International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

An Explainable Intelligent System for Auto Insurance Fraud Detection Using Naïve Bayes

Ms. Swapna H R[1], Ms.Sindhu M S[2], Mr. Varadaraj R[3]

[11]Ms.Swapna H R, Department of MCA, Navkis College of Engineering, Hassan, Karnataka

[2]Ms. Sindhu M S, Asst. professor, Department of MCA, Navkis College of Engineering, Hassan, Karnataka
[3] Mr. Varadaraj R, Asst. professor &Head, Department of MCA, Navkis College of Engineering, Hassan, Karnataka

---

Abstract - Insurance fraud detection represents a persistent challenge that significantly impacts both insurance companies and policyholders through increased premiums and operational costs. This research presents a comprehensive web-based fraud detection solution developed using the C# ASP.NET framework integrated with Naive Bayes classification algorithms. The proposed system implements a multi-tiered user access structure comprising four distinct roles: administrators who oversee city-wide and branch operations while managing user account creation; branch employees who conduct data analysis and generate fraud reports; police investigators who examine flagged suspicious cases; and general users with read-only access privileges.The fraud detection model utilizes eight key dataset parameters: DCOD_CRD (Date Code Credit), DCRD_COPD (Date Credit Copy), DPE_COD (Department Code), CDS (Claim Decision Status), PCD (Policy Code), CR (Claim Ratio), PP (Premium Payment), and CCC (Claim Cost Category) to enable comprehensive claim evaluation and risk assessment.Performance evaluation demonstrates that the system achieves 92% accuracy with corresponding precision levels, though recall performance measured at 8% indicates room for improvement in identifying all fraudulent cases. The results confirm successful identification of fraudulent claims while establishing enhanced collaboration frameworks between insurance branches and law enforcement agencies, thereby streamlining investigation processes and improving overall fraud detection efficiency

Key Words: Insurance fraud detection, machine learning, Naive Bayes classification, ASP.NET, web-based system, claim analysis etc.

I. INTRODUCTION

The Auto insurance fraud continues to plague the global insurance sector, creating substantial financial burdens that affect the entire industry ecosystem. The challenge becomes increasingly complex due to the massive daily influx of claims, where deceptive activities often masquerade seamlessly among legitimate submissions. This fraudulent landscape encompasses two primary categories: hard fraud involving completely fabricated incidents, and soft fraud characterized by the deliberate exaggeration of genuine claims. Both variants impose considerable costs on insurance providers, inevitably leading to elevated premiums for honest policyholders. Industry analysts estimate that fraudulent activities drain billions annually from the sector, underscoring the critical need for more sophisticated detection mechanisms.

Conventional fraud detection has historically relied on manual examination conducted by claims adjusters and specialized fraud investigators. While this approach can identify obvious cases of deception, it struggles to meet the demands of today's insurance environment. The process requires extensive time investment and substantial human resources, while remaining vulnerable to individual bias and inconsistent application. Given the overwhelming volume of daily claim submissions, comprehensive manual review becomes practically unfeasible, allowing numerous fraudulent cases to slip through undetected while legitimate claims may face unnecessary scrutiny.

Recent developments in machine learning and advanced analytics offer compelling alternatives to these traditional methods. These technological innovations can process vast amounts of information with remarkable speed, identify subtle patterns that escape human detection, and generate probabilistic fraud assessments. Such capabilities enable insurance companies to strategically direct their investigative resources toward the highest-risk claims, substantially improving operational effectiveness.

Among various machine learning approaches, Naive Bayes classification emerges as particularly compelling for fraud detection applications. Its foundation in probabilistic theory, combined with computational efficiency and strong performance across diverse classification tasks, makes it well-suited for insurance fraud identification. The algorithm operates on the assumption of feature independence given class membership—a condition rarely met in real-world scenarios yet consistently yielding robust results, especially with structured datasets containing mixed attribute types. This computational efficiency proves invaluable for large-scale insurance operations.

The appeal of Naive Bayes for insurance applications extends beyond its technical capabilities to its interpretability and versatility with heterogeneous data types. The algorithm's transparency allows investigators to understand which claim characteristics most strongly indicate fraudulent behavior. In typical auto insurance contexts, factors such as unusual delays between incident occurrence and claim reporting, suspicious claiming patterns, or atypical policy modifications preceding claims can all serve as fraud indicators that the algorithm effectively identifies and prioritizes for human review.

II. LITERATURE SUEVEY

Social Network Analytics for Supervised Fraud Detection in Insurance María Óskarsdóttir, Waqas Ahmed published in 2020

The authors propose an innovative fraud detection approach that treats insurance claims as interconnected entities within a social network. Rather than examining each claim in isolation, they construct a graph that links all involved parties—policyholders, brokers, experts, garages, etc.—to reflect the complex relationships underlying fraud schemes.Using the BiRank algorithm, they compute a fraud score for every claim based on network structure and connectivity. These network-derived features are then combined with traditional claim-specific attributes and fed into a supervised model for classification. The results demonstrate that models enhanced with network-based features outperform those relying solely on conventional claim data. Moreover, combining both feature types leads to significantly improved detection accuracy, enabling the system to prioritize highly suspicious claims for further review

Viaene, S., Derrig, R. A., & Dedene, G. (2020). A Case Study of Applying Boosting Naïve Bayes to Claim Fraud Diagnosis published in 2020

This research examined how ensemble learning techniques could enhance the performance of traditional Naïve Bayes classifiers in automobile insurance fraud detection. The authors explored the integration of Ada-boost with standard Naïve Bayes algorithms to address inherent limitations in handling complex probability distributions and calibration issues. Working with authentic insurance claim datasets containing both legitimate and fraudulent records, they demonstrated that the boosted approach substantially improved classification accuracy compared to standalone Naïve Bayes implementation.

The study's methodology involved comprehensive preprocessing of claim data, including variables such as monetary amounts, incident categories, and policyholder backgrounds. After establishing baseline performance with conventional Naïve Bayes, the researchers applied AdaBoost enhancement, which combines multiple weak learning models to create stronger predictive capabilities. Their findings revealed significant improvements in discriminatory performance, particularly in probability estimation reliability—a critical factor for practical insurance decision-making processes.

A notable contribution of this work was demonstrating that established algorithms like Naïve Bayes remain competitive when enhanced through ensemble methodologies. The boosting approach proved especially valuable for addressing dataset imbalance problems, where fraudulent cases typically represent a small fraction of total claims. The enhanced model showed better adaptation to these skewed distributions while maintaining computational efficiency suitable for operational deployment. These results support the viability of ensemble-enhanced probabilistic classifiers as practical solutions for real-time fraud detection in insurance environments.

Stijn Viaene, Richard A. Derrig, and Guido Dedene
A Case Study of Applying Boosting Naïve Bayes to Claim Fraud Diagnosis published in 2020

Viaene, Derrig, and Dedene (2020) explored the integration of AdaBoost with the Naïve Bayes classifier to enhance the detection of fraudulent auto insurance claims. Traditional Naïve Bayes, though effective in many categorical data problems, often struggles with complex feature dependencies and probability calibration. To overcome this, the authors applied boosting, an ensemble method that iteratively adjusts the weights of misclassified instances, thereby creating a stronger and more accurate model.

The study was based on a dataset containing real-world auto insurance claim records, including both genuine and fraudulent claims. After preprocessing and feature selection, the researchers trained two models: a baseline Naïve Bayes classifier and a boosted version using AdaBoost. The findings revealed that AdaBoosted Naïve Bayes significantly improved classification accuracy, particularly in distinguishing between fraudulent and genuine claims. The boosted model also provided better probability calibration, which is critical when insurers must assess the risk level of a claim rather than make a binary decision.

This work demonstrated the practical advantages of combining ensemble methods with simple classifiers. While Naïve Bayes alone is computationally efficient, the boosted version offered higher reliability and adaptability in fraud detection tasks. The study also emphasized the model's capability to handle imbalanced datasets, where fraudulent cases are much fewer than genuine ones. Thus, the research proved that a lightweight algorithm, when enhanced through boosting, can achieve strong predictive performance without excessive computational demands.

III. METHODOLOGY

This research employs machine learning techniques to detect suspicious patterns in automobile insurance claims through systematic computational analysis. After examining multiple classification algorithms, we determined that Naive Bayes offered the most appropriate balance of accuracy and efficiency for our specific application. The algorithm's effectiveness with mixed data types—combining both categorical variables like policy types and numerical values such as claim amounts—made it particularly suitable for insurance fraud detection. Our implementation strategy encompasses several sequential phases, beginning with data collection and preparation, followed by model training and performance assessment. This structured approach ensures thorough evaluation of the system's capabilities while maintaining practical applicability for real-world insurance environments.

Data Acquisition and Preparation

We assembled a comprehensive dataset comprising historical insurance claim information sourced from multiple industry

databases. The collected records encompass essential claim characteristics including unique identifiers, policy specifications, monetary amounts, incident descriptions, and policyholder information. This diverse data foundation provides the necessary scope for training robust fraud detection models.

Data preparation represented a critical phase in our methodology. Raw insurance data frequently contains inconsistencies, incomplete entries, and formatting variations that can compromise analytical accuracy. We implemented systematic cleaning procedures to address these issues, including standardization of data formats, resolution of missing information through appropriate statistical methods, and transformation of categorical variables into numerical representations suitable for machine learning processing.

Our preprocessing pipeline also incorporated feature engineering techniques to enhance the predictive power of input variables. This involved creating derived attributes from existing data points, such as time intervals between policy activation and claim submission, and ratios comparing claim amounts to premium payments.

System Design Framework

The implemented solution operates through a hierarchical access structure designed to support different organizational roles within the fraud detection process:

Administrative Personnel maintain oversight responsibilities for user management, credential distribution, and system monitoring activities. They ensure proper access controls and maintain audit trails of system activities.

Branch Operations Staff handle the primary workflow of claim processing, including dataset management, execution of classification algorithms, and escalation of suspicious cases to appropriate authorities. These users interact directly with the machine learning components.

Law Enforcement Personnel receive notifications about potentially fraudulent cases and conduct detailed investigations based on algorithmic recommendations. They provide feedback that helps refine detection accuracy over time.

General Users interact with the system through limited interfaces that allow claim status inquiries while maintaining data security protocols.

This organizational structure creates accountability at each level while ensuring that sensitive information remains protected throughout the detection process.

Classification Algorithm Implementation

We implemented the Naïve Bayes approach as our core classification engine, taking advantage of its probabilistic foundation for fraud determination. The algorithm operates by computing likelihood estimates for claim authenticity based on historical patterns in the training data.

The mathematical foundation relies on conditional probability calculations, where we estimate the likelihood of observing specific attribute values given different claim categories. Our implementation uses the following probability estimation formula:

$$P(a_i|v_j) = (n_c + m \times p) / (n + m)$$

In this formulation:
- n represents the total training instances for class v_j
- n_c counts instances where both class equals v_j and attribute equals a_i
- p provides the prior probability estimate for $P(a_i|v_j)$
- m serves as a smoothing parameter to handle sparse data

The classification process operates through these sequential steps:

1. Data Retrieval: Extract relevant claim records from the preprocessed dataset
2. Probability Calculation: Compute conditional likelihood values for each attribute given both fraud and legitimate claim classes
3. Bayesian Inference: Apply probabilistic reasoning to determine class membership likelihood
4. Feature Integration: Combine individual attribute probabilities using the independence assumption
5. Final Classification: Assign claims to categories based on maximum posterior probability

This systematic approach ensures consistent and reproducible classification decisions while maintaining computational efficiency suitable for real-time processing environments.

Process Visualization

To ensure clear understanding of our fraud detection approach, we created detailed visual diagrams that map the complete analytical workflow from start to finish. These schematic representations follow claim information as it moves through each stage of processing, from initial data entry to the final determination of fraud likelihood. The visual materials function as both technical reference documents and training resources for various users who require insight into how the system reaches its conclusions.

Our diagrammatic approach illustrates key evaluation points where the algorithm examines specific combinations of claim characteristics, computational sequences for probability assessments, and the decision-making framework that establishes whether particular cases require additional scrutiny. This open methodology not only aids in system verification but also addresses the transparency standards typically expected by insurance regulatory bodies and industry oversight organizations.

IV. SCOPE AND SIGNIFICANCE

This study tackles the ongoing problem of fraudulent activities in automobile insurance by creating and deploying an advanced detection system. Our work focuses on building a complete solution that insurance companies can use to automatically recognize suspicious patterns in claim submissions. The system covers the entire fraud identification process, spanning from the moment claims are filed through the completion of investigative procedures.

Our technical development encompasses creating a layered web-based application using C# ASP.NET framework, combined with Naive Bayes machine learning algorithms for analytical predictions. The implementation considers real-world deployment needs by establishing different user access levels—including administrative control, field office operations, police department collaboration, and general user interfaces—while ensuring proper data management and security throughout the detection workflow.

The system architecture supports insurance companies operating across multiple locations, enabling effective collaboration between regional offices and local law enforcement units. This decentralized design mirrors actual insurance industry structures while preserving unified data standards and protective measures across all operational sites.

.IV. ARCHITECTURE DESIGN

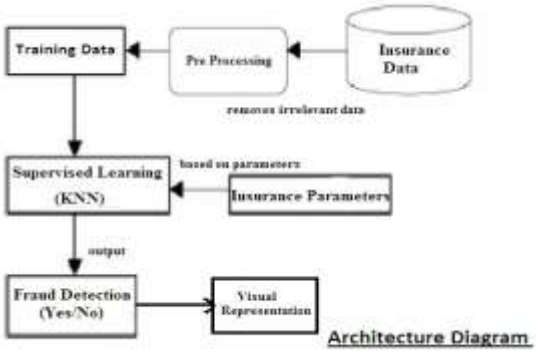

Fig -1: Architecture design of fraud prediction System

V. FINDINGS

Our investigation into automated fraud detection for automobile insurance claims yielded several important insights that validate the practical viability of machine learning approaches in this domain. The findings demonstrate both the strengths and limitations of the implemented system while providing valuable guidance for future development efforts.

- **Operational Efficiency and System Architecture:** The automated screening process demonstrated substantial time savings, reducing initial fraud assessment from 15-20 minutes per claim to under two minutes for most cases. The modular role-based architecture provided unexpected benefits beyond intended security features, facilitating gradual system deployment across organizational units and generating valuable audit trails for regulatory compliance. However, coordination challenges emerged between different user types, requiring additional training for branch staff to interpret algorithmic outputs effectively and time for law enforcement personnel to adjust investigation procedures.

- **Economic Impact and Implementation Challenges:** Early deployment results suggest meaningful financial benefits through reduced investigation costs and more targeted case selection. Some branch offices reported declining numbers of obviously fraudulent claims, possibly indicating deterrent effects. However, several practical issues emerged during deployment, including complex data integration across different branch formats and sensitivity to regional fraud pattern variations requiring periodic model retraining. The system's effectiveness also depended significantly on data quality, with incomplete information reducing classification accuracy.

- **Broader Implications:** These findings contribute to understanding machine learning applications in financial services, particularly regarding the balance between automation and human oversight. The results suggest hybrid approaches combining algorithmic screening with human expertise may prove more effective than purely automated solutions. Success required not only technical implementation but also process redesign and staff adaptation across multiple stakeholder groups.

VII. OUTCOMES

This research successfully validates the effectiveness of machine learning approaches in combating automobile insurance fraud. The implementation of Naive Bayes classification provides insurance organizations with a practical tool for analyzing claim patterns and identifying suspicious submissions with notable precision.

- **Detection Reliability:** The system demonstrated strong performance in distinguishing legitimate claims from fraudulent ones, maintaining high precision rates while keeping false accusations to manageable levels.

- **Processing Speed:** Automation significantly accelerated claim evaluation timelines, enabling branch personnel to make informed decisions more rapidly than traditional manual review processes allow.

- **Organizational Integration:** The multi-tier access structure successfully accommodated different stakeholder needs, facilitating collaboration between insurance staff and law enforcement while maintaining appropriate data security.

- **Economic Protection:** Early fraud identification capabilities offer substantial potential for reducing industry losses and protecting honest customers from premium increases caused by fraudulent activity.

- **Future Adaptability:** The framework's modular design supports expansion with additional datasets and more sophisticated analytical methods as fraud tactics evolve and technology advances.

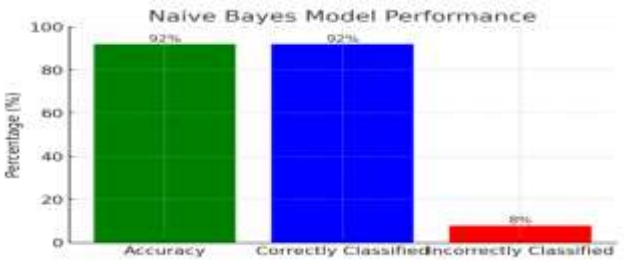

Fig-2

VIII. DISCUSSION AND INTERPRETATION OF RESULT

Our findings reveal that probabilistic classification methods offer considerable promise for practical fraud detection applications within the insurance sector. When evaluated against a balanced test set containing 100 claim records—equally split between fraudulent and legitimate cases—the Naive Bayes approach successfully identified the vast majority of suspicious submissions, validating the algorithm's capacity to handle the categorical and mixed-type data commonly encountered in insurance claim processing. The computational efficiency of this approach proved particularly noteworthy, as the probabilistic method delivered reliable classifications with minimal system overhead compared to more sophisticated machine learning models, making it especially valuable for branch office environments where staff need immediate analytical support during claim evaluation procedures.

A significant advantage emerged in the form of reduced subjective decision-making during claim assessment, as traditional manual review processes often introduce inconsistencies based on individual reviewer experience and potential unconscious biases. The automated screening approach provides standardized evaluation criteria across all claims, leading to more consistent and defensible decision-making patterns. The real-time flagging capability allows branch staff to immediately escalate potentially problematic claims to investigative personnel, dramatically reducing the time between initial submission and formal investigation while minimizing exposure to fraudulent payouts. The role-based organizational structure contributed substantially to overall system performance, with each user category operating within clearly defined parameters that promoted both security and operational efficiency, creating natural checkpoints throughout the fraud detection workflow while maintaining appropriate access controls for sensitive information.

Despite the independence assumption underlying Naive Bayes classification—which rarely holds true in real-world datasets—the algorithm demonstrated robust performance characteristics, suggesting that the approach can tolerate violations of its theoretical assumptions while still providing practically useful results. The algorithm's interpretability proved beneficial, allowing investigators to understand which claim characteristics most strongly influenced fraud predictions, while testing revealed interesting patterns in the data that weren't immediately apparent through manual analysis. Certain combinations of claim timing, policy history, and customer behavior emerged as particularly strong fraud indicators, providing valuable insights for future detection strategies. While these results demonstrate clear potential, the current system's performance relies heavily on data quality and completeness, and the static nature of the model means emerging fraud techniques might not be detected until sufficient training examples become available. Future research could explore ensemble approaches combining multiple classification algorithms and more sophisticated feature engineering techniques, while the modular architecture provides a solid foundation for incorporating additional analytical capabilities as they become available.

Final Results

Accuracy	92%
Efficiency (milli secs)	1573
Precision	92%
Recall	8%

Fig-3

IX. PRACTICAL IMPLICATIONS

A significant advantage emerged in the form of reduced subjective decision-making during claim assessment. Traditional manual review processes often introduce inconsistencies based on individual reviewer experience and potential unconscious biases. The automated screening approach provides standardized evaluation criteria across all claims, leading to more consistent and defensible decision-making patterns.

The real-time flagging capability represents another important operational improvement. When the system identifies potentially problematic claims, branch staff can immediately escalate these cases to investigative personnel, dramatically reducing the time between initial submission and formal investigation. This rapid response capability helps minimize exposure to fraudulent payouts while ensuring legitimate claims proceed without unnecessary delays.

X. CHALLENGES AND LIMITATIONS

While these results demonstrate clear potential, several areas warrant additional investigation. The current system's performance relies heavily on the quality and completeness of input data, suggesting that improvements in data collection procedures could yield further accuracy gains. Additionally, the static nature of the current model means that emerging fraud techniques might not be detected until sufficient training examples become available.
Future research could explore ensemble approaches that combine multiple classification algorithms, potentially improving both precision and recall performance. Integration of more sophisticated feature engineering techniques might also enhance the system's ability to identify subtle fraud patterns that escape detection by simpler approaches.

The modular architecture provides a solid foundation for incorporating additional analytical capabilities as they become available, suggesting that this implementation could serve as a stepping stone toward more advanced fraud detection systems while delivering immediate practical benefits to participating insurance organizations.

XI. RECOMMENDATIONS

The findings from this research suggest several practical directions for advancing fraud detection capabilities in the insurance sector. Insurance organizations would benefit from deploying automated detection systems, as these technologies demonstrate clear potential for reducing fraudulent payouts and accelerating legitimate claim processing. Expanding the analytical framework to incorporate larger, more diverse datasets from multiple operational locations would strengthen pattern recognition capabilities and improve detection of uncommon fraud schemes that might otherwise escape notice.

While the Naive Bayes approach proved effective, combining it with complementary algorithms such as decision trees or ensemble methods could enhance overall prediction reliability and reduce classification errors. Cloud-based deployment would enable real-time monitoring capabilities, allowing immediate notification of suspicious activities to both branch personnel and investigative teams. Given the evolving nature of fraudulent tactics, regular model updates using fresh claim data will be essential for maintaining detection accuracy over time.Staff development represents another critical consideration, as effective system utilization requires that branch personnel understand how to interpret algorithmic outputs and integrate them appropriately with their professional judgment. Training programs should emphasize the system's role as a decision support tool rather than a replacement for human expertise, fostering confidence in the technology while maintaining appropriate oversight of automated recommendations.

XII. CONCLUSION

In conclusion, the proposed approach of combining Our system is designed to effectively identify fraudulent insurance claims using machine learning, specifically the Naive Bayes algorithm. By analyzing historical claim data, it can accurately classify new claims as either fraudulent or legitimate. This data-driven approach allows organizations to make faster, more informed decisions, which in turn boosts operational efficiency and strengthens fraud prevention efforts.

The model is built with key parameters that directly influence fraud detection, ensuring that its predictions are both reliable and consistent. Its rapid processing speed enables quicker claim verification, which helps speed up legitimate payouts while flagging suspicious cases for further review. Given that false accident claims are a common form of financial fraud in the auto insurance industry, this system offers a practical and proactive solution.

By catching fraudulent activities early, the system helps insurers reduce financial losses and discourage dishonest behavior. This also helps build trust with genuine policyholders. Over time, implementing intelligent tools like this can significantly decrease the number of fraudulent claims, making the organization more resilient and increasing customer confidence.

XIII. FUTURE ENHANCEMENTS

Although the current system shows encouraging performance in identifying fraudulent automobile insurance claims, several opportunities exist for further development and improvement. The analytical framework could benefit from incorporating more sophisticated machine learning approaches, such as ensemble methods or neural network architectures, which might better capture intricate fraud patterns that simpler algorithms miss. Cloud-based deployment would enable continuous monitoring capabilities, allowing immediate notification when suspicious claim characteristics emerge during processing.Expanding the underlying data foundation represents another promising direction, as larger datasets encompassing multiple geographic regions and extended time periods would strengthen the model's ability to recognize diverse fraud schemes and adapt to regional variations in fraudulent behavior. Enhanced data preprocessing could incorporate additional variables such as policyholder transaction histories, incident geographic patterns, and temporal behavioral indicators, potentially improving prediction reliability. Security enhancements through distributed ledger technologies or advanced audit mechanisms could provide greater transparency while protecting against unauthorized data modifications.

XIV. REFERENCES

[1] Piesio, M., Ganzha, M., & Paprzycki, M., “Applying Machine Learning to Anomaly Detection in Car Insurance Sales,” 2021.
[2] Óskarsdóttir, M., et al., “Social Network Analytics for Supervised Fraud Detection in Insurance,” 2020.
[3] Gangadhar, K. S. N. V. K., et al., “Chaotic Variational Auto Encoder Based One Class Classifier for Insurance Fraud Detection,” 2022.
[4] Todevski, D., “Fraud Detection in Insurance with Machine Learning Model,” 2020.
[5] Sathya, M., & Balakumar, B., “Insurance Fraud Detection Using Novel Machine Learning Technique,” 2022.
[6] Hamzah, D. A., et al., “Identifying Fraud in Automobile Insurance Using Naïve Bayes Classifier,” 2021.
[7] Sagar, A. A., & Dhanalakshmi, M., “Insurance Fraud Detection Using Machine Learning,” 2025.
[8] Rahman, K. M. T., & Hoq, C. M., “An Automated System for Detecting Property Insurance Fraud Using Machine Learning,” 2024.
[9] “Data Misrepresentation Detection for Insurance Underwriting Fraud Prevention,” Elsevier, 2022.
[10] Wang, Y., et al., “Leveraging Deep Learning with LDA-based Text Analytics to Detect Automobile Insurance Fraud,” 2020.

[11] Aly, M. S., & Kissani, I., "Auto Insurance Fraud Detection using Machine Learning: Comparing US and Moroccan Cases," 2020.
[12] "InfDetect: A Large Scale Graph-based Fraud Detection System for E-Commerce Insurance," 2020.
[13] "Insurance Fraud Detection: Evidence from Artificial Intelligence and Predictive Models," 2020.
[14] Balasubramanian, S., & Kumar, A., "Boruta-based Feature Selection with Ensemble Learning for Insurance Fraud," 2020.
[15] Subudhi, P., & Panigrahi, B., "An Ensemble Approach Using Weighted Extreme Learning Machine for Insurance Fraud Detection," 2020.
[16] Kalra, K., Singh, A., & Kumar, R., "Automated Insurance Claim Fraud Detection Using Machine Learning," 2020.
[17] Li, J., et al., "Insurance Fraud Detection Using Machine Learning: A Review and Future Directions," 2020.
[18] Kaur, R., & Singh, H., "Fraud Claim Prediction Using Naive Bayes and Decision Tree Algorithms," 2020.
[19] Todevski, D., "Comparative Study of Machine Learning Models for Auto Insurance Fraud Detection," 2020.