iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VII **Month of publication:** July 2025

DOI: https://doi.org/10.22214/ijraset.2025.73494

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

Towards Smart Fraud Detection: Integrating Machine Learning and Anomaly Detection in Vehicle Insurance Claims

Akula Sai Venkat Jashwanth[1], Dr. K. Santhi Sree[2]
[1]*Post Graduate Student, MCA, Department of Information Technology, Jawaharlal Nehru Technological University, Hyderabad, India*
[2]*Professor, Department of Information Technology, Jawaharlal Nehru Technological University, Hyderabad, India*

Abstract: ***Vehicle insurance fraud is a persistent challenge in the insurance sector, leading to substantial financial losses and operational inefficiencies. This research presents an intelligent and hybrid framework for fraud detection in vehicle insurance claims by combining supervised machine learning models with anomaly detection techniques. The proposed solution integrates a robust preprocessing pipeline that handles data cleaning, encoding, scaling, feature selection, and anomaly filtering using Isolation Forest and Autoencoders to remove suspicious data before model training. A strategic feature selection process was employed to retain the top 20 predictive features, ensuring model interpretability and performance. Multiple models including Random Forest, XGBoost, and Logistic Regression were trained and evaluated, with a meta-model ensemble delivering the final prediction. Additionally, the system is deployed as a secure web application featuring admin authentication, real-time predictions, and model explainability through SHAP and LIME visualizations. The results demonstrate that this hybrid approach significantly enhances fraud detection accuracy and provides transparency in decision-making. This solution has the potential to streamline claims processing and support insurance companies in reducing fraud-related risks.***

Keywords: Vehicle Insurance Fraud, Machine Learning, Anomaly Detection, Isolation Forest, Autoencoder, Feature Selection, Meta-Modeling, SHAP, LIME, Fraud Detection Pipeline, Insurance Claim Prediction, Ensemble Learning, Data Preprocessing, Secure Web Application, Explainable AI (XAI).

I. INTRODUCTION

The rapid growth of the vehicle insurance industry has been accompanied by a parallel rise in fraudulent claims, posing serious threats to insurers' financial stability and operational efficiency. Detecting fraud in insurance claims is a complex task due to the diversity of data, evolving fraud patterns, and the subtlety with which fraudulent activities are often disguised. Traditional rule-based systems are increasingly proving inadequate in identifying sophisticated and dynamic fraud behavior. As a result, insurance companies are turning towards intelligent, data-driven approaches for more accurate and scalable fraud detection.

With the advent of machine learning and anomaly detection, there is now the potential to build models that not only learn patterns from historical claim data but also detect outliers and unusual behavior that may indicate fraudulent activity. Integrating these technologies allows for a proactive approach to fraud prevention rather than reactive claim investigation. Furthermore, incorporating explainability through tools such as SHAP and LIME helps ensure that predictions made by these models are transparent and can be trusted by end-users.

This research introduces a hybrid fraud detection pipeline tailored for vehicle insurance claims, combining machine learning algorithms with anomaly detection techniques and interpretability tools. The pipeline is embedded in a secure, user-friendly web application that enables real-time fraud prediction while offering administrative access control and prediction transparency.

A. Objective

The primary objective of this research is to develop and deploy a smart, scalable, and explainable fraud detection system for vehicle insurance claims using a combination of machine learning and anomaly detection. The goal is to:

- Build a robust preprocessing pipeline to handle data quality and feature engineering.
- Detect and remove anomalous or suspicious data using Isolation Forest and Autoencoder techniques.
- Train multiple classification models and integrate the best performers into an ensemble meta-model.

©IJRASET: All Rights are Reserved

- Provide prediction transparency using SHAP and LIME.
- Deploy the system as a secure web application with real-time fraud detection and explainable outputs for administrative users.

This work aims to support insurance companies in minimizing fraud-related losses while promoting trust and accountability through explainable AI.

II. LITERATURE SURVEY

MachinyaTongesai et al. [1] proposed a machine learning-based fraud detection system focused on insurance claims, utilizing the XGBoost algorithm due to its high performance in terms of precision, recall, and execution time. Their study emphasized the importance of preprocessing and feature selection to enhance fraud identification while aiming to reduce financial losses and streamline claims processing. Despite its strong predictive capability, the model's limitations include class imbalance, dependency on computational resources, and lack of consideration for contextual or behavioral data. Moreover, the single-source dataset used in the study restricts the model's applicability to diverse insurance markets.

In a multi-algorithmic approach, Arif Ismail Alrais [2] designed a fraud detection framework that leverages traditional classifiers such as Random Forest, Logistic Regression, K-Nearest Neighbors (KNN), and XGBoost. The study explored the role of exploratory data analysis and algorithmic comparison in improving the efficiency of insurance fraud detection. While the model achieved reasonable accuracy, its reliance on an outdated dataset (1994–1996) and imbalanced class distribution posed challenges in generalizability and training efficiency. Additionally, computational constraints limited the depth of experimentation, and findings lacked applicability to modern, real-time claim data.

Chamal Gomes et al. [3] investigated the use of unsupervised deep learning models, namely Autoencoders (AEs) and Variational Autoencoders (VAEs), to detect anomalous insurance claims. Their model introduced a variable importance technique to help interpret which features influenced fraud detection most significantly. The research demonstrated the viability of fraud detection without relying on labeled datasets, which is particularly useful in cases of limited supervision. However, the approach was hindered by the use of outdated data, difficulty in model validation due to unsupervised learning, and substantial computational demands. Furthermore, the system lacked real-time adaptability to evolving fraud patterns.

Adopting a sequential anomaly detection strategy, Hugo Cedervall and Anton Hansson [4] developed an unsupervised LSTM-based autoencoder model to identify fraudulent claim sequences in Swedish home insurance data. Their solution eliminated the need for labeled data and achieved interpretable, near state-of-the-art results in detecting suspicious activity. Nonetheless, the model's assumptions - particularly that all anomalies imply fraud - raised concerns about false positives. Additionally, the study was confined to a narrow insurance domain and faced limitations in terms of scalability and ease of interpretation for non-technical stakeholders.

III. METHODOLOGY OF PROPOSED SYSTEM

A. *Proposed System*

The proposed system is a smart, modular fraud detection framework for vehicle insurance claims. It integrates machine learning, anomaly detection, and explainable AI (XAI) to improve prediction accuracy and trust. Designed for real-world deployment, the system emphasizes scalability, transparency, and ease of use.

At a high level, the system performs data preprocessing, anomaly filtering, model training using multiple classifiers, and combines them via stacking. Explainable AI tools such as SHAP and LIME are integrated into a user-friendly web application to offer interpretability. The application allows secure login, real-time data input, prediction display, and graphical explanation of decisions. This framework serves as a reliable and interpretable tool for identifying fraudulent claims in real-time.

B. *Dataset Description*

The dataset utilized in this study comprises 1,000 vehicle insurance claim records, each characterized by 39 distinct features that encapsulate both personal and incident-related information. The primary objective of this dataset is to enable the classification of claims as either fraudulent or genuine, as indicated by the target variable fraud_reported.The dataset includes a mix of numerical, categorical, and datetime features, spanning multiple dimensions:

1) Customer Profile Information:

Attributes such as age, policy_number, policy_bind_date, state, insured_zip, insured_sex, insured_education_level, insured_occupation, and insured_relationship offer insights into the demographic and policy-related details of the insured individuals.

2) Vehicle and Policy Details:

Fields like auto_make, auto_model, auto_year, policy_annual_premium, umbrella_limit, and deductible provide information about the insured vehicle and the structure of the insurance policy.

3) Incident Information:

Features such as incident_date, incident_type, collision_type, incident_severity, authorities_contacted, incident_location, incident_hour_of_the_day, and number_of_vehicles_involved detail the nature and context of the reported incident.

4) Claim-Related Data:

Variables like total_claim_amount, injury_claim, property_claim, and vehicle_claim capture the financial aspects of the claims.

5) Behavioral& Miscellaneous:

Other attributes such as hobbies, capital-gains, capital-loss, bodily_injuries, and witnesses are included to potentially expose behavioral patterns or red flags relevant to fraud.

The dataset exhibits a balanced distribution of numerical (int64, float64), categorical (object), and temporal (datetime64[ns]) datatypes. This diversity necessitates careful preprocessing, including encoding, scaling, and feature engineering.

To prepare the data for modeling, missing values were addressed, categorical features were encoded using a combination of Label Encoding, Ordinal Encoding, and Target Encoding, and numerical features were standardized using MinMaxScaler. Additionally, domain knowledge was used to derive engineered features that improve the fraud detection capability of the model.

This real-world-style dataset enables the development of a robust and generalizable fraud detection pipeline by simulating complex fraud patterns that insurance companies frequently encounter.

C. System Architecture

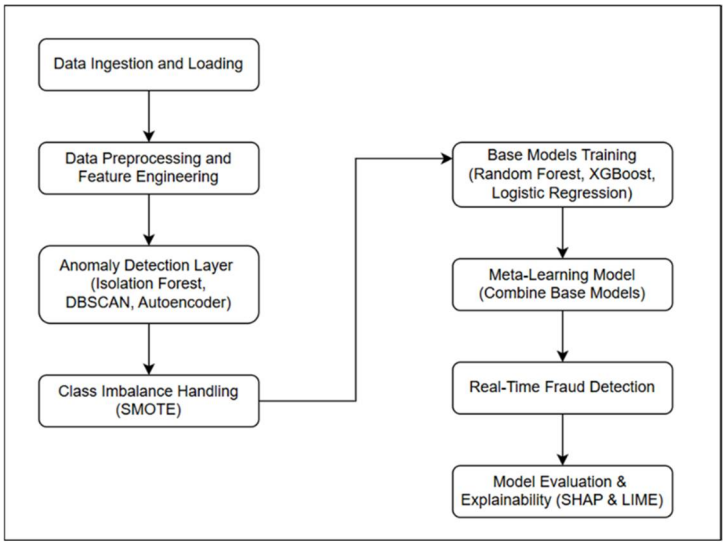


Figure-1: System Architecture of the Vehicle Insurance Fraud Detection System

The proposed system follows a modular and layered architecture designed to enhance fraud detection accuracy in vehicle insurance claims. Each component of the pipeline plays a vital role in building a robust, interpretable, and real-time prediction system. The workflow can be described as follows:

1) Data Ingestion and Loading

The process begins with importing raw insurance claim data from various sources. This includes structured data such as customer profiles, claim history, and policy details. Ensuring that the data is accurately ingested and efficiently stored is crucial for maintaining the integrity of downstream processes.

2) Data Preprocessing and Feature Engineering

Before model training, the raw data is cleaned, transformed, and enriched with engineered features. Preprocessing steps include handling missing values, encoding categorical variables, and standardizing numerical features. Feature engineering is applied to extract meaningful patterns and relationships that improve the predictive power of the model.

3) Anomaly Detection Layer (Isolation Forest, DBSCAN, Autoencoder)

A critical layer in the architecture focuses on detecting abnormal or suspicious claim patterns. This is achieved using a hybrid approach that integrates:

- Isolation Forest: Efficiently isolates anomalies by constructing random trees.
- DBSCAN: A clustering-based method that groups similar data points and identifies outliers.
- Autoencoder: A neural network that learns a compressed representation and flags deviations in reconstruction. These techniques help to flag potentially fraudulent claims even before classification, providing an extra layer of insight.

4) Class Imbalance Handling (SMOTE)

Fraudulent cases are often rare in real-world datasets, leading to class imbalance. To mitigate this, SMOTE (Synthetic Minority Oversampling Technique) is used to generate synthetic examples of the minority class, ensuring that the model doesn't become biased towards the majority (non-fraudulent) class.

5) Base Models Training (Random Forest, XGBoost, Logistic Regression)

Multiple base classifiers are trained to independently learn fraud patterns. These include:

- Random Forest for robust decision trees,
- XGBoost for gradient boosting and high performance,
- Logistic Regression for baseline probabilistic interpretation.

Training diverse models allows the system to capture a wide variety of fraud behaviors.

6) Meta-Learning Model (Combine Base Models)

The outputs of the base models are then fed into a meta-learning model, which learns to combine their predictions for better overall performance. This ensemble technique enhances generalization and reduces model-specific biases.

7) Real-Time Fraud Detection

Once the meta-model is trained, it is deployed to evaluate incoming claims in real-time. This ensures that suspicious claims are flagged immediately, allowing for swift action by investigators or automated systems.

8) Model Evaluation and Explainability (SHAP & LIME)

Finally, the system includes explainability tools to make the model's decisions interpretable:

- SHAP (SHapley Additive exPlanations) provides global and local feature impact scores.
- LIME (Local Interpretable Model-agnostic Explanations) offers localized explanations for individual predictions.

These tools build trust with stakeholders and help validate the rationale behind flagged frauds.

D. Methodology

The proposed system follows a systematic, multi-stage methodology combining advanced machine learning techniques with anomaly detection algorithms to enhance the accuracy of insurance fraud prediction. The methodology is designed to address challenges like data imbalance, noise, and feature redundancy, and consists of the following phases:

1) Data Preprocessing and Cleaning

The raw dataset is first inspected to handle missing values, inconsistent entries, and irrelevant or redundant features. Categorical features are encoded using appropriate encoding techniques such as Label Encoding, Ordinal Encoding, and One-Hot Encoding, depending on the nature and number of categories. Continuous variables are scaled using StandardScaler to normalize the range and improve model convergence.

2) Feature Engineering

To enrich the dataset, new meaningful features are derived from existing ones. These include:

- Claim Ratio: Ratio of total claim amount to annual premium.
- Vehicle Age: Computed from the vehicle's manufacturing year.
- Policy Tenure: Time elapsed since the policy bind date.

These engineered features enhance the model's ability to learn fraud patterns more effectively.

3) Feature Selection

The next step involves selecting the top 20 most influential features using a Random Forest model's feature importance scores. This reduces dimensionality, improves training efficiency, and minimizes noise.

4) Anomaly Detection

To detect outliers that may skew model performance, two unsupervised anomaly detection techniques are applied:

- Isolation Forest: Identifies anomalies by randomly partitioning data and isolating outliers in fewer steps.
- Autoencoder: A neural network-based model that learns a compressed representation of the data. High reconstruction error indicates potential anomalies.

Data points flagged as anomalous by both models are excluded from further training.

5) Data Balancing Using SMOTE

Due to class imbalance in fraud vs. non-fraud cases, Synthetic Minority Oversampling Technique (SMOTE) is applied to generate synthetic examples of the minority class. This helps the model learn minority class patterns more effectively, improving recall and F1-score.

6) Model Training

Multiple machine learning classifiers - Random Forest, XGBoost, LightGBM, and CatBoost - are trained on the preprocessed and balanced dataset. Each model is evaluated based on accuracy, precision, recall, and F1-score.

7) Meta Model Selection

The best-performing model (Random Forest) is selected based on its performance metrics and saved for deployment. This model is used in the final web application for fraud prediction.

8) Model Explainability

To enhance transparency and trust in the predictions:

- LIME (Local Interpretable Model-Agnostic Explanations) is used in the web app to show feature contributions for individual predictions.
- This aligns with the principles of Explainable AI (XAI) and helps end-users and domain experts understand the reasoning behind a classification.

This structured methodology ensures a robust, accurate, and explainable fraud detection system that can be practically applied in insurance claim processing workflows.

IV. EXPERIMENTAL ANALYSIS AND RESULTS

A. Key Features

1) Integrated Machine Learning Pipeline
2) Advanced Anomaly Detection
3) Feature Engineering and Selection
4) Explainable AI (XAI) Integration
5) Secure and Functional Web Application
6) Multi-Model Comparison
7) Scalable Design

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

8) Data Preprocessing Pipeline
9) Confidence-based Predictions
10) Real-world Dataset Utilization

B. Results

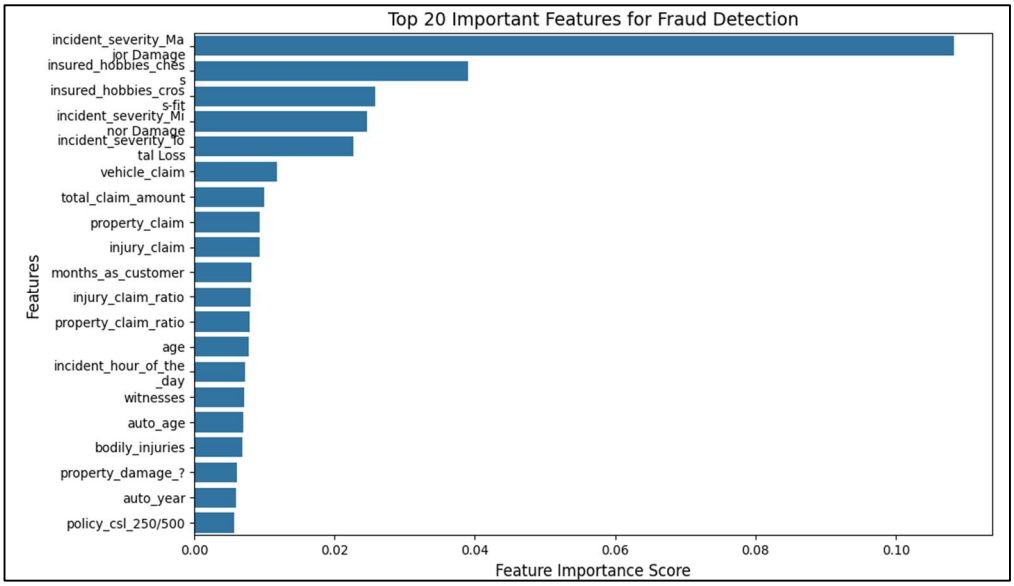


Figure-2: Top 20 features of the dataset after Feature Selection

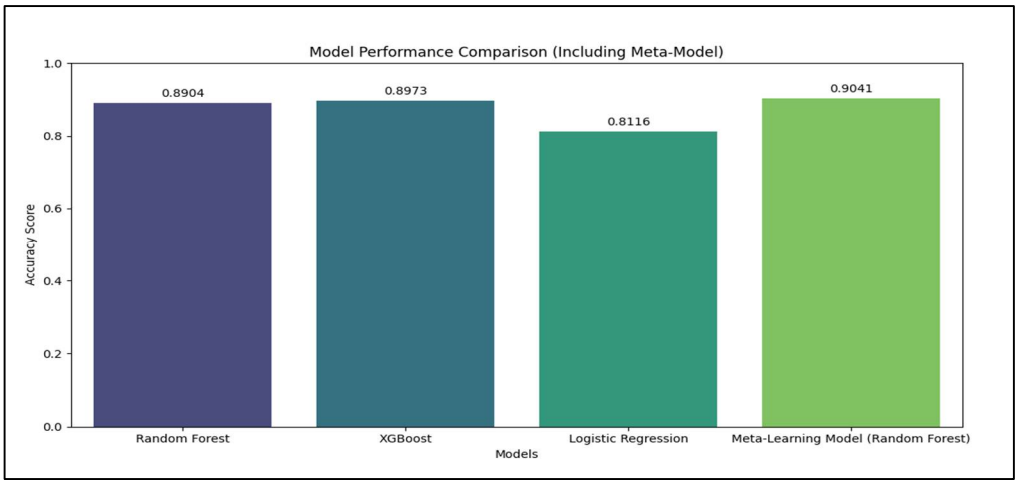


Figure-3: Comparison of Model Accuracies. Meta-Learning Model outperforms other base models

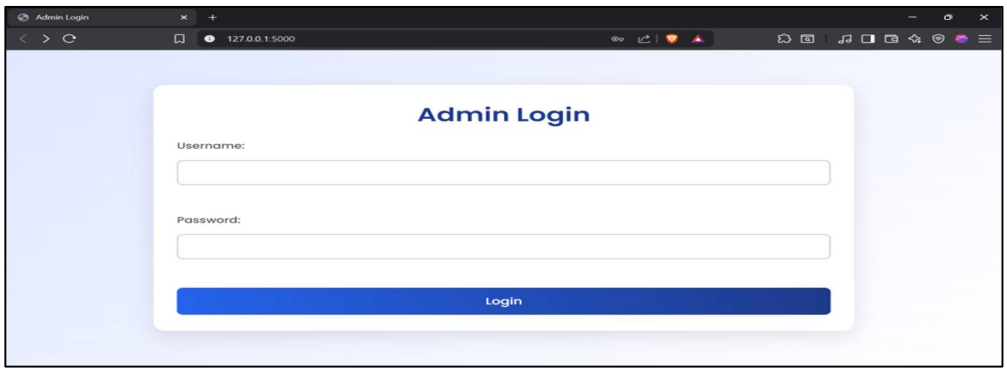


Figure-4: Insurance Company Administrator Login Page

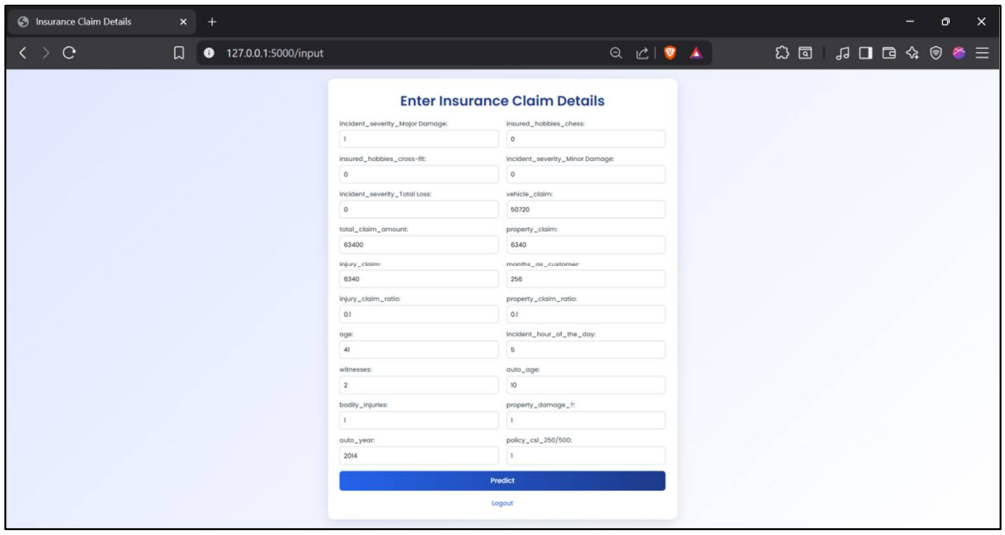


Figure-5: Form to enter the insurance claim details

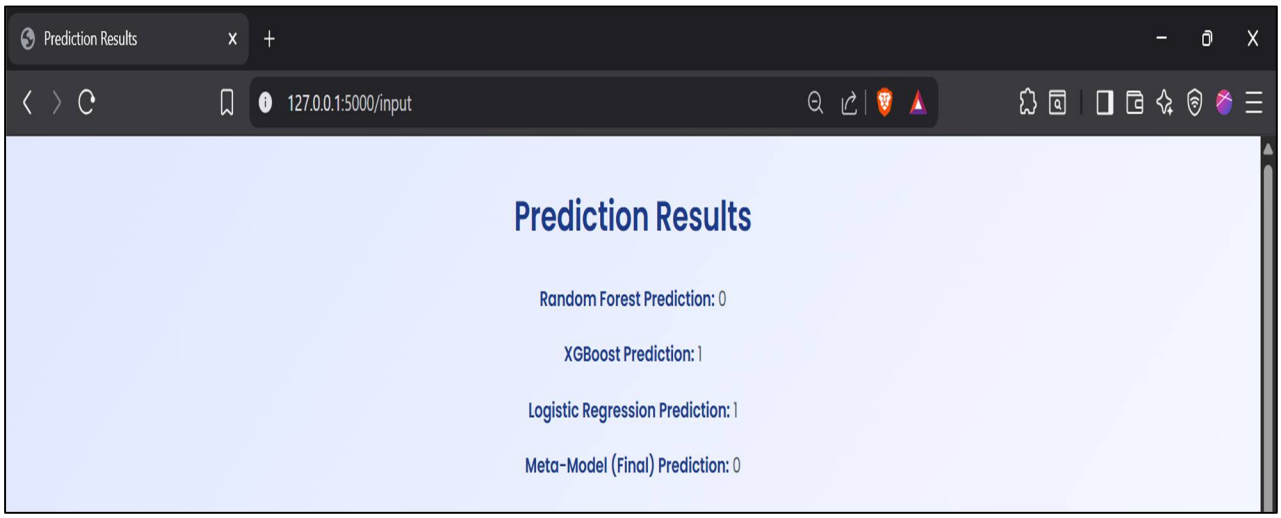


Figure-6: Result Page displaying the predictions made by individual and meta classifiers

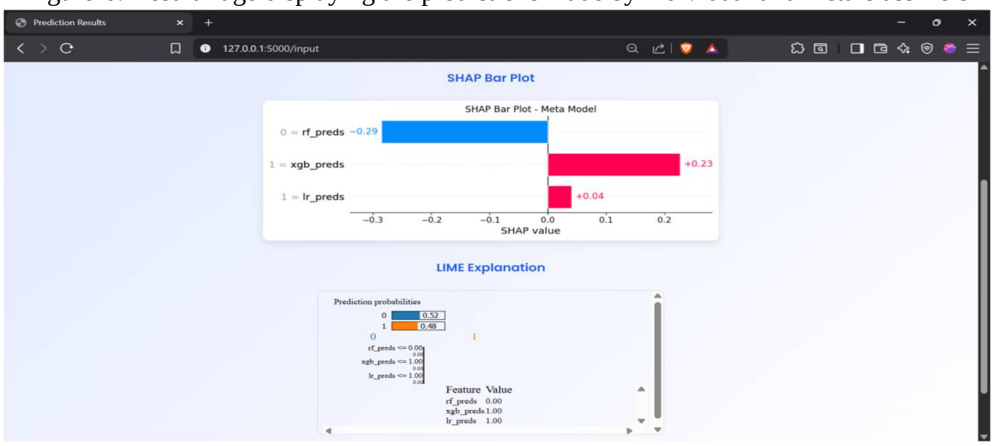


Figure-7: SHAP and LIME plots for the prediction

The final prediction results of the fraud detection system were obtained using three base models - Random Forest, XGBoost, and Logistic Regression - and a meta-model that combines their predictions. In one test case, the Random Forest and Logistic Regression models predicted class 0 (non-fraud), while XGBoost predicted class 1 (fraud). The meta-model, which considers the outputs of these individual models, ultimately predicted class 0. This ensemble decision highlights how the final output is influenced not by a single model but by the collective evidence provided by all base learners.

To explain this prediction, SHAP (SHapley Additive exPlanations) values were analyzed. The SHAP bar plot revealed that the Random Forest model had the most significant influence on the meta-model's decision, contributing a SHAP value of -0.29, which strongly pushed the prediction toward class 0. In contrast, XGBoost had a SHAP value of +0.23, indicating moderate support for class 1. Logistic Regression contributed a minor negative SHAP value of -0.04, also supporting class 0. Overall, the SHAP explanation confirmed that the Random Forest model's output played a dominant role in determining the final prediction as non-fraud.

Complementing the SHAP analysis, the LIME (Local Interpretable Model-agnostic Explanations) visualization further clarified the rationale behind the meta-model's decision. LIME showed that the model predicted class 0 with a high confidence score of 0.81. Features such as rf_preds = 0 and lr_preds = 0 were shown to contribute positively toward the prediction of class 0, while xgb_preds = 1 supported class 1 but with lesser impact. This local interpretability provided by LIME affirms that the final decision of the meta-model aligns well with the collective influence of its base models, particularly favoring the non-fraudulent class.

V. LIMITATIONS AND FUTURE SCOPE

Despite the effectiveness of the proposed fraud detection framework, several limitations need to be considered. Firstly, the system has been trained on a single dataset related to vehicle insurance claims, which may not capture the full diversity of fraud patterns found across different regions or insurance companies. This limitation may affect the model's generalizability in broader real-world applications. Secondly, the current implementation depends solely on structured data with a fixed set of features. It does not yet leverage unstructured or multimodal data such as claim narratives, audio conversations, or vehicle images, which are often rich sources of fraud indicators. Another constraint is that the system performs binary classification - labeling claims as either fraudulent or legitimate - without accounting for partial or ambiguous fraud cases that might exist on a spectrum. Additionally, the model is static and does not include an automated retraining mechanism, making it vulnerable to performance degradation over time due to evolving fraud tactics or changes in customer behavior. While explainability tools such as LIME are integrated, their outputs may still be too technical for non-expert users to interpret effectively without additional guidance or training.

Looking ahead, several opportunities exist to enhance and extend the system. Future versions could integrate real-time claim data and behavioral telemetry from vehicles to support more responsive and context-aware fraud detection. The inclusion of multimodal data sources, such as text from claim reports, voice logs, or image evidence, combined with natural language processing and computer vision techniques, would significantly increase the robustness and accuracy of predictions.

Moreover, implementing adaptive learning strategies - such as periodic retraining using fresh data or feedback from insurance investigators - can help the system stay current with emerging fraud trends. Instead of a strict binary output, future iterations could introduce a fraud risk scoring system that classifies claims based on likelihood or severity, offering greater flexibility for decision-making. From a deployment perspective, the web application can be expanded with secure API endpoints, multi-user access controls, and cloud-based hosting to support real-time production use. Furthermore, integration with regulatory platforms or national fraud registries could foster better collaboration and early detection of organized fraud schemes. These future enhancements aim to not only improve technical performance but also align the system more closely with the practical needs and constraints of the insurance industry.

VI. CONCLUSION

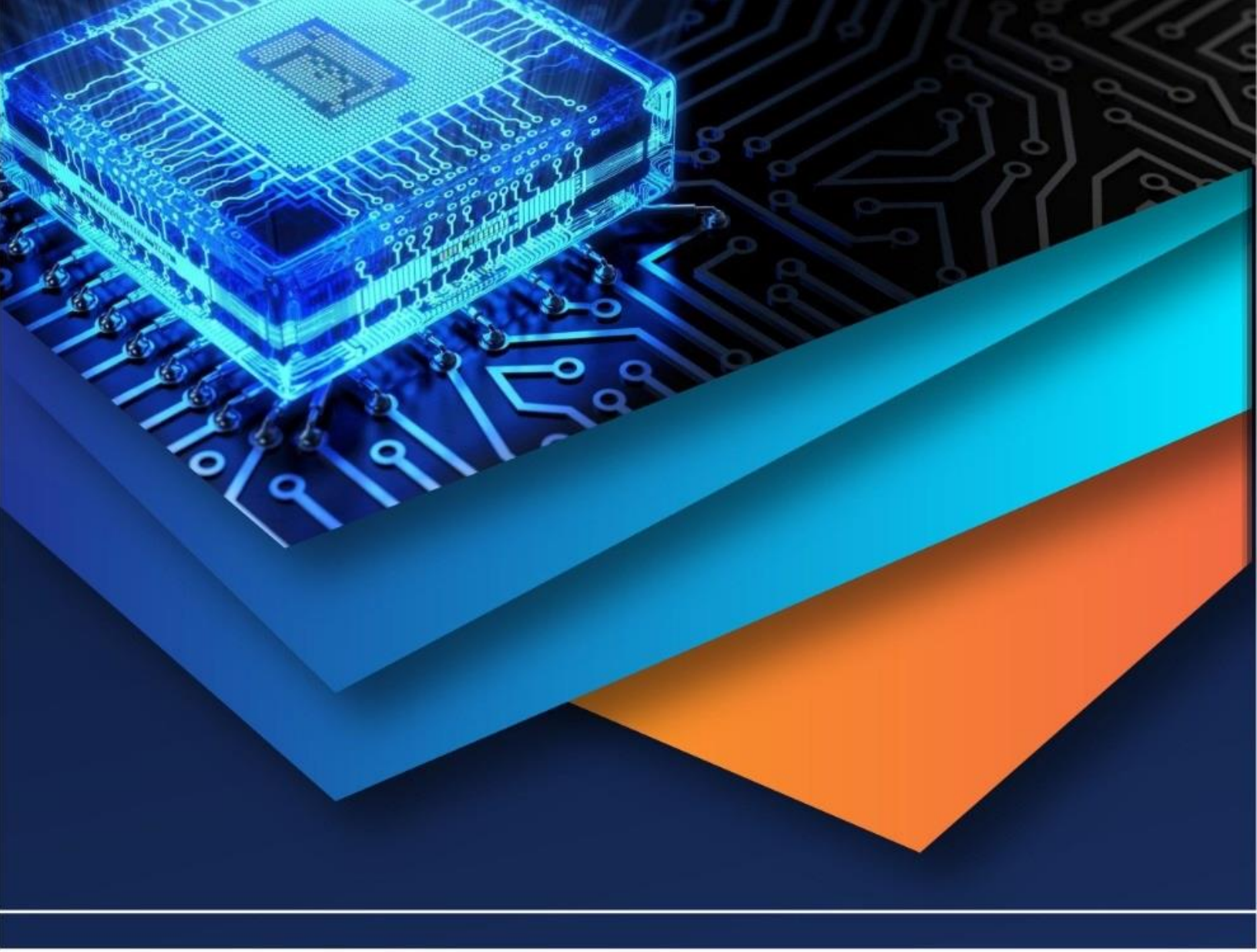
The proposed vehicle insurance fraud detection system presents a holistic and intelligent approach that combines traditional machine learning models, anomaly detection techniques, and explainable AI tools to effectively identify fraudulent claims. By addressing both the accuracy and interpretability challenges commonly associated with fraud prediction systems, the framework ensures reliable performance while remaining transparent and user-friendly. The integration of anomaly detection using Isolation Forest and Autoencoder allows the system to capture subtle or previously unseen fraud patterns, enhancing the quality of the training data for the supervised models. The use of ensemble techniques like stacking further improves predictive performance by leveraging the strengths of multiple classifiers. Additionally, explainability is achieved through SHAP and LIME, enabling stakeholders to understand and trust the system's predictions. The deployment of the system through a secure, minimalistic web interface ensures ease of use and real-time applicability, making it suitable for operational environments.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Overall, the system not only meets the current requirements of fraud detection in vehicle insurance but also lays a strong foundation for future expansion and scalability. It serves as a significant step towards more intelligent, interpretable, and practical fraud detection solutions in the insurance industry. By balancing predictive power with transparency and usability, this work contributes meaningfully to the ongoing efforts to combat financial fraud and protect stakeholders across the insurance ecosystem.

REFERENCES

[1] T. Machinya, G. Mbizo, and K. Zvarevashe, "Insurance Fraud Detection using Machine Learning," in Proceedings of the 2022 1st Zimbabwe Conference of Information and Communication Technologies (ZCICT), Nov. 2022, DOI:https://doi.org/10.1109/ZCICT55726.2022.10046034
[2] Arif Ismail Alrais, "Fraudulent Insurance Claims Detection Using Machine Learning." M.S. thesis, Dept. of Graduate Programs & Research, Rochester Institute of Technology, Dubai, 2022. Available:https://repository.rit.edu/cgi/viewcontent.cgi?article=12510&context=theses
[3] C. Gomes, Z. Jin, and H. Yang, "Insurance fraud detection with unsupervised deep learning," Journal of Risk and Insurance, vol. 88, no. 3, pp. 591–624, Sept. 2021, DOI:https://doi.org/10.1111/jori.12359
[4] H. Cedervall and A. Hansson, "Insurance Fraud Detection using Unsupervised Sequential Anomaly Detection." M.S. thesis, Dept. of Computer and Information Science, Linköping University, Linköping, Sweden, 2022. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1633422/FULLTEXT01.pdf
[5] X. Song, "Research on the application of Bayesian network and deep learning in insurance fraud detection," in *Proc. 2024 3rd Int. Conf. Smart City Challenges Outcomes Urban Transform. (SCOUT)*, Bhubaneswar, India, Jul. 26–27, 2024, doi: 10.1109/SCOUT64349.2024.00029.
[6] N. Ramesar, S. Ramoudith, N. Sharma, and P. Hosein, "A cost-minimization approach to automobile insurance fraud detection," in *Proc. 2023 IEEE Int. Conf. Technol. Manag., Oper. Decis. (ICTMOD)*, Rabat, Morocco, Nov. 22–24, 2023, doi: 10.1109/ICTMOD59086.2023.10438120.
[7] R. Achary, C. J. Shelke, and V. K. Shrivastava, "Insurance claim fraud detection using Benford's method and machine learning," in *Proc. 2025 6th Int. Conf. Intell. Commun. Technol. Virtual Mobile Netw. (ICICV)*, Tirunelveli, India, Jun. 17–19, 2025, doi: 10.1109/ICICV64824.2025.11085965.
[8] M. Shanthini, "Stacking classifier-based automated insurance fraud detection system," in *Proc. 2022 IEEE Conf. Interdiscip. Approaches Technol. Manag. Social Innov. (IATMSI)*, Gwalior, India, Dec. 21–23, 2022, doi: 10.1109/IATMSI56455.2022.10119435.
[9] R. Chauhan, R. Negi, and D. K. Verma, "Analysis of machine learning algorithms for insurance fraud detection," in *Proc. 2023 IEEE 11th Region 10 Humanitarian Technol. Conf. (R10-HTC)*, Rajkot, India, Oct. 16–18, 2023, doi: 10.1109/R10-HTC57504.2023.10461774.
[10] R. Agarwal, D. Kalsi, P. Jain, P. Gupta, and R. Goel, "Car insurance fraud detection using machine learning models," in *Proc. 2025 Int. Conf. Next Gener. Inf. Syst. Eng. (NGISE)*, Ghaziabad, India, Mar. 28–29, 2025, doi: 10.1109/NGISE64126.2025.11085234.
[11] A. Singhal, N. Singhal, Divya, and K. Sharma, "Machine learning methods for detecting car insurance fraud: Comparative analysis," in *Proc. 2023 3rd Int. Conf. Intell. Technol. (CONIT)*, Hubli, India, Jun. 23–25, 2023, doi: 10.1109/CONIT59222.2023.10205670.
[12] B. Sharan, M. Hassan, V. D. Vani, V. H. Raj, G. Nijhawan, and P. P. Pawar, "Machine learning-based fraud detection system for insurance claims in IoT environment," in *Proc. 2024 Int. Conf. Adv. Comput., Commun. Appl. Informat. (ACCAI)*, Chennai, India, May 9–10, 2024, doi: 10.1109/ACCAI61061.2024.10601786.
[13] Z. Chen, Y. Yu, and C. Liao, "International perspectives on rural medical insurance fraud detection: A machine learning approach with SHAP interpretability," in *Proc. 2025 IEEE 7th Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, Guangzhou, China, May 9–11, 2025, doi: 10.1109/CISCE65916.2025.11065331.
[14] A. Jain, T. Dutta, and T. S. Kumar, "Fraud detection and analysis system for car insurance claim using random forest classifier," in *Proc. 2023 Int. Conf. Recent Adv. Sci. Eng. Technol. (ICRASET)*, B G Nagara, India, Nov. 23–24, 2023, doi: 10.1109/ICRASET59632.2023.10420001.
[15] S. K., P. A. S., A. D., S. R., and Y. J., "Healthcare insurance fraud detection powered by blockchain and machine learning: An analysis and framework," in *Proc. 2024 IEEE Int. Conf. Smart Power Control Renewable Energy (ICSPCRE)*, Rourkela, India, Jul. 19–21, 2024, doi: 10.1109/ICSPCRE62303.2024.10675307.
[16] V. R. Saddi, B. Gnanapa, S. Boddu, and J. Logeshwaran, "Fighting insurance fraud with hybrid AI/ML models: Discuss the potential for combining approaches for improved insurance fraud detection," in *Proc. 2023 4th Int. Conf. Commun., Comput. Ind. 6.0 (C2I6)*, Bangalore, India, Dec. 15–16, 2023, doi: 10.1109/C2I659362.2023.10431155.
[17] B. K. Sethi, D. Singh, and P. K. Sarangi, "Medical insurance fraud detection based on blockchain and deep learning approach," in *Proc. 2022 Int. Conf. Disruptive Technol. Multi-Disciplinary Res. Appl. (CENTCON)*, Bengaluru, India, Dec. 22–24, 2022, doi: 10.1109/CENTCON56610.2022.10051253.
[18] A. P. Satone and S. Daronde, "Assessing the effectiveness of machine learning algorithms in vehicle insurance fraud prediction: A comparative approach," in *Proc. 2025 Int. Conf. Electron. Renewable Syst. (ICEARS)*, Tuticorin, India, Feb. 11–13, 2025, doi: 10.1109/ICEARS64219.2025.10941222.
[19] V. R. Saddi, S. Boddu, B. Gnanapa, N. Jiwani, and T. Kiruthiga, "Leveraging big data and AI for predictive analysis in insurance fraud detection," in *Proc. 2024 Int. Conf. Integr. Circuits Commun. Syst. (ICICACS)*, Raichur, India, Feb. 23–24, 2024, doi: 10.1109/ICICACS60521.2024.10498658.
[20] S. Vyas and S. Serasiya, "Fraud detection in insurance claim system: A review," in *Proc. 2022 2nd Int. Conf. Artif. Intell. Smart Energy (ICAIS)*, Coimbatore, India, Feb. 23–25, 2022, doi: 10.1109/ICAIS53314.2022.9742984.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)