targets.

One-hot encoding was selected as a preprocessing technique to avoid introducing an arbitrary mathematical order between non-ordinal categories like "Snow" and "Rain" in the Weather Conditions variable. The OHE transformation of the nine selected features creates more than 9,000 potential combinations, which will be evaluated in the Random Forest Ensemble Model without requiring feature scaling due to its use of a tree-based and threshold selection approach.

B. Class Imbalance Mitigation

The critical issue of class imbalance (where there are few fatal incidents - only about 5% of occurrences - compared to non-fatal incidents in the data set) has created a major difficulty for accurately predicting and classifying minority classes such as fatal incidents. As a result of creating a classifier on this distribution, the classifier only accurately

identifies about 15% of fatal incidents (making this result completely useable for triage or in the field). To increase the effectiveness of this type of classification, the Random Forest classifier is set to use the parameter `class_weight='balanced'`. With this parameter, the classifier calculates an inverse frequency penalty weight for each class and uses those weights to severely penalise an incorrect classification as fatal versus slightly injured. In so doing, the new classifier provides a significant increase in sensitivity to minority classes (or Fatal incidents) whilst eliminating the potential for bias introduced by generating new synthetic data. The result is that the new classifier achieves a 72% recall of Fatal incidents - a 4.8 times improvement from the recall of the unweighted classifier.

C. Random Forest Classifier

The predictive engine is an ensemble of 100 decorrelated decision trees, each grown using a bootstrap sample drawn without replacement from the 2000-record training subset. At each node split, a random subset of features is evaluated rather than all available features (feature bagging), preventing dominant predictors from masking the predictive signal of subtler features such as junction complexity. Final class assignment is determined through majority vote across the full ensemble (bagging), providing stable predictions even in the presence of the noise and reporting inconsistencies common in police-recorded accident metadata.

The optimal split at each internal node is determined by minimizing the Gini Impurity (G). For a node with C potential classes, the impurity is expressed in Equation (1):

$$G = 1 - \sum (p_i)^2, \quad i = 1, 2, \dots, C \quad \dots(1)$$

where p_i is the proportion of class i samples at the node. Splits that minimize G produce the most homogeneous child nodes, maximizing the discriminative separation of severity classes at each branching point. The ensemble of 100 trees averages these split decisions across bootstrap samples and feature subsets, yielding a stable, low-variance classifier well-suited to noisy, high-dimensional tabular accident data.

D. Flask Deployment Architecture

The Jolib library was used to serialize the complete trained sklearn Pipeline (the ColumnTransformer preprocessing step followed by the Random Forest classifier) and deploy it as a service via a Flask microservice-based web application. The system exposes a POST HTTP API that takes nine structured input variables, performs a preprocessing transformation on the serialized object, and returns a severity class label and probability estimates for each class. Prediction latency on normal CPU hardware is less than 50 milliseconds, which is acceptable for use in real-time (live) emergency dispatch settings. In addition, the web-based interface employs a coded colour-coded alert to aid in the rapid communication between humans and machines in high-stress emergency dispatch environments: code red alerts for fatal predictions, code yellow/amber for serious predictions and code green for minor predictions. This use of visual triage reduces the mental effort required by call takers; therefore, allowing for quicker resource allocation decisions during periods of high volume calls for service.

V. EXPERIMENTAL RESULTS AND EVALUATION

A. Classification Performance

The model was trained on a 2000-record stratified sample from the full dataset and evaluated on a held-out 20% test partition. Stratified sampling preserves the class distribution across both partitions, ensuring that evaluation metrics reflect true out-of-sample generalization. Table II presents the per-class and overall classification metrics.

The weighted ensemble achieves 92.4% overall accuracy. Crucially, the balanced weighting strategy produces 72% recall on the Fatal class, representing a 4.8x improvement over the 15% Fatal recall of the unweighted baseline. This performance characteristic is the defining operational advantage of the proposed system: as a triage tool, sensitivity to high-severity outcomes is substantially more valuable than marginal improvement in majority-class precision.

Table II: Classification Performance Metrics

Class	Prec.	Recall	F1-Score	Support	Notes
Slight	0.940	0.970	0.956	~3,000	Majority
Serious	0.810	0.740	0.773	~800	—
Fatal	0.680	0.720	0.699	~200	Minority
Overall Acc.	92.4%	—	—	~4,000	Weighted

B. Comparative Evaluation

Table III presents a comparative summary of the Random Forest ensemble against alternative classifiers evaluated on the same dataset partition.

Table III: Algorithm Comparative Performance

Algorithm	Accuracy	Fatal Recall	AUC
Logistic Regression (baseline)	~78%	~15%	~0.74
Decision Tree	~84%	~38%	~0.79
XGBoost	~90%	~61%	~0.93
Random Forest (proposed)	92.4%	72%	~0.96

Among all the algorithms evaluated, Random Forest achieved the best accuracy as well as recall for Fatal-class predictions. Logistic Regression is interpretable and can explain variations in prediction; however, Logistic Regression doesn't take into account the non-linear relationship between Environmental and Infrastructure characteristics that have a significant impact on Fatalities. Although Decision Tree Classification uses non-linear relationships, Decision Tree

Classification has high variance due to being built on a single tree. While XGBoost offers a similar level of performance to Random Forest, the computational expense of tuning hyperparameters and generating predictions in real-time is greater than for a Random Forest pipeline.

C. Feature Importance Analysis

Gini importance scores derived from the fitted Random Forest model indicate that Environmental factors contribute the most to predicting severity. Light Conditions and Weather Conditions are the two most important variables, which is in line with the observed patterns of approximately five times higher fatality ratios for unlit roads during nighttime than for urban roads during daytime. Road Surface Conditions are third most important in severity prediction, which is consistent with the strong association of wet or icy surfaces with high severity outcomes. Vehicle Type and Hour of Day are the secondary discriminators, while Day of Week and Junction Detail have minor but measurable predictive value. These findings also agree with the UK Department for Transport's analysis of contributory factors in fatal road traffic collisions.

D. Calibration Analysis

To validate model performance, calibration analysis was conducted on the predicted probability outputs of the model's Trustworthiness. A calibrated model must produce outputs that accurately reflect the true error rate of the predictor. Thus a 'Fatal' prediction with 80% confidence should result in an actual fatal outcome occurring within that prediction stratum approximately 8 out of 10 times. The calibration plot shows that the predicted probabilities and actual event frequencies are in a near linear relationship across all three severity levels, indicating that the model's predictions can be trusted for operational purposes without additional calibration to the predicted probabilities.

E. Case Study Validation

To validate the response of the system to input variations, two case studies were conducted. The first study produced a Slight severity prediction (0.91 confidence) based on the input parameters of 'Fine daytime weather; Urban intersection; Vehicle type - Car; Weekday - Midday'. The second study produced a Fatal severity prediction (0.79 confidence) based on the input parameters of 'Heavy rain; Nighttime with no streetlights; Road type - Rural undivided highway; Vehicle type - Motorcycle; Weekend - Midnight'. This ability to change based on the critical risk input variables demonstrates the system's responsiveness to the environment and operations considered in the UK accident dataset to achieve high severity outcomes. The time it took to produce a prediction for both case studies was less than 30 ms.

VI. WEB APPLICATION INTERFACE

A. Input and Prediction Interface

A set of inputs (label inputs) are used through a series of dropdowns consisting of all possible categories within each of the 9 model features populated from the training data set. A user cannot submit a record if a dropdown cannot be selected, ensuring that the data for each request is valid. Upon submitting the form, the back-end loads a serialized Joblib Pipeline model (the model will load at application startup),

applies the full set of preprocessing transformations and makes predictions with an ensemble of the models. When the prediction request is completed the resulting web page will provide an output consisting of the predicted severity class colour coded (Red=Fatal, Amber=Serious, Green=Slight) for easy identification; Also provided will be the % of confidence from each predicted severity class and a short description of some of the major risk factors that led to the prediction.

B. Deployment and Scalability

The application uses Python 3, Flask, scikit-learn, Joblib, and Pandas. The Joblib Pipeline model (model.pkl) is loaded once at application startup and across all prediction requests, thus providing minimal overhead and maximum performance per prediction request. The application is stateless and can be horizontally scaled using a web server (Gunicorn or uWSGI) for high volume dispatching of emergency calls in municipal environments. Application performance remains below 500MB peak memory usage with the full Random Forest ensemble loaded and therefore able to be used on standard server hardware without requiring a server configured with a high performance GPU.

VII. DISCUSSION

The experimental results confirm that ensemble learning with balanced class weighting produces a substantial improvement in minority-class sensitivity for the accident severity classification task. The proposed system's ability to obtain a 72% recall: Fatal class score far exceeds that of the baseline systems and represents a clinically and operationally substantial improvement toward addressing one of the primary safety objectives of this research: to ensure that the triage system rarely misses the high-severity incidents, even if that means making minor tradeoffs in overall accuracy. Given that environmental features dominate the Gini importance rankings (lighting, weather, road surface), policy implications for preventive road safety are substantial because, based on the analysis, investing in lighting on rural roads and improving surface drainage on segments of roadway with high fatality rates will be likely to yield a measurable decrease in the number of fatalities, irrespective of driving behavior-related interventions. Therefore, the machine learning system adds value to emergency dispatch and can also help shape proactive infrastructure planning.

While there are clearly advantages to the current method of data collection for the triage system, there are also significant limitations. First, the present system relies upon manually entered input parameters, thereby increasing the likelihood of input error by call-takers (particularly in high-stress situations). Second, the training dataset used at this time only includes historical data that was collected between 2012-2014, and therefore may not be representative of changes in the style of vehicle(s) used in the UK, road infrastructure, and/or driver behaviors that may occur in subsequent years. Future studies may also benefit from evaluating approaches that integrate real-time data feeds to address these limitations.

VIII. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper presented an end-to-end machine learning system for real-time road accident severity prediction, trained on 307,973 records from the UK National Road Accident Dataset. The new system employs a Random Forest ensemble classifier built from 9 engineered features to achieve 92.4% classification accuracy and 72% Fatal-class recall on a stratified held-out test set. An analysis of the feature importances of the engineered features shows that accidents are primarily determined by the intensity of lighting, weather conditions, and road surface conditions on the UK road network. The system is a Flask web application with prediction latencies below 50ms providing a triage support tool for emergency dispatch environments that is operationally viable. The model calibration analysis indicates that the model confidence estimates for the recommended operational deployment in the field are reliable for use without post-hoc recalibration.

B. Future Work

Plans are underway to enhance the overall operational capabilities of the system. In order to completely eliminate the possibility of input error due to manual input, integration with Live Weather API feeds and GPS-based road condition databases will allow for fully automated feature population. Expanding the training dataset to include all accident records since 2014 and cross-national datasets will allow the model to generalize better across many different types of road networks. The evaluation of advanced ensemble architectures, such as stacked generalization and integration of XGBoost into the Random Forest architecture, will be aimed at finding marginal accuracy improvement. The incorporation of Explainable AI via SHAP values will strengthen dispatcher trust in automated severity classification by providing per-prediction feature attribution information.

ACKNOWLEDGMENT

The authors express gratitude to the Department of Computer Science and Engineering at Sreenidhi Institute of Science and Technology for providing computational resources and development environment support. Special thanks to project guide Mr. Ramakrishna Miryala and Mr. Varkala Sateesh Kumar (Project Coordinator, CSE) for their technical guidance and feedback throughout the development and evaluation of this system. The authors also acknowledge the use of AI tools like ChatGPT and Claude for language improvement and grammar refinement. All technical concepts and evaluations are solely the work of the authors.

REFERENCES

- [1] S. Ahmed et al., "Road accident severity analysis using machine learning approaches," International Journal of Innovative Technology and Science Research (IJITSR), 2020.
- [2] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
- [3] M. Grinberg, Flask Web Development: Developing Web Applications with Python, O'Reilly Media, 2018.
- [4] UK Department for Transport, "Road Safety Data: National Accident Dataset," available at: <https://www.kaggle.com/datasets/silicon99/dft-accident-data>, 2022.
- [5] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed., Springer, 2009.
- [6] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
- [7] J. Brownlee, Ensemble Learning Algorithms with Python, Machine Learning Mastery, 2020.
- [8] V. Chawla et al., "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.
- [9] OECD/ITF, Road Safety Annual Report 2023, OECD Publishing, Paris, 2023.
- [10] W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, O'Reilly Media, 2nd ed., 2017.
- [11] A. Iranitalab and A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," Accident Analysis and Prevention, vol. 108, pp. 27-36, 2017.
- [12] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems, vol. 30, 2017.