Machine Learning for Vehicle Insurance Fraud Detection:

A High-Recall Random Forest Approach

By: Rusi Rothschild

**Abstract**

Vehicle insurance fraud is a multi-billion dollar problem, costing the U.S. industry over $40 billion annually and forcing policyholders to pay hundreds more each year in inflated premiums. As fraud schemes become more sophisticated, outdated detection methods fail to keep up—making it harder for insurers to spot suspicious claims before payouts are made. This study aims to build a machine learning model capable of proactively identifying fraudulent claims and uncovering the features most predictive of fraud. Using a dataset of 15,420 claims—with only 6% labeled as fraudulent—a range of classification models were applied, including logistic regression, decision trees, random forest, and XGBoost. The data underwent preprocessing and feature engineering, and class imbalance was addressed through stratified sampling, SMOTE, and class weighting. Evaluation focused on recall, given the high cost of missed fraud. The final Random Forest model achieved a recall of 0.97 and an AUC of 0.81, outperforming other models in fraud detection while maintaining interpretability. SHAP analysis and feature importance rankings consistently identified key fraud signals such as fault attribution, policy type, vehicle price, deductible amount, age, and recent address changes. These findings suggest that a well-calibrated ensemble model, grounded in domain-relevant features, can improve fraud screening. By flagging high-risk claims early, insurers can better allocate investigative resources, reduce false payouts, and protect policyholders. This research contributes to the growing call for transparent, high-recall fraud detection systems—and lays a foundation for future work integrating external data sources, real-time deployment, and collaborative modeling across insurers.

**Introduction**

Insurance fraud is a widespread and persistent challenge in the financial services sector, particularly within the field of vehicle insurance. It contributes to billions in global losses each year and drives up premiums for honest policyholders. As fraudulent behavior becomes increasingly complex, traditional rule-based and manual detection methods are proving insufficient. These outdated systems often fail to catch sophisticated schemes in time, allowing fraud to propagate undetected (Odeyemi et al., 2024).

Vehicle insurance fraud is typically committed through staged collisions, falsified accident reports, inflated repair costs, or repeated use of images from previous damage claims (Maiano et al., 2023). Some fraudsters even rely on recycled or tampered vehicle parts and collude with repair shops or medical providers to generate illegitimate payouts (Schrijver et al., 2024).

Existing research spans diverse areas, including machine learning (ML) for fraud prediction, behavioral psychology of offenders, and ethical implications of automated systems. However, a gap remains in understanding which fraud indicators are consistently most effective across different models and operational contexts. This paper investigates that gap by exploring how ML tools, human behaviors, and systemic challenges affect vehicle fraud detection — with a focus on identifying and prioritizing the most predictive fraud signals.

ML models have become a cornerstone of fraud detection due to their scalability, adaptability, and pattern-recognition capabilities. Studies demonstrate that classifiers such as Random Forest (RF), Support Vector Machines (SVM), Logistic Regression (LR), and especially ensemble methods like XGBoost and AdaBoost, consistently outperform manual methods in detecting vehicle fraud (F. et al., 2023). In a blockchain-integrated system, XGBoost achieved 7% greater accuracy than decision trees (Dhieb et al., 2020), while Yücel (2022) found that artificial neural networks (ANN) trained with composite variables reached 76.56% accuracy.

Beyond supervised learning, hybrid approaches show promise. Muthura and Matheka (2023) applied a combined K-Means and SVM model to uncover fraud in health insurance — a method adaptable to auto claims. Maiano et al. (2023) introduced a deep learning-based image analysis system that identified repeat damage claims by comparing uploaded vehicle photos, illustrating how unstructured data can supplement structured models. While these approaches are effective, they are often applied without clear consensus on which features or fraud signals should be prioritized. Improved transparency and consistency in identifying high-impact indicators could help streamline model design across insurance providers.

ML models have become a cornerstone of fraud detection due to their scalability, adaptability, and pattern-recognition capabilities. Studies demonstrate that classifiers such as Random Forest (RF), Support Vector Machines (SVM), Logistic Regression (LR), and especially ensemble methods like XGBoost and AdaBoost, consistently outperform manual methods in detecting vehicle fraud (F. et al., 2023). In a blockchain-integrated system, XGBoost achieved 7% greater accuracy than decision trees (Dhieb et al., 2020), while Yücel (2022) found that artificial neural networks (ANN) trained with composite variables reached 76.56% accuracy.

Beyond supervised learning, hybrid approaches show promise. Muthura and Matheka (2023) applied a combined K-Means and SVM model to uncover fraud in health insurance — a method adaptable to auto claims. Maiano et al. (2023) introduced a deep learning-based image analysis system that identified repeat damage claims by comparing uploaded vehicle photos, illustrating how unstructured data can supplement structured models.

While these approaches are effective, they are often applied without clear consensus on which features or fraud signals should be prioritized. Improved transparency and consistency in identifying high-impact indicators could help streamline model design across insurance providers.

Even the most advanced algorithms cannot succeed without strong data infrastructure and ethical safeguards. Many insurers operate in silos, with limited data sharing and fragmented fraud tracking (Power & Power, 2015). This makes it harder to identify patterns across providers or track serial fraudsters. As Schrijver et al. (2024) observed, a heavy reliance on labeled, structured data also limits exploration of newer techniques like graph-based models or text mining.

Ethical concerns are mounting as well. Mullins et al. (2021) and Khaleel et al. (2023) highlighted issues of algorithmic bias, lack of transparency, and misuse of customer data. These risks are especially acute in AI-based systems that deny claims or flag individuals for investigation without clear justification. To address this, Odeyemi et al. (2024) emphasized the role of explainable AI (XAI) and federated learning — tools that improve both transparency and data privacy. Technological and ethical progress must go hand in hand. Models must not only detect fraud effectively but do so in ways that are fair, interpretable, and trusted by users.

Insurance companies can significantly improve fraud detection accuracy by prioritizing a small set of key fraud indicators — such as agent type, accident area, fault attribution, presence of witnesses, and whether a police report was filed — rather than relying solely on broad, generalized models.

**Methods**

This study aimed to detect fraudulent vehicle insurance claims using a supervised machine learning approach. Given the binary nature of the target variable, this was structured as a classification problem. A variety of classification algorithms were tested, including logistic regression, decision tree, and random forest models. The methodology emphasized model interpretability and robustness, with particular attention paid to handling class imbalance and evaluating recall, given the cost of undetected fraud.

The dataset used for this project was sourced from Kaggle and contains 15,420 entries and 33 columns, each representing an individual vehicle insurance claim. The target variable, FraudFound\_P, is binary and indicates whether a claim was classified as fraudulent (1) or not (0). The dataset includes a wide range of features related to policy details, customer demographics, claim timing, vehicle information, and incident specifics. Notable columns include AccidentArea, PolicyType, VehiclePrice, DriverRating, and WitnessPresent, among others. The structure consists of a mix of categorical and numerical data, with categorical fields dominating. Personally identifiable information was removed, and ID-like fields such as PolicyNumber and RepNumber were treated as non-predictive. The diversity of the dataset allows for meaningful exploration of potential fraud patterns across different types of claims.

Before modeling, several data preprocessing steps were carried out to clean and prepare the dataset. Since the dataset was fully complete with no missing values, imputation was not required. Categorical variables were encoded using different techniques based on their nature: ordinal categories (such as Month, VehiclePrice, and Days\_Policy\_Accident) were mapped to ordered integers using custom mappings, while nominal variables without intrinsic order (like Make and PolicyType) were label-encoded. Binary variables, including AccidentArea and WitnessPresent, were manually mapped to 0 and 1. The dataset did not require scaling or normalization due to the dominance of categorical variables, and no major outliers were detected in the numerical features after visual inspection. These steps ensured the data was in a consistent and usable format for machine learning algorithms.

Feature engineering focused on enhancing model performance and improving interpretability. Several features were engineered or recoded to better represent their underlying meaning. Ordinal variables were encoded with care to preserve their natural order, such as converting months and weekdays into numerical sequences. In addition, inconsistent entries in certain variables (e.g., MonthClaimed values of 0) were flagged and handled appropriately. High-cardinality identifiers like PolicyNumber and RepNumber were retained for tracking purposes but excluded from model training due to their lack of predictive value. No dimensionality reduction techniques were applied at this stage, as the number of features remained manageable and categorical encodings preserved distinct groupings. Overall, the engineered dataset retained all relevant claim-level features while ensuring model-ready formatting.

Multiple classification models were evaluated to assess predictive performance and interpretability. These included logistic regression, decision tree classifier, random forest, and XGBoost. Logistic regression was chosen for its simplicity and transparency, serving as a baseline. Decision trees allowed for interpretable rule-based predictions, while random forest was selected for its robustness to overfitting and its ability to handle mixed data types. XGBoost, a gradient boosting framework, was also implemented based on its strong preliminary performance in LazyPredict and its reputation for effectively handling imbalanced, tabular data.

Model implementation and evaluation were conducted using Python 3.10 with the scikit-learn, imbalanced-learn, and XGBoost libraries. LazyPredict was used to benchmark a wide range of models and identify strong candidates for more detailed experimentation. For final modeling, hyperparameter tuning and feature selection were explored on the top-performing algorithms.

The dataset was split into training and testing subsets using an 80/20 stratified split to preserve class proportions. For initial evaluations, models were trained on the imbalanced data using class\_weight adjustments, while SMOTE was also applied in additional experiments to generate synthetic examples of fraudulent claims. Logistic regression was trained on one-hot encoded and scaled features, while tree-based models used label-encoded features.

Hyperparameters for random forest and XGBoost were tuned using randomized search and grid search with cross-validation to estimate model generalizability. Threshold adjustments were also applied to maximize recall while maintaining acceptable precision, given the high cost of missing fraud cases.

Model performance was evaluated using standard classification metrics, including precision, recall, F1-score, and ROC-AUC. Due to the significant class imbalance and the consequences of undetected fraud, recall for the positive class (fraud) was prioritized. The F1-score was reported to balance precision and recall, especially given the potential for high false positive rates. ROC curves and AUC values were used to assess overall discriminative ability across thresholds.

For the XGBoost and random forest models, SHAP analysis was applied to interpret feature contributions and validate the most important predictors identified during model training. This interpretability step supported transparency and domain relevance for fraud detection.

The entire analysis was conducted using Python 3.10 in a Jupyter Notebook environment hosted on Google Colab. The modeling process was supported by commonly used libraries in the data science workflow. Pandas and NumPy were used for data manipulation and exploration, while Scikit-learn provided tools for preprocessing, modeling, and evaluation. Imbalanced-learn was used for applying SMOTE to address class imbalance. Matplotlib and Seaborn were used for data visualization, and SHAP was utilized for interpreting model predictions. LazyPredict was used during early stages to benchmark model performance. All work was performed on a cloud-hosted CPU runtime with 12 GB of RAM, suitable for training medium-scale machine learning models.

To support reproducibility, a consistent random seed (random\_state=42) was applied across all randomized processes, including train-test splits, SMOTE resampling, and model initialization. The full workflow was contained in a single, annotated Jupyter Notebook, allowing for step-by-step execution of the entire pipeline — from data loading to model evaluation. While the dataset is publicly available on Kaggle, all code, preprocessing logic, and model configurations are self-contained and can be rerun in the same Colab environment. [Future updates, including final XGBoost model tuning and feature importance analysis, will be added to the same reproducible notebook structure.]

**Results**

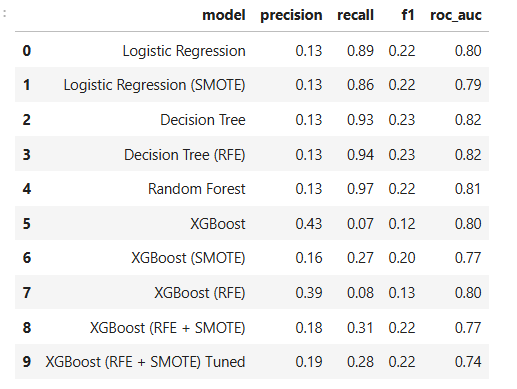
The primary goal of this analysis was to develop and evaluate machine learning models capable of reliably identifying fraudulent vehicle insurance claims, with a focus on maximizing fraud detection (high recall) while maintaining interpretability and acceptable levels of false positives. A secondary objective was to identify which features had the greatest impact on predicting fraud, in order to provide actionable insights for future investigations.

The dataset included 15,420 individual insurance claim records, spanning 33 features capturing demographic, policy, claim, and vehicle details. The target variable, *FraudFound\_P*, was highly imbalanced, with only about 6% of claims labeled as fraudulent (Appendix Figure A1). Descriptive statistics confirmed a median policyholder age of approximately 40, with a right-skewed distribution toward younger drivers. Most claims originated from urban accident areas, with Mondays and January showing peak frequencies for both accidents and claims (Appendix Figure A2). The deductible variable was constant across nearly all observations at $400, suggesting no discriminatory predictive value, while other variables such as vehicle price, driver rating, and policy type showed greater variability. These patterns were visualized using histograms and bar plots, helping to identify class imbalances and categorical skews (see Appendix Figures A2).

To address the class imbalance, a stratified train-test split was applied, with SMOTE resampling explored during model development. However, preliminary experiments showed that oversampling sometimes reduced recall — a critical measure in fraud detection — so the final models prioritized recall through threshold adjustments and class weighting rather than oversampling alone.

A standard logistic regression model achieved a recall of 0.89 and a precision of 0.13 for the fraud class, resulting in an F1-score of 0.22 and a ROC AUC of 0.80, with these results summarized in Table A1 below. To address class imbalance, a second logistic regression model was trained on SMOTE-resampled data. This SMOTE-based model showed a slightly lower recall of 0.86 with the same precision, and a similar F1-score of 0.22 with a ROC AUC of 0.79. The SMOTE-based version performed marginally worse in terms of recall than the non-SMOTE model, making it less favorable overall (Appendix Figure A3). While both logistic regression configurations improved fraud recall compared to a naive baseline, their recall was still not sufficient to prioritize fraud detection at the highest possible level.

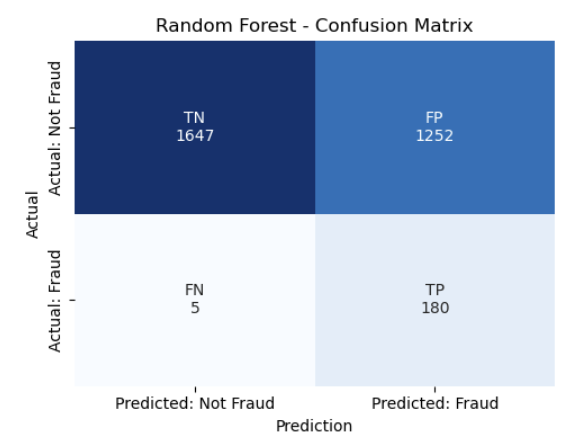
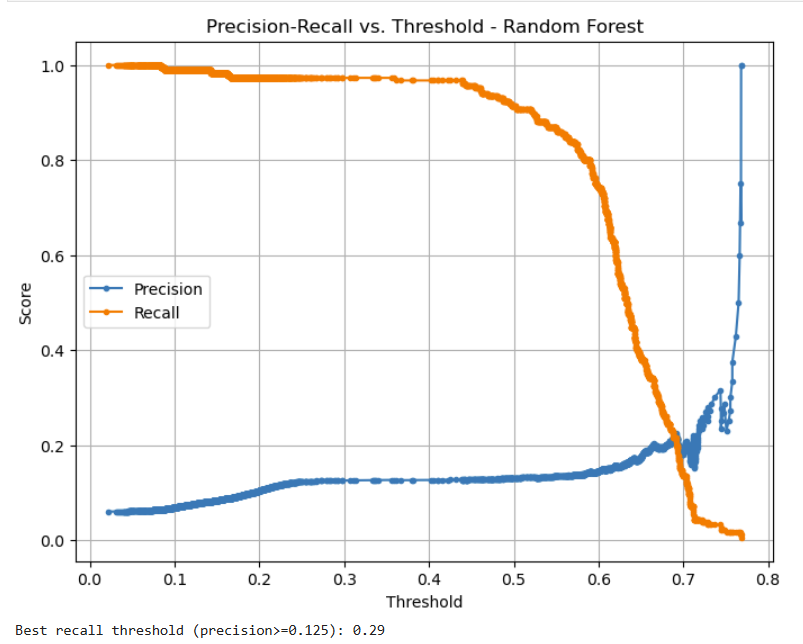
Subsequent testing with a decision tree classifier, enhanced through recursive feature elimination (RFE), yielded much stronger results. This model achieved a recall of 0.94 for fraudulent claims, with a precision of 0.13 and a ROC AUC around 0.82. Although the decision tree with RFE demonstrated higher ROC AUC than the Random Forest, its slightly lower recall meant it could still miss some fraud cases. Since maximizing fraud detection was the project’s primary objective, the Random Forest—achieving a recall of 0.97 and comparable precision—was ultimately selected despite its slightly lower ROC AUC of 0.81. Its higher recall ensured fewer fraudulent claims would go undetected, as reflected in the comparative performance shown in Table A1 below.

For XGBoost, I developed and evaluated four different versions, including standard, RFE-based, SMOTE-enhanced, and hyperparameter-tuned models. While these XGBoost configurations showed consistently strong ROC AUC scores and impressive precisions, their recall for the fraud class remained too low to justify adoption, as shown in Table A1, given the critical importance of detecting fraud in this setting. As a result, despite XGBoost’s overall high classification potential, the inability to sufficiently identify fraudulent claims rendered it ****impractical for this domain objective (Appendix Figure A5).

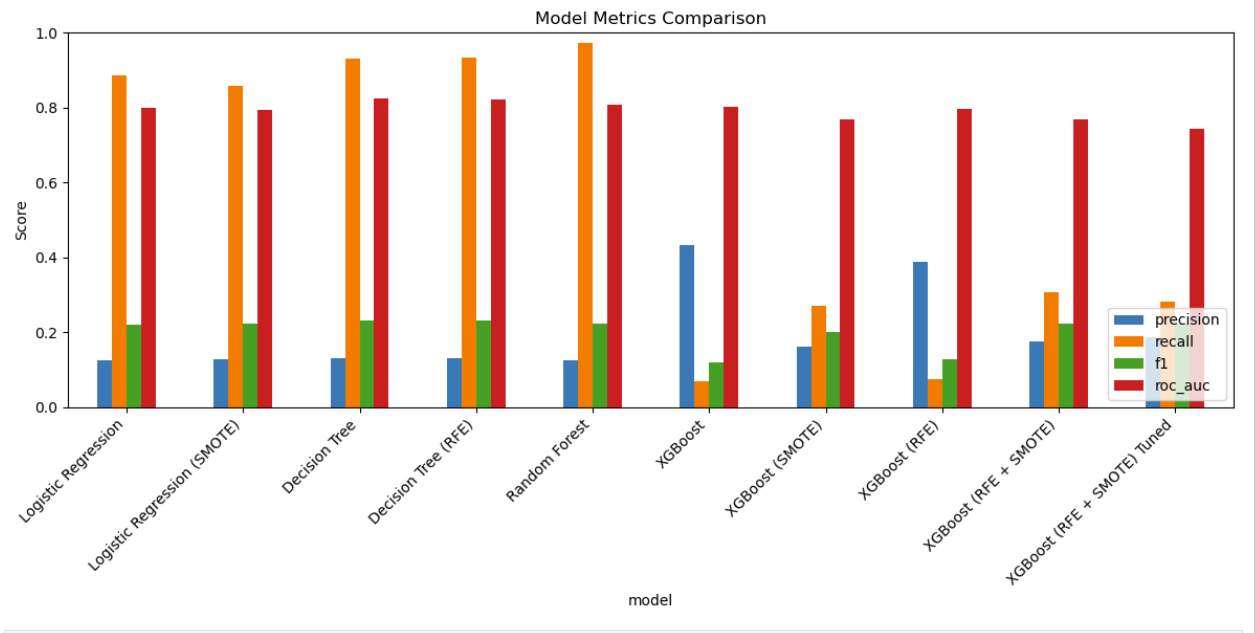
**Table A1** – *Model performance summary table* *of all tested*

*models based on precision, recall, F1-score, and ROC AUC.*

The final Random Forest model used 100 estimators, a maximum depth of 6, and a reduced decision threshold of 0.3 to favor recall, resulting in a precision of 0.13, recall of 0.97, F1-score of 0.22, and ROC AUC of 0.81. Confusion matrix -as shown in Figure A6- and precision-recall curve visualizations – shown in Figures A7- confirmed this trade-off, showing a high true positive rate for fraud detection, with an acceptable burden of false positives given the cost of missed fraud cases. Random Forest’s combination of simplicity, interpretability, and high recall ultimately made it the selected final model. The results of all model comparisons are displayed in a clustered bar chart shown in Figure A4, and a multi-model ROC curve (Appendix Figure A5), providing a clear comparative overview of precision, recall, F1-score, and ROC AUC.

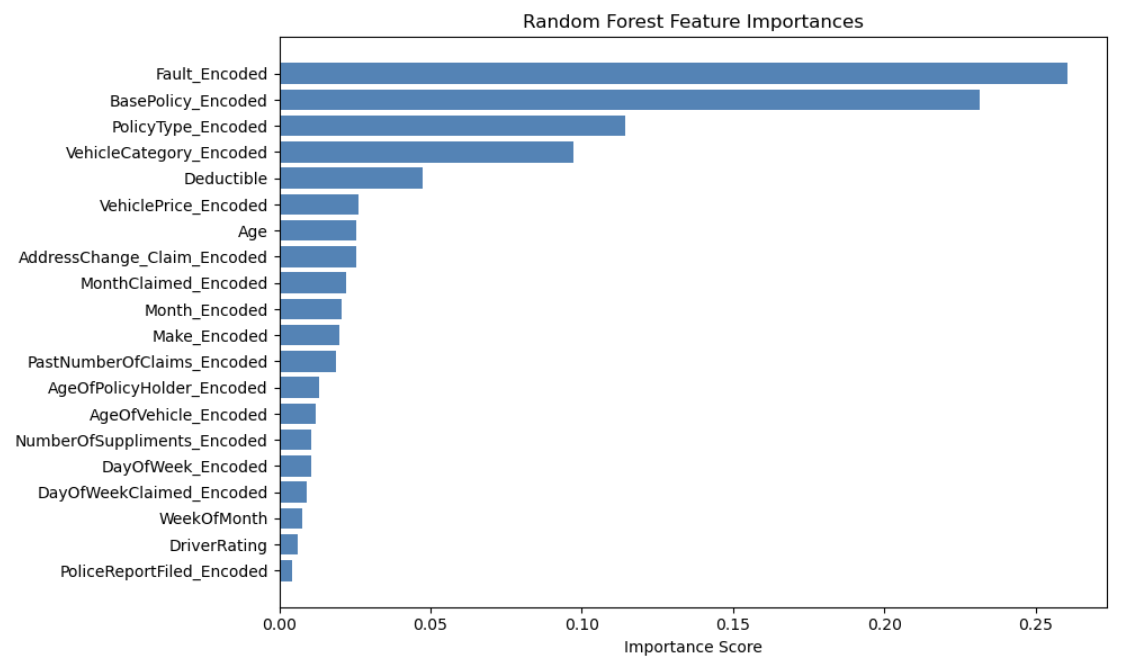
 

**Figure A6** – *Confusion matrix for Random Forest* **Figure A7** – *Precision-recall vs. threshold plot for Random Forest*



**Figure A4** *- Clustered bar chart comparing the precision, recall, F1, and ROC AUC for each model*

Beyond model-level performance, the analysis also focused on identifying key predictors of fraud. Feature importance derived from the Random Forest, visualized in Figure A8, highlighted Fault, BasePolicy, Vehicle Category, Deductible, Vehicle Price, Age, and Address Change as the strongest contributors to fraud classification. These features align with domain expectations: for example, policyholders changing their address shortly before or after filing a claim can be a fraud signal, while fault attribution and higher vehicle values tend to correlate with suspicious claims. To deepen interpretability, a SHAP summary analysis on the XGBoost model was performed, confirming Fault, BasePolicy, and Age as the most influential variables driving fraud predictions (Appendix Figure A9). These consistency checks bolster confidence in the stability and relevance of the identified predictors.



**Figure A8** – *Feature importance plot from Random Forest*

Diagnostic visualizations, including ROC curves, confusion matrices, and precision-recall curves (Appendix Figures A5, A6, and A7) helped evaluate model performance and trade-offs. These plots showed that the Random Forest model had a high true positive rate for fraud with a moderate false positive rate, suggesting good generalization without clear overfitting. The slightly lower ROC AUC for Random Forest compared to the Decision Tree indicated a minor trade-off in overall discrimination, but this was acceptable because Random Forest achieved the highest recall of 0.97 using a decision threshold of 0.3. Overall, these results suggest that Random Forest is a practical and robust choice for highlighting claims most likely to be fraudulent, supporting investigators in prioritizing their reviews.

**Discussion**

This study set out to develop a machine learning model capable of detecting fraudulent vehicle insurance claims—a problem that continues to cost the industry billions annually and burdens honest policyholders with rising premiums. The Random Forest classifier, optimized for high recall and supported by careful feature engineering and evaluation, emerged as the most effective model. Achieving a recall of 0.97 and an F1-score of 0.22 at a lowered threshold of 0.3, the model prioritized catching nearly every fraudulent claim. This was essential given the real-world cost of undetected fraud. While the precision of 0.13 indicates a higher rate of false positives, this was considered a reasonable trade-off in a context where human investigators are better equipped to manage occasional over-flagging than to miss costly instances of fraud.

One of the most striking outcomes of this project was the consistency of fraud signals across models. Features such as Fault attribution, Base Policy type, Vehicle Category, Vehicle Price, Age, Deductible amount, and recent Address Change repeatedly surfaced as top predictors of fraudulent behavior. These variables not only align with known patterns in fraud detection—like attempts to exploit policy gaps or file claims after suspicious address moves—but also offer actionable intelligence for insurers. Importantly, these findings were reinforced through SHAP analysis on an XGBoost model, which, despite lower recall, provided deeper interpretability of how these features influenced individual predictions. The fact that multiple models highlighted the same key features strengthens confidence in the relevance of these signals and suggests that targeted scrutiny of claims involving these characteristics could improve investigative efficiency.

The model’s strong performance highlights the effectiveness of ensemble learning for handling imbalanced classification tasks, particularly when coupled with domain-specific preprocessing. While logistic regression offered transparency and ease of deployment, it lacked the recall needed for a high-stakes detection problem. Decision trees, though interpretable, performed modestly. XGBoost showed strong overall metrics but fell short in recall, highlighting the need for context-driven metric prioritization in model selection. The final Random Forest model balanced performance with practical applicability, offering insurers a tool that could be realistically deployed to flag high-risk claims early in the assessment pipeline.

That said, the project does have several limitations worth noting. The dataset originated from a single source and time period, which may limit how well the model performs on newer claims or data from other insurance providers. The fraud labels were binary—simply marked as either fraud or not fraud—without any detail about the type or severity of fraud involved. This makes it difficult to understand what kinds of fraud the model is best at detecting. Additionally, although class balancing methods like SMOTE were tested to help the model learn from the small number of fraud cases, these techniques did not always improve results and sometimes reduced performance. The model was also built using only structured, pre-filled data such as claim timing, vehicle type, and policy details. It did not include more complex inputs like claim descriptions or photos, which could offer important context for detecting fraud. Finally, the model was trained and evaluated on a single dataset, so its performance on claims from other insurers or time periods remains untested.

Despite these constraints, the practical implications of the findings are compelling. In a real-world claim processing environment, deploying a Random Forest model trained on these fraud indicators could transform how insurers prioritize incoming cases. Claims with high predicted fraud probability could be fast-tracked to investigative teams, improving resource allocation and reducing payout delays. By using machine learning as a pre-screening tool rather than a final decision-maker, insurers can preserve human oversight while expanding their reach and precision. The model’s interpretability also supports regulatory compliance and stakeholder trust—two essential factors in the deployment of AI in financial services.

This work contributes to the broader research landscape by reinforcing the value of combining domain knowledge with ensemble learning for fraud detection. Unlike some studies that rely solely on performance metrics, this project emphasizes explainability and real-world feasibility. Moreover, it aligns with recent calls in the literature for greater transparency in feature selection, encouraging insurers to focus on a core set of behavioral and policy-related indicators rather than overly complex or hard-to-understand models. In contrast to prior work using black-box neural networks or experimental architectures, this study shows that a carefully tuned, interpretable ensemble model can deliver both accuracy and trust.

Looking ahead, several directions remain for future exploration. Expanding the dataset with more diverse or recent fraud cases could improve robustness. Integrating external data—such as credit scores, geographic data, or social media signals—might capture additional fraud cues. Applying privacy-preserving collaboration techniques could allow insurers to work together without sharing sensitive data, enhancing fraud detection at an industry level. Moreover, real-time deployment in production systems would enable dynamic model updates and continuous learning from new claim outcomes. A future focus on fairness and explainability will also be critical, especially if such models begin influencing claim decisions that affect customer payouts.

In conclusion, this research demonstrates the real potential of machine learning in enhancing the detection of vehicle insurance fraud. By uncovering a reliable set of predictive features and balancing recall with interpretability, the final model offers a concrete, scalable solution that can support fraud investigators while protecting honest policyholders. The findings highlight both the promise and the responsibility of using AI in sensitive, high-impact domains—and lay a foundation for smarter, fairer, and more efficient fraud prevention systems.

**References**

Dhieb, N., Ghazzai, H., Besbes, H., & Massoud, Y. (2020). A secure AI-driven architecture for automated

insurance systems: Fraud detection and risk measurement. *IEEE Access, 8*, 67555–67565. <https://doi.org/10.1109/ACCESS.2020.2983300>

F, A. A., A, O. O., S, A. O., A, A. J., I, O. H., & O, O. (2023). Prediction of automobile insurance fraud claims using machine learning. *The Scientific Temper, 14*(3). <https://doi.org/10.58414/scientifictemper.2023.14.3.29>

Firdaus, R., Xue, Y., Gang, L., & Sibt e Ali, M. (2022). Artificial intelligence and human psychology in online transaction fraud. *Frontiers in Psychology, 13*. <https://doi.org/10.3389/fpsyg.2022.947234>

Kadiu, F. (2015). Insurance fraud, case of Albania. *Mediterranean Journal of Social Sciences, 6*(3). <https://doi.org/10.5901/mjss.2015.v6n3s1p21>

Khaleel, M. F., Abdulkareem, Z. M., & Hamodi, Y. I. (2023). Developing ethical usage guidelines for customers of artificial intelligence and big data analytics: Ethical applications within realm of big data analytics. *Journal of Artificial Intelligence, Machine Learning and Neural Network, 41*, 39–50. <https://doi.org/10.55529/jaimlnn.41.39.50>

Maiano, L., Montuschi, A., Caserio, M., Ferri, E., Kieffer, F., Germanò, C., Baiocco, L., Ricciardi Celsi, L., Amerini, I., & Anagnostopoulos, A. (2023). A deep-learning–based antifraud system for car-insurance claims. *Expert Systems with Applications, 231*, 120644. <https://doi.org/10.1016/j.eswa.2023.120644>

Mall, S., Ghosh, P., & Shah, P. (2018). Management of fraud: Case of an Indian insurance company. *Accounting and Finance Research, 7*(3). <https://doi.org/10.5430/afr.v7n3p18>

Moradi, M., & Fateminejad, S. M. (2019). Sharing and analyzing data to reduce insurance fraud. *Journal of Management and Accounting Studies, 5*(3), 96–100. <https://doi.org/10.24200/jmas.vol5iss03pp96-100>

Mullins, M., Holland, C. P., & Cunneen, M. (2021). Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market. *Patterns, 2*(10), 100362. <https://doi.org/10.1016/j.patter.2021.100362>

Muthura, B. N., & Matheka, A. (2023). A hybrid model for detecting insurance fraud using K-Means and Support Vector Machine algorithms. *Open Journal for Information Technology, 6*(2), 51–61. <https://doi.org/10.32591/coas.ojit.0602.05143m>

Odeyemi, O., Mhlongo, N. Z., Nwankwo, E. E., & Soyombo, O. T. (2024). Reviewing the role of AI in fraud detection and prevention in financial services. *International Journal of Science and Research Archive, 11*(1). <https://doi.org/10.30574/ijsra.2024.11.1.0279>

Power, D. J., & Power, M. L. (2015). Sharing and analyzing data to reduce insurance fraud. *MWAIS 2015 Proceedings*, 19. <http://aisel.aisnet.org/mwais2015/19>

Schrijver, G., Sarmah, D. K., & El-Hajj, M. (2024). Automobile insurance fraud detection using data mining: A systematic literature review. *Intelligent Systems with Applications, 21*, 200340. <https://doi.org/10.1016/j.iswa.2024.200340>

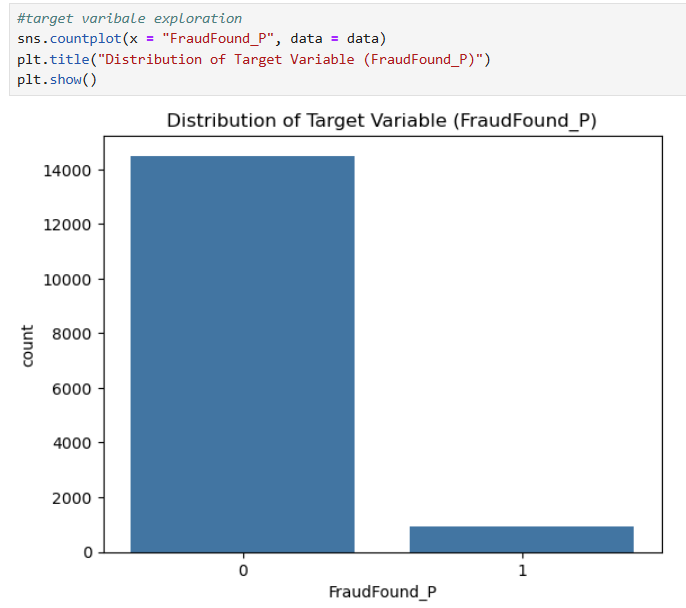
Sekh, A. A., Dogra, D. P., Kar, S., Roy, P. P., & Prasad, D. K. (2020). ELM-HTM guided bio-inspired unsupervised learning for anomalous trajectory classification. *Cognitive Systems Research, 63*, 68–79. <https://doi.org/10.1016/j.cogsys.2020.04.003>

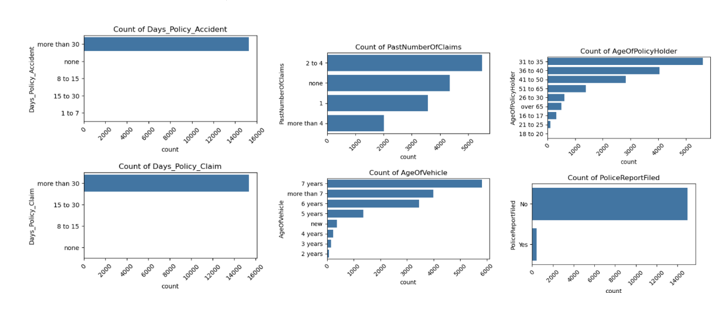
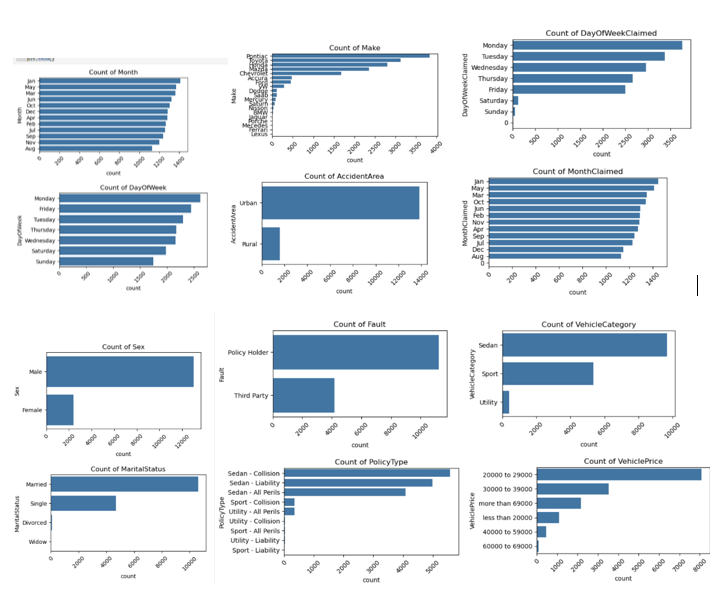
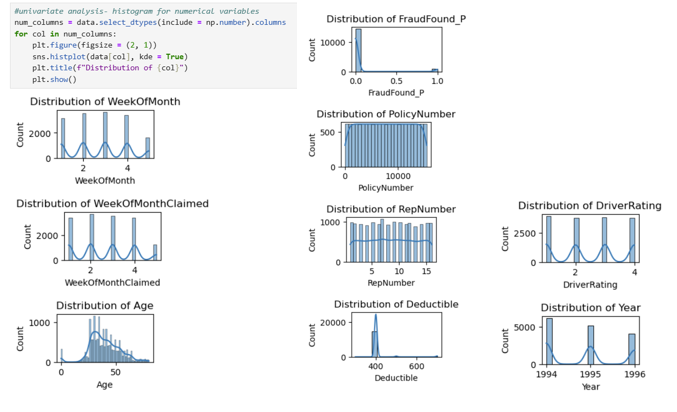
Videnović, S., & Hanić, A. (2021). Internal fraud committed by employees in insurance sector. *Tokovi Osiguranja, 37*(4), 81–97. <https://doi.org/10.5937/tokosig2102081v>

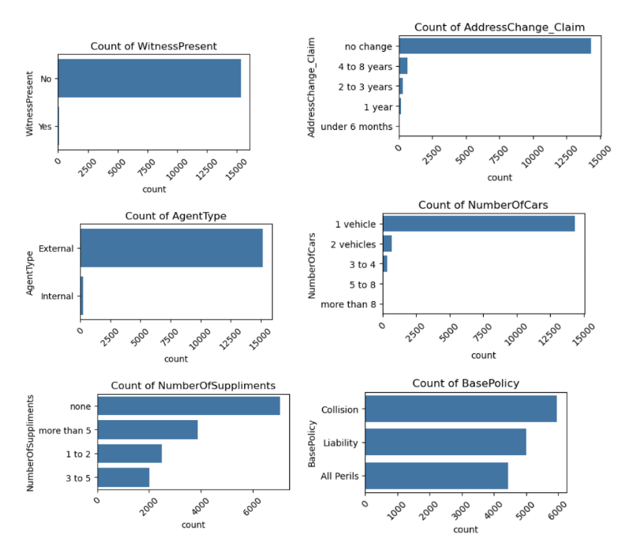
Yücel, A. (2022). A novel data processing approach to detect fraudulent insurance claims for physical damage to cars. *Journal of New Results in Science, 11*(2). <https://doi.org/10.54187/jnrs.1129440>

**Appendix**

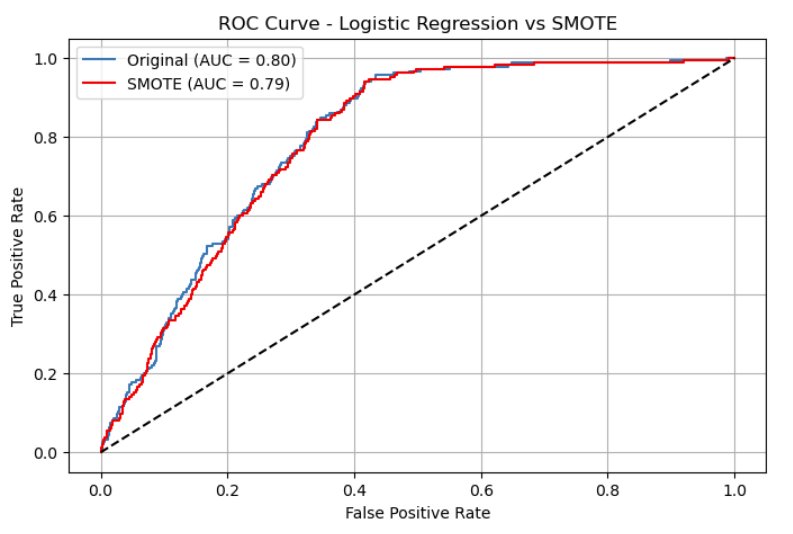
**Figure A1** – Target variable distribution (FraudFound\_P)



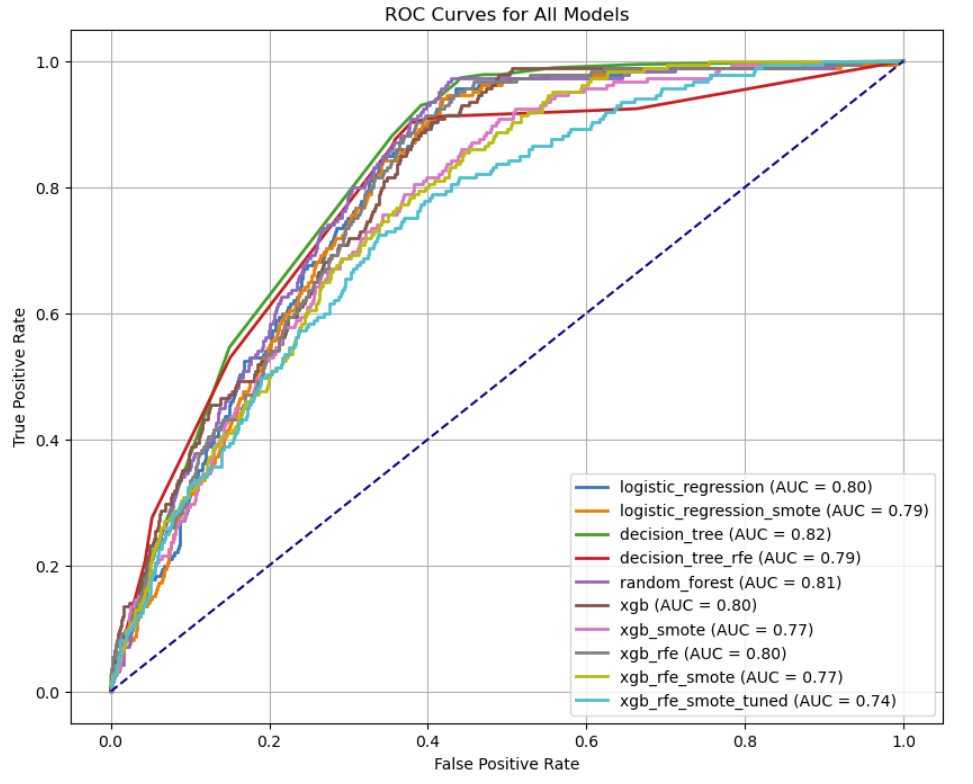
**Figure A2** – Histograms and bar plots of feature distributions



**Figure A3** – ROC curve comparing standard Logistic Regression vs Logistic Regression with SMOTE



**Figure A5** - ROC curves comparing all models, showing their true positive rates and area under the curve (AUC)



**Figure A9** – SHAP summary plot (XGBoost)  
